



HAL
open science

Conditions de Monge, Transport Optimal et Pont Relationnel : propriétés, applications et extension du couplage d'indétermination

Pierre Bertrand

► **To cite this version:**

Pierre Bertrand. Conditions de Monge, Transport Optimal et Pont Relationnel : propriétés, applications et extension du couplage d'indétermination. Statistiques [math.ST]. Sorbonne Université, 2021. Français. NNT : 2021SORUS368 . tel-03602842

HAL Id: tel-03602842

<https://theses.hal.science/tel-03602842>

Submitted on 9 Mar 2022

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Sorbonne Université

ED 386 : École doctorale de Sciences Mathématiques de Paris Centre
Laboratoire de Probabilités, Statistique et Modélisation

Thèse de doctorat de Mathématiques

Conditions de Monge, Transport Optimal et Pont Relationnel *Propriétés, applications et extension du couplage d'indétermination*

par M. Pierre BERTRAND

dirigée par Michel Broniatowski et Jean-François Marcotorchino

présentée et soutenue publiquement le 30 novembre 2021
devant un jury composé de :

M. Julien AH-PINE	MdC, Université Lyon 2	Examinateur
M. Gérard BIAU	Professeur, Sorbonne Université	Examinateur
M. Michel BRONIATOWSKI	Professeur, Sorbonne Université	Directeur
Mme Patricia CONDE-CESPEDES	MdC, ISEP	Examinatrice
M. Jean-François MARCOTORCHINO	Professeur, ISUP	Directeur
M. Gilbert SAPORTA	Professeur, CNAM	Examinateur

Rapporteurs :

M. Gilbert SAPORTA	Professeur, CNAM
M. Konstatinos ZOGRAFOS	Professor, University of Ioannina

Et c'est tout à fait normal. C'est
comme ça que la réalité se
défend : en laissant les enfants
devenir des hommes. Tu
m'aimes tout autant, je sais. [...]
Tu m'aimes tout autant, mais
pas de la même façon. Tu n'as
plus l'imagination qu'il faudrait
pour me donner à vivre.

Romain Gary, *Les enchanteurs*,
1973

À mon enfance

Remerciements

Au grand jour de la mémoire habituelle, les images du passé pâlisent peu à peu, s'effacent, il ne reste plus rien d'elles, nous ne les retrouverons plus. Ou plutôt nous ne le retrouverions plus, si quelques mots [...] n'avaient été soigneusement enfermés dans l'oubli[...]

Marcel Proust, *À l'ombre des jeunes filles en fleurs*, 1918

Nous y voici. La page la plus lue. Je disais un jour à Michel que certains papiers de recherche semblent écrits par l'auteur pour l'auteur. Cette page-ci est écrite pour d'autres, pour remercier, témoigner, ou d'une amitié ou d'une aide, ou d'une origine ou de tout à la fois. Elle est remplie d'anecdotes, suivant l'épigraphe mentionné et selon lequel c'est dans celles-ci que la mémoire s'ancre le plus solidement. Commençons.

Naturellement, mes premiers remerciements vont à Jean-François Marcotorchino. Il m'a pris sous son aile, a initié cette thèse, m'encourageant à sortir de la tranquillité relative d'un poste d'ingénieur. Il m'a proposé un sujet large, une autoroute, de laquelle dériver à loisir, c'était dans une salle de réunion de Thales. Plus tard, s'adaptant à mes disponibilités, nous nous retrouvions et discussions de maths dans son appartement. Toujours très bien reçu, je m'étonnais souvent d'être invité à rester dîner. Je remercie d'ailleurs son épouse Colette pour sa gentillesse et ses attentions. Si ma thèse a pu constituer un loisir c'est grâce à l'intelligence du sujet défini par Jean-François, à l'affection qu'il m'a témoignée et à sa disponibilité infinie. Je ne l'ai jamais nommé autrement que Marco, lui demandant la permission une fois pour toute ; c'était il y a quatre ans.

Sous l'initiative de Marco justement, nous rencontrâmes Michel Broniatowski pour co-encadrer le projet. C'était une excellente idée et le trio ainsi formé était parfait. Outre que nous avançons, je crois pouvoir dire que nous nous amusons tous. De son bureau

de Paris 6 à son appartement, Michel s'est toujours plié à mes horaires de bureau en me recevant tard le soir. Sur certains sujets de statistiques, il m'a proposé des cours particuliers improvisés qui transformeraient n'importe quel élève en docteur en maths. L'habitude aidant, le temps passant, vinrent des rituels : après le travail nous prenions parfois un *jus dans un troquet* ou nous mangions les fameuses pâtes de Michel. Enfin, toujours cette phrase pour conclure nos avancées mathématiques : *je crois que nous tenons quelque chose* ; je peux dire aujourd'hui qu'elle me manquera. Si je suis promu docteur c'est bien grâce au temps que m'a consacré Michel. Je ne saurai jamais lui témoigner assez de reconnaissance pour ses idées mathématiques, sa culture, sa bienveillance et, également, pour son humour.

Il y a deux ans, Gilles Celeux a participé à mon évaluation de mi-thèse, je le remercie pour son appréciation et ses commentaires bienveillants. Only a few people actually have read, let us say more than one third of this report, among them belong the two reviewers : Gilbert Saporta and Konstantinos Zografos. I thank you both for dedicating some of your valuable time to this task, for your kind comments and of course for your overall approval. Julien Ah-Pine, Gérard Biau, Patricia Conde-Céspedes et Gilbert Saporta (à nouveau) ont accepté de faire partie de mon jury, c'est un immense honneur et je les en remercie sincèrement.

Viennent maintenant mes proches. Mes parents, sans qui je n'écrirais pas ces lignes. Maman, que d'aventures depuis les tables de multiplication à Kourou, tu m'as encouragé, parfois réconforté et toujours aimé au travers d'innombrables attentions : ce goût du travail, je te le dois. Papa, tu m'as souvent réconforté et tant inspiré, toujours devant avec panache, comme dans cette barque coulant, dans cet avion survolant l'Amazonie ou sur le voilier : si la thèse est originale c'est certainement dû au goût de la liberté que tu m'as transmis. Vivi, merci de jouer le rôle du chien dans la chanson si belle de Reggiani. Mamie Paule, te voici réapparaître dans ces lignes, je garde près de moi un billet qui t'appartiendra toujours, la certitude que "oui, on peut" et tant de souvenirs depuis la plage où nous nous baignions jusqu'aux restaurants chinois que nous partageons. Papy et Mamie, si j'ai autant de plaisir à retourner, chaque été, en Lozère c'est bien grâce à vous deux et à tous les bons moments passés ensemble. Tintin, depuis toujours, tu as une place à part comme en témoigne ton surnom qui doit dater de mes premiers mots et que je suis heureux d'écrire dans ce document. A tes côtés, Tatïe, m'a toujours si bien accueilli que j'avais l'impression d'être son fils quand Simon, pour parfaire le tableau apparaissait comme un petit frère.

À mes amis les plus proches je dédie chacun une anecdote. Lilian, merci pour ce voyage en Corse et notre course improbable pour la navette retour. Romain, Pointu, depuis déjà dix ans nous ne savons arrêter de discuter sans que le soleil se lève. Benjamin, il y a jusqu'à ton sourire qui n'est pas commun mais j'aurais du mal à m'en passer. Charline, ta gentillesse et ton dévouement il y a deux ans je ne les oublierai jamais : aujourd'hui je confirme que tu n'es pas que bête. Laure, ces karaokés à des heures improbables il n'y

a que toi pour les organiser. Nicolas, merci pour cette vodka inoubliable et ton soutien bienveillant depuis nos cours à la Mine. Paul-Arthur, merci pour notre recette fétiche et cette expression *chiche* à l'origine de tant de moments magiques. Pierre-Antoine, ces lignes prouvent que j'ai su oublier mon absence dans tes remerciements. Étienne, égoïstement, je suis heureux que tu sois défroqué. Herbert, merci de partager avec moi ces balades en moto. RJ, merci pour tous nos CCA.

Merci à mes déjà vieux amis de Cachan : Alain, Damien, Jessica, Laurent, Ludovic et Tristan pour tous les RU partagés pendant quatre années ; c'est un peu aussi pour vous rejoindre que je deviens docteur. Merci à mes amis de Thiers : Cécile, Johnny, Pierre-Antoine, Rémi et Théo pour toutes nos cassosseries. Merci également à mes amis de la BNP : Paulo pour m'avoir accueilli à l'origine et conseillé de faire du tennis à une époque creuse, Tom pour toutes les vies que nous avons sauvées, Mathis pour me faire goûter le rêve audonien et Alexis pour tes conseils de lecture.

Merci à mes anciens voisins, Javier et Ivane pour quelques falafels salvateurs durant la rédaction du manuscrit et Caroline pour avoir supporté Brassens à des horaires quelque peu tardifs.

Enfin, merci à Baptiste de renvoyer mon coup droit trop croisé et à Florence de m'entraîner à rater des volées ; nous savons maintenant que tout se corrige en se détendant.

Vous tous, je vous remercie pour votre soutien dans les passages creux, les bons moments passés ensemble, les échanges, les discussions, les crises, les galères, les joies et les découvertes. Vous enrichissez ma vie et en motivez l'intérêt.

Table des matières

1	Introduction	1
2	Construction de deux couplages canoniques	7
2.1	Problèmes historiques	8
2.1.1	Problème de Monge	8
2.1.2	Problème de Monge-Kantorovich	10
2.1.3	Applications	12
2.2	Deux modélisations : l'indépendance et l'indétermination	14
2.2.1	Écart à l'indépendance : l'entropie d'Alan Wilson	14
2.2.2	Écart à l'indétermination : déviation quadratique	17
2.3	Optimisation sans hypothèse sur les marges	22
2.4	Deux fonctions de couplage canoniques	24
2.4.1	Fonction de coût et divergence	24
2.4.2	Justification structurelle des deux divergences	27
2.4.3	Un écart relatif	29
3	Propriétés de l'indétermination	31
3.1	Probabilité de construction de π^+	32
3.2	Corrélation et indétermination	35
3.3	Conditions de Monge discrètes	39
3.3.1	Monge - indétermination	40
3.3.2	Caractérisation structurelle de l'indétermination	41
3.3.3	Log Monge - indépendance	43
3.3.4	Caractérisation structurelle de l'indépendance	44
3.3.5	Résumé du parallèle	45
3.4	Indétermination et analyse relationnelle mathématique	45
3.4.1	Introduction	45
3.4.2	Formules de passage	49
3.4.3	Codage linéaire du type de relation	50
3.4.4	Interprétation relationnelle de l'indétermination	50
3.4.5	Égalité accords désaccords probabiliste	53
3.4.6	Exemple de jugements	55
3.5	Tirage selon l'indétermination	55

3.5.1	Procédé constructif	55
3.5.2	Graphiques du tirage selon l'indétermination	57
3.5.3	Temporalité	58
3.5.4	Limitation des collisions	59
4	Applications de l'indétermination	63
4.1	Problème dit de l'espion	64
4.2	Problème du partitionnement de tâches	69
4.3	Clustering de graphes	72
4.3.1	Introduction	72
4.3.2	Formalisation	77
4.3.3	Modularité originelle – indépendance	79
4.3.4	Modularité déformée – indétermination	80
4.3.5	Résumé de l'application expérimentale à des graphes variés	80
4.3.6	Comparaison dans le cadre Gilbert	81
4.3.7	Graphes spécifiques	83
4.3.8	Un seuil commun	86
5	Coefficients de corrélation	89
5.1	Codage relationnel	90
5.1.1	Critère de Rand	90
5.1.2	Critère de Janson-Vegelius	91
5.1.3	Critère du χ_2	93
5.1.4	Critère de Tchuprow	95
5.1.5	Critère de Lerman-Chah-Gini	96
5.1.6	Dualité indépendance-indétermination	97
5.1.7	Comparaison des valeurs de critères	98
5.1.8	Forme générique des critères	101
5.2	Focus sur le Janson-Végélius	104
5.2.1	Histoire	104
5.2.2	Modèle des deux juges	107
5.2.3	Valeur moyenne	109
5.3	Un produit scalaire commun	114
5.3.1	Encadrement	115
5.3.2	Maximisation	117
5.3.3	Diagonalisation	121
5.3.4	Minimisation	131
6	Cas continu	137
6.1	Indétermination continue	137
6.1.1	Fonction de couplage cas continu	137
6.1.2	Un cas continu du problème de déviation quadratique	138
6.1.3	Tirage selon l'indétermination : cas continu	141
6.1.4	Caractérisation des marges éligibles	142
6.2	Une copule d'indétermination définissable localement	143
6.2.1	Définition formelle	144

6.2.2	Lien entre marges et copule	144
6.2.3	Une copule d'indétermination locale	146
6.2.4	Ecart maximal entre les deux copules	150
6.2.5	Conclusions	152
6.3	Indétermination et minimisation de la vraisemblance moyenne	153
6.3.1	Définition de la vraisemblance moyenne	153
6.3.2	Évanouissement de la corrélation sous l'indétermination	154
6.3.3	Vraisemblance moyenne et collisions	154
6.4	Statistique de test basée sur la vraisemblance moyenne	155
6.4.1	Définition de t_n	155
6.4.2	Quelques notions et lemmes usuels	156
6.4.3	Théorème de Sanov	159
6.4.4	Analyse de la statistique de test t_n	160
7	Conclusion	167
A	Minimization with respect to divergences and applications	171
A.1	Introduction	171
A.2	Projection using divergences	172
A.3	Mimisation of couple matchings	173
A.4	Applications	175
A.4.1	Guessing problem	175
A.4.2	Task Partitioning	178
B	Independence versus Indetermination : basis of two canonical clustering criteria	181
B.1	Introduction	181
B.2	Parallel discovery of two dual couplings	183
B.2.1	Some few words about Optimal Transport	184
B.2.2	The Alan Wilson's Entropy Model : role of independence	185
B.2.3	The minimal trade model : role of indetermination	186
B.2.4	Expected difference between coupling	188
B.2.5	Structural Justification based upon an axiomatic result of Imre Csiszar	189
B.3	Monge properties : a justification of the \oplus/\otimes notation	190
B.3.1	Monge property – Indetermination	190
B.3.2	Log-Monge property – Independence	192
B.4	Logical indetermination and Condorcet's voting equilibrium	193
B.5	Application to network clustering	196
B.5.1	Introduction	196
B.5.2	Summary of an application to various networks	198
B.5.3	Gilbert Experimental Tests	199
B.6	Conclusions	201

C	Logical indetermination coupling :a method to minimize drawing matches and its applications	203
C.1	Introduction	203
C.2	From discrete transport problem to coupling function	204
C.2.1	The Alan Wilson's entropy model : role of "independence"	207
C.2.2	The minimal trade model : role of <i>indetermination</i>	208
C.3	Discrete <i>indetermination</i> and associated properties	209
C.3.1	Monge property	209
C.3.2	A dependence to avoid matches	211
C.3.3	Constructive definition of an <i>indetermination</i> matrix	212
C.3.4	Drawing under <i>indetermination</i>	213
C.4	Applications of <i>indetermination</i> structure	215
C.4.1	Deviation from <i>indetermination</i> : Janson Vegelius coefficient	215
C.4.2	Clustering problem dedicated to graphs	217
C.4.3	Guessing or spy problem	219
C.4.4	Tasks partitioning	224
C.5	Continuous <i>indetermination</i>	227
C.5.1	Coupling function	227
C.5.2	Constructive method for adapted margins	229
C.6	Conclusion	230
D	Continuous indetermination and average likelihood minimization	233
D.1	Introduction	233
D.2	Continuous indetermination	235
D.2.1	Coupling function	235
D.2.2	Constructive method for adapted margins	237
D.2.3	A specific indetermination copula	238
D.2.4	A local indetermination copula	241
D.2.5	Maximal spread between the two copulas	243
D.2.6	Conclusions about an indetermination (or indeterminacy) copula	244
D.3	Indetermination and average likelihood	245
D.3.1	Indetermination prevents correlation extraction	246
D.3.2	Average likelihood as a continuous couple matching notion	247
D.3.3	Statistical test based on average likelihood	247
D.4	Conclusion	253
D.5	Proof of proposition 61	253
D.5.1	relation between (F, G) and (R, S)	253
D.5.2	Two bounds of λ	255
D.5.3	λ positive	255
D.5.4	Transfer of hypothesis D.2	256

Introduction

Nous partons des travaux du mathématicien de renom Gaspard Monge et notamment de son fameux article [73] introduisant le problème dit "du transport de masse". Dans ce domaine, des avancées spectaculaires ont été faites après la seconde guerre mondiale par le prix Nobel russe d'économie Leonid Kantorovich au travers de l'article [55] définissant le célèbre problème MKP pour "Monge-Kantorovich Problem".

Cédric Villani, médaille Fields 2010, a formulé une théorie autour de ces travaux, théorie qu'il a enrichie et définie par la publication de deux livres majeurs : [95] et [94] traitant autant de sujets théoriques que de leurs applications possibles. Ce travail a permis d'unifier des sujets comme "the geometry of dissipative evolution equations, the reshaping of measurable metric spaces with reduced Ricci's curvature" et d'élargir la théorie classique des espaces métriques avec une "reduced sectional curvature"...

Au vu de ces premières lignes, il semble présomptueux de retourner travailler sur une théorie parcourue par les plus brillants mathématiciens des siècles passés. C'est que nous traiterons d'un sujet à la lisière discrète de la théorie mentionnée et relancé par l'engouement actuel autour des réseaux sociaux : "la modularisation de graphes et la reconnaissance de loi de probabilité". Nous disposerons ainsi des techniques utilisées par ces illustres prédécesseurs en reliant nos travaux à ceux de Gaspard Monge, Maurice Fréchet et Antoine Caritat de Condorcet.

En travaillant ainsi autour du fameux MKP sur lequel des statisticiens comme des économistes se sont penchés, nous aboutirons à l'élaboration de nouvelles mesures de corrélation entre deux partitions données et à des méthodes d'évaluation de celles existantes.

Ce parcours sera l'occasion de présenter et de détailler un couplage peu connu dit d'indétermination. Après l'avoir extrait de plusieurs problèmes anciens, nous établirons diverses propriétés le concernant et proposerons des interprétations permettant de mieux le cerner.

Précisément, bien que non historique, nous pouvons estimer que le point de départ de notre travail est un résultat de Csiszár dans [29] qui limite, sous des hypothèses raisonnables les divergences utilisables dans un problème d'optimisation à deux possibles : la divergence de Kullback-Leibler et la divergence \mathbb{L}_2 . Cette limitation éclaire une observation pratique, l'engouement historique pour les deux problèmes d'optimisation correspondants ; nous les rappellerons dans le chapitre 2 : le problème d'Alan Wilson basé sur l'entropie et le Minimal Trade Model (MTM) basé sur \mathbb{L}_2 .

Les deux problèmes, sous un regard différent expriment la même chose. L'objectif est de trouver la loi de probabilité sur pq éléments qui se rapproche le plus de l'uniforme tout en vivant dans un espace $\mathcal{L}(\mu, \nu)$ contraint défini par deux marges : μ loi donnée sur p éléments et ν donnée sur q éléments. C'est seulement la méthode pour mesurer l'écart à l'uniforme qui varie selon les deux divergences déjà mentionnées. Au final, on projette via deux écarts canoniques sur un sous-espace des lois de probabilités sur pq éléments défini par deux marges imposées.

Une forme close, dépendant évidemment des marges, existe pour chacun des deux problèmes d'optimisation. Ce lien entre les marges et la loi de probabilité π^* résultante, appelé couplage, conduit à définir deux couplages selon la mesure utilisée, ou bien l'indépendance $\pi^\times = C^\times(\mu, \nu)$ ou bien l'indétermination $\pi^+ = C^+(\mu, \nu)$ selon le schéma synthétique suivant :

$$\mathcal{U} \xrightarrow{KL} \mathcal{L}(\mu, \nu) = \pi^\times \quad \Bigg| \quad \mathcal{U} \xrightarrow{\mathbb{L}^2} \mathcal{L}(\mu, \nu) = \pi^+$$

Le premier couplage rejoint l'indépendance couramment utilisée et, si son intérêt est évident, nous ne saurions découvrir des propriétés sinon les rappeler. Le second, dont le nom remonte aux travaux de Jean-François Marcotorchino [64] apparaît dans des documents s'intéressant au MTM mais sans avoir jusqu'alors suscité un intérêt suffisant pour une étude spécifique ; ce sera l'objet essentiel du présent document.

Contenu du document

Ainsi, le chapitre 3 regroupe des propriétés de π^+ expliquant notamment et le choix de la notation \oplus et le nom *indétermination*. Nous y rappelons le lien déjà mentionné dans [68] entre chacun des couplages et les matrices dites de Monge, proposant la caractérisation des matrices de contingence représentatives de chacun. Nous y ajoutons une décomposition des matrices de contingence d'indétermination en une loi arbitraire et une loi uniforme qui s'accompagne d'une méthode pour tirer suivant ce couplage. Expliquée par cette décomposition mais exprimée dès la divergence utilisée, nous montrons une propriété de réduction optimale des collisions sous l'indétermination. Ainsi, à marges fixées (conformément au problème MTM), si l'on cherche à réduire le nombre d'égalités entre deux tirages successifs, la meilleure méthode consiste à suivre l'indétermination.

Le chapitre 4 se base sur cette capacité à réduire les collisions pour faire apparaître l'indétermination dans deux applications : le cryptage des données et le partitionnement de tâches. Dans le premier cas, un émetteur envoie un message crypté corrélé avec le clair et observé par un espion qui tente d'en déduire le message clair ; une collision est alors la réussite de l'espion au milieu du canal de communication : il a trouvé le message clair. Dans le second cas, une affectation de p tâches à q bureaux est réalisée suivant une loi de probabilité π couplant les marges qui représentent pour l'occasion la distribution des tâches et les prérogatives des bureaux. Les bureaux réalisant dans un ordre arbitraire les tâches affectées, une collision est la réalisation de la bonne tâche. Certes on ne veut alors pas en minimiser les occurrences mais choisir $\pi = \pi^+$ fournit ici une borne inférieure.

Le couplage d'indétermination est intimement lié à l'analyse relationnelle mathématique (ARM), que nous introduirons en détail dans le document et qui permet notamment

d'encoder différemment n tirages de deux variables aléatoires à p et q modalités remplaçant les matrices de contingence. Nous passons ainsi d'une notation $n_{u,v}$ de taille pq comptant les individus ayant la modalité u pour la première variable et v pour la seconde à deux matrices X et Y (une par variable) de taille n^2 remplies de 0 et de 1 suivant la relation :

$$X_{i,j} = \delta_i \text{ a la même modalité que } j$$

La transposition de coefficients de corrélation classiques comme le χ_2 ou le critère de *Rand* d'un espace à l'autre (contingence / relationnel) permet alors de rendre triviales des propriétés de ces derniers qui peuvent être dures à démontrer dans l'espace des contingences. Nous aurons ainsi l'occasion dans le chapitre 5 d'interpréter *Rand* comme un écart à l'indétermination dans un cadre particulier et d'observer qu'un critère moins connu dit de Janson-Vegelius et introduit dès l'article [52] le généralise. Nous analyserons en détail ce dernier pour estimer son comportement limite typiquement lorsque le nombre de modalités augmente.

Avant celle-ci, l'ARM a déjà engendré plusieurs thèses et notamment [74], [1] et [25]. Cette dernière est dédiée au clustering de graphes et reprend l'heuristique de Girvan-Newman qu'elle généralise à des modularités basées sur des critères locaux variés. De fait, la modularité originelle compare le voisinage d'un sommet avec ce qu'il aurait dû être sous une hypothèse d'indépendance entre la distribution des arêtes et celle des classes. L'écart entre la réalité et l'hypothèse admet autant de mesures que de critères de corrélation menant à des expérimentations multiples, voir par exemple l'article [21] sur le sujet. Nous nous limitons à deux critères locaux : l'écart à nos deux couplages canoniques menant à deux "modularités" M^\times et M^+ . Nous remarquerons d'ailleurs que la première n'est autre que la modularité classique. Après rappel des résultats expérimentaux de [25], nous nous placerons dans le cadre particulier des graphes de Gilbert pour extraire une comparaison théorique entre les deux critères. Cette comparaison montre un comportement globalement similaire. Pour autant, nous terminons l'application aux graphes en proposant quelques exemples sur lesquels les critères varient totalement.

Outre le clustering de graphe, l'ARM trouve une application naturelle dans l'analyse de corrélation entre deux variables aléatoires ; en témoigne entre autres l'article [8]. Par ailleurs et comme déjà mentionné, tous les critères disposent d'une écriture dans les deux espaces : contingence et relationnel. Le travail de [74] après une transposition de chaque critère classique dans l'espace relationnel, consiste à remarquer, critère par critère qu'il se réécrit comme un écart à l'un de nos deux couplages ; [29] apporte un éclairage supplémentaire au résultat. Le rappel dans le chapitre 5 des critères principaux sous leur forme contingentielle et relationnelle nous conduit à les condenser sous une forme générale regroupant l'ensemble. De cette forme générale nous extrayons un membre commun : le produit scalaire entre les codages relationnels des tirages dont nous mesurons la corrélation : $(X|Y)$.

Ce coefficient commun dispose évidemment d'une écriture dans l'espace des contingences mais c'est bien son écriture relationnelle qui nous aide à l'étudier en détail. Ainsi, nous montrons qu'il se décompose en un produit scalaire de matrices de blocs diagonaux et une permutation, permutation qui regroupe en réalité toute l'information de la corrélation. Nous focalisant sur le lien entre corrélation et permutation, nous montrons que plusieurs permutations engendrent la même corrélation et ajoutons qu'une très large majorité des permutations engendre des corrélations faibles. Nous relierons finalement le

nombre de permutations engendrant une corrélation données avec la valeur de $(X|Y)$.

En revenant à l'écriture contingentielle de $(X|Y)$, nous constatons qu'il s'agit de la fonction de coût d'un problème d'optimisation entière. Alors, nous appuyant sur la relation inverse "nombre de permutations - valeur de $(X|Y)$ ", nous proposons un algorithme de résolution aléatoire dudit problème.

Le document se termine par le chapitre 6, l'extension de la notion d'indétermination dans le domaine continu et la recherche d'une copule d'indétermination qui fait écho à l'article [51]. Celui-ci propose une méthode de détection d'anomalies à partir de l'observation d'un vecteur à plusieurs variables (typiquement une centaine). On suppose qu'une anomalie a lieu lorsque la réalisation du vecteur est rare, reste que sur cent variables, chaque réalisation est unique sauf à observer énormément d'événements, comme effet de la malédiction de la dimension. La solution proposée consiste à estimer chacune des distributions des marges puis à les coupler selon l'indétermination (étendue pour l'occasion à plus de deux variables), permettant de classer la rareté des événements observés. Dans l'article, il est fait mention d'une copule d'indétermination, or, si coupler selon l'indétermination est simple, nous montrons qu'il est impossible de définir une copule d'indétermination comme il existe une copule d'indépendance : la dépendance aux marges est trop importante. Nous montrons également que, localement, des marges différentes partagent la même copule d'indétermination.

Après écriture et définition d'un couplage d'indétermination continu, nous cherchons à retrouver pour celui-ci la propriété clef remarquée dans ce document de réduction des collisions. Or, si le cas discret nous permet d'utiliser l'égalité pour définir une collision, nous ne saurions attendre de deux tirages successifs dans un intervalle qu'ils tombent sur le même. Afin de transposer la notion de collision nous définissons une vraisemblance moyenne entre deux lois continues et montrons qu'elle recouvre bien le cas discret.

De même que l'indétermination et les collisions entretenaient un rapport particulier dans le cas discret, nous montrons que l'indétermination continue minimise la vraisemblance moyenne à marges fixées. Nous profitons de cette propriété pour définir une statistique de test distinguant l'indétermination par rapport à un voisinage de l'indépendance. Le document se conclut sur l'étude du test introduit en estimant la pente de Bahadur associée ; il s'agit d'une notion pour évaluer la probabilité de ne pas rejeter l'hypothèse nulle quand on aurait dû.

Regroupement des articles soumis

Le document contient également quatre annexes qui correspondent à trois articles de recherche soumis dans des journaux et un article accepté dans une conférence. Ils reprennent une majorité des résultats de la thèse sous un format plus concis. Nous détaillons ci-dessous les correspondances.

1. Annexe A : Article accepté à *Geometric Science of Information*, 2021. Il s'agit d'un résumé reprenant la propriété de réduction des collisions remarquée dans le chapitre 3 et les applications au problème de l'espion et du partitionnement de tâches proposées dans le chapitre 4.
2. Annexe B : Article soumis au journal *Advances in Data Analysis and Classification* en Juillet 2020 et révisé en Mars 2021. Cet article reprend la construction du couplage d'indétermination et d'indépendance présente dans le chapitre 2 et justifie

que l'on puisse se restreindre à ces deux-ci grâce au lien que nous avons établi avec le travail de Csiszár ([29]). En transposant chaque couplage dans le cadre du clustering de graphes, nous proposons de nous restreindre également aux deux critères qu'ils génèrent retrouvant ainsi notamment la célèbre modularité. Nous y menons une étude de leur proximité générale et des différences particulières conformément aux travaux conduits dans le chapitre 4.

3. Annexe C : Article soumis au journal *Information and Inference* en Décembre 2020. Cet article reprend la propriété de réduction des collisions, l'association historique matrices de Monge / indétermination ainsi que la décomposition d'un tirage selon l'indétermination présentées dans le chapitre 3. Deux coefficients de corrélation canoniques (le χ_2 et le JV) illustrant la dichotomie indétermination / indépendance y sont introduits ; ils correspondent à une mince partie du chapitre 5. Les deux applications au problème de l'espion et du partitionnement de tâches du chapitre 4 y sont décrites bien plus en détail que dans l'annexe A. L'article contient également une caractérisation élégante des marges éligibles à un couplage d'indétermination ainsi que la mention du couplage continu d'indétermination présenté dans le chapitre 6.
4. Annexe D : Article soumis au journal *Journal of Multivariate Analysis* en Mai 2021. Cet article regroupe l'ensemble du chapitre 6 à savoir la transposition historique dans le cas continu de la notion de couplage d'indétermination, la nouvelle généralisation de la notion de collision au travers de la vraisemblance moyenne, la définition d'un test statistique basé sur cette vraisemblance moyenne et l'étude du test associé. La seule différence majeure est l'omission d'un théorème de Sanov spécifique.

Construction de deux couplages canoniques

Notre-Dame-des-Fleurs fait ici
son entrée solennelle par la
porte du crime, porte dérobée,
qui donne sur un escalier noir
mais somptueux.

Jean Genet,
Notre-Dame-des-Fleurs, 1939

Ce chapitre est l'occasion de proposer une revue historique des problèmes de transport optimal. Certes nous n'aurons pour objectif une quelconque exhaustivité mais plutôt nous pensons donner un aperçu afin d'inscrire notre travail dans la lignée ou plutôt comme la correction d'une étape oubliée. De fait, en sus d'introduire la double formulation classique Monge et Monge-Kantorovitch, nous repasserons rapidement dans un domaine discret, domaine dans lequel nous proposerons d'étudier deux problèmes précis. Or, c'est le désengagement de la recherche pour une des deux solutions qui justifie la naissance du travail présenté dans ces pages, comme la soudaine réapparition d'une formule délaissée.

Il s'agit du point d'entrée des deux papiers que nous avons soumis [10] et [11], il s'agit également du point d'entrée dans ce manuscrit. A l'origine, un problème d'optimisation avec des contraintes sur les marges d'une fonction de coût classique s'appliquant à une probabilité : l'entropie, livre une formule close qui n'est autre que la tout aussi classique indépendance.

Une réécriture du problème montre que les marges définissent un espace sur lequel projeter une fonction de probabilité initiale, un "prior guess" sous la divergence de Kullback-Leibler et nous amène à la lecture de l'article [29] traitant précisément ce problème. Ajoutant des contraintes structurelles sur la loi de probabilité projetée, Csizar montre que le choix des mesures de qualité de la projection se cantonne à deux divergences : celle de Kullback-Leibler et \mathbb{L}^2 .

Revenant des divergences aux fonctions de coût, nous nous empressons de conduire un travail symétrique en résolvant le problème associé à cette seconde divergence. On découvre alors une loi de probabilité couplant les marges imposées selon ce que nous appellerons tout au long du document *indétermination*. Si l'étude de ce couplage, de ses propriétés comme de ses applications sera repoussé aux chapitres suivants, nous rappelons

en détail le résultat principal de l'article [29] et mesurons déjà l'écart entre les deux couplages comme une évaluation de l'impact du changement de divergence.

2.1 Problèmes historiques

2.1.1 Problème de Monge

Comme mentionné en introduction, Gaspard Monge fut le premier, en 1781 à s'intéresser au problème qu'il appelait alors "remblais et déblais", nous reportons ici la première page de son mémoire, comme un illustre et lointain point de départ.

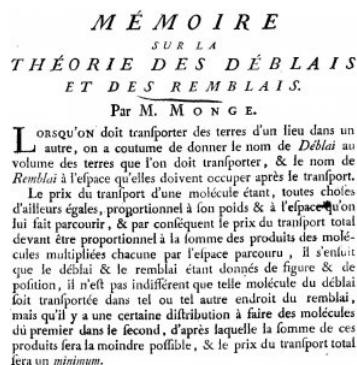


FIGURE 2.1 – Première page du mémoire de Monge

Il sera intéressant d'observer que nombre de propriétés étaient déjà pressenties dans ledit mémoire. Notamment, le respect des conditions de Monge par les matrices que nous manipulerons tout au long de ce document. Pour l'instant, remanions la formulation énoncée par ce que nous appellerions aujourd'hui un *abstract*. La question initiale est assez simple : comment déplacer avec le moins d'effort une pile de sable d'un lieu à un autre (voir la Figure 2.2) ?

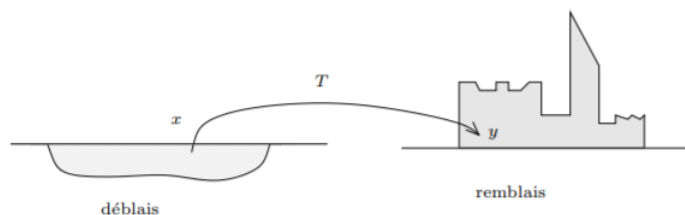


FIGURE 2.2 – Première description du problème de Monge

Introduisons-le sous sa formalisation actuelle : on imagine que le sable occupe une région A d'un espace \mathcal{X} (typiquement \mathbb{R}^n , \mathbb{R}^3 dans la Figure 2.2) et que sa destination est représentée par une région B (le bâtiment sur la figure) d'un espace \mathcal{Y} . On suppose bien sûr que les volumes sont identiques.

Suivant la description de Monge revue, nous cherchons une fonction de transfert $\mathcal{T} : A \rightarrow B$, (\mathcal{T} pour "Transport"), où $\mathcal{T}(x) \in B$ représente la destination dans B de l'élément de A initialement localisé en $x \in A$. Le travail à effectuer et à minimiser s'écrit donc :

Problème 1 (Problème de Monge original).

$$\min_{\mathcal{T}} F(\mathcal{T}) = \int_{\mathcal{X}} (x - \mathcal{T}(x)) dx$$

Il faut ainsi minimiser selon les fonctions de transfert \mathcal{T} conservant les masses durant le passage de A à B : nous sommes en train de minimiser le travail à fournir. La contrainte de conservation, elle, ancre le problème dans le réel et le complique autant qu'elle en justifie l'intérêt.

Généralisons : soient \mathcal{X} et \mathcal{Y} deux espaces séparables et convexes de \mathbb{R}^n représentant l'origine et la destination dans le même espace. On note $\mathbb{P}(\mathcal{X})$ l'espace des mesures de Borel sur \mathcal{X} normées à 1. Le tas de sable est représenté par une mesure $\mu \in \mathbb{P}(\mathcal{X})$ et le trou à remplir par une mesure $\nu \in \mathbb{P}(\mathcal{Y})$. On a supposé que les trous avaient le même volume ce qui s'écrit :

$$0 < \mu(\mathcal{X}) = \nu(\mathcal{Y}) \leq 1$$

Ainsi, μ et ν sont deux mesures de Borel avec la même masse totale que l'on s'empresse de fixer à 1. Rajoutons alors une fonction de coût sur l'espace produit : $\mathbf{C} : \mathcal{X} \times \mathcal{Y} \rightarrow \mathbb{R}^+$. Quelques exemples de fonctions de coût sont données ici :

- Coût linéaire : $\mathbf{C}(x, y) = |x - y|$; (modèle classique)
- Coût quadratique : $\mathbf{C}(x, y) = (x - y)^2$; (modèle classique Euclidien)
- Coût polynomial : $\mathbf{C}(x, y) = (x - y)^p$ ($p > 2$); (moins fréquent)
- $\mathbf{C}(x, y) = h(|x - y|)$ est aussi possible pour peu que h soit convexe.

Ces derniers exemples supposent x et y dans le même espace pour que l'opération *moins* ait un sens. La fonction de transfert est alors une opération assignant une correspondance "un-pour-un" $\mathcal{T} : \mathcal{X} \rightarrow \mathcal{Y}$ (dans notre cas, $\mathcal{X} = \mathcal{Y}$).

Comme \mathcal{T} doit satisfaire la contrainte de conservation de la masse, on constate l'égalité entre la mesure image de μ par \mathcal{T} ($\mu_{\mathcal{T}}$, définie juste après) et ν qui représente la destination. Utilisant la mesure image et la notion même de mesure, on s'affranchit des notations de l'époque mais on simplifie l'écriture du problème.

Définition 1 (Mesure image).

Soit \mathcal{T} une fonction de transfert Borélienne, la mesure image de μ par \mathcal{T} est la mesure de Borel notée $\mu_{\mathcal{T}}$ et définie par son action sur chaque Borelien B de \mathcal{Y} :

$$\mu(\mathcal{T}^{-1}(B))$$

Une fonction de transfert Borélienne $\mathcal{T} : \mathcal{X} \rightarrow \mathcal{Y}$ est alors dite de transport (entre μ et ν) si $\nu = \mu_{\mathcal{T}}$. Le problème de Monge généralisé consiste enfin à trouver parmi ces fonctions celle qui minimise le coût. En introduisant la fonction de coût \mathbf{C} , on l'écrit :

Problème 2 (Problème de Monge généralisé).

$$\inf_{\mathcal{T}} F(\mathcal{T}) = \int_{\mathcal{X}} \mathbf{C}(\mathbf{x}, \mathcal{T}(\mathbf{x})) d\mu(x)$$

sous la contrainte :

$$\mu_{\mathcal{T}} = \nu$$

Une solution à ce problème (s'il y en a une) s'appelle une fonction de transfert optimale ou une solution de Monge. Remarquons d'abord qu'une telle solution peut ne pas exister : c'est le cas par exemple si μ est une masse de Dirac δ_a alors que ν n'en est pas une. De plus, même si des fonctions de transfert existent, l'optimal peut ne pas exister. L'exemple de la masse de Dirac, peut sembler dégénéré. Malheureusement, même pour des mesures avec des densités continues, le problème reste complexe. Prenons un exemple pour nous en convaincre.

Supposons μ et ν absolument continues par rapport à la mesure de Lebesgue avec pour densités respectives f et g : $d\mu(x) = f(x) dx$ et $d\nu(y) = g(y) dy$ où f et g sont des fonctions intégrables sur \mathcal{X} et \mathcal{Y} respectivement. Dans ce cas, $\mathcal{T} : \mathbb{R}^n \rightarrow \mathbb{R}^n$ satisfait la contrainte du Problème 1 si et seulement si \mathcal{T} est injectif de Classe C^1 et \mathcal{T} définit un changement de variable admissible :

$$f(x) = g(\mathcal{T}(x)) |\det(J(\mathcal{T}(x)))| \quad (2.1)$$

Ainsi, l'hypothèse forte que \mathcal{T} est un C^1 difféomorphisme mène à l'équation 2.1 bien loin d'être linéaire. De plus, l'hypothèse d'un difféomorphisme ne tient pas en général.

Enfin, on peut remarquer que le problème de Monge est très rigide dans sa formulation : toute la masse de x doit être associée à la cible $\mathcal{T}(x)$ sans permettre une séparation.

En réalité, le problème de Monge (Problème 2) est beaucoup plus compliqué que le problème de Monge Kantorovich que nous introduisons dans la section suivante et n'admet souvent aucune solution sinon en imposant des contraintes notamment sur la fonction de coût.

2.1.2 Problème de Monge-Kantorovich

Pour faire face aux difficultés du problème de Monge, une technique, fréquente dans la résolution de problèmes d'optimisation consiste à étendre ou relaxer le domaine de recherche lui-même. C'est par exemple le cas pour la méthode dite de "Séparateurs à vaste marge". En 1942, Kantorovitch propose de relaxer le problème de Monge en admettant les solutions séparant la masse (voir [55]). Il propose en fait deux choses :

1. une version relaxée du problème de Monge
et
2. un principe de dualité *que nous ne traiterons pas ici*.

Plus précisément, il introduit la formulation suivante du problème qui devait prendre le nom de problème de Monge-Kantorovich (le fameux MKP) :

Problème 3 (Problème de Monge-Kantorovitch (MKP)).

$$P(\pi^*) = \min_{\pi \in \Pi(\mu, \nu)} \int_{\mathcal{X} \times \mathcal{Y}} \mathbf{C}(x, y) d\pi(x, y)$$

Ici, $\Pi(\mu, \nu)$ est l'ensemble des "plans de transfert" (une généralisation des fonctions de transfert dans le cas bi-dimensionnel) i.e l'ensemble des mesures de probabilité Boréliennes sur l'espace produit $\mathcal{X} \times \mathcal{Y}$ respectant les marges μ et ν . Cela implique :

$$\pi(A, \mathcal{Y}) = \mu(A); \pi(\mathcal{X}, B) = \nu(B), \text{ pour tout Borélien } A \in \mathcal{X} \text{ et } B \in \mathcal{Y} \quad (2.2)$$

On remarque que $\Pi(\mu, \nu)$ est un espace convexe compris dans l'ensemble des mesures Boréliennes de masse finie sur l'ensemble produit $\mathcal{X} \times \mathcal{Y}$. De plus, la nouvelle fonction à optimiser apparaît linéaire en π de sorte que, sous des hypothèses raisonnables sur la fonction de coût \mathbf{C} , des arguments de compacité apportent l'existence d'au moins une solution optimale notée $\pi^* \in \Pi(\mu, \nu)$.

Dans la plupart des applications de la théorie du transport optimal, la fonction de coût $\mathbf{C} : \mathcal{X} \times \mathcal{Y} \rightarrow \mathbb{R}^+$ est semi-continue inférieurement (Théorème de Kellerer [57] en 1984, ou son élargissement de 2009 par Mathias Beiglböck, Christian Leonard et Walter Schachermayer [9]). Pour autant, le problème 2 peut être défini pour toute fonction de coût Borélienne à valeurs dans \mathbb{R}_+^* . Deux contraintes doivent alors être satisfaites par \mathcal{T} :

1. $\mu_{\mathcal{T}} = \nu$,
2. $\int_{\mathcal{X}} h(\mathcal{T}(x)) d\mu(x) = \int_{\mathcal{Y}} h(y) d\nu(y)$, pour toute fonction h continue

Notons \mathcal{M}^a l'ensemble des plans admissibles. Cette fois, une solution π^* du problème 3 ne doit pas forcément être engendrée par un mapping un-vers-un $\mathcal{T} \in \mathcal{M}^a$. Ainsi, la nouvelle construction ne garantit que l'existence d'une solution dite faible ou relâchée au problème de Monge originel 2.

Comme suggéré dans le papier de Lawrence Evans [33] sur le problème 3, un bon moyen de comprendre l'approche duale introduite par Kantorovich's consiste à revenir à une version discrète. C'est cette version, formulée sur un espace fini, qui nous intéressera par la suite.

Nous conserverons du problème 3 et l'approche de probabilité croisée π qui nous mènera à la définition de couplage et la restriction à un espace défini par deux marges à respecter.

Par analogie avec l'équation 2.2, on suppose disposer d'un ensemble $\mathcal{U} = \{1, 2, 3 \dots p\}$ d'origines (par exemple p usines) et d'un ensemble $\mathcal{V} = \{1, 2, \dots q\}$ de destinations (par exemple q magasins). On suppose également pouvoir associer à chaque transfert de u à v un "coût de transport" positif ou nul $\mathbf{C}(u, v, \pi)$ dépendant éventuellement du transfert π lui-même. On suppose enfin connaître la distribution initiale $\mu = \{\mu_1, \mu_2, \dots \mu_p\}$ dans les origines (les productions de chaque site) et la distribution finale $\nu = \{\nu_1, \nu_2, \dots \nu_q\}$ attendue dans les destinations (les ventes par magasin). Ces quantités jouent le rôle des marges données pour une loi de probabilité bi-dimensionnelle $\pi_{u,v}$ (jouant le rôle de $\pi(x, y)$ dans l'équation 2.2 pour le cas continu). Ces marges doivent satisfaire l'égalité des masses selon l'équation :

$$\sum_{u=1}^p \mu_u = \sum_{v=1}^q \nu_v = 1$$

On cherche alors la solution $\pi_{u,v}^*$; $u = 1, 2, 3 \dots p$; $v = 1, 2, 3 \dots q$ du problème d'optimisation discret suivant :

Problème 4 (Version Discrète du MKP).

$$\min_{\pi} \sum_{u=1}^p \sum_{v=1}^q C(u, v, \pi) \pi_{u,v}$$

sous les contraintes :

$$\pi_{u,\cdot} = \sum_{v=1}^q \pi_{u,v} = \mu_u ; \forall u \in \{1, \dots, p\}$$

$$\pi_{\cdot,v} = \sum_{u=1}^p \pi_{u,v} = \nu_v ; \forall v \in \{1, \dots, q\}$$

$$\pi_{\cdot,\cdot} = \sum_{u=1}^p \sum_{v=1}^q \pi_{u,v} = 1$$

$$\pi_{u,v} \geq 0 ; \forall (u, v) \in \{1, \dots, p\} \times \{1, \dots, q\}$$

C'est très clairement l'analogue discret du problème 3 et il est intimement lié à la notion de table de contingence comme celle apparaissant dans la figure 2.1. De fait, si la notion de table de contingence est statistique et constitue un mode de stockage de tirages croisés entre deux variables, sa forme même permet également de représenter efficacement la loi de probabilité π que l'on souhaite affiner. La partie suivante sur les applications renforcera le double rôle de cette table.

2.1.3 Applications

Dans les paragraphes précédents, nous avons présenté quelques extraits de la littérature autour des problèmes 2 et 3. La raison de l'introduction de cette thématique, fascinante et difficile à la fois, était de pouvoir ensuite dériver des résultats dans un cadre discret peut-être moins connu mais tout autant utile autour du *Minimum Trade Model* et de ses variantes *Spatial Interaction Models* et *International Trade Exchanges Models*.

Une première approche intuitive est proposée par Bernard Fustier dans [40] où des formulations discrètes sont énoncées. Nous adoptons une approche un peu différente mais consistante avec les méthodes habituelles.

Suivant les domaines d'application proposés dans [40] justement, on part d'une matrice représentant les échanges monétaires relatifs à p zones (pays, zones géographiques) dans \mathcal{U} et à q autres zones dans \mathcal{V} . On peut facilement poser $p = q = M$ si l'on s'intéresse aux échanges entre les M différents pays (164 à ce jour) enregistrés à l'OMC (Organisation Mondiale du Commerce). Les échanges sont alors représentés par un tableau rectangulaire (ou carré) représentant les origines-destinations où le terme général noté $n_{u,v}$ apparaissant à l'intersection de la ligne u et de la colonne v indique le montant commercial échangé de la région u vers la région v souvent exprimé en \$. Formellement, on travaille avec la table croisant \mathcal{U} avec \mathcal{V} : ($|\mathcal{U}| = p$, $|\mathcal{V}| = q$) selon la figure 2.1.

En analyse statistique, une telle table est appelée matrice de contingence quand les quantités sont des entiers selon les notations :

- $n_{u,v}$: masse échangée de u vers v
- $n_{u,\cdot}$: masse totale exportée depuis u
- $n_{\cdot,v}$: masse totale importée par v

	\mathcal{V}	1	...	v	...	q	
\mathcal{U}	1	$n_{1,1}$					$n_{1,\cdot}$
	...						
	u			$n_{u,v}$			$n_{u,\cdot}$
	p					$n_{p,q}$	$n_{p,\cdot}$
		$n_{\cdot,1}$		$n_{\cdot,v}$		$n_{\cdot,q}$	$n_{\cdot,\cdot}$

TABLE 2.1 – Tableau de contingence

— $n_{\cdot,\cdot} = N$: masse totale échangée dans le système

La masse totale exportée doit être égale à la masse totale reçue; on retrouve la contrainte de préservation de la masse chère aux problèmes 2 et 3.

$$\sum_{u=1}^p n_{u,\cdot} = \sum_{v=1}^q n_{\cdot,v} = n_{\cdot,\cdot} = N$$

Si l'on divise chaque valeur de la table par la somme totale $n_{\cdot,\cdot}$, on retombe sur une loi de probabilité :

$$\begin{aligned} n_{u,v} &\rightarrow \frac{n_{u,v}}{n_{\cdot,\cdot}} = \pi_{u,v} \\ n_{u,\cdot} &\rightarrow \frac{n_{u,\cdot}}{n_{\cdot,\cdot}} = \mu_u \\ n_{\cdot,v} &\rightarrow \frac{n_{\cdot,v}}{n_{\cdot,\cdot}} = \nu_v \end{aligned}$$

Le problème initial d'échange (monétaire) se transforme ainsi en un problème consistant à exprimer une loi de probabilité P_π sur le produit $\mathcal{U} \times \mathcal{V}$ ayant les marges adéquates μ et ν ; sa densité étant $\pi_{u,v}$: précisément le problème 4.

Les affinités commerciales entre pays, zones ou entités qui transparaissent dans les valeurs de la table 2.1 proviennent de nombreux facteurs tels que la proximité géographique ou l'héritage historique. Aussi, constater, modéliser puis comprendre ces affinités, mesurer leurs déviations par rapport à la moyenne est le but des modèles d'échange globaux.

D'autres applications suivent une modélisation similaire comme les modèles d'interactions spatiales. Dans ce cas, à la place de masses monétaires, les valeurs représentent le nombre de personnes voyageant (ou s'expatriant) d'un secteur à l'autre. Ce peut aussi être les trajets communs domicile-travail. Seule la nature des quantités et les interprétations associées changent, la modélisation, elle, reste identique.

On citera ici comme finalité possible : comprendre et prévoir les mouvements. Les modèles sous-jacents doivent toujours vérifier les contraintes de préservation de masse, de marges fixes, tandis que le domaine varie et parcourt l'économie, la politique, le géographie, la sociologie, les transports, l'urbanisme ou encore la démographie.

Parmi une grande liste de papiers présentant le contexte, citons : une conférence au Collège de France de G. Carlier- I. Ekeland [22] et A. Rudneff-Chupin [84], deux documents traitant des concepts autour du transport optimal. D'autres auteurs ont publié

dans ce domaine des articles purement applicatifs comme B. Fustier [40] sur les mouvements économiques, mais aussi des articles plus théoriques sur les interactions spatiales comme ceux de A. Wilson [100] et F. Willekens [97].

Comme nous le verrons par la suite, l'une des difficultés inextricables est de choisir la bonne fonction de coût apparaissant dans le problème 4 et adaptée au domaine que l'on traite.

C'est précisément sur le choix de cette fonction que notre travail apporte un éclairage nouveau. Si nous commençons par étudier le problème pour deux différentes versions de ce paramètre (l'entropie dans la sous-section 2.2.1 et la déviation quadratique dans la sous-section 2.2.2), nous nous efforçons de motiver ces deux choix comme canoniques dans la section 2.4. Fort de cette limitation à deux fonctions de coût, nous nous écarterons rapidement des domaines d'application tangibles du problème de Monge-Kantorovitch pour nous placer dans un cadre purement statistique autour de la dépendance entre deux variables avec pour motivation d'étudier les propriétés de chaque couplage canonique ; ce sera l'objet des chapitres suivants.

2.2 Deux modélisations : l'indépendance et l'indétermination

Les problèmes que l'on vient d'introduire apparaissent déjà, sous leur forme discrète dans des applications plus anciennes autour de *l'ajustement d'une matrice de contingence à marges fixées*. Selon le problème optimisé, des critères de dépendance entre variables peuvent en être extraits comme le critère du Chi2. Autour de ce problème historique, les séminaires d'Edwards Deming et Frederick Stephan [30], les travaux de Maurice Fréchet [38], [39], et les papiers : Yves Tillé [91], Sue McNeil et Chris Hendrickson [70], Sir Alan G. Wilson [98], [99] et [100], Frans Willekens [97], Seven Erlander et Neil F. Stewart [32] nous aident à définir les outils à employer. Ainsi, les travaux précités montrent que nous pouvons nous réduire à étudier quelques fonctions de coûts seulement et notamment à cause des structures intrinsèques qu'elles renferment. Reste que cette restriction est motivée par des observations ou expérimentales ou historiques. Nous appliquerons ultérieurement le travail de justification des divergences majeures issu de l'article [29] et dont le nom préfigure déjà le résultat : une limitation à l'entropie et à la déviation quadratique. Avant de plonger nos problèmes dans l'espace des divergences pour les justifier, nous continuons en définissant disons "au hasard", deux fonctions de coût.

2.2.1 Écart à l'indépendance : l'entropie d'Alan Wilson

D'abord proposée par Sir Alan Wilson en 1969 pour *modéliser les interactions spatiales* mais cohérente avec le *système des échanges international* ou *l'ajustement des matrices de contingences à marges fixées*, l'**entropie des flux** d'Alan Wilson est introduite dans plusieurs publications notamment [98], est développée dans [99] et enfin précisée dans son travail [100]. Une justification fondamentale de cette approche (comprendre de la fonction de coût utilisée) est le contexte suivant : dans un système de particules élémentaires, on cherche une distribution $\pi_{u,v}$ (la fréquence des flux normalisée) pour maximiser l'entropie dite de Boltzmann et de Shannon.

Introduisons formellement ce problème dit *Program of Spatial Interaction System*.

Problème 5 (Alan Wilson - PSIS).

$$\begin{aligned} \max_{\pi} \quad & \sum_{u=1}^p \sum_{v=1}^q \pi_{u,v} \ln(\pi_{u,v}) \\ \text{sous les contraintes :} \quad & \sum_{v=1}^q \pi_{u,v} = \mu_u, \forall u \\ & \sum_{u=1}^p \pi_{u,v} = \nu_v, \forall v \\ & \sum_{u=1}^p \sum_{v=1}^q \pi_{u,v} = 1 \\ & \pi_{u,v} \geq 0, \forall (u, v) \end{aligned}$$

Si l'on n'a aucune hypothèse sur le système, les contraintes s'effacent sinon une somme à 1 et des valeurs positives de sorte que le principe de Laplace de raison insuffisante donne comme solution une égalité de chacune des valeurs soit la loi uniforme \mathbb{U} définie ci-dessous.

Définition 2 (Loi uniforme). *Sur l'espace produit $\mathcal{U} \otimes \mathcal{V}$ on définit la loi uniforme \mathbb{U} par :*

$$\mathbb{U}_{u,v} = \frac{1}{pq}, \quad \forall (u, v)$$

La distribution solution est donc uniformément répartie. C'est le plus grand degré de désordre pour l'entropie de Boltzmann et Shannon et plus généralement, on souhaite que toute fonction de coût naturelle conduise à cette solution sans contrainte sur les marges : nous aurons l'occasion de renchérir sur cette remarque.

Lorsque l'on dispose d'informations a priori sur le système à savoir les flux entrants et sortants de chaque élément, le degré de désordre décroît drastiquement puisque l'on impose à la somme des colonnes et des lignes de respecter les marges μ_u et ν_v . Calculons maintenant la solution optimale décrite dans le théorème suivant.

Théorème 1.

La solution du Problème 5 est $\pi_{u,v}^ = \pi_{u,v}^\times = \mu_u \nu_v$ de sorte que $n_{u,v}^* = \frac{n_u \cdot n_v}{N}$.*

On en déduit que ce problème mesure un écart à l'indépendance, couplage habituel et reconnu dans π^\times .

Démonstration. Donnons ici une démonstration courte permettant d'arriver à la solution du problème. On introduit les multiplicateurs de Lagrange : λ_u pour la première contrainte, ω_v pour la seconde et θ pour la troisième. Le Lagrangien se formule comme

suit :

$$\begin{aligned}
\mathbf{L}(\pi, \lambda, \omega, \theta) &= - \sum_{u=1}^p \sum_{v=1}^q \pi_{u,v} \ln(\pi_{u,v}) \\
&+ \sum_{u=1}^p \lambda_u \left(\mu_u - \sum_{v=1}^q \pi_{u,v} \right) \\
&+ \sum_{v=1}^q \omega_v \left(\nu_v - \sum_{u=1}^p \pi_{u,v} \right) \\
&+ \theta \left(\sum_{u=1}^p \sum_{v=1}^q \pi_{u,v} - 1 \right)
\end{aligned}$$

L'entropie à optimiser est concave. On cherche donc ici un maximum et les conditions du premier ordre vont se révéler suffisante pour l'obtenir. Comme les coefficients de Lagrange sont définis à une constante près, on choisit $\theta = 1$ de sorte que les conditions du premier ordre s'écrivent :

$$\frac{\partial \mathbf{L}(\pi, \lambda, \omega, \theta)}{\partial \pi_{u,v}} = -\ln(\pi_{u,v}) - 1 + \lambda_u + \omega_v + 1 = 0 \quad (2.3)$$

$$\frac{\partial \mathbf{L}(\pi, \lambda, \omega, \theta)}{\partial \lambda_u} = \mu_u - \sum_{v=1}^q \pi_{u,v} = 0 \quad (2.4)$$

$$\frac{\partial \mathbf{L}(\pi, \lambda, \omega, \theta)}{\partial \omega_v} = \nu_v - \sum_{u=1}^p \pi_{u,v} = 0 \quad (2.5)$$

$$\text{de 2.3 on extrait} \quad \pi_{u,v} = e^{\lambda_u + \omega_v} \quad (2.6)$$

$$\text{de 2.4 et 2.6 on obtient} \quad \sum_{v=1}^q \pi_{u,v} = e^{\lambda_u} \sum_{v=1}^q e^{\omega_v} \quad (2.7)$$

$$\text{de manière symétrique, avec 2.5 et 2.6} \quad \sum_{u=1}^p \pi_{u,v} = e^{\omega_v} \sum_{u=1}^p e^{\lambda_u} \quad (2.8)$$

En sommant 2.6 selon (u, v) , il vient $1 = \sum_{u,v} e^{\lambda_u + \omega_v} = \sum_u e^{\lambda_u} \sum_v e^{\omega_v}$.
Maintenant, avec 2.7 et 2.8 on dérive :

$$1 = \frac{\mu_u}{e^{\lambda_u}} \frac{\nu_v}{e^{\omega_v}}$$

Enfin, (en utilisant 2.6) apparaît le résultat escompté. □

En conclusion : de la maximisation de l'entropie, on obtient une solution qui s'exprime sous la forme de l'indépendance que ce soit comme une loi de probabilité ou bien comme une ventilation d'une valeur globale N dans le cas d'une table de contingence. En supposant l'indépendance des lignes et des colonnes, la ventilation habituelle donne la même formule. On observe ici que le degré de désordre, comme anticipé, n'est pas total : l'intensité des flux est proportionnelle aux poids des partenaires de l'échange monétaire international si l'on reprend l'application la plus classique.

2.2.2 Écart à l'indétermination : déviation quadratique

Le *Minimal Trade Model* (voir [89], [64] et [20]) impose les mêmes contraintes que dans le problème 5 mais en modifiant la structure de la fonction à optimiser de façon à proposer une autre ventilation des flux origines-destinations $n_{u,v}$.

Optimisation en dimension deux

On pose comme toujours $\pi_{u,v} = n_{uv}/N$. Dans notre nouveau problème 6, le critère mesure la déviation quadratique des valeurs des cellules au cas de "non information" déjà mentionné à savoir la loi uniforme \mathbb{U} . On minimise donc l'information restante tout en respectant les marges. Par ailleurs, le développement de la ventilation dans l'équation 2.11 justifie le terme *Minimal Trade Model* : on cherche à minimiser le carré des échanges. Formellement, cela donne :

Problème 6 (Minimal Trade Model).

$$\begin{aligned} \min_{\pi} \quad & \sum_{u,v} \left\{ \pi_{u,v} - \frac{1}{pq} \right\}^2 \\ \text{sous les contraintes :} \quad & \sum_v \pi_{u,v} = \mu_u, \forall u \\ & \sum_u \pi_{u,v} = \nu_v, \forall v \\ & \sum_{u,v} \pi_{u,v} = 1 \\ & \pi_{u,v} \geq 0, \forall (u, v) \end{aligned}$$

Comme on le voit, seule la fonction de coût a changé et si la fonction de coût précédente était justifiée par l'application aux interactions entre particules, celle présentée ici est justifiée comme écart au cas de non-information. Le choix de l'écart quadratique étant quant à lui habituel en tant qu'il a le bon goût d'avoir une dérivée linéaire et d'être toujours positif.

Une condition supplémentaire doit être ajoutée aux marges données a priori pour que la solution calculée existe comme loi de probabilité. Cette condition 2.9 introduite dans [64] est une inégalité garantissant la positivité des termes $\pi_{u,v}^*$. Nous étudierons la forme et la fréquence des couples de marges (μ, ν) la respectant dans la section 3.1.

$$p \min_u \mu_u + q \min_v \nu_v \geq 1 \rightarrow p \min_u n_{u,\cdot} + q \min_v n_{\cdot,v} \geq N \quad (2.9)$$

Dans cette section, nous supposerons, que la condition 2.9 est systématiquement respectée. C'est une donnée du problème car, s'agissant par exemple des masses monétaires propres à chaque pays, nous n'avons aucune action sur les marges μ_u et ν_v . Par ailleurs, nous remarquons que, puisque la matrice $\pi_{u,v}$ représente des fréquences, l'avant-dernière contrainte du Problème 6 joue un rôle pour garantir $\pi \leq 1$.

Remarquons enfin que dans le cas du tableau de contingence proprement dit, il s'agit de tirages et la valeur associée $n_{u,v}$ doit donc être entière ce qui complexifie grandement

le problème. Normalement, une fois divisé par la somme des cases $n_{.,.}$, on obtient les $\pi_{u,v}$ mais ceux-ci sont très particuliers : ils sont rationnels de même dénominateur en plus de sommer à 1. Le relâchement de cette contrainte de rationalité conduit naturellement au Problème 6.

Intéressons-nous à la résolution dudit problème avec le théorème suivant.

Théorème 2.

La solution du Problème 6 est

$$\pi_{u,v}^* = \pi_{u,v}^+ = \frac{\mu_u}{q} + \frac{\nu_v}{p} - \frac{1}{pq}. \quad (2.10)$$

Ainsi, il mesure un écart à l'indétermination introduite formellement juste après.

Définition 3 (Indétermination).

Lorsque les marges respectent l'hypothèse 2.9, on peut coupler selon l'indétermination par la formule :

$$\pi^+(u, v) = \frac{\mu_u}{q} + \frac{\nu_v}{p} - \frac{1}{pq}$$

Nous étudierons en détail cette expression et montrerons qu'elle renferme de nombreuses propriétés et s'interprète comme un couplage garantissant l'égalité des voix pour et contre entre les deux tirages μ et ν .

Démonstration.

Résolvons enfin ce problème en démontrant le théorème 2.

On commence par développer la fonction de coût pour obtenir une égalité intéressante que nous utiliserons par la suite.

$$\sum_{u,v} \left(\pi_{u,v} - \frac{1}{pq} \right)^2 = \sum_{u,v} \pi_{u,v}^2 - \frac{1}{pq} \quad (2.11)$$

On voit ainsi que l'effet de la constante s'efface totalement de sorte que la fonction du Lagrangien peut s'écrire :

$$\begin{aligned} \mathbf{L}(\pi, \lambda, \omega, \theta) &= \sum_{u=1}^p \sum_{v=1}^q \pi_{u,v}^2 \\ &- \sum_{u=1}^p \lambda_u \left(\mu_u - \sum_{v=1}^q \pi_{u,v} \right) \\ &- \sum_{v=1}^q \omega_v \left(\nu_v - \sum_{u=1}^p \pi_{u,v} \right) \\ &- \theta \left(\sum_{u=1}^p \sum_{v=1}^q \pi_{u,v} - 1 \right) \end{aligned}$$

La fonction qui nous intéresse étant convexe, on recherche un minimum ce qui engendre le système d'équations suivant :

$$\frac{\partial \mathbf{L}(\pi, \lambda, \omega, \theta)}{\partial \pi_{u,v}} = 2\pi_{u,v} - \lambda_u - \omega_v - \theta = 0 \quad (2.12)$$

$$\frac{\partial \mathbf{L}(\pi, \lambda, \omega, \theta)}{\partial \lambda_u} = \mu_u - \sum_{v=1}^q \pi_{u,v} = 0 \quad (2.13)$$

$$\frac{\partial \mathbf{L}(\pi, \lambda, \omega, \theta)}{\partial \omega_v} = \nu_v - \sum_{u=1}^p \pi_{u,v} = 0 \quad (2.14)$$

Les multiplicateurs de Lagrange étant définis à une constante près, on suppose $\sum_v \omega_v = 0$ et on somme 2.12 sur v pour obtenir $2\mu_u = \sum_{v=1}^q 2\pi_{u,v} = q\lambda_u + q\theta$ puis

$$\lambda_u + \theta = \frac{2\mu_u}{q}, \forall u \quad (2.15)$$

De 2.14 on obtient

$$2\nu_v = \sum_{u=1}^p 2\pi_{u,v} = \sum_{u=1}^p (\lambda_u + \omega_v + \theta) = \sum_{u=1}^p \left(\frac{2}{q}\mu_u + \omega_v \right) = \frac{2}{q} + p\omega_v$$

pour en déduire

$$\omega_v = \frac{2\nu_v}{p} - \frac{2}{pq}, \forall v \quad (2.16)$$

En remplaçant dans 2.12 $\lambda_u + \theta$ et ω_v par leurs valeurs respectives, données par 2.15 et 2.16 on obtient :

$$\pi_{u,v}^* = \frac{\mu_u}{q} + \frac{\nu_v}{p} - \frac{1}{pq}, \forall (u, v) \quad (2.17)$$

C'est le moment d'utiliser notre hypothèse 2.9 pour constater que les valeurs π^* ainsi exprimées sont bien positives et conclure. On reviendra sur cette hypothèse dans les chapitres suivants, l'idée étant simplement ici de traduire le problème 2 comme un écart à l'indétermination tout juste introduite. \square

Optimisation en dimension arbitraire

Nous partons ici d'un constat simple : la solution du problème 6 ne permet pas de coupler un nombre arbitraire de variables mais seulement deux. Or, nous le verrons dans une partie ultérieure consacrée aux applications : il est primordial d'avoir une formule couplant un nombre arbitraire. Pour cela, nous proposons le problème 7 et espérons évidemment généraliser la formule de π^+ ; ce sera le cas comme indiqué dans le théorème 3.

Problème 7 (Minimal Trade Model (cas multidimensionnel)).

On se donne n distributions aléatoires $\mu^k, k = 1 \dots n$ sur p^k modalités respectivement. π est une distribution de probabilités sur l'ensemble produit $\otimes_{k=1}^n \{1, \dots, p^k\}$. Le problème est alors

$$\min_{\pi} \sum_{u_1, u_2, \dots, u_n} \left\{ \pi_{u_1, u_2, \dots, u_n} - \prod_{i=1}^n \frac{1}{p^i} \right\}^2$$

sous les contraintes **H1**

avec **H1** les contraintes suivantes :

$$\begin{aligned}\pi_{\cdot, \dots, \cdot, u_k, \cdot, \dots, \cdot} &= \mu_{u_k}^k, \forall 1 \leq k \leq n, \forall u_k \\ \pi_{\cdot, \dots, \cdot} &= 1 \\ \pi_{u_1, u_2, \dots, u_n} &\geq 0, \forall (u_1, \dots, u_n)\end{aligned}$$

Théorème 3.

La solution du Problème 7 est

$$\pi_{u_1, u_2, \dots, u_n}^* = \pi_{u_1, u_2, \dots, u_n}^+ = \sum_{k=1}^n \frac{\mu_{u_k}^k}{\prod_{i=1, i \neq k}^n p^i} - \frac{n-1}{\prod_{i=1}^n p^i}$$

Remarque 1 (Généralisation de l'indétermination).

La formule liée au théorème 3 généralise évidemment la forme précédente d'indétermination puisque si $n = 2$, on retrouve le théorème 2.

De plus, c'est bien la connaissance du résultat dans le cas bidimensionnel qui nous permet de trouver le résultat général puisque nous allons proposer une preuve par récurrence.

Démonstration.

Comme indiqué dans la remarque précédente, nous allons appuyer la preuve sur le résultat du théorème 2 afin de ne pas avoir à refaire les calculs. Ainsi, nous proposons une preuve par récurrence sur n , l'initialisation ($n=2$) étant fournie par ledit théorème nous proposons simplement une preuve de l'étape de récurrence.

Pour ce faire, nous constatons que π vit dans le simplexe de $\mathbb{R}^{\prod_{k=1}^n p^k}$ que nous nommerons $S_n(p^1, \dots, p^n)$ (attention, n ne donne pas la dimension puisque $S_{n-1}(p^1, \dots, p^{n-1})$ perd p^n dimensions). Nous pouvons néanmoins remarquer que se donner $\pi \in S_n(p^1, \dots, p^n)$ revient à se donner la marge ν de π sur les $\prod_{k=1}^{n-1} p^k$ premières variables, la marge de π sur les p^n dernières variables et à coupler ces deux marges. Ainsi, nous décomposons le problème d'optimisation en deux.

Introduisons un deuxième jeu de contraintes **H2** enrichissant **H1** en fixant la marge ν que l'on vient d'introduire :

$$\pi_{u_1, \dots, u_{n-1}, \cdot} = \nu_{u_1, \dots, u_{n-1}}, \forall (u_1, \dots, u_{n-1})$$

Le problème se réécrit :

$$\begin{array}{ll} \min_{\nu} & \min(\pi) \\ \text{sous les contraintes} & \text{sous les contraintes} \\ \text{H1} & \text{H2} \end{array} \quad \sum_{u_1, u_2, \dots, u_n} \left\{ \pi_{u_1, u_2, \dots, u_n} - \prod_{i=1}^n \frac{1}{p^i} \right\}^2$$

Dans **H2**, les contraintes sur les marges entre 1 et $n - 1$ sont portées par μ puisque $\pi_{\cdot, \dots, \cdot, u_k, \cdot, \dots, \cdot} = \nu_{\cdot, \dots, \cdot, u_k, \cdot, \dots, \cdot}, \forall 1 \leq k \leq n - 1$. De sorte que **H2** s'écrit ici :

$$\begin{aligned}\pi_{(\cdot, \dots, \cdot), u_n} &= \mu_{u_n}^n, \forall u_n \\ \pi_{(u_1, \dots, u_{n-1}), \cdot} &= \nu_{u_1, \dots, u_{n-1}}, \forall (u_1, \dots, u_{n-1}) \\ \pi_{(\cdot, \dots, \cdot)} &= 1 \\ \pi_{(u_1, u_2, \dots), u_n} &\geq 0, \forall (u_1, \dots, u_n)\end{aligned}$$

L'ajout des parenthèses permet de montrer que l'on se place dans le cadre du théorème 2 avec deux marges. Par ailleurs, la fonction de coût se réécrit comme toujours sous la forme $\Sigma\pi^2$ sans décentrage. Aussi, on sait que l'argmin du problème intérieur est

$$\pi^*(\nu)_{u_1, \dots, u_n} = \frac{\nu_{u_1, \dots, u_{n-1}}}{p^n} + \frac{\mu_{u_n}^n}{\prod_{i=1}^{n-1} p^i} - \frac{1}{\prod_{i=1}^n p^i}$$

de sorte que l'on a à optimiser :

$$\begin{aligned} \min_{\nu} \quad & \sum_{u_1, u_2, \dots, u_n} \{\pi^*(\nu)_{u_1, u_2, \dots, u_n}\}^2 \\ \text{sous les contraintes} \quad & \\ & H1 \end{aligned}$$

En développant, on obtient

$$\begin{aligned} \sum_{u_1, u_2, \dots, u_n} \pi^*(\nu)_{u_1, \dots, u_n}^2 &= \sum_{u_1, u_2, \dots, u_n} \left(\frac{\nu_{u_1, \dots, u_{n-1}}}{p^n} + \frac{\mu_{u_n}^n}{\prod_{i=1}^{n-1} p^i} - \frac{1}{\prod_{i=1}^n p^i} \right)^2 \\ &= \sum_{u_1, u_2, \dots, u_n} \nu_{u_1, \dots, u_{n-1}}^2 + Cste \end{aligned}$$

ν optimise donc le problème sous sa forme avec $(n-1)$ variables ainsi $\nu_{u_1, u_2, \dots, u_{n-1}}^* = \sum_{k=1}^{n-1} \frac{\mu_{u_k}^k}{\prod_{i=1, i \neq k}^{n-1} p^i} - \frac{n-2}{\prod_{i=1}^{n-1} p^i}$ Et enfin :

$$\begin{aligned} \pi_{u_1, u_2, \dots, u_n}^* &= \pi^*(\nu^*)_{u_1, u_2, \dots, u_n} \\ &= \frac{\sum_{k=1}^{n-1} \frac{\mu_{u_k}^k}{\prod_{i=1, i \neq k}^{n-1} p^i} - \frac{n-2}{\prod_{i=1}^{n-1} p^i}}{p^n} + \frac{\mu_{u_n}^n}{\prod_{i=1}^{n-1} p^i} - \frac{1}{\prod_{i=1}^n p^i} \\ &= \sum_{k=1}^n \frac{\mu_{u_k}^k}{\prod_{i=1, i \neq k}^n p^i} - \frac{n-1}{\prod_{i=1}^n p^i} \\ &= \pi_{u_1, u_2, \dots, u_n}^+ \end{aligned}$$

□

La majorité du document restera, notamment pour des besoins de clarté, consacrée au cas du couplage de deux marges.

Le choix de la fonction de coût dans le problème de transport discret 4 suffit donc à passer du couplage extrêmement connu, utilisé et interprété qu'est l'indépendance π^\times (formule 1) à celui tout juste introduit d'indétermination π^+ (formule 3). Cette proximité entre les deux couplages, nous la retrouverons tout au long du document et travaillerons essentiellement sur les propriétés du second. Pour commencer, l'introduction des matrices de Monge dans le chapitre qui suit sera l'occasion de justifier la notation qui paraît ici arbitraire du "+" et du "×" pour l'appellation des deux optimaux.

En attendant, nous commençons par relâcher l'hypothèse 2.9 dans la section suivante qui minimise le problème 6 dans un cadre plus général.

2.3 Optimisation sans hypothèse sur les marges

Nous reprenons ici les calculs d'optimisation du problème 6. En anticipant un peu, nous observons qu'il s'agit de minimiser le numérateur d'un critère dit de Janson-Végélius que nous définirons plus tard dans la définition 20. L'originalité de cette section et son apport est l'absence de la fameuse hypothèse (2.9) sur μ et ν permettant à π^+ d'être positif. Nous souhaitons comprendre l'impact de cette omission sur la forme de la solution optimale.

Problème 8 (MTM sous condition de positivité).

$$\begin{aligned} \min_{\pi} \quad & \sum_{u,v} \left\{ \pi_{u,v} - \frac{1}{pq} \right\}^2 \\ \text{subject to :} \quad & \\ & \sum_v \pi_{u,v} = \mu_u \\ & \sum_u \pi_{u,v} = \nu_v \\ & \sum_{u,v} \pi_{u,v} = 1 \\ & \pi_{u,v} \geq 0 \quad (\text{hypothèse auparavant garantie par (2.9)}) \end{aligned}$$

Auparavant, ce que nous faisons consistait à calculer le minimum sans l'hypothèse de positivité, à trouver π^+ et à constater qu'il était positif sous l'hypothèse 2.9. Cette fois, nous utilisons Karush-Kuhn-Tucker (KKT) pour résoudre sans hypothèse sur μ et ν .

Propriété 1.

La solution du problème 8 est de la forme $\pi_{u,v}^* = (\alpha_u + \beta_v - k)\delta_{(u,v) \in \mathcal{E}^+}$ avec \mathcal{E}^+ un sous-ensemble de $\mathcal{E} = \mathcal{U} \otimes \mathcal{V}$ sur lequel π^* est positif.

Démonstration.

On applique KKT pour obtenir la fonction suivante regroupant les contraintes :

$$\begin{aligned} L_{\pi,\lambda,\omega,\theta,\eta} &= \sum_{u,v} \pi_{u,v}^2 + \sum_u \lambda_u \left(\mu_u - \sum_v \pi_{u,v} \right) + \sum_v \omega_v \left(\nu_v - \sum_u \pi_{u,v} \right) \\ &+ \theta \left(1 - \sum_{u,v} \pi_{u,v} \right) + \sum_{u,v} (-2\eta_{u,v}) (-\pi_{u,v}) \end{aligned}$$

Petite remarque : dans l'écriture du KKT, les contraintes associées aux inégalités (ici le dernier lot de contraintes $\pi_{u,v} \geq 0$) s'écrivent sous une forme $g_i(x) \leq 0$. En l'occurrence, $g_{u,v}(\pi) = -\pi_{u,v}$ et le coefficient associé est $\epsilon_{u,v} \geq 0$ dans le théorème. Nous avons transformé ce dernier en $-2\eta_{u,v}$ avec $\eta_{u,v} \leq 0$ pour simplifier les calculs.

En calculant le gradient selon $\pi_{u,v}$ qui doit être nul on trouve pour tout (u, v) :

$$2\pi_{u,v} = \lambda_u + \omega_v + \theta + 2\eta_{u,v} \tag{2.18}$$

Avec les conditions de Problème 8 auxquelles nous ajoutons :

$$\begin{aligned}\eta_{u,v}\pi_{u,v} &= 0 \\ \eta_{u,v} &\leq 0\end{aligned}$$

Ainsi, π et η sont alternativement nuls sur l'ensemble produit \mathcal{E} . Découpons le quadrillage global en deux : $\mathcal{E}^+ = \{(u,v) \in \mathcal{E} \text{ t.q. } \pi_{u,v} > 0\}$ et $\mathcal{E}^0 = \{(u,v) \in \mathcal{E} \text{ t.q. } \pi_{u,v} = 0\}$. L'hypothèse de positivité de π nous permet d'affirmer qu'il s'agit d'une partition (*i.e.* \mathcal{E}^- est vide).

En sommant sur v l'équation 2.18 :

$$2\mu_u = q\lambda_u + \sum_v \omega_v + q\theta + \sum_v 2\eta_{u,v}$$

Si l'on suppose $\sum_v \omega_v = 0$ puisque la somme des variables de KKT est définie à une constante près :

$$\lambda_u + \theta + \frac{\sum_v 2\eta_{u,v}}{q} = \frac{2\mu_u}{q} \quad (2.19)$$

En sommant maintenant sur u l'équation 2.18 puis en remplaçant grâce à l'équation 2.20 :

$$\begin{aligned}2\nu_v &= \sum_u (\lambda_u + \theta + 2\eta_{u,v} + \omega_v) \\ &= \sum_u \left(2\frac{\mu_u}{q} + 2\eta_{u,v} - 2\sum_v \frac{\eta_{u,v}}{q} \right) + p\omega_v \\ &= \frac{2}{q} + 2\eta_{\cdot,v} - \frac{2\eta_{\cdot,\cdot}}{q} + p\omega_v\end{aligned}$$

Soit :

$$\omega_v = \frac{2\nu_v}{p} - \frac{2}{pq} + \frac{2\eta_{\cdot,v}}{p} - \frac{2\eta_{\cdot,\cdot}}{pq}$$

Si l'on reporte tout dans l'équation de départ 2.18 :

- si $\pi_{u,v} \neq 0$ alors $\pi_{u,v} = \pi_{u,v}^+ + \frac{\eta_{u,\cdot}}{q} + \frac{\eta_{\cdot,v}}{p} - \frac{\eta_{\cdot,\cdot}}{pq}$
- si $\pi_{u,v} = 0$ alors $\eta_{u,v} < 0$

Ainsi la forme de l'optimal $\pi_{u,v}$ est identique à celle de π^+ : $\alpha_u + \beta_v - k$ pour les éléments non nuls de sorte que l'on obtient bien la formule annoncée :

$$\pi_{u,v}^* = (\alpha_u + \beta_v - k)\delta_{(u,v) \in \mathcal{E}^+}$$

□

Remarque 2.

η et π jouent des rôles symétriques, l'un prenant la partie positive et l'autre la partie négative de la même expression.

La forme de l'optimal reste donc additive tout comme celle de π^+ mais ne peut se développer sur tout le support.

Il est d'ailleurs étonnant que l'optimal ne recouvre pas tout le support, la forme de la fonction de coût semble indiquer qu'il faille répartir au maximum. Pour mieux comprendre ce qui peut arriver, nous proposons un exemple dans la section suivante.

Nous regrettons de n'avoir pu résoudre exactement le problème en proposant des valeurs pour les différentes variables. En dehors de la forme de la solution, les valeurs exactes de celle-ci nous échappent. C'est que nous ne disposons plus simplement des marges fixes pour nous aider. En effet, l'équation habituelle $\sum_{1 \leq v \leq q} \pi_{u,v} = \mu_u$ devient $\sum_{1 \leq v \leq q \text{ t.q. } (u,v) \in \mathcal{E}^+} \pi_{u,v} = \mu_u$ de sorte que chaque sommation a lieu sur des sous-ensembles :

- $\mathcal{E}^+(u) = \{1 \leq v \leq q \text{ t.q. } (u, v) \in \mathcal{E}^+\}$
- $\mathcal{E}^+(v) = \{1 \leq u \leq p \text{ t.q. } (u, v) \in \mathcal{E}^+\}$

Ces sous-ensembles de sommation, dépendant respectivement de la ligne et de la colonne, aucune forme close n'a pu être extraite des conditions de marge.

Un optimal au support restreint

Nous essayons de comprendre comment un optimal peut recouvrir seulement une partie du support. Plaçons nous dans un cadre élémentaire : deux marges de taille 2.

Si l'on présente $\pi_{u,v}$ solution du problème sur un ensemble produit de taille 4, le choix d'une case suffit à tout déterminer (voir la table 2.1 dans un cadre 2×2). Posons donc $x = \pi_{1,1}$. Il nous faut minimiser la somme des carrés soit :

$$x^2 + (\mu_1 - x)^2 + (\nu_1 - x)^2 + (\mu_2 - \nu_1 + x)^2$$

Sachant que le déséquilibre des marges est un gage de non-respect de l'hypothèse 2.9, nous utilisons des marges identiques loin de l'uniforme sur un ensemble de taille 2 : $\mu = \nu = (0.1, 0.9)$ ce qui donne comme fonction de coût :

$$x^2 + (0.1 - x)^2 + (0.1 - x)^2 + (0.8 + x)^2$$

Comme tout doit être positif, $0 \leq x \leq 0.1$. On remarque bien ici le problème : donner à x une valeur positive entraîne une petite diminution (on gagne sur les deux $(0.1 - x)^2$ mais nous fait perdre beaucoup sur le $(0.8 + x)^2$) cela est engendré par le grand déséquilibre au sein des marges. Lorsque le quadrillage est de dimension supérieure, c'est le même effet qui permet à l'optimal des valeurs nulles sur les cases des lignes et colonnes à faibles marges. Et ce pour permettre aux croisement des grosses marges de prendre une valeur élevée mais minimale. Si l'on s'amuse ici à développer : $4x^2 + 1.2x + 0.65$ est évidemment minimum pour $x = 0$.

2.4 Deux fonctions de couplage canoniques

2.4.1 Fonction de coût et divergence

Les paragraphes précédent ont vu l'introduction de deux lois de probabilités solutions de deux problèmes d'optimisation. Dans cette section nous allons voir ces deux solutions comme des couplages et justifier leur caractère canonique.

Étant donné deux marges, la méthode classique et naturelle de couplage consiste à utiliser l'indépendance et à poursuivre les calculs. Ou que ce soit ce qui la justifie ou que son habitude soit trop ancrée, elle apparaît naturellement dans nos applications à la vie courante. Dans les travaux scientifiques, l'approche est similaire. Ainsi, l'utilisation du χ^2 ne mesure-t-elle rien d'autre qu'un écart à l'indépendance.

À l'origine de sa suprématie est l'expérience, l'empirique. Si nous souhaitons modéliser un lancer de dés, nous couplerons habituellement les lois de chacun selon l'indépendance, intuitivement.

Bien que très intuitif, ce n'est certes pas (et de loin) le seul couplage éligible. De fait, comme introduit par Sklar dans [87], toute copule génère un couplage que l'on peut appliquer à n'importe quelles marges. La transposition dans le cas discret revient à remplir une table de contingence, tout remplissage respectant les marges donnera un couplage valide comme mentionné dans la définition 4.

Définition 4 (Fonction de couplage).

Une fonction de couplage C est une méthode pour construire une table de contingence à pq éléments en respectant deux marges données en argument. $C(\mu, \nu)$ est ainsi une loi de probabilité sur $\mathcal{U} \times \mathcal{V}$ dont les marges sont μ et ν , formellement :

$$C : S_p \times S_q \rightarrow S_{pq} \\ (\mu, \nu) \mapsto C(\mu, \nu) \in \mathcal{L}_{\mu, \nu}$$

avec

$$\mathcal{L}_{\mu, \nu} = \{\pi \in S_{pq} \text{ tels que les marges de } \pi \text{ soient } \mu \text{ et } \nu\}$$

Dans [10] nous avons introduit la notion de *fonction de couplage* associée à la résolution d'un problème d'optimisation ; évidemment comme ceux déjà rencontrés :

Problème 9 (Version discrète).

$$\min_{\pi} \sum_{u=1}^p \sum_{v=1}^q \mathbf{M}(\pi_{u,v}) \\ \text{sous les contraintes :} \\ \sum_{v=1}^q \pi_{u,v} = \mu_u; \forall u \in \{1, \dots, p\} \\ \sum_{u=1}^p \pi_{u,v} = \nu_v; \forall v \in \{1, \dots, q\} \\ \pi_{u,v} \geq 0; \forall (u, v) \in \{1, \dots, p\} \times \{1, \dots, q\}$$

Le choix de la fonction \mathbf{M} dépend de l'application. Les choix varient : concentrer π autour de l'uniforme est une motivation classique en ce qu'elle réduit la quantité d'information transmise par les marges dans le couplage résultant de la solution.

Un problème comme ci-dessus est essentiellement défini par sa fonction de coût \mathbf{M} , tandis que les marges (μ, ν) apparaissent comme des paramètres. Aussi, on espère que la solution d'un problème 4 P particulier s'écrive $C^P(\mu, \nu)$.

D'où l'idée d'associer à un problème d'optimisation une fonction de couplage exprimant la forme générale des solutions.

Définition 5 (Association fonction de couplage et problème d'optimisation).

Pour un problème P suivant la structure précédente, on définit la fonction de couplage C^P par : $C^P(\mu, \nu) = \pi^*(P, \mu, \nu)$ pourvu que π^* , solution unique de P , existe pour les marges μ et ν .

L'existence d'une solution dépend de la fonction de coût \mathbf{M} tout comme des marges. Ce sujet est discuté dans de nombreux papiers, nous ne le détaillerons pas ici.

Dans notre papier [10] nous avons exprimé les deux fonctions de coût liées aux sections précédentes comme des divergences de π par rapport à l'uniforme. Ce faisant, nous avons pu justifier grâce aux résultats du papier [29] que les deux formules (équations (2.20) et (2.21)) étaient les deux seules menant à des fonctions de couplage structurées ; le sens de cet adjectif sera précisé dans les lignes qui viennent. Commençons par rappeler les deux coûts des sections précédentes.

$$\mathbf{M}^\times : x \mapsto x \log(pqx) \quad (2.20)$$

$$\mathbf{M}^+ : x \mapsto pq \left(x - \frac{1}{pq} \right)^2 \quad (2.21)$$

Sous le formalisme des divergences nous uniformisons l'écriture de ces deux équations.

Déjà, en utilisant le résultat des théorèmes 2 et 1, on observe une même solution lorsque les contraintes de marge sont relâchées, à savoir, la loi uniforme \mathbb{U} . Cette coïncidence n'est pas anodine puisque nous allons voir que les deux coûts sont des divergences (définition 6) de π vivant dans un espace contraint par rapport à \mathbb{U} .

Définition 6 (Divergence avec paramètre).

Étant donné une fonction positive $\phi : \mathbb{R}^+ \mapsto \mathbb{R}^+$, convexe en 1 et vérifiant $\phi(1) = 0$, et deux lois de probabilités discrètes m et n sur pq élément, on définit la divergence D_ϕ de m par rapport à n :

$$D_\phi(m|n) = \sum_{u=1}^p \sum_{v=1}^q n_{u,v} \phi \left(\frac{m_{u,v}}{n_{u,v}} \right) = \mathbb{E}_n \left[\phi \left(\frac{m_{u,v}}{n_{u,v}} \right) \right]$$

par ailleurs, nous posons $0 * \phi(x) = 0$ pour tout x .

En notant comme déjà indiqué \mathbb{U} la loi uniforme sur $\mathcal{U} \otimes \mathcal{V}$ et en choisissant $\phi : x \mapsto x \log(x) - x + 1$ il vient :

$$\begin{aligned} D_{KL}(\pi | \mathbb{U}) &= \sum_{u=1}^p \sum_{v=1}^q \frac{1}{pq} [pq\pi_{u,v} \log(pq\pi_{u,v}) - pq\pi_{u,v} + 1] \\ &= \sum_{u=1}^p \sum_{v=1}^q \pi_{u,v} \log(pq\pi_{u,v}) \\ &= \sum_{u=1}^p \sum_{v=1}^q \mathbf{M}^\times(\pi_{u,v}) \end{aligned}$$

de sorte que le problème 5 associé à la fonction de coût \mathbf{M}^\times se réécrit comme la minimisation de D_{KL} qui n'est autre que la divergence de Kullback-Leibler habituelle.

De même, en choisissant $\phi : x \mapsto (x - 1)^2$ (la déviation du Chi-deux) :

$$\begin{aligned}
D_2(\pi | \mathbb{U}) &= \sum_{u=1}^p \sum_{v=1}^q \frac{1}{pq} (1 - pq\pi_{u,v})^2 \\
&= pq \sum_{u=1}^p \sum_{v=1}^q \left(\pi_{u,v} - \frac{1}{pq} \right)^2 \\
&= \sum_{u=1}^p \sum_{v=1}^q \mathbf{M}^+(\pi_{u,v})
\end{aligned} \tag{2.22}$$

Au final les deux coûts \mathbf{M}^\times et \mathbf{M}^+ se distinguent l'un de l'autre par la manière dont la divergence par rapport à \mathbb{U} est estimée. Par ailleurs, les contraintes sur π définissent un sous-espace $\mathcal{L}_{\mu,\nu}$ du simplexe S_n auquel π doit appartenir.

2.4.2 Justification structurelle des deux divergences

La présentation précédente nous a permis de lier l'optimisation d'un problème de transport associé à une notion dite de fonction de couplage tout en fournissant deux exemples. Outre qu'il soit toujours intéressant de proposer des exemples, le choix des deux n'est pas anodin. De fait, il est justifié par un travail de Imre Csiszár dans [29] intitulé "Why least squares and maximum entropy?". Dans ce document, Csizar montre que la recherche de divergences vérifiant des propriétés additionnelles intuitives restreint le choix aux deux mentionnés : la différence des carrés ou l'entropie c'est-à-dire exactement les deux coûts \mathbf{M}^+ et \mathbf{M}^\times . Précisons son approche.

Intégrons nos problèmes de transport dans le cadre requis par Csizar. Nous avons déjà remarqué que les deux problèmes 5 et 6 reviennent à minimiser une divergence entre π et l'uniforme tout en contraignant π à respecter des marges, c'est à dire à vivre dans un espace éligible $\mathcal{L}_{\mu,\nu}$ lui même à l'intérieur du simplexe S_n . Dans le problème 5, la fonction de distance est l'entropie de Kullback-Leibler, dans le problème symétrique 6 il s'agit de la norme \mathbb{L}_2 .

Une question plus générale exprimée par Csizar est : "Comment adapter une estimation u_0 pour qu'elle respecte un ensemble de contraintes?". Disons que u_0 vit dans un espace S_n dans \mathbb{R}^n , $n \in \mathbb{N}$ tandis que les contraintes définissent un sous-espace $L \in \mathcal{L}$ (\mathcal{L} étant l'ensemble des sous-espaces définis par un jeu de contraintes affines voir [29] pour plus de détails). La formalisation opérée dans l'article passe par la définition d'une *règle de projection* Π , c'est-à-dire une fonction prenant en argument un sous-espace $L \in \mathcal{L}$ et lui attribuant une méthode Π_L qui projette toute estimation $u_0 \in S_n$ vers un vecteur dans L :

$$\begin{aligned}
\Pi : \quad \mathcal{L} &\rightarrow (S_n \rightarrow S_n) \\
L &\mapsto \Pi_L : (u_0 \mapsto \Pi_L(u_0) \in L)
\end{aligned}$$

L'article ajoute alors une suite de propriétés naturelles pour une *règle de projection* que nous rapportons ci-après :

- *consistance* : si $L' \subset L$ et $\Pi_L(S_n) \subset L'$ alors $\Pi_{L'} = \Pi_L$; exprimant ainsi que si le résultat d'une projection dans un espace en englobant un plus petit est systématiquement dans le plus petit alors les projections sur les deux espaces sont égales.
- *séparation contrainte unique* : si L et L' sont définis par une unique contrainte et ne sont pas égaux alors les deux projections sont toujours différentes $\forall u_0 \in S_n, \Pi_L \neq \Pi_{L'}$ (à moins que u_0 soit à la fois dans L et L'). Prenons un exemple pour mieux comprendre. Restreignons-nous à \mathbb{R}^2 , et minimisons $\|\cdot\|_2$ en projetant sur deux droites. Le résultat sera différent sauf si les deux contiennent l'estimation (en l'occurrence 0).
- *continuité* : Π est continu par rapport à sa variable $L \in \mathcal{L}$; cela revient à dire qu'il est continument défini par rapport à chaque paramètre des contraintes affines.
- *invariance d'échelle* : $\Pi_{\lambda L}(\lambda u) = \lambda u$ pour tout réel λ et tout $u_0 \in S_n$.
- *local* : pour tout sous-ensemble $J \subset \{1, \dots, n\}$, $(\Pi_L)_J = (\Pi_{L'})_J$ dès lors que $L_J = L'_J$ où la notation L_J signifie que l'on ne garde que les contraintes agissant sur les coordonnées du sous-ensemble J d'entiers et où $(\Pi_L)_J$ est la restriction du vecteur résultant de l'application de Π_L aux coordonnées dans J (c'est donc un vecteur de $\mathbb{R}^{|J|}$). Cette propriété exprime que le résultat de l'application de Π_L sur un jeu de coordonnées ne dépend que des contraintes sur ces coordonnées.
- *transitivité* : pour tout sous-ensemble $L' \subset L$, $\Pi_{L'} = \Pi'_L \circ \Pi_L$. Comprendre donc que la projection dans un espace englobant un autre espace n'affecte pas le résultat final si l'on projette ensuite dans le petit espace.

Toutes les propriétés mentionnées apparaissent intuitives, nonobstant une écriture parfois fastidieuse, et légitimes pour définir une règle de projection. Le résultat majeur du papier [29] est qu'astreindre la règle de projection à respecter ces propriétés revient exactement à restreindre leur forme à deux schémas généraux :

- $\Pi_L : u \rightarrow \operatorname{argmin}_{v \in L} [\sum_{i=1}^n \alpha_i (v_i - u_i)^2]$ pour un vecteur fixé $\alpha \in \mathbb{R}^n$.
- $\Pi_L : u \rightarrow \operatorname{argmin}_{v \in L} [\sum_{i=1}^n \alpha_i h_\beta(v_i | u_i)]$ pour un vecteur fixé $\alpha \in \mathbb{R}^n$ et avec h_β des fonctions spécifiques, définies dans le papier et rejoignant l'entropie lorsque $\beta = 1$

A ce stade nous avons déjà drastiquement limité les possibilités de fonctions de coût. Liées à une projection de l'uniforme, elle sont composées de projections \mathbb{L}_2 ou de fonctions h_β similaires à l'entropie. L'ajout d'une dernière propriété sur la solution optimale hors contraintes, à savoir que celle-ci respecte le principe de proportionnalité restreint enfin les choix à $\alpha = \beta = 1$. En détail, on suppose que la mise à jour de la masse totale disponible (par exemple dans le cadre d'une application monétaire) se traduit par une mise à jour proportionnelle de chaque composante du vecteur solution.

Pour revenir à nos deux problèmes, l'estimation est la loi uniforme, les sous-espaces $L \subset S_n$ sont les $\mathcal{L}_{\mu, \nu}$ définis par les contraintes de marge μ et ν . Aussi, si l'on suppose les propriétés de la règle de projection respectées, les deux fonctions de coûts et donc les deux fonctions de couplages associés proposés constituent une vision exhaustive.

Pour résumer, avec les notations habituelles, nous introduisons deux fonctions de couplage canoniques notées :

$$\pi^\times = \mu \otimes \nu = C^{\text{Problem 5}}(\mu, \nu) = C^\times(\mu, \nu) \text{ pour l'indépendance; } \quad (2.23)$$

$$\pi^+ = \mu \oplus \nu = C^{\text{Problem 6}}(\mu, \nu) = C^+(\mu, \nu) \text{ pour l'indétermination. } \quad (2.24)$$

2.4.3 Un écart relatif

En tant qu'elles optimisent toutes les deux un problème visant à projeter la loi uniforme sur le même espace, on s'attend à ce que l'écart soit minime dans un sens à définir. Nous apportons ici une mesure de cette proximité. Plus précisément, nous supposons que μ et ν suivent la loi de Dirichlet qui ne représente rien d'autre que la loi uniforme sur les loi de probabilités ; nous en rappelons ici la forme.

Définition 7 (Loi de Dirichlet).

La densité de la loi de Dirichlet \mathcal{D}_p qui exprime une loi uniforme sur les lois de probabilités à p éléments s'exprime comme suit :

$$f(\mu_1, \dots, \mu_p) \prod_{k=1}^p d\mu_k = \frac{1}{B(p)} \prod_{k=1}^p \mu_k^0 \prod_{k=1}^p d\mu_k = \frac{1}{B(p)} \prod_{k=1}^p d\mu_k$$

où B est la fonction Bêta multinomiale.

Ayant exprimé une densité pour μ et ν (remplacer p par q), nous appliquons les deux fonctions de couplage C^+ et C^\times . Comme distance, nous prenons :

$$\Delta_p = \mathbb{E}_{(\mu, \nu) \sim \mathcal{D}_p \otimes \mathcal{D}_q} \left[\sum_{u=1}^p \sum_{v=1}^q [(\mu \otimes \nu)_{u,v} - (\mu \oplus \nu)_{u,v}]^2 \right]$$

puis suivons les égalités :

$$\begin{aligned} \Delta_p &= \mathbb{E}_{(\mu, \nu) \sim \mathcal{D}_p \otimes \mathcal{D}_q} \left[\sum_{u=1}^p \sum_{v=1}^q \left[\left(\mu_u - \frac{1}{p} \right) \left(\nu_v - \frac{1}{q} \right) \right]^2 \right] \\ &= \mathbb{E}_{\mu \sim \mathcal{D}_p} \left[\sum_{u=1}^p \left(\mu_u - \frac{1}{p} \right)^2 \right] \mathbb{E}_{\nu \sim \mathcal{D}_q} \left[\sum_{v=1}^q \left(\nu_v - \frac{1}{q} \right)^2 \right] \\ &= pq \mathbb{E}_{\mu \sim \mathcal{D}_p} \left[\left(\mu_1 - \frac{1}{p} \right)^2 \right] \mathbb{E}_{\nu \sim \mathcal{D}_q} \left[\left(\nu_1 - \frac{1}{q} \right)^2 \right] \end{aligned}$$

Aussi, nous notons qu'il nous faut calculer la variance de \mathcal{D}_p ; s'agissant d'une loi bien connue, nous en connaissons les moments et notamment :

Propriété 2 (Variance de la loi de Dirichlet).

$$\mathbb{V}_{X \sim \mathcal{D}_p}[X] = \frac{p-1}{p^2(p+1)}$$

Cette dernière proposition 2 implique entre autres que les marges se concentrent autour de leur centres respectifs $\frac{1}{p}$ et $\frac{1}{q}$ dès que p ou q augmente. Ayant remarqué que les deux couplages sont égaux si μ ou ν est uniforme, un argument de continuité nous fait prédire que Δ_p converge vers 0 si $p+q$ augmente. C'est exactement ce qui a lieu suivant l'expression :

$$\Delta_p = \frac{1}{pq} \left(\frac{p-1}{p+1} \cdot \frac{q-1}{q+1} \right) \leq \frac{1}{pq}$$

Lorsque p ou q augmente et tandis que les marges se concentrent vers l'uniforme correspondant, les fonctions de couplages se rapprochent rapidement en espérance suivant la vitesse de $\frac{1}{pq}$.

Nous profitons de cette section dédiée à la différence moyenne entre l'indépendance et l'indétermination pour montrer un cas d'égalité de ces deux notions lorsque l'une des deux marges est uniforme. Cette remarque est évidente avec la proposition suivante.

Propriété 3 (Ecart Indépendance Indétermination).

$$\begin{aligned}\pi_{u,v}^\times - \pi_{u,v}^+ &= \mu_u \nu_v - \frac{\mu_u}{q} - \frac{\nu_v}{p} + \frac{1}{pq} \\ &= \left(\mu_u - \frac{1}{p}\right) \left(\nu_v - \frac{1}{q}\right)\end{aligned}$$

Remarque 3. *La forme obtenue dans la proposition 3 montre que les deux couplages sont égaux si et seulement si l'une des deux marges est uniforme.*

Propriétés de l'indétermination

— Ô buffet du vieux temps, tu
sais bien des histoires,
Et tu voudrais conter tes contes,
et tu bruis
Quand s'ouvrent lentement tes
grandes portes noires.

Arthur Rimbaud, *Le Buffet*,
1870

L'objectif de ce chapitre est de mieux comprendre le couplage d'indétermination introduit à plusieurs reprises dans ce document. En effet, le couplage d'indépendance est obtenu de façon similaire, et est immédiatement relié à un tirage naturel en tant qu'on le retrouve avec un peu d'intuition dans la modélisation d'expériences. C^+ n'a pas cette propriété et son interprétation est donc un exercice en soi. Pour le résoudre, nous présentons ici plusieurs approches.

Nous commençons par rappeler une propriété structurelle vérifiée par toute matrice d'*indétermination* ainsi que son pendant pour l'*indépendance*. Chacune des relations structurelle caractérise le couplage et en justifie la notation : ou additive ou multiplicative. La relation additive que nous précisons, liée à l'indétermination remonte à une remarque de Gaspard Monge dans le document [73] datant de la fin du XVIII^e siècle.

Dans une seconde partie, nous réintroduisons un encodage dit relationnel pour définir une notion de cas "pour" et "contre" conduisant à une égalité formelle. Nous poursuivrons cette approche en l'appliquant dans un cas 2×2 auquel il est simple d'associer une valeur morale : les jugements judiciaires. Cet encodage sera réutilisé dans le chapitre 5 lié aux critères de corrélation dans lequel nous insisterons davantage sur la dualité espace des matrices de contingence / espace des matrices relationnelles.

En analysant la formule du couplage d'*indétermination*, et en s'appuyant sur son respect de la formule de Monge, nous établissons une décomposition de toute matrice pq d'indétermination en une ligne arbitraire de taille q et une collection de p écarts additionnels. Cette écriture s'accompagne d'une méthode de tirage et nous permet de comprendre comment un tel couplage réduit les collisions. De fait, l'écriture même du problème 6 tend à réduire les collisions (comprendre, la probabilité que deux tirages de

suite soient égaux) et il est fascinant de découvrir comment cette propriété s'inscrit dans la forme de la solution.

La majorité des propriétés que nous regroupons ici est également répertoriée dans l'article [11] comme des pré-requis pour résoudre des problèmes connus d'affectations ou d'encryptage ; nous reviendrons sur ces applications dans le chapitre 4. Dans le cadre du présent document nous détaillons sensiblement plus chacune des propriétés et ajoutons quelques expériences.

Avant toute chose, nous allons estimer la vraisemblance d'un couplage selon l'indétermination. En effet, contrairement à l'indépendance, toutes les marges ne sauraient se coupler selon la formule associée. Ne sachant pour l'instant produire une expérience le réalisant, nous allons tirer au hasard parmi les lois des marges et calculer la proportion acceptant l'indétermination comme couplage. De surcroît nous proposons une méthode pour modifier deux marges quelconques afin de les plonger dans l'espace des couples vérifiant l'hypothèse (2.9).

3.1 Probabilité de construction de π^+

Comme il s'agit d'une pure curiosité mathématique (des éléments peu probables uniformément peuvent arriver car les expériences ne sauraient tirer selon l'uniforme) nous allons simplifier au maximum les calculs en restreignant le cadre, l'objectif est simplement d'avoir une idée de l'impact des restrictions imposées par les contraintes sur les marges.

Théorie

On a vu que si l'on se donne μ une loi de probabilité sur p valeurs discrètes, appartenant donc au simplexe S_p , il n'est pas toujours possible de construire π^+ couplant μ avec elle-même. Il faut pour cela que le couple (μ, μ) vérifie l'hypothèse (2.9) qui s'écrit ici :

$$p \min \mu + q \min \mu \geq 1$$

soit :

$$\min \mu \geq \frac{1}{2p}$$

Le cadre étant restreint, nous voulons estimer la probabilité qu'un tel événement arrive. Pour cela, on se donne donc une loi de probabilité naturelle sur les lois de probabilité sur p éléments : la loi uniforme. Les lois de probabilité vivent dans un le simplexe S_p et, au sein de ce sous-espace de $[0, 1]^{p-1}$, il faut calculer la mesure de Lebesgue de l'ensemble valide.

Propriété 4.

La proportion des μ valides est $\frac{1}{2^{p-1}}$.

Démonstration.

On impose des bornes restreintes aux intégrales construisant μ :

$$\int_{\frac{1}{2p}}^{1-\frac{p-1}{2p}} \int_{\frac{1}{2p}}^{1-\frac{p-2}{2p}-x_1} \dots \int_{\frac{1}{2p}}^{1-\frac{1}{2p}-\sum_{i=1}^{p-2} x_i} dx_1 \dots dx_{p-1}$$

Avec le changement de variable : $x_1 \leftarrow x_1 + \frac{1}{2p}$

$$\int_0^{\frac{1}{2}} \int_{\frac{1}{2p}}^{1-\frac{p-1}{2p}-x_1} \dots \int_{\frac{1}{2p}}^{1-\frac{2}{2p}-\sum_{i=1}^{p-2} x_i} dx_1 \dots dx_{p-1}$$

Si l'on continue avec les changements de variables successifs : $x_i \leftarrow x_i + \frac{1}{2p}$

$$\int_0^{\frac{1}{2}} \int_0^{\frac{1}{2}-x_1} \dots \int_0^{\frac{1}{2}-\sum_{i=1}^{p-2} x_i} dx_1 \dots dx_{p-1}$$

C'est donc exactement la définition d'une loi de probabilité qui sommerait à $\frac{1}{2}$ à la place de 1 d'où le résultat puisque chaque composante est alors multipliée par $\frac{1}{2}$ et la dernière est imposée par la somme à 1. \square

Remarque 4.

En réalité, le résultat précédent n'est pas surprenant. On dispose d'une méthode constructive pour avoir un μ valide. L'expression donnée par l'hypothèse (2.9) montre que μ_u est supérieur à $\frac{1}{2p}$ pour tout u . On en déduit : $\mu_u = \frac{1}{2p} + \frac{r_u}{2}$ où r est une loi de probabilité quelconque sur p éléments.

Jusqu'ici, nous avons supposé coupler μ avec elle-même, ce qui simplifiait les calculs mais restreignait inutilement le problème. De fait, dans le reste du document, nous souhaitons coupler μ avec une deuxième loi ν . Suivant les mêmes hypothèses que précédemment, on suppose tirer μ et ν uniformément parmi les lois de probabilités, donc parmi S_p et S_q respectivement. En posant

$$\alpha = \min_u p * \mu_u, \tag{3.1}$$

le respect de l'hypothèse (2.9) se traduit par l'existence d'un réel α positif comme mentionné dans la proposition 5.

Propriété 5 (Construction marges éligibles, cas discret).

On peut coupler les marges (μ, ν) selon l'indétermination si et seulement si il existe un réel α positif tel que :

$$\begin{aligned} \forall 1 \leq u \leq p, \mu_u &\geq \frac{\alpha}{p} \\ \forall 1 \leq v \leq q, \nu_v &\geq \frac{1-\alpha}{q} \end{aligned}$$

Une première remarque, α étant le minimum d'un ensemble de p éléments sommant à 1, le numérateur de la seconde inégalité est bien dans $[0, 1]$, imposant donc que ν soit toujours supérieur à un réel $\beta = 1 - \alpha$ dans $[0, 1]$. Une seconde remarque concerne l'introduction de la variable p dans la définition, elle est là pour ne pas briser la symétrie entre μ et ν , assurant que toutes les valeurs α de $[0, 1]$ sont éligibles sans regard pour p ou q .

Reprenant la remarque 4, on en déduit, l'existence de deux loi de probabilités r et s quelconques sur p et q éléments telles que :

$$\begin{aligned}\forall 1 \leq u \leq p, \mu_u &= \frac{\alpha}{p} + (1 - \alpha)r_u \\ \forall 1 \leq v \leq q, \nu_v &= \frac{1 - \alpha}{q} + \alpha s_v\end{aligned}$$

À α fixe, la proportion éligible de l'espace S_p des μ est donc $(1 - \alpha)$, celle des ν dans S_q est α . Les deux lois étant tirées indépendamment, la proportion éligible est le produit des deux. Au final, la proportion éligible dans l'espace S_{p+q} est donnée par :

$$\int_{\alpha=0}^1 \alpha^{p-1} (1 - \alpha)^{q-1} d\alpha = \frac{(p-1)!(q-1)!}{(p+q-2)!} \quad (3.2)$$

Les résultats d'éligibilité se résument dans la proposition suivante :

Propriété 6 (Proportion valide).

Si μ est tiré dans le simplexe S_p uniformément, la probabilité que le couple (μ, μ) respecte la condition 2.9 est $\frac{1}{2^{p-1}}$. Dans ce cas, il existe une loi de probabilité r dans le simplexe S_p telle que μ vérifie :

$$\forall u, \mu_u = \frac{1}{2p} + \frac{r_u}{2}. \quad (3.3)$$

Si l'on tire également et indépendamment une loi ν dans le simplexe S_q , la probabilité que le couple (μ, ν) respecte la condition 2.9 est $\frac{(p-1)!(q-1)!}{(p+q-2)!}$. Dans ce cas, il existe un réel α , une loi de probabilité r dans le simplexe S_p et une loi s dans le simplexe S_q tels que :

$$\forall u, \mu_u = \frac{\alpha}{p} + (1 - \alpha)r_u \quad (3.4)$$

$$\forall v, \nu_v = \frac{1 - \alpha}{q} + \alpha s_v \quad (3.5)$$

De surcroît les écritures précédentes caractérisent le respect de la condition.

Remarque 5 (Des formes différentes).

Nous notons que l'expression de la proportion éligible est très variable selon que l'on s'intéresse au couplage de μ avec lui-même ou avec une loi ν indépendante. Ce n'est pas l'écart entre p et q qui le justifie (d'autant moins que la formule couplée ne rattrape pas la première si $p = q$) mais bien le côté indépendant. μ impose à lui-même $\alpha = 1 - \alpha$ générant un seul cas, à rebours de l'intégration donnant la formule générale.

Simulation

La proposition précédente permet de construire des mesures de probabilité valides à condition de savoir construire une mesure de probabilité. Pour cela, nous prenons au hasard $p - 1$ piquets dans $[0, 1]$ que nous trions. Nous rajoutons 0 et 1. Ensuite, nous fixons μ_i comme la distance entre le piquet $i - 1$ et le suivant.

Oublions les formes de la proposition 6 et simulons des μ quelconques pour regarder le nombre à simuler avant d'en trouver un tel que (μ, μ) vérifie la condition (2.9). En

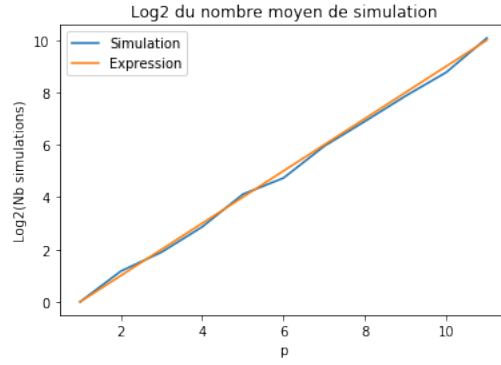


FIGURE 3.1 – Nombre moyen de simulations avant de trouver un μ valide

répétant pour chaque valeur de p , 1000 fois l'opération et en prenant la moyenne, on obtient la courbe suivante :

De la même manière, pour chaque valeur de (p, q) apparaissant dans la figure 3.2, nous simulons 1000 fois des couples (μ, ν) jusqu'à en obtenir un vérifiant la condition 2.9 et notons le nombre moyen nécessaire.

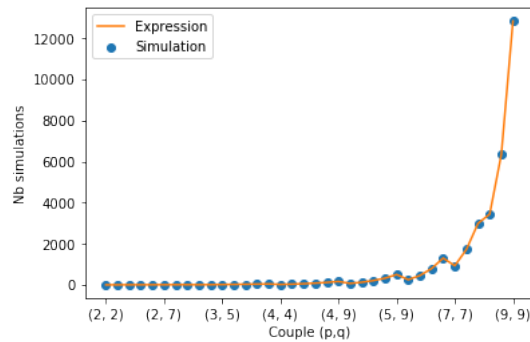


FIGURE 3.2 – Nombre moyen de simulations avant de trouver un (μ, ν) valide

Les deux figures 3.1 et 3.2 confirment en sus de la remarque constructive, les calculs effectués : les couples respectant les conditions nécessaires pour être couplés selon C^+ sont rares. Pour autant, la proposition 6 nous permet de les générer effectivement : toute loi de probabilité génère, une fois fixé la valeur de α , une loi éligible.

Dans le reste de ce chapitre nous nous employons à comprendre ce que signifie la réalisation de ce couplage, lorsque celle-ci a un sens.

3.2 Corrélation et indétermination

Pour commencer, et comme une approche convenue en statistiques face à un couplage inconnu, nous mesurons ici la corrélation associée à un couplage d'indétermination. Il est communément appelé corrélation la valeur de l'indice de Pearson dont nous rappelons la formule dans la définition 8. Un point intéressant est que ce dernier indice est un écart

à l'indépendance. On admet donc implicitement que l'indépendance représente le cadre minimal d'information entre les deux variables. D'autres indices de mesure comme l'indice de Spearman rappelé dans la définition 9 appliquent la même heuristique mais avec une gradation différente (voir par exemple sur le sujet des corrélations le livre [50]).

Appliquer un indice d'écart à l'indépendance revient donc à mélanger deux mondes : l'un basant la corrélation minimale sur l'indépendance et l'autre sur l'indétermination suivant, comme toujours, le choix de la fonction de coût associée au problème d'optimisation (problème 5 ou problème 6). On peut se poser la question de la légitimité d'application d'un critère du premier monde à un couplage du second. Deux arguments nous confortent dans ce projet : premièrement, le critère de Rand est fréquemment appliqué sur un couplage pour estimer si les deux variables sont liées ce qui représente le mélange inverse puisque Rand est, comme nous le verrons, lié à l'indétermination ; secondement, le mélange permet seul d'estimer l'écart.

Définition 8 (Pearson).

Le coefficient de Pearson associé à un tirage $((U_1, V_1), \dots, (U_n, V_n))$ est donné par la formule

$$\rho_P(U, V) = \frac{\sum_{i=1}^n (U_i - \bar{U})(V_i - \bar{V})}{\sqrt{\sum_{i=1}^n (U_i - \bar{U})^2 \sum_{i=1}^n (V_i - \bar{V})^2}}$$

évidemment, il s'agit, sur un tirage, d'une estimation de $\frac{\text{cov}(U, V)}{\sigma_U \sigma_V}$.

Définition 9 (Spearman).

Le coefficient de Spearman associé à un tirage $((U_1, V_1), \dots, (U_n, V_n))$ est donné par la formule

$$\rho_S(U, V) = \frac{\sum_{i=1}^n (rg(U)_i - \overline{rg(U)})(rg(V)_i - \overline{rg(V)})}{\sqrt{\sum_{i=1}^n (rg(U)_i - \overline{rg(U)})^2 \sum_{i=1}^n (rg(V)_i - \overline{rg(V)})^2}}$$

où la fonction de rang $rg(\cdot)$ associe à chaque élément son rang dans le tirage. Evidemment, il s'agit, là encore, sur un tirage, d'une estimation de $\frac{\text{cov}(rg(U), rg(V))}{\sigma_{rg(U)} \sigma_{rg(V)}}$.

Remarque 6.

On constate que spearman est le coefficient de pearson appliqué aux variables de rang.

Dans cette section on s'intéresse à l'application des deux critères de corrélation dans un cadre discret et théorique puis numérique. Afin de pouvoir mener les calculs, le cadre théorique sera le couplage d'indétermination de deux binomiales. Pour le cadre numérique, nous pouvons nous permettre de générer des variables quelconques.

On se donne deux variables binomiales $U \sim B(p)$ et $V \sim B(q)$ que l'on couple selon l'indétermination de sorte que l'on a la formule habituelle :

$$\mathbb{P}^+((U, V) = (a, b)) = \frac{1}{2}\mathbb{P}(U = a) + \frac{1}{2}\mathbb{P}(V = b) - \frac{1}{4}$$

Propriété 7.

La corrélation de Pearson du couplage précédent suit la formule :

$$\rho_P^+(U, V) = \frac{-4pq + 2(p + q) - 1}{4\sqrt{p(p-1)}\sqrt{q(q-1)}} \quad (3.6)$$

avec

$$\frac{1}{2} \leq p + q \leq \frac{3}{2} \quad |p - q| \leq \frac{1}{4}$$

Démonstration.

Tout d'abord, la nécessité d'avoir chacune des valeurs de probabilité positives conduit aux restrictions sur (p, q) proposées dans l'énoncé. Sous ces hypothèses,

$$\begin{aligned} \mathbb{E}_+(UV) &= \frac{p+q}{2} - \frac{1}{4} \\ \mathbb{E}(U)\mathbb{E}(V) &= pq \\ \mathbb{V}(U) &= \sqrt{p(p-1)} \\ \mathbb{V}(V) &= \sqrt{q(q-1)} \end{aligned}$$

Le résultat dérive directement de ces quatre égalités. □

Nous joignons dans la figure suivante le graphe de la corrélation pour constater qu'elle varie entre 0 et -0.3 environ.

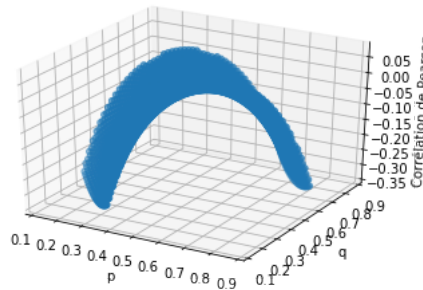


FIGURE 3.3 – Variation de la corrélation de deux binomiales couplées selon l'indétermination

Remarque 7.

Nous savons déjà que lorsque p ou q vaut $\frac{1}{2}$, la corrélation est nulle puisque les deux couplages indépendance et indétermination sont identiques.

Propriété 8.

$$-\frac{1}{3} \leq \rho_P^+(U, V) \leq 0.$$

Démonstration.

La réécriture du numérateur $-(2p-1)(2q-1)$ donne déjà le signe sur le pavé de \mathbb{R}^2 nous intéressant.

On sait de plus que la valeur extrême est à chercher aux limites de l'intervalle.

Les deux conditions sur (p, q) donnent comme valeurs limites $(p, q) = (\frac{1}{4}, \frac{3}{4})$ et $(p, q) = (\frac{3}{4}, \frac{1}{4})$ ce qui fournit une valeur minimale de $-\frac{1}{3}$ pour $\rho_P^+(U, V)$. □

Les calculs de cette section nous permettent de constater que l'indétermination propose toujours une corrélation (dans le sens commun du terme) négative. Et que cette

dernière, proche de 0, ne peut excéder 1/3 en valeur absolue. Cette proximité des deux couplages n'est pas étonnante puisque les deux constituent des méthodes pour aplatir une distribution sur le pavé en respectant les marges. Elle s'exprimera à nouveau dans le chapitre 4 dédié aux application de l'indétermination, typiquement au clustering de graphes ou encore dans le chapitre 5 consacré aux multiples coefficients de corrélation ; le comportement de critères dérivés des deux notions étant similaire.

Passons maintenant à un cadre expérimental et comparons deux distributions afin de voir si l'indétermination se distingue du hasard complet sous les coefficients ρ_P ou ρ_S ; nous comparons :

- la distribution des corrélations d'un couplage selon l'indétermination de marges aléatoires
- la distribution des corrélations d'une distribution totalement aléatoire

Nous suivons plus précisément l'algorithme 1

Algorithm 1 Calculer la distribution de la corrélation sur l'indétermination et le random

Input nombre de tirages R , nombre de points par tirage T , taille des marges (p, q)

Output $(\rho_P^+, \rho_S^+, \rho_P, \rho_S)_{1 \leq \dots \leq R}$

$L_{cor} = []$

for $r = 1 \dots R$ **do**

$\pi \leftarrow$ matrice aléatoire de taille (p, q) sommant à 1 (selon la méthode des piquets consistant à tirer $pq - 1$ points dans $[0,1]$ à les trier, puis à calculer les écarts)

$\mu \leftarrow$ vecteur aléatoire de taille p sommant à 1 (selon la méthode des piquets).

$\nu \leftarrow$ vecteur aléatoire de taille q sommant à 1 (selon la méthode des piquets).

$\pi^+ \leftarrow \mu \oplus \nu$

$(U, V) \leftarrow$ tirage de taille T selon π

$(\rho_P, \rho_S) \leftarrow$ indice de Pearson et de Spearman de (U, V)

$(U^+, V^+) \leftarrow$ tirage de taille T selon π^+

$(\rho_P^+, \rho_S^+) \leftarrow$ indice de Pearson et de Spearman de $(U, V)^+$

$L_{cor}.append((\rho_P^+, \rho_S^+, \rho_P, \rho_S))$

end for

Cet algorithme parcourt donc les valeurs des deux indices de corrélation sous un couplage d'indétermination et pour une distribution quelconque. Les résultats sont affichés dans les figures 3.4 et 3.5. On constate une concentration plus marquée de Pearson sous l'indétermination autour de 0. Ceci est normal et réunit les deux mondes introduits tout au long du document. Certes les valeurs divergent mais elles poursuivent le même but de minimisation de la corrélation ; quelle que soit l'expression de celle-ci.

Un point plus étonnant, justement alors que nous voudrions rapprocher les deux mondes est la distribution du coefficient de Spearman. On n'observe en effet aucune différence flagrante permettant d'identifier un couplage d'indétermination par rapport à une distribution quelconque.

La corrélation entre deux variables couplées selon l'indétermination est donc généralement assez faible, a minima si l'on se base sur le critère de Pearson. Cela rejoint l'approche générale de comparaison entre les deux couplages : ils sont proches tout en ayant des propriétés similaires. Se cantonner à la corrélation ou aux approches classiques en statistiques ne nous permettrait pas de cerner l'indétermination. Certes nous pourrions conclure que

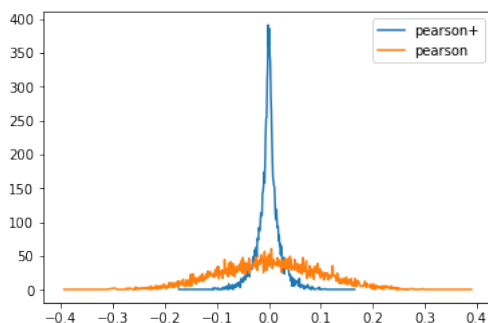


FIGURE 3.4 – Distribution du coefficient de Pearson, sous l'indétermination et sans couplage particulier

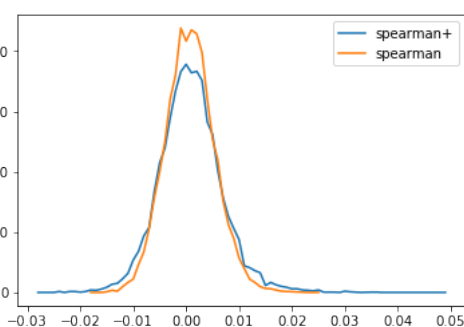


FIGURE 3.5 – Distribution du coefficient de Spearman, sous l'indétermination et sans couplage particulier

ce couplage est proche de l'indépendance mais omettrions des propriétés particulières. Pour faire émerger ces dernières, il nous faut utiliser les notations de l'analyse relationnelle comme proposé dans la section 3.4. Déjà dans la section qui suit, nous remarquons une propriété structurelle de ce couplage, propriété qui justifiera la notation \oplus qui lui est associée.

3.3 Conditions de Monge discrètes

Nous suivons les traces du célèbre mémoire de Monge [73] dont nous rapportons ici une figure particulière 3.6. A deux reprises, dans la partie centrale et dans une partie excentrée (en haut à gauche), un schéma à quatre points avec des droites apparaît. Rappelons que dans le problème de Monge initial 1, un point d'origine correspond à une unique destination. C'est ce qui est ici représenté par les droites : A est transporté en a ou en b . C'est évidemment la droite pleine qu'il faut considérer comme précisé par Monge dans cet énoncé : *Les routes suivies par deux points quelconques ne peuvent se rencontrer entre leurs extrémités*. Cette idée l'amène à introduire ce que nous appelons maintenant les matrices de Monge qui encodent la propriété énoncée.

Ce chapitre est l'occasion d'étudier les liens entre les matrices de Monge et les deux couplages introduits dans le chapitre précédent. Pour ce faire, nous introduisons ici les

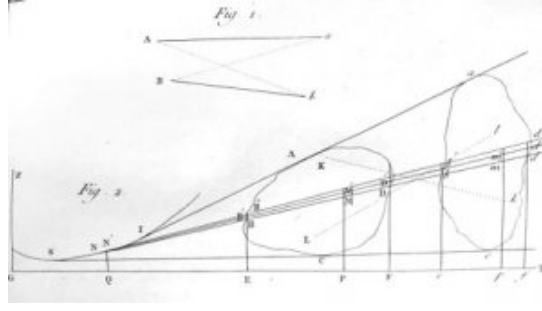


FIGURE 3.6 – Figure introductive historique des matrices de Monge

conditions de Monge et de Log-Monge sur des matrices quelconques. Il est intéressant de remarquer que l'ordre naturel est ici inversé : l'indétermination est retrouvée avant l'indépendance comme réalisation des matrices de Monge originelles. C'est que, du point de vue des matrices de Monge, l'indétermination apparaît plus simplement : l'indépendance requiert en effet l'utilisation supplémentaire du logarithme comme nous le précisons par la suite.

3.3.1 Monge - indétermination

Définition

Pour définir les propriétés de Monge, nous suivons le travail très complet effectué par Rainer Burkard, Bettina Klinz et Rüdiger Rudolf et regroupé dans l'article [19] de 66 pages. Ils étudient la propriété des matrices *de Monge*, d'abord observée par Monge puis redécouverte par Alan Hoffman dans [49] qui lui donna son nom.

Définition 10 (Matrices de Monge et Anti-Monge).

Une matrice de pq nombres réels $c_{u,v}$ est dite de Monge si elle satisfait la propriété :

$$c_{u,v} + c_{u',v'} \leq c_{u',v} + c_{u,v'} \quad \forall 1 \leq u \leq u' \leq p, 1 \leq v \leq v' \leq q$$

Réciproquement, elle est dite Anti-Monge si :

$$c_{u,v} + c_{u',v'} \geq c_{u',v} + c_{u,v'} \quad \forall 1 \leq u \leq u' \leq p, 1 \leq v \leq v' \leq q$$

La vérification des deux propriétés, lorsqu'elle survient, donne lieu à une égalité.

$$c_{u,v} + c_{u',v'} = c_{u',v} + c_{u,v'} \quad \forall 1 \leq u \leq u' \leq p, 1 \leq v \leq v' \leq q$$

Cette dernière propriété (l'égalité) est très particulière et intéressante. On verra qu'elle correspond au cas d'indétermination ou d'indécision qui joue un rôle important dans la théorie du vote de Condorcet.

Cellules adjacentes

En réalité, on se rend compte que la propriété 10 est vérifiée si et seulement si la propriété est vraie sur les cellules adjacentes :

$$c_{u,v} + c_{u+1,v+1} \leq c_{u+1,v} + c_{u,v+1} \quad \forall 1 \leq u \leq p, 1 \leq v \leq q \quad (3.7)$$

Génération cumulative

Les matrices de Monge peuvent être générées de façon cumulative. On se donne une matrice $\pi_{i,j}$ de nombres positifs d'ordre (p, q) qui vérifie les conditions de Monge. Alors, on montre facilement que la matrice suivante les vérifie aussi.

$$\Pi_{u,v} = \sum_{i=1}^u \sum_{j=1}^v \pi_{i,j} \quad \forall 1 \leq u \leq p, 1 \leq v \leq q \quad (3.8)$$

Sous ce formalisme, il est évident que les matrices π et Π traduisent les notions de fonction de distribution et de densité dans le cas discret.

Digression historique

Le type d'approche proposé par la formule 3.8 et son lien avec la formule 3.7 sont déjà présents dans le papier de 1951 de Maurice Fréchet [38] [page 8, formules (12), (13) et (14)]. Détaillons ici son approche.

Il commence par définir la fonction $\Pi_{u,v}$ comme distribution du couple (A, B) (probabilité d'avoir $A \leq u, B \leq v$) puis il pose :

$$\Pi_{u,v} = \frac{1}{N} \sum_{s=1}^u \sum_{t=1}^v n_{s,t} = \sum_{s=1}^u \sum_{t=1}^v \pi_{s,t} \quad (3.9)$$

où les $n_{s,t}$ sont des cellules d'une table de contingence remplie d'entiers positifs. Il remarque alors l'égalité

$$n_{u,v} = N[\Pi_{u-1,v-1} - \Pi_{u-1,v} - \Pi_{u,v-1} + \Pi_{u,v}] \quad (3.10)$$

A ce point, M. Fréchet est dans le contexte exact de la propriété 3.7 puisqu'il traite des cellules adjacentes. Par l'équation 3.9, on voit qu'il suit la propriété 3.8. Ainsi, son $\Pi_{u,v}$ est une matrice Anti-Monge puisque les $n_{u,v}$ sont positifs. Dans son papier et en raisonnant simplement sur l'équation 3.10, M. Fréchet a pu déduire ses fameuses bornes de Fréchet et introduire la notion de copule que nous aurons l'occasion d'utiliser ultérieurement.

3.3.2 Caractérisation structurelle de l'indétermination

Nous commençons par montrer une égalité caractérisant les matrices respectant la propriété 10.

Lemme 1.

Si une matrice respecte le cas d'égalité de la condition de Monge alors on a (voir [64]) l'égalité suivante.

$$\sum_{u,u'=1}^p \sum_{v,v'=1}^q [c_{u,v} - c_{u',v} - c_{u,v'} + c_{u',v'}]^2 = 4pq \left[\sum_{u=1}^p \sum_{v=1}^q c_{u,v}^2 - \sum_{u=1}^p \frac{c_{u,\cdot}^2}{q} - \sum_{v=1}^q \frac{c_{\cdot,v}^2}{p} + \frac{c_{\cdot,\cdot}^2}{pq} \right] = 0$$

où

$$c_{u,\cdot} = \sum_{v=1}^q c_{u,v}$$

$$c_{\cdot,v} = \sum_{u=1}^p c_{u,v}$$

$$c_{\cdot,\cdot} = \sum_{u=1}^p \sum_{v=1}^q c_{u,v}$$

Cette égalité est fort utile et permet de démontrer facilement la propriété suivante :

Propriété 9 (Monge - Indétermination).

Une matrice de probabilité de Monge est forcément issue d'un couplage sous la forme d'indétermination.

Démonstration.

Dans le cas d'une matrice de probabilité on ajoute les contraintes habituelles présentes par exemple dans le problème 4. Une telle matrice de Monge $\pi_{u,v}$ satisfait toujours les hypothèse du Lemme 1 dont la conclusion se réécrit :

$$\sum_{u=1}^p \sum_{v=1}^q \pi_{u,v}^2 - \sum_{u=1}^p \frac{\mu_u^2}{q} - \sum_{v=1}^q \frac{\nu_v^2}{p} + \frac{1}{pq} = 0$$

En suivant le même développement que pour montrer l'égalité 2.11, on obtient :

$$\sum_{u=1}^p \sum_{v=1}^q \left[\pi_{u,v} - \frac{\mu_u}{q} + \frac{\nu_v}{p} - \frac{1}{pq} \right]^2 = \sum_{u=1}^p \sum_{v=1}^q \pi_{u,v}^2 - \sum_{u=1}^p \frac{\mu_u^2}{q} - \sum_{v=1}^q \frac{\nu_v^2}{p} + \frac{1}{pq} = 0 \quad (3.11)$$

La conclusion est alors immédiate :

$$\pi = \pi^+$$

□

Nous proposons ici une seconde preuve, plus simple ne faisant intervenir que la définition 10.

Démonstration.

En sommant sur u' et v' l'égalité de la définition 10 on obtient tout de suite :

$$\sum_{u'} \sum_{v'} c_{u,v} + c_{u',v'} - c_{u',v} - c_{u,v'} = pq c_{u,v} + c_{\cdot,\cdot} - q c_{\cdot,v} - p c_{u,\cdot} = 0;$$

soit exactement

$$c_{u,v} = \frac{c_{u,\cdot}}{q} + \frac{c_{\cdot,v}}{p} - \frac{c_{\cdot,\cdot}}{pq}.$$

□

Nous résumons tous les résultats précédents sur les matrices de Monge dans le théorème 4.

Théorème 4 (Caractérisation des matrices de Monge).

Soit $\pi_{u,v}$ une matrice de probabilités alors les propriétés suivantes sont équivalentes.

1. π est Monge et Anti-Monge.

2. $\sum_{u=1}^p \sum_{v=1}^q \pi_{u,v}^2 = \sum_{u=1}^p \sum_{v=1}^q \left[\frac{\mu_u}{q} + \frac{\nu_v}{p} - \frac{1}{pq} \right]^2$

3. $\pi_{u,v} = C^+(\mu, \nu)_{u,v} = \pi_{u,v}^+ = \frac{\mu_u}{q} + \frac{\nu_v}{p} - \frac{1}{pq}$

4. π minimise le problème 6 pour les marges (μ, ν) .

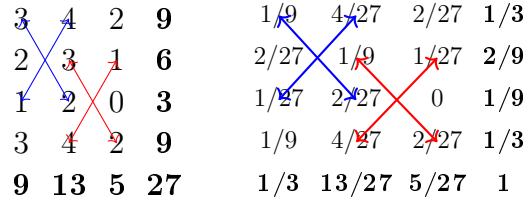


FIGURE 3.7 – Exemple d’un couplage d’indétermination
Forme statistique et probabilité

5. Toutes les sous-tables $\{u, v, u', v'\}$ de taille 2×2 extraites de π ont la même somme sur leur diagonale et leur anti-diagonale.

La dernière propriété du théorème 4 est illustrée par la figure 3.7. En effet, si l’on prend les flèches rouges et bleues on doit aboutir au même résultat : 0. Prenons la forme statistique :

$$\begin{aligned} \text{flèches bleues} & : 3 + 2 - 1 - 4 = 0 \\ \text{flèches rouges} & : 3 + 2 - 4 - 1 = 0 \end{aligned}$$

L’égalité reste aussi vraie dans le cas de la forme probabilité puisque l’on passe à celle-ci en divisant par la somme de la matrice (ici 27).

Introduisons ici un exemple (figure 3.7) que nous utilisons comme étalon des propriétés sur l’indétermination.

Les flèches illustrent la propriété de l’équation 3.7 qui, on le sait, caractérise les matrices de Monge. Cette égalité a déjà été vérifiée à l’instant ; on a pu remarquer qu’il s’agissait d’une vérification additive justifiant la notation + du couplage d’indétermination. Nous allons maintenant constater que la même équation avec un signe \times caractérise l’indépendance et justifie cette fois le second signe utilisé. Pour y parvenir, introduisons les matrices de Log-Monge.

3.3.3 Log Monge - indépendance

Nous présentons ici une propriété semblable aux conditions de Monge : les conditions de Log-Monge qui peuvent d’ailleurs se traduire comme le respect des conditions originelles de Monge pour une matrice Logarithme à définir. Cette condition ne peut concerner que des matrices aux coefficients positifs. Étant donné nos applications (les matrices de probabilité), l’hypothèse revient à dire que la distribution atteint tout l’espace.

Définition

Définition 11 (Matrices Log-Monge et Anti-Log-Monge).

Une matrice pq de nombres réels strictement positifs $c_{u,v}$ est dite Log-Monge si elle satisfait la condition suivante :

$$\ln(c_{u,v}) + \ln(c_{u',v'}) \leq \ln(c_{u',v}) + \ln(c_{u,v'}) \quad \forall 1 \leq u \leq u' \leq p, 1 \leq v \leq v' \leq q$$

3	13/3	5/3	9	1/9	13/81	5/81	1/3
2	26/9	10/9	6	2/27	26/243	10/243	2/9
1	13/9	5/9	3	1/27	13/243	5/243	1/9
3	13/3	5/3	9	1/9	13/81	5/81	1/3
9	13	5	27	1/3	13/27	5/27	1

FIGURE 3.8 – Exemple d'un couplage d'indépendance
Forme statistique et probabilité

De manière symétrique elle est dite *Anti-Log-Monge* si ses coefficients vérifient l'inégalité :

$$\ln(c_{u,v}) + \ln(c_{u',v'}) \geq \ln(c_{u',v}) + \ln(c_{u,v'}) \quad \forall 1 \leq u \leq u' \leq p, 1 \leq v \leq v' \leq q$$

Le cas d'égalité sera appelé *Log-Monge* pour faciliter la lecture et conduire le travail symétrique à celui mené dans le cas des matrices de Monge de la définition 10.

De Log-Monge à Monge

Remarque 8 (De Log-Monge à Monge).

On vérifie facilement que c satisfait les conditions proposées dans la définition 11 si et seulement si $\ln(c)$ vérifie la condition équivalente (comprendre avec le même sens d'inégalité) dans la définition 10 où le logarithme est pris élément par élément.

3.3.4 Caractérisation structurelle de l'indépendance

Le cas d'égalité donne des résultats très intéressants en tant qu'il traduit la structure dite d'indépendance très utilisée en statistique et probabilités. En utilisant le passage de la remarque 8, on peut facilement montrer le théorème 5 pendant du théorème 4.

Théorème 5 (Caractérisation des matrices Log-Monge).

Soit $\pi_{u,v}$ une matrice de probabilité sans terme nul, alors les propriétés suivantes sont équivalents.

1. $\pi_{u,v}$ est une matrice Log-Monge.
2. $\pi_{u,v} = C^\times(\mu, \nu)_{u,v} = \pi_{u,v}^\times = \mu_u \nu_v$
3. π minimise le problème 5 pour les marges (μ, ν) .
4. Toutes les sous-tables $\{u, v, u', v'\}$ de taille 2 extraites de π ont le même produit sur leur diagonale et leur anti-diagonale.

Introduisons ici, comme dans la partie sur les matrices de Monge, un exemple (figure 3.8) que nous utilisons comme étalon des propriétés sur l'indépendance.

Les valeurs sont ici des fractions car l'idée est de respecter les mêmes marges pour les deux matrices de manière à illustrer les problèmes d'optimisation à marges fixées susmentionnés.

On vérifie immédiatement la dernière propriété du théorème 5 de manière identique au cas Monge :

$$\text{flèches bleues} : 3 * 13/9 - 1 * 13/3 = 0$$

$$\text{flèches rouges} : 1/9 * 5/81 - 1/9 * 5/81 = 0$$

L'égalité reste aussi vraie dans le cas de la forme probabilité puisque l'on passe à celle-ci en divisant par la somme de la matrice (toujours 27) et inversement.

3.3.5 Résumé du parallèle

Nous proposons dans la figure A.1 un résumé de l'obtention et des propriétés des deux couplages.

Les deux couplages s'obtiennent initialement par optimisation d'un problème discret à marges fixées dans lequel le choix de la fonction de coût conduit à l'un ou à l'autre. Nous avons vu que la justification des deux fonctions était semblable de sorte que l'on ne peut a priori justifier l'engouement obtenu par l'indépendance au détriment de l'indétermination.

De même, l'introduction des deux formes de matrices de Monge montre que les propriétés visibles sur un couplage se retrouvent dans l'autre sous une forme semblable. Là encore, l'utilisation prépondérante de l'indépendance n'est pas expliquée.

Ce choix de l'indépendance provient de la facilité de son interprétation ; réaliser un couplage selon l'indépendance est évident et naturel : on l'explique, on le comprend. Réaliser un couplage selon l'indétermination (dont la formule est rappelée dans le théorème 2) n'est proposé dans aucun article. Aussi, nous travaillerons principalement à décrire ce couplage peu connu et à proposer diverses interprétations permettant de mieux l'appréhender.

Par ailleurs, nous notons que beaucoup de critères statistiques mesurant la corrélation entre variables reposent sur un écart à l'indépendance ; nous aurons l'occasion d'en lister quelques-uns par la suite. S'agissant d'un couplage qui garantit une distribution aussi uniforme que possible à marges fixées (en tant que solution du problème 5), c'est une construction légitime. Pour autant, l'indétermination s'exprimant elle aussi comme solution du problème 6 quasi identique, nous pourrions nous attendre, nonobstant sa non-interprétabilité, à voir apparaître des critères d'écart à l'indétermination. Nous montrerons qu'il en existe en effet de nombreux qui, après réécriture, s'écrivent comme tel. Ces réécritures nécessiteront le passage dans l'espace relationnel de nos formules continuelles ; nous introduisons formellement cet espace dans la partie suivante.

3.4 Indétermination et analyse relationnelle mathématique

3.4.1 Introduction

L'analyse relationnelle mathématique¹, notée ARM, repose sur une modélisation particulière des réalisations de variables aléatoires ou variables statistiques catégorielles,

1. Il existe également une *théorie de l'Analyse Relationnelle* dans le domaine des Sciences Humaines, plutôt orientée psychologie, elle a trait à l'étude des structures induites par les relations interpersonnelles

quelle que soit leur valeur. Les variables sont codées selon une relation binaire ne conservant que les relations en oubliant la modalité. Cette méthode permet de comparer deux-à-deux des variables évoluant dans une large variété de données et pour lesquelles les individus qui les réalisent peuvent être :

- les clients dans le marketing ;
- les textes en linguistique ;
- les patients en médecine.

Formellement, on commence par se donner deux variables aléatoires : U sur p modalités, V sur q modalités. Un jeu de données est constitué de la réalisation des deux variables aléatoires sur n individus : $((U_1, V_1), \dots, (U_n, V_n))$. Typiquement U code la ville de résidence et V le dernier achat réalisé sur un site quelconque.

S'agissant de données réelles, souvent des individus, on suppose évidemment que la réalisation de U sur l'individu i influence la valeur de V_i .

L'encodage le plus fréquent dans la littérature scientifique pour résumer un jeu de données est la matrice dite de contingence qui consiste à regrouper les individus ayant le même couple de réalisation et à les compter. Ainsi :

Définition 12 (Encodage contingence).

Donnons-nous (U_1, \dots, U_n) et (V_1, \dots, V_n) deux tirages de taille n de $U \sim \mu$ et $V \sim \nu$ respectivement. Nous définissons une matrice de contingence de taille $p \times q$ (illustrée par la figure 2.1) en posant :

$$n_{u,v} = \#\{i/1 \leq i \leq n \ \& \ U_i = u \ \& \ V_i = v\}$$

Remarque 9 (Notations somme).

Nous ajoutons la modalité "." union des modalités pour que la notation $n_{u,\cdot} = \sum_{v=1}^q n_{u,v}$ soit cohérente, ainsi évidemment que les deux autres notations $n_{\cdot,v}$ et $n_{\cdot,\cdot}$.

Remarque 10.

L'écriture de contingence s'applique également dans le cas d'une seule variable, menant alors à un vecteur. Nous proposons directement l'écriture double car celle-ci nous servira tout au long du document.

L'analyse relationnelle mathématique propose un encodage différent basé uniquement sur la relation entre les individus. Deux individus sont ici dits *en relation* selon s'ils partagent la même modalité. On code alors cette information via la matrice binaire de la Figure 3.1 : la case à l'intersection de la ligne i et de la colonne j vaut 1 si les individus i et j sont en relation (comprendre ont la même modalité) et 0 sinon. Ce codage est résumé par la formule 3.12 :

$$X_{i,j} = \delta_i \text{ est en relation avec } j \tag{3.12}$$

Plus formellement, nous avons la définition 12.

ou sociales entre des individus ou au sein de communautés d'intérêt ; on dirait aujourd'hui "réseau social". C'est la raison pour laquelle nous appelons ici, celle qui nous concerne, Analyse Relationnelle Mathématique des Données. Cela dit, l'Analyse Relationnelle Mathématique des Données s'intéresse bien sûr aux modélisations de relations et donc pourrait être utilisée de façon profitable par la théorie de l'Analyse Relationnelle en Sciences Humaines comme moyen spécifique d'analyse.

	A	B	C	D	E
A	1	1	0	0	0
B	1	1	0	0	0
C	0	0	1	0	0
D	0	0	0	1	1
E	0	0	0	1	1

TABLE 3.1 – Exemple d’un encodage relationnel

	A	B	C	D	E
A	1	1	1	0	0
B	1	1	1	0	0
C	1	1	1	1	1
D	0	0	1	1	1
E	0	0	1	1	1

TABLE 3.2 – Exemple d’un Encodage Relationnel dans le cas d’une modalité multiple

Définition 13 (Encodage relationnel).

Donnons-nous (u_1, \dots, u_n) et (v_1, \dots, v_n) deux tirages de taille n de $U \sim \mu$ et $V \sim \nu$ respectivement. Nous définissons alors deux matrices relationnelles X et Y (une par variable) de taille $n \times n$ par :

$$X_{i,j} = \mathbb{1}_{u_i=u_j}, \quad \forall 1 \leq i, j \leq n$$

$$Y_{i,j} = \mathbb{1}_{v_i=v_j}, \quad \forall 1 \leq i, j \leq n$$

Prenons un premier exemple de codage pour cinq individus $\{A, B, C, D, E\}$ distribués selon trois nationalités. La variable U codant cette nationalité réalise la modalité *français* pour les individus A et B , *espagnol* pour C et *anglais* pour D et E .

La matrice de la figure 3.1 suit la règle juste définie pour l’exemple initial des cinq individus. Cette matrice binaire relationnelle contient la même information que la représentation vectorielle de la variable consistant à décrire la modalité de chaque individu ; de fait, nous proposerons l’algorithme 2 de codage retour. L’avantage est qu’elle est bien plus flexible. Notamment, des situations complexes ou paradoxales comme, dans notre exemple, la multi-nationalité sont plus simples à encoder. Ici, si l’individu C est citoyen des trois pays, nous devons juste rajouter des 1 selon la matrice suivante (3.2) tandis que le codage classique aurait amené à introduire une quatrième modalité pour {français, espagnol, anglais}.

En plus de sa capacité à prendre en compte des relations complexes, ce format peut simplifier des calculs dans des domaines divers : Clustering, Ranking, Théorie du vote, Classification, Modularisation de graphes. Ainsi, des écritures quadratiques dans l’espace contingenciel pourront avoir une traduction linéaire dans l’espace de l’analyse relationnelle ; nous y reviendrons ultérieurement. Enfin, comme toute réécriture, elle permet de voir émerger des dualités et de découvrir des égalités entre structures ; nous aurons également l’occasion d’y revenir.

Remarque 11 (Espace de stockage).

A première vue, le codage relationnel en n^2 où n est le nombre d'individus paraît gourmand en stockage. Plusieurs éléments permettent de relativiser cette assertion.

Tout d'abord, la matrice est creuse puisque composée essentiellement de 0 ce qui permet d'optimiser son stockage.

Surtout, contrairement à l'écriture de contingence, et comme montré dans [27], ce stockage est linéaire dans le nombre de variables ce qui se révèle rapidement intéressant. Supposons par exemple que nous détaillions des caractéristiques d'individus. Prenons leur âge, leur formation, leur sexe, leur pays de résidence, leur taille, leur poids, etc. ... En écriture de contingence, nous allons obtenir un stockage de taille le produit du nombre de modalité des variables. Aussi, si chacune en a 10, l'ordre de stockage sera : 10^{nb} de variables. En écriture relationnelle, chaque variable donne lieu à une matrice de taille $n * n = 10^{20}$ maximum ; difficile en effet de constater plus de réalisations que la taille de la population mondiale. On obtient ainsi un stockage en nb de variables $* 10^{20}$.

Cet exemple montre bien que l'ajout d'une variable se fait sans effort dans le cadre relationnel alors qu'elle coûte de plus en plus cher dans le cadre de l'écriture des contingences.

De façon extrême, comme chaque variable a au moins deux modalités, dès que leur nombre dépasse le log de la taille de la population, le codage relationnel est profitable.

Le modèle relationnel de stockage des données est présenté sous une forme statistique. Pour autant, nous aurons besoin de proposer une construction de ces matrices dans un cadre probabiliste.

On se donne ainsi U variable aléatoire à valeurs dans $\{0, 1\}^p$ et V variable aléatoire à valeurs dans $\{0, 1\}^q$. Elles représentent les cases cochées sur deux QCM avec multi-choix de taille respective p ou q . On suppose :

- $U \sim \mu$
- $V \sim \nu$
- $U_n = (U_1, \dots, U_n)$ iid sous μ
- $V_n = (V_1, \dots, V_n)$ iid sous ν

Définition 14 (Matrices relationnelles - cas probabiliste).

Les matrices relationnelles sont $X_{i,j}^n = 1_{U_i U_j \neq 0}$ et $Y_{i,j}^n = 1_{V_i V_j \neq 0}$. Cela revient à dire que la case (i, j) vaut 1 dès que i et j ont une case cochée en commun.

Pour montrer que le codage relationnel regroupe bien toutes les informations de réalisation, nous proposons un algorithme 2 de retour au vecteur des réalisations U depuis sa matrice relationnelle X . On peut de plus supposer p inconnu, ce nombre de classes visible évoluant alors avec le nombre n de tirages observé.

Les modalités sont alors renommées de 1 à p , ce dernier nombre augmentant au fur et à mesure que l'on en observe de nouvelles. Grâce à l'algorithme 2, nous savons que nous pouvons générer X comme U indifféremment. Il sera intéressant d'étudier la loi image d'une loi proba sur l'autre. Pour ce faire, la diagonalisation de critères de corrélation que nous présenterons dans une autre partie sera nécessaire.

Algorithm 2 Calculer U depuis X

```
(U) ← []  
for i = 0...n do  
  if ∃j < i, Xi,j = 1 then  
    Un ← Uj  
  else  
    Un ← #set(U) + 1  
  end if  
end for
```

3.4.2 Formules de passage

Nous ne reprendrons pas ici l'ensemble des résultats obtenus dans le cadre de l'ARM mais pouvons citer quelques papiers autour du sujet et notamment [68], [64], [71], [67], [66], [3] et [2].

Parmi les résultats, nous sélectionnons une base de formules de passage que nous utiliserons beaucoup par la suite. Nous pouvons les regrouper sous le terme *principe de transfert relationnel* car elles participent grandement à la compréhension des propriétés générées par les fonctions de coût déjà introduites. Ce principe mentionné initialement par Sir M.G. Kendall en 1948 pour des corrélations sur des modèles de ranking dans le papier [58] a été généralisé à l'ensemble des relations par J.-F. Marcotorchino en 1984 (voir [64]) pour s'appliquer à tout type de relation binaire. Il s'agit de linéariser des expressions au moins quadratiques de l'espace de contingence traditionnel dans un sous espace des relations binaires : l'espace relationnel tout juste introduit.

Nous donnons ici les exemples les plus utiles pour la suite de nos travaux :

Propriété 10 (Principe de transfert relationnel).

On se donne U et V deux variables catégorielles de dimensions respectives p et q dont la pq matrice de contingence sur les N individus est représentée comme dans la Figure 2.1 par son terme général $n_{u,v}$ tandis que leur encodage relationnel est respectivement de terme général binaire $x_{i,j}$ et $y_{i,j}$ regroupés dans deux matrices semblables à celle de la figure 3.1. Les formules suivantes sont alors vérifiées et permettent de passer d'une représentation à l'autre.

1. $\sum_{u=1}^p \sum_{v=1}^q n_{u,v}^2 = \sum_{i=1}^N \sum_{j=1}^N x_{i,j} y_{i,j}$
2. $\sum_{u=1}^p n_{u,\cdot}^2 = \sum_{i=1}^N \sum_{j=1}^N x_{i,j}$
3. $\sum_{v=1}^q n_{\cdot,v}^2 = \sum_{i=1}^N \sum_{j=1}^N y_{i,j}$
4. $\sum_{u=1}^p \sum_{v=1}^q \frac{n_{u,v}^2}{n_{u,\cdot} n_{\cdot,v}} = \sum_{i=1}^N \sum_{j=1}^N \frac{x_{i,j} y_{i,j}}{x_{i,\cdot} y_{\cdot,j}}$
5. $\sum_{u=1}^p \sum_{v=1}^q n_{u,v} n_{u,\cdot} n_{\cdot,v} = \sum_{i=1}^N \sum_{j=1}^N \frac{x_{i,\cdot} + x_{\cdot,j}}{2} y_{i,j} = \sum_{i=1}^N \sum_{j=1}^N \frac{y_{i,\cdot} + y_{\cdot,j}}{2} x_{i,j}$

Le principe 10 ci-dessus permet de linéariser les expressions de contingence quadratiques et difficiles à optimiser apparaissant dans les problèmes des chapitres précédents. Une fois sous leur forme relationnelle, la fonction de coût apparaît alors souvent comme un produit scalaire ; nous aurons l'occasion de le constater par la suite.

3.4.3 Codage linéaire du type de relation

Les outils de traduction présentés dans la proposition 10 sont utilisés pour réécrire les fonctions de coûts portant sur les $n_{u,v}$ en fonction des $x_{i,j}$. De surcroît, il nous faut transformer les contraintes afin de les exprimer dans ces nouvelles variables. Il est aisé de noter que le codage relationnel peut s'appliquer à toute relation binaire entre éléments d'un ensemble quelconque. Les propriétés de ladite relation doivent alors elles-mêmes être traduites dans cet espace. A titre d'exemple, nous introduisons deux ensembles de contraintes pour deux types de relations :

— X_{order} une relation d'ordre total qui est donc Réflexive, Antisymétrique, Complète et Transitive

— X_{equiv} une relation d'équivalence qui est donc Réflexive, Symétrique et Transitive

Les propriétés de ces deux relations peuvent toutes être codées linéairement dans l'espace relationnel comme le montrent les deux propositions suivantes.

Propriété 11 (Codage des propriétés de X_{order}).

$$\begin{array}{ll} x_{i,j} \in \{0,1\} & \text{Binaire} \\ x_{i,i} = 1, \forall i & \text{Réflexive} \\ x_{i,j} + x_{j,i} = 1 \forall (i,j) & \text{Antisymétrique et Complète} \\ x_{i,j} + x_{j,k} - x_{i,k} \leq 1 \forall (i,j,k) & \text{Transitive} \end{array}$$

Remarque 12 (Interprétation de la relation d'ordre).

Dans le cas d'une relation d'ordre, les termes $\{x_{i,j}\}$ où $x_{i,j} = 1$ peuvent se comprendre comme i est classé avant j ou bien i est préféré à j . Cette dernière interprétation est particulièrement intéressante lors de l'application à la théorie du vote.

Propriété 12 (Codage des propriétés de X_{equiv}).

$$\begin{array}{ll} x_{i,j} \in \{0,1\} & \text{Binaire} \\ x_{i,i} = 1, \forall i & \text{Réflexive} \\ x_{i,j} - x_{j,i} = 0 \forall (i,j) & \text{Symétrique} \\ x_{i,j} + x_{j,k} - x_{i,k} \leq 1 \forall (i,j,k) & \text{Transitive} \end{array}$$

Remarque 13 (Interprétation de la relation d'équivalence).

Dans le cas d'une relation d'équivalence, les termes $\{x_{i,j}\}$ où $x_{i,j} = 1$ peuvent signifier que i est similaire à j ou encore que les deux sont dans la même classe ; l'application canonique est celle, susmentionnée, de réencodage d'une matrice de contingence.

On remarque que l'on peut ici facilement sélectionner des propriétés sur un codage et découvrir ainsi des relations composant un nombre arbitraire de ces équations linéaires.

3.4.4 Interprétation relationnelle de l'indétermination

Une des notions clefs introduites dans les paragraphes précédents est le couplage d'indétermination. Comme mentionné, son étude dans la littérature souffre de quelques

	$n_{1,1}$	$n_{1,2}$...	$n_{1,q}$	$n_{2,1}$...	$n_{2,q}$...	$n_{p,q}$
$n_{1,1}$	11	10	...	10	01	...	00	...	00
$n_{1,2}$	10	11	...	10	00	...	00	...	00
\vdots
$n_{1,q}$	10	10	...	11	00	...	01	...	01
$n_{2,1}$	01	00	...	00	11	...	10	...	00
\vdots
$n_{2,q}$	00	00	...	01	10	...	11	...	01
\vdots
$n_{p,q}$	00	00	...	01	00	...	01	...	11

TABLE 3.3 – Encodage relationnel simultané de deux variables

lacunes. En particulier, on ne connaît pas d'expérience menant à un tel couplage. A la lumière de l'analyse relationnelle, on peut clarifier ce concept et justifier son nom au passage.

Nous partons de la formule 3.11 qui traduit l'indétermination introduite formellement et s'écrit :

$$pq \sum_{u=1}^p \sum_{v=1}^q \pi_{u,v}^2 - p \sum_{u=1}^p \mu_u^2 - q \sum_{v=1}^q \nu_v^2 + 1 = 0$$

En utilisant le principe de transfert 10, cette expression se transforme en :

$$pq \sum_{i=1}^p \sum_{j=1}^q x_{i,j} y_{i,j} - p \sum_{i=1}^p \sum_{j=1}^q x_{i,j} - q \sum_{i=1}^p \sum_{j=1}^q y_{i,j} + \sum_{i=1}^p \sum_{j=1}^q 1 = 0 \quad (3.13)$$

Nous voyons naturellement apparaître une expression relationnelle de l'indétermination et allons l'utiliser pour mieux la cerner. Avant de continuer, il nous faut introduire un artefact permettant de représenter simultanément les codages relationnels de deux relations différentes.

Tout d'abord, nous rappelons qu'un codage relationnel sur N individus est une matrice binaire $N \times N$, X (selon la formule 3.12 et la Figure 3.1). Ainsi, quelle que soit la dimension (classiquement p et q) des variables U et V , les deux codages relationnels auront la même taille. On les couple alors selon l'illustration 3.3 et la définition 15.

Définition 15 (Couplage relationnel).

On couple deux matrices relationnelles X et Y sur N individus par la matrice Z suivante revenant à concaténer les valeurs de X et Y .

$$\begin{cases} z_{i,j} = 10 & \text{si } (i,j) \in \mathcal{C}_{uv} \times \mathcal{C}_{uv'} \\ z_{i,j} = 01 & \text{si } (i,j) \in \mathcal{C}_{uv} \times \mathcal{C}_{u'v} \\ z_{i,j} = 11 & \text{si } (i,j) \in \mathcal{C}_{uv} \times \mathcal{C}_{uv} \\ z_{i,j} = 00 & \text{si } (i,j) \in \mathcal{C}_{uv} \times \mathcal{C}_{u'v'} \end{cases}$$

La matrice Z ainsi définie encode sur une seule table, les deux matrices relationnelles. S'ensuit une définition d'un accord/désaccord entre U et V sur chaque couple d'individu

(i, j) : au croisement (i, j) , si z est de la forme $\epsilon\epsilon$, c'est un accord. Si au contraire il est de la forme $\epsilon(1 - \epsilon)$, c'est un désaccord.

On calcule alors les quantités suivantes représentant les accords et désaccords entre X et Y (ou bien entre U et V) :

$$Q_{\epsilon_x \epsilon_y} = \sum_{i=1}^N \sum_{j=1}^N \delta_{\epsilon_x, x_{i,j}} \delta_{\epsilon_y, y_{i,j}} \quad \epsilon_x \in \{0, 1\} \text{ et } \epsilon_y \in \{0, 1\}$$

Ces quantités comptent les couples d'individus pour lesquels $x_{i,j} = \epsilon_x$ et $y_{i,j} = \epsilon_y$. Ainsi $A_1 = Q_{1,1}$ et $A_0 = Q_{0,0}$ représentent les situations où X et Y sont d'accord alors que $Q_{0,1}$ et $Q_{1,0}$ comptent les configurations où X et Y sont en désaccord.

Le nombre d'accords et de désaccords possibles est loin d'être identique selon les valeurs des ϵ .

Nous avons en effet les pondérations :

- Le nombre d'accords de type (1,1) (comprendre suivant la définition 15 que les deux individus i et j sont dans la même classe pour les deux variables U et V) est pq .
- Le nombre d'accords de type (0,0) (les deux individus sont dans des classes différentes pour les deux variables) est $pq(p-1)(q-1)$.
- Le nombre de désaccords de type (0,1) (les deux individus sont dans des classes différentes pour la première variable et dans la même classe pour la seconde variable) est $pq(p-1)$.
- Le nombre de désaccords de type (1,0) (les deux individus sont dans la même classe pour la première variable et dans des classes différentes pour la seconde variable) est $pq(q-1)$.

Avec une somme à p^2q^2 , ces pondérations représentent tous les cas de la figure 3.3 et conduisent à poser le critère d'indétermination suivant que nous allons montrer égal à l'indétermination déjà rencontrée.

Définition 16 (Indétermination : égalité des accords et désaccords).

On considère deux matrices relationnelles de taille $N \times N$ représentant deux variables catégorielles U et V . Elles sont dites en couplage d'indétermination si la valeur de leurs accords pondérés vaut celle de leurs désaccords pondérés selon la formule :

$$\frac{A_1}{pq} + \frac{A_0}{pq(p-1)(q-1)} = \frac{Q_{0,1}}{pq(q-1)} + \frac{Q_{1,0}}{pq(p-1)} \quad (3.14)$$

Montrons maintenant l'égalité des deux définitions d'indétermination.

Propriété 13 (Indétermination égalité).

L'indétermination de la définition 16 est celle précédemment introduite dans la définition 3

Démonstration. Le travail est presque déjà entièrement fait. Nous traduisons la nouvelle égalité "accords-désaccords" multipliée par $p(p-1)q(q-1)$:

$$(p-1)(q-1)A_1 + A_0 = (p-1)Q_{0,1} + (q-1)Q_{1,0}$$

Si l'on remplace alors les quantités par leurs valeurs relationnelles, il vient :

$$(p-1)(q-1) \sum_{i=1}^N \sum_{j=1}^N x_{i,j} y_{i,j} + \sum_{i=1}^N \sum_{j=1}^N (1-x_{i,j})(1-y_{i,j}) - (p-1) \sum_{i=1}^N \sum_{j=1}^N x_{i,j}(1-y_{i,j}) - (q-1) \sum_{i=1}^N \sum_{j=1}^N (1-x_{i,j})y_{i,j} = 0$$

En développant et simplifiant, on trouve l'égalité 3.13 à laquelle nous nous étions arrêtés avant d'introduire le couplage relationnel. Nous retrouvons donc bien l'indétermination introduite en premier lieu selon la deuxième caractérisation du Théorème 4. \square

3.4.5 Égalité accords désaccords probabiliste

Nous proposons ici de reprendre l'égalité classique de la proposition 16 dans un cadre probabiliste. En effet, cette égalité s'écrit dans un contexte de n tirages, contexte dans lequel on peut définir une indétermination statistique ; si celle-ci a lieu, la matrice de contingence associée divisée par le nombre de tirage fournit évidemment une loi de probabilité qui a le bon goût de vérifier l'indétermination probabiliste.

Reste que, la réciproque est fautive : si une probabilité π couple ses marges selon l'indétermination, un tirage de n valeurs selon π a peu de chances de réaliser l'indétermination statistique (un tirage illustre rarement parfaitement les probabilités sous-jacentes).

Aussi, nous souhaitons définir une égalité "accords-désaccords" à tester sur une loi de probabilité et non plus sur un ensemble de tirages. On s'inspire pour cela largement de l'égalité de la proposition 16, égalité qui, nous l'avons vu dans ladite proposition caractérise l'indétermination statistique :

$$\frac{XY}{pq} + \frac{\bar{X}\bar{Y}}{p(p-1)q(q-1)} = \frac{X\bar{Y}}{pq(q-1)} + \frac{\bar{X}Y}{p(p-1)q}$$

Ainsi, on se donne deux tirages indépendants (u_i, v_i) et (u_j, v_j) selon π et l'on nomme (comme de manière générale) *accord* lorsque les deux marges sont égales ou différentes ensemble ; le cas contraire étant un *désaccord*. On peut définir l'espérance de leur différence pondérée :

$$\frac{\mathbb{E}_{\pi \otimes \pi}(X_{i,j}Y_{i,j})}{pq} + \frac{\mathbb{E}_{\pi \otimes \pi}(\bar{X}_{i,j}\bar{Y}_{i,j})}{p(p-1)q(q-1)} - \frac{\mathbb{E}_{\pi \otimes \pi}(X_{i,j}\bar{Y}_{i,j})}{pq(q-1)} - \frac{\mathbb{E}_{\pi \otimes \pi}(\bar{X}_{i,j}Y_{i,j})}{p(p-1)q} \quad (3.15)$$

On constate alors dans la proposition suivante que cette différence est nulle si et seulement si la probabilité sous-jacente couple ses marges selon l'indétermination.

Théorème 6 (Interprétation de π^+).

π , loi de probabilité sur pq variables catégorielles suit le couplage par indétermination de ses marges si et seulement l'équation 3.15 vaut 0.

Démonstration.

Soit donc π une loi de probabilité sur pq variables catégorielles ; elle est définie par ses

valeurs $\pi_{u,v}$, $1 \leq u \leq p$ et $1 \leq v \leq q$. On tire les valeurs de deux individus indépendants selon $\pi : (u_i, v_i)$ et (u_j, v_j) .

Regardons maintenant les accords et désaccords avec les notations de l'analyse relationnelle : X, Y sont deux matrices symétriques de taille $2 * 2$ (il y a deux individus) valant 1 sur la diagonale et de valeur $X_{i,j} = X_{j,i} = \mathbb{1}_{u_i=u_j}$ et $Y_{i,j} = Y_{j,i} = \mathbb{1}_{v_i=v_j}$. X et Y représentent respectivement les égalités de tirage entre individus selon la variable de marges U et selon l'autre marge V .

Nous allons maintenant voir que l'équation 3.15 vaut 0 précisément lorsque π est égale au couplage de ses marges selon la formule d'indétermination :

$$\pi_{u,v} = \frac{\pi_{u,\cdot}}{q} + \frac{\pi_{\cdot,v}}{p} - \frac{1}{pq} \quad (3.16)$$

On calcule en effet si l'on tire deux individus selon π indépendamment.

$$\begin{aligned} - \mathbb{E}_{\pi \otimes \pi}(X_{i,j} Y_{i,j}) &= \sum_{u_i, v_i} \sum_{u_j, v_j} \pi_{u_i, v_i} \pi_{u_j, v_j} \mathbb{1}_{u_i=u_j \& v_j=v_j} = \sum_{u,v} \pi_{u,v}^2 \\ - \mathbb{E}_{\pi \otimes \pi}(\overline{X_{i,j} Y_{i,j}}) &= \sum_{u_i, v_i} \sum_{u_j, v_j} \pi_{u_i, v_i} \pi_{u_j, v_j} \mathbb{1}_{u_i \neq u_j \& v_i \neq v_j} = \sum_{u,v} \pi_{u,v} (1 - \pi_{u,\cdot} - \pi_{\cdot,v} + \pi_{u,v}) \\ - \mathbb{E}_{\pi \otimes \pi}(X_{i,j} \overline{Y_{i,j}}) &= \sum_{u_i, v_i} \sum_{u_j, v_j} \pi_{u_i, v_i} \pi_{u_j, v_j} \mathbb{1}_{u_i=u_j \& v_i \neq v_j} = \sum_{u,v} \pi_{u,v} (\pi_{u,\cdot} - \pi_{u,v}) \end{aligned}$$

En remplaçant dans l'équation 3.15, on obtient :

$$\begin{aligned} &\frac{\sum_{u,v} \pi_{u,v}^2}{pq} + \frac{\sum_{u,v} \pi_{u,v} (1 - \pi_{u,\cdot} - \pi_{\cdot,v} + \pi_{u,v})}{p(p-1)q(q-1)} \\ &= \frac{\sum_{u,v} \pi_{u,v} (\pi_{u,\cdot} - \pi_{u,v})}{pq(q-1)} + \frac{\sum_{u,v} \pi_{u,v} (\pi_{\cdot,v} - \pi_{u,v})}{p(p-1)q} \end{aligned}$$

En réduisant au même dénominateur il vient :

$$\begin{aligned} \iff &(p-1)(q-1) \sum_{u,v} \pi_{u,v}^2 + \sum_{u,v} \pi_{u,v} (1 - \pi_{u,\cdot} - \pi_{\cdot,v} + \pi_{u,v}) \\ &= (p-1) \sum_{u,v} \pi_{u,v} (\pi_{u,\cdot} - \pi_{u,v}) + (q-1) \sum_{u,v} \pi_{u,v} (\pi_{\cdot,v} - \pi_{u,v}) \end{aligned}$$

En regroupant :

$$\iff pq \sum_{u,v} \pi_{u,v}^2 - p \sum_u \pi_{u,\cdot}^2 - q \sum_v \pi_{\cdot,v}^2 + 1 = 0$$

Soit finalement en utilisant la deuxième caractérisation du théorème 4 :

$$\iff \pi_{u,v} = \frac{\pi_{u,\cdot}}{q} + \frac{\pi_{\cdot,v}}{p} - \frac{1}{pq}$$

On a donc bien montré que $\pi = \pi^+$ équivaut à une égalité de l'espérance des *accords* et *désaccords* normalisés sur deux tirages indépendants. \square

Ce déroulement montre que l'indétermination se caractérise aussi dans un cadre probabiliste par une égalité dont l'écriture et *a fortiori* l'interprétation n'est permise que dans le cadre relationnel. Nous aurons l'occasion de justifier les noms *accords* et *désaccords* dans la sous-section 3.4.6. Pour l'instant, levons une ambiguïté entre les deux formes statistique et probabiliste de notre égalité.

De fait, le théorème 6 en regard de la définition 16 paraît étonnant : si tous les $X_{i,j}$ sont d'espérance nulle, comme $X_{i,i} = 1$, la définition 16 du cadre statistique ne peut être vérifiée. C'est qu'un tirage à n éléments comporte des liens non mentionnés ici, ces mêmes liens, remarqués dans le chapitre 5 et qui nous amèneront à introduire des permutations générant X : les $X_{i,j}$ sont tout sauf indépendants. Ainsi, si pris séparément les $X_{i,j}$ sont d'espérance nulle, la matrice X n'a pas pour espérance l'identité.

3.4.6 Exemple de jugements

Dans cette sous-section, nous repassons justement dans le cadre statistique et appliquons le formalisme de l'analyse relationnelle aux décisions judiciaires. Chaque individu est de fait condamné ou non (pour un vol par exemple) (première variable) tandis qu'il est coupable ou non (seconde variable).

On dispose ainsi d'une distribution μ sur la première variable et ν sur la seconde. Lors du couplage sur un phénomène comme celui-ci, on attribue naturellement à chacun des cas une valeur morale :

- 00 : relâcher un innocent, parmi les voix pour ;
- 01 : relâcher un coupable, parmi les voix contre ;
- 10 : condamner un innocent, parmi les voix contre ;
- 11 : condamner un coupable, parmi les voix pour.

L'optimisation d'un type de voix contre se fait toujours au détriment du second type, c'est le type de société qui place le niveau de tolérance. Quel que soit ce seuil le pire cas que nous puissions obtenir est d'avoir au sein de la population autant d'individus "pour" que "contre" (si l'on passe le cas d'égalité, renverser toutes les décisions de condamnation permet immédiatement de faire mieux que le cas d'égalité).

Ce dernier correspond au cas où l'on a $00 + 11 = 10 + 01$ soit en probabilités $\pi_{0,0} + \pi_{1,1} = \pi_{0,1} + \pi_{1,0}$. En utilisant la caractérisation des matrices de Monge donnée par 4, on voit immédiatement que cela correspond au cas d'indétermination.

C'est justement de cette propriété caractéristique des matrices de Monge et donc du couplage que nous étudions que nous partirons dans la section suivante pour révéler un procédé constructif.

3.5 Tirage selon l'indétermination

3.5.1 Procédé constructif

Le couplage C^+ n'étant pas l'indépendance, il induit par définition une dépendance entre les deux marges. Nous détaillons celle-ci dans la présente section. Notamment, nous exprimons sa capacité à réduire les collisions, notion que nous définirons précisément mais qui exprime essentiellement l'égalité entre deux tirages successifs.

Nous utilisons la notion de matrice de Monge qui caractérise l'*indétermination* elle-même associée à la projection \mathbb{L}^2 et qui peut s'écrire :

$$\pi_{u,v}^+ - \pi_{u',v}^+ = \pi_{u,v'}^+ - \pi_{u',v'}^+, \forall 1 \leq u \leq p, \forall 1 \leq v \leq q \quad (3.17)$$

Cette dernière écriture exprime une différence constante entre deux lignes, quelle

que soit la colonne sur laquelle on l'observe. De cette remarque nous déduisons une décomposition valable pour toute matrice d'indétermination π .

Nous supposons que la marge définie pour tout u , $1 \leq u \leq p$ par $\mu_u = \pi_{u,\cdot}$ est minimale en $u = 1$. Cela simplifie les notations.

En ce cas l'écart observé dans l'équation 3.17 est toujours positif entre $u = 1$ et les autres lignes. Formellement pour tout (u, v) , nous posons :

$$\delta_u = \pi_{u,v}^+ - \pi_{1,v}^+ = \frac{\mu_u - \mu_1}{p} \geq 0 \quad (3.18)$$

avec δ_u indépendant de v .

Une matrice d'indétermination π^+ se décompose ainsi en une distribution arbitraire sur q colonnes $\pi_{1,v}^+$, $\forall 1 \leq v \leq q$ et $(p-1)$ réels positifs δ_u chacun égal à $\frac{\mu_u - \mu_1}{p}$.

Réciproquement, nous fournissons ci-dessous une méthode pour construire une matrice d'indétermination.

1. Pour la première ligne, nous fixons une distribution arbitraire sur q colonnes : $(\Pi_{1,v}^+)_{1 \leq v \leq q}$.
2. Pour toute autre ligne u , nous posons $\Pi_{u,v}^+ = \Pi_{1,v}^+ + \frac{\Delta_u}{q}$, $\forall 1 \leq v \leq q$ avec Δ_u un réel quelconque tel que $\Pi_{u,v}^+$ reste positif (limité donc inférieurement par l'opposé du minimum de $\Pi_{1,v}^+$).
3. Enfin, pour obtenir une loi de probabilité (sommant donc à 1), nous utilisons $T = \sum_{u=1}^p \sum_{v=1}^q \Pi_{u,v}^+$ et définissons : $\pi_{u,v}^+ = \frac{\Pi_{u,v}^+}{T}$ pour tout (u, v) ainsi que $\delta_u = \frac{\Delta_u}{T}$ pour tout u .

Au final, il y a autant de loi d'indétermination sur pq éléments que de choix de première ligne $(\pi_{1,v}^+)_{1 \leq v \leq q}$ et d'incrémentes $(\delta_1 = 0, \delta_2, \dots, \delta_p)$.

Remarque 14 (Des lignes au colonnes).

Dans la construction précédente, les lignes et les colonnes jouent des rôles différents. Évidemment, toute propriété est symétrique de sorte que nous pourrions aisément transférer chaque étape de u à v .

Remarque 15 (δ_u exprime la différence entre les marges).

Induite par la construction précédente, la marge μ sur $\{1, \dots, p\}$ est :

$$\mu_u = \pi_{u,\cdot}^+ = \mu_1 + \delta_u, \forall 1 \leq u \leq p.$$

Aussi, la division par q que nous avons ajoutée dans la définition permet d'interpréter exactement δ_u comme la différence entre μ_u et μ_1 .

Nous supposons que les δ_u sont positifs ; d'après la remarque 15 cela revient à dire que la première ligne est minimale.

Au plus δ_u est choisi important, au plus la ligne u est fréquente et au plus les tirages sur la marge v à u fixé seront proches de l'uniforme. Ce comportement est cohérent avec la propriété de réduction des collisions que nous retrouverons dans la section 3.5.4. Nous renvoyons à cette section pour la définition d'une collision et en reprenons les notations. Une collision entre deux tirages indépendants (U_1, V_1) et (U_2, V_2) a lieu lorsque les deux couples sont égaux. La stratégie sous-jacente au couplage de réduction des collisions est la suivante :

- Premier cas : u_1 est tiré avec μ_{u_1} faible, donc δ_{u_1} faible. Il est peu probable de tirer un u_2 identique et C^+ se permet une marge sur v à $u = u_1$ fixé aussi déséquilibrée que nécessaire pour respecter les marges.
- Second cas : u_2 est tiré avec μ_{u_2} important, donc δ_{u_1} important. Il est probable de tirer un u_2 identique et la collision ne peut être évitée qu'en jouant sur la marge v à $u = u_1$. $v|u_1$ est alors rendu proche de l'uniforme par δ_{u_1} .

Une marge uniforme μ revient à dire que $\delta = 0$. Dans ce cas, toutes les lignes sont égales à la première et l'on a pour tout (u, v) :

$$\begin{aligned}\pi_{u,v}^+ &= \pi_{1,v}^+ \\ &= \frac{\pi_{\cdot,v}}{q} \\ &= \pi_{u,\pi_{\cdot,v}}\end{aligned}$$

ce qui nous ramène à l'*indépendance*. On peut en effet aisément vérifier que si μ est uniforme et quel que soit ν , $C^+(\mu, \nu) = C^\times(\mu, \nu)$, ce qui explique le résultat.

Au contraire, plus les deux marges μ et ν s'écartent de l'uniforme, plus $\pi^+ = C^+(\mu, \nu)$ et $\pi^\times = C^\times(\mu, \nu)$ seront éloignés ; nous l'avons notamment remarqué dans [10] et rapporté dans la section 2.4.3.

Propriété 14 (Ecart maximal sur une ligne).

L'écart maximal sur une ligne $(\pi_{v|u}^+)_v$ ne peut excéder $\frac{1}{p}$ ce qui signifie que p s'immisce dans la définition d'une ligne sur les $\{1, \dots, q\}$ colonnes.

Démonstration.

Tous les $(\pi_{v|u}^+)_v$ présentent les mêmes écarts entre leurs éléments : on ne fait qu'ajouter δ_u partout. Aussi, l'écart maximal est donné par l'écart maximal possible sur $(\pi_{v|u=1}^+)_v$ (le minimum des μ_u). Comme la somme sur cette marge de valeurs positives est de μ_1 exactement, le déséquilibre maximal possible est de μ_1 entre deux éléments de la même ligne. Par ailleurs, s'agissant du minimum des μ_u qui sont des éléments positifs sommant à 1, on voit facilement : $\mu_1 \leq \frac{1}{p}$ ce qui suffit pour conclure. \square

3.5.2 Graphiques du tirage selon l'indétermination

Prenons la construction d'indétermination proposée dans la section précédente. Un déséquilibre sommant à μ_1 est donné au départ : $\pi_{1,v}^+$, $1 \leq v \leq q$. Les δ_u sont ensuite calculés par cohérence avec les écarts $\mu_u - \mu_1$

Réaliser un tirage selon la loi π^+ s'opère de la manière suivante :

1. On tire un $U = u$ selon μ ;
2. On joue avec une pièce pipée selon l'écart : $I = \text{Bern}\left(\frac{\mu_u - \mu_1}{\mu_u}\right)$;
3. Si la sortie est 0, on tire $V = v$ selon $\pi_{1,v}$ et si la sortie est 1, on tire $V = v$ selon la loi uniforme sur $\{1, \dots, q\}$.

Propriété 15 (Construction d'un tirage selon l'indétermination).

La construction tout juste introduite permet bien de tirer suivant l'indétermination.

Démonstration.

Montrons que les probabilités respectent la formule de l'indétermination.

$$\begin{aligned}
 \mathbb{P}(U = u, V = v) &= \mathbb{P}(V = v|U = u)\mathbb{P}(U = u) \\
 &= \mathbb{P}(V = v|U = u, I = 0)\mathbb{P}(U = u)\mathbb{P}(I = 0) \\
 &\quad + \mathbb{P}(V = v|U = u, I = 1)\mathbb{P}(U = u)\mathbb{P}(I = 1) \\
 &= \pi_{1,v}\mu_u \frac{\mu_1}{\mu_u} + \frac{1}{q}\mu_u \frac{\mu_u - \mu_1}{\mu_u} \\
 &= \pi_{1,v}\mu_1 + \frac{\mu_u - \mu_1}{q}
 \end{aligned}$$

Remarquons alors que la somme sur u des probas que nous venons de calculer fait $\mathbb{P}(V = v)$ de sorte que

$$\mathbb{P}(V = v) = \nu_v = p\pi_{1,v}\mu_1 + \frac{1 - p\mu_1}{q}$$

Ou encore :

$$\frac{-\mu_1}{q} = \frac{\nu_v}{p} - \pi_{1,v}\mu_1 - \frac{1}{pq}$$

En remplaçant, il vient finalement la formule attendue :

$$\mathbb{P}(U = u, V = v) = \frac{\mu_u}{q} + \frac{\nu_v}{p} - \frac{1}{pq}$$

□

Le procédé constructif décrit dans la section précédente et permettant de créer une expérience en deux temps générant un couplage d'indétermination amène au graphique 3.10 présenté juste après.

L'histogramme de cette figure 3.10 permet de se rendre compte de l'évolution des distributions. On a classé les probabilités de μ pour avoir $\mu_i \leq \mu_{i+1}, \forall i$. Ainsi, on voit que plus μ_u augmente et plus les tirages selon v sont équilibrés. En effet, $\pi_{1,v}$ apporte un déséquilibre relatif (il peut lui-même être uniforme) mais la partie complémentaire, est tirée uniformément.

La proportion affichée est, bien sûr, tout à fait factice et arbitraire mais elle présente bien l'évolution du tirage sur la seconde variable en fonction de la classe obtenue sur la première variable.

Par ailleurs, l'ordonnancement des μ_u en accord avec celui de la proportion de tirages uniformes montre aussi que les parties les plus équilibrées sont celles sur les marges les plus importantes et donc les plus probables.

3.5.3 Temporalité

Une nouvelle interprétation du couplage dit d'indétermination peut être vue à partir de la construction 15. On constate en effet, si l'on s'affranchit de la partie probabilité qui suppose que tout somme à 1 que l'on peut introduire une notion de temporalité, naturelle et représentée par la première variable u ; appelons-la t pour plus de cohérence.

Définition 17.

On se donne $\Pi_{1,v}, 1 \leq v \leq q$ et $(\delta_t)_{t \in \mathbb{N}}$ positifs alors $\Pi_{t,v} = \Pi_{1,v} + \sum_{u=1}^t \delta_u$ donne un tirage respectant l'indétermination

Démonstration.

Il suffit d'appliquer la proposition 15 en adaptant pour passer d'une loi de probabilité aux tirages : ici $\Pi_{t,v}$ (noter la majuscule) somme au nombre de tirages réalisés à l'instant t . \square

Avec ces notations, on modélise typiquement une distribution d'individus se répartissant uniformément après le temps $t = 2$ entre différentes catégories mais de façon purement arbitraire au temps 1. Imaginons une évacuation d'une salle de concert disposant (évidemment) de plusieurs sorties. Mesurons toutes les minutes le nombre de personnes sorties par chaque porte. Au début de l'incident, disons pendant le temps 1, les premiers individus vont se répartir arbitrairement entre les différentes portes créant $\Pi_{1,\cdot}$. Ensuite, on peut s'attendre, à en avoir autant sortant de chaque porte toutes les minutes. On sera ainsi, par construction, dans le cadre d'une réalisation statistique d'indétermination. L'indétermination, si l'on veut pousser le parallèle sur le nom, ne s'applique donc qu'au temps 1, le reste étant déterministe et parfaitement réparti.

Mieux, le parallèle avec l'évacuation d'une salle de concert montre que l'attente que l'on a d'une évacuation uniforme des individus après un temps d'adaptation est une mesure de la qualité de disposition des portes. Si une porte d'évacuation est peu empruntée, il eût été préférable de la placer ailleurs. Ainsi, l'écart à l'indétermination, associé à des simulations, apparaît comme un critère de construction envisageable.

3.5.4 Limitation des collisions

Le processus d'indétermination est à rebours de celui d'indépendance, les deux ne se rejoignant que lorsque les marges sont uniformes. Ainsi, l'indétermination concède un déséquilibre important donné par $\pi_{1,v}$ sur les u de faible probabilité tandis que plus la marge est importante et plus $\pi_{1,v} + \delta_u$ sera près de l'uniforme. Au contraire, l'indépendance applique le même déséquilibre proportionnellement sur chaque marge. Le but de chaque couplage étant de se rapprocher de l'uniforme, l'indétermination sacrifie les petites marges au profit des grandes quand l'indépendance distribue proportionnellement.

Dans le cas discret, le problème 6 s'écrit comme la projection d'une estimation \mathbb{U} (comme déjà indiqué, l'optimal si l'on supprime les contraintes de marge). La distance \mathbb{L}^2 (Equation (2.20)) est utilisée pour projeter \mathbb{U} sur l'espace $\mathcal{L}_{\mu,\nu}$ des mesures de probabilité respectant μ et ν . Le théorème 2 livre la forme de la solution optimale $\pi^+ = C^+(\mu, \nu)$.

π^+ introduit forcément une dépendance entre ses marges étant donné que la fonction de couplage n'est pas C^\times .

De surcroît, si l'on utilise maintenant l'approche basée sur l'entropie de Kullback-Leibler pour projeter, on obtient alors l'indépendance. En effet, le problème 5 s'exprime lui aussi comme une projection de \mathbb{U} sur le même sous-espace $\mathcal{L}_{\mu,\nu}$ mais avec la distance juste citée. La solution étant alors, selon le théorème 1 : $\pi^\times = C^\times(\mu, \nu)$ c'est à dire l'absence de dépendance, par définition.

Pour résumer, avec $\mathcal{L}(\mu, \nu)$ l'espace des mesures de probabilité ayant pour marges μ et ν on a :

$$\mathbb{U} \xrightarrow{KL} \mathcal{L}(\mu, \nu) = \pi^\times \quad \Bigg| \quad \mathbb{U} \xrightarrow{\mathbb{L}^2} \mathcal{L}(\mu, \nu) = \pi^+$$

Regardons maintenant en détail la fonction de coût liée à l'*indétermination* en interprétant légèrement le problème qu'elle résout. Si l'on retire les constantes (celles-ci nous aidant à unifier les deux problèmes comme des projections sous divergence) on remarque que la fonction de coût :

$$\mathbf{M}^+ = pq \sum_{u,v} \left\{ \pi_{u,v} - \frac{1}{pq} \right\}^2$$

s'écrit également :

$$\mathbf{M}^+ = \sum_{u=1}^p \sum_{v=1}^q \pi_{u,v}^2.$$

Lorsque nous minimisons \mathbf{M}^+ , la formulation du problème vise précisément à réduire les collisions de couples suivant la Définition 18 d'une collision.

Définition 18 (Collision).

Une variable aléatoire $W = (U, V)$ suivant une probabilité π , on se donne deux tirages indépendants $W_1 = (U_1, V_1)$ et $W_2 = (U_2, V_2)$. On dit qu'il y a collision lorsque $W_1 = W_2$ (comprendre que les deux coordonnées des deux couples tirés sont égales).

De fait, la probabilité de collision entre deux tirages successifs indépendants de W , n'est rien d'autre que

$$\sum_{U_1, U_2=1}^p \sum_{V_1, V_2=1}^q \pi_{U_1, V_1} \pi_{U_1, V_1} \pi_{U_2, V_2} \delta_{U_1=U_2} \delta_{V_1=V_2} = \sum_{U_1, U_2=1}^p \sum_{V_1, V_2=1}^q \pi_{U_1, V_1} = \mathbf{M}^+.$$

Une interprétation de l'indétermination, déjà exprimée dans le problème 6 mais demandant de s'affranchir de la notation utile pour une écriture à l'aide de divergence, est donc de minimiser les collisions de couples et ce, à marges fixées. Le chapitre dédié aux applications sera l'occasion d'interpréter dans deux problèmes pratiques les collisions pour faire émerger l'*indétermination* comme un extrême ou de solution ou de pire cas.

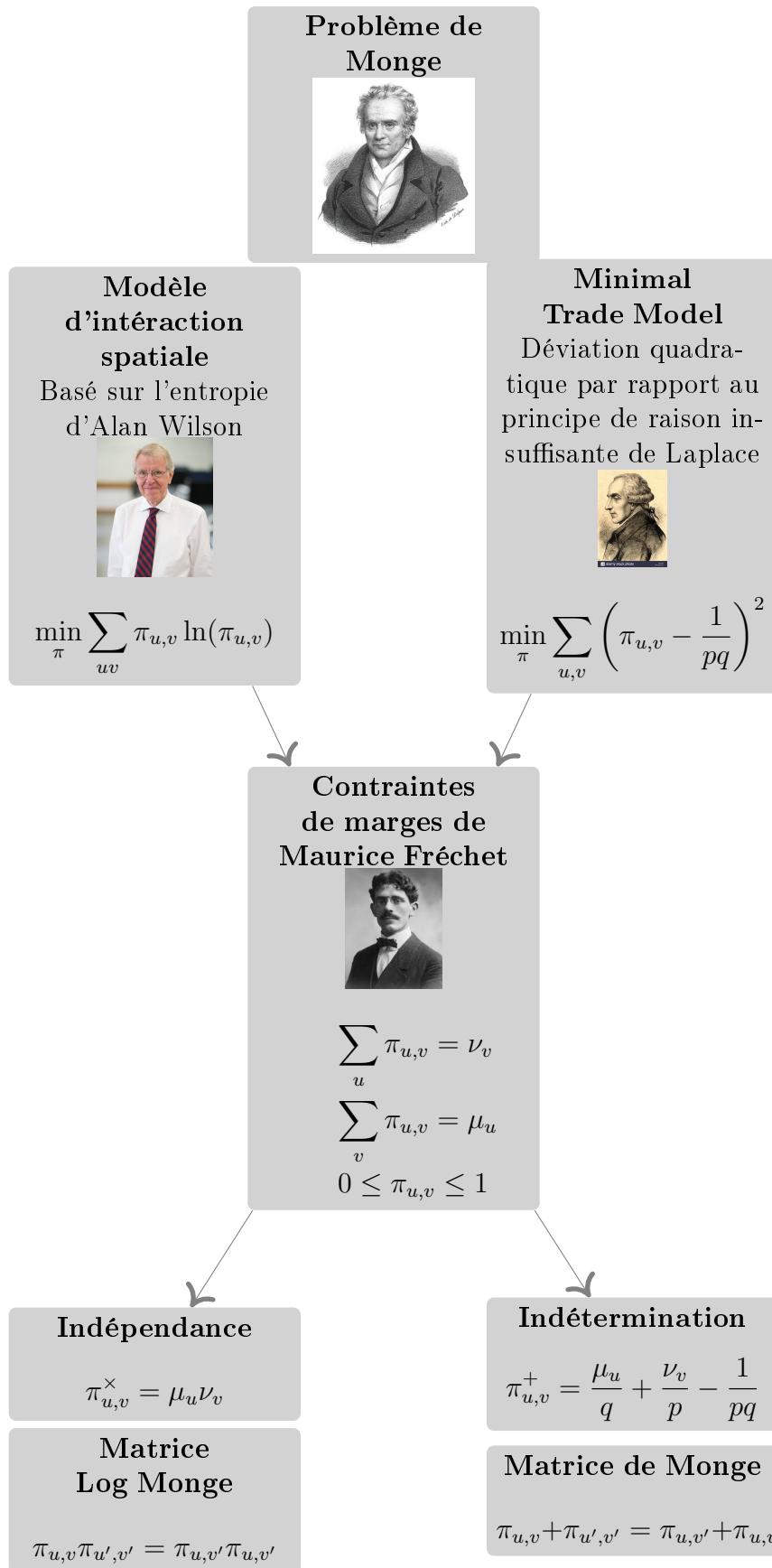


FIGURE 3.9 – Résumé de la symétrie indépendance / indétermination

61

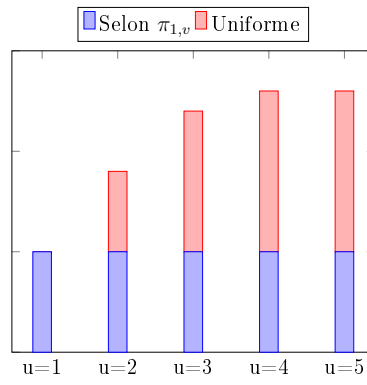


FIGURE 3.10 – Distribution des types de tirages de v selon la valeur de u

Applications de l'indétermination

L' avait l' don, c'est vrai, j'en
conviens,
L'avait l' génie,
Mais sans technique, un don
n'est rien
Qu'un' sal' manie...

Georges Brassens, *Le mauvais
sujet repent*, 1953

Le chapitre 3 s'est arrêté sur une interprétation de la fonction de coût D_2 que minimise l'indétermination comme une évaluation du nombre de collisions. Le couplage minimise ainsi la probabilité que deux tirages successifs soient égaux. Le présent chapitre, suivant notre article [11], est l'occasion d'interpréter de deux manières une collision, ou bien la réussite d'un espion ou bien la réalisation d'une tâche, pour observer que le couplage d'*indétermination* apparaît naturellement dans deux problèmes exposés dans les articles [69] et [18].

Plus précisément, lors d'une communication cryptée entre Alice et Bob, un espion Charlie propose un message clair selon la distribution des clairs associés au crypté observé. Une collision représente ici une réussite de Charlie : le clair proposé est celui émis. Or, Alice a la main sur le couplage crypté / clair et nous montrons que le meilleur choix pour contrer Charlie consiste à suivre l'indétermination.

De même, dans le partitionnement de tâches, celles-ci sont affectées aléatoirement à des bureaux selon un couplage arbitraire ; les bureaux quant à eux réalisent les tâches attribuées au hasard. Une collision a lieu lorsque la tâche cible est réalisée en premier par le bureau d'attribution. Nous montrons que la pire affectation consiste à suivre l'indétermination qui fournit donc un lot de bornes.

Une seconde partie du chapitre s'intéresse au clustering de graphe, problème dont la bibliographie, que nous résumerons, démontre l'actualité. Nous avons exploré l'application de l'indétermination dans ce domaine dans notre article [10]. Ainsi, nous remarquons que l'heuristique classique dite de Girvan-Newman se fonde sur une notion de modularité qui n'est rien d'autre qu'une agrégation d'un critère local de comparaison à l'indépendance. Au vu des similitudes entre les deux couplages, nous nous intéressons à une modularité

originale basée sur un critère local d'écart à l'indétermination. Cette généralisation de l'heuristique de Girvan-Newman n'est pas inédite. De fait, Patricia Conde-Cespedes dans sa thèse [25] généralise déjà à une liste de critères (repris dans le chapitre 5); chaque critère de corrélation fournissant un algorithme original, elle les compare sur différents graphes.

Dans ce document, nous nous limitons aux deux critères canoniques et, nous affranchissant d'une expérimentation déjà réalisée d'ailleurs dans [25], nous étudions dans un premier temps la similitude entre les deux critères dans le cadre de graphes aléatoires suivant Gilbert, dans un second temps, nous insistons sur les différences pouvant émerger sur des graphes spécifiques. Une dernière section retrouve ce que d'autres auteurs ont pu appeler la limite de résolutions et remarque qu'elle est identique pour les deux critères.

4.1 Problème dit de l'espion

En cryptographie, un message u d'un alphabet fini \mathcal{U} de taille p est envoyé par Alice à destination de Bob tandis qu'un espion, souvent appelé Charlie, essaie de l'intercepter ou de le deviner. Une stratégie classique pour qu'Alice converse efficacement avec Bob consiste à encoder le message ou par l'usage d'un couple (clef privée, clef publique) ou par un encodage symétrique requérant une clef secrète partagée par nos deux protagonistes.

La littérature scientifique autour des méthodes de cryptage est diverse, dépendant évidemment de la situation. Pour autant le standard AES pour Advanced Encryption Standard est privilégié, le détail de la méthode est disponible dans de nombreux articles.

Le cryptage vise à contrecarrer l'interception du message par Charlie : s'il le capte, il n'obtiendra qu'un charabia illisible. Aussi, Charlie est supposé avoir accès au message crypté v qui vit dans un alphabet \mathcal{V} de taille q et qui est donc une fonction du message clair u .

Le problème 10 de l'espion introduit ci-après remonte à l'article [69] et formalise la situation exposée. Remarquons néanmoins qu'en cryptographie, Charlie décode une séquence de messages tandis que le problème dit de l'espion n'introduit qu'un seul message u à décoder. De plus, la version initiale que nous présentons est très limitée puisque Charlie devine le message de l'émetteur sans aucune information reliée à celui-ci. Une seconde version verra l'introduction d'une variable V observée et qui, corrélée à U pourra (entre autres) représenter le message crypté.

Formalisons la situation originale : U est une variable aléatoire à valeur dans \mathcal{U} suivant la loi de probabilité $\mathbb{P}_U = \mu$. L'émettrice Alice génère une séquence de messages indépendants sous μ .

Problème 10 (Problème original de l'espion ou "Guessing Problem").

Quand Alice envoie un message $U = u$ à Bob, l'espion Charlie tente d'en deviner le contenu u . Pour cela il dispose d'une suite infinie de questions formatées pour chaque tentative \tilde{u} : "Le message envoyé était-il \tilde{u} ?" et pour lesquelles la réponse est toujours binaire : "oui/non".

Définition 19 (Stratégie Originale).

Une stratégie $S = \sigma$ de Charlie est définie par un ordre sur \mathcal{U} représentant la suite des tentatives qu'il effectuera : le premier essaie, le second, jusqu'au p -ième. L'ordre peut être déterministe ou aléatoire et est donc soumis à une loi de probabilité que nous notons \mathbb{P}_S .

Par ailleurs, pour chaque position, $i \in [1, p]$, $\sigma[i]$ est l'élément dans \mathcal{U} correspondant au i -ème essai.

L'article [69] propose une mesure de la performance de chaque stratégie σ de Charlie. Il s'agit de calculer le ρ -moment de G qui compte le nombre d'essais requis pour que Charlie trouve le message u envoyé. Nous ajouterons ultérieurement une seconde mesure de la performance.

Définition 20 (Mesure de la performance).

La fonction $G(\sigma, u)$ compte le nombre de questions requises pour obtenir un "oui" dans le problème 10 et ce lorsque l'ordre proposé par Charlie est $S = \sigma$ et que le message d'Alice est $U = u$. Même si u est fixé, il peut s'agir d'une variable aléatoire typiquement si S en est une.

De fait, $G(S, U)$ est une variable aléatoire dont la définition formelle est :

$$G(\sigma, u) = \sum_{i=1}^p i \mathbb{1}_{\sigma[i]=u}$$

Nous définissons également l'efficacité d'une stratégie S comme étant le ρ -moment de $G(S, U)$ sous le couplage indépendant de la stratégie $S \sim P_S$ et du message envoyé $U \sim P_U$.

$$\|G(S, U)\|_\rho = [\mathbb{E}_{(S, U) \sim \mathbb{P}_S \otimes \mathbb{P}_U} (G(S, U)^\rho)]$$

Appliquons la définition précédente sur un exemple. On se donne un alphabet $\mathcal{U} = \{a, b, c, d\}$, duquel est extrait un message $u = c$ tandis que la stratégie σ choisie par Charlie est l'ordre (b, c, a, d) (comprendre que son premier essai est b son second c et ainsi de suite). Il s'ensuit :

$$\begin{aligned} G(\sigma, u) &= \sum_{i=1}^p i \mathbb{1}_{u=\sigma[i]} \\ &= 1 \cdot \mathbb{1}_{u=b} + 2 \cdot \mathbb{1}_{u=c} + 3 \cdot \mathbb{1}_{u=a} + 4 \cdot \mathbb{1}_{u=d} \\ &= 2 \cdot \mathbb{1}_{u=c} \\ &= 2 \end{aligned}$$

Il a été montré dans le même article [69] un résultat naturel : pour peu que $\mathbb{P}_U = \mu$ soit connu, la meilleure stratégie consiste à proposer les éléments de l'alphabet dans l'ordre décroissant des probabilités associées suivant donc une stratégie déterministe : proposer en premier le message le plus fréquent, en second le deuxième plus fréquent et ainsi de suite de sorte que l'on a

$$\mu_{\sigma[p]} \leq \dots \leq \mu_{\sigma[1]}.$$

De surcroît, les auteurs démontrent une borne inférieure sur le nombre moyen de questions valable quelle que soit la stratégie utilisée comme exposé dans le Théorème 7.

Théorème 7 (Efficacité maximale).

En moyenne, le nombre minimal de questions pour résoudre le problème 10 est :

$$\min_S \|G(S, U)\|_\rho \geq (1 + \log(p))^{-\rho} \left[\sum_{u \in \mathcal{U}} \mathbb{P}(U = u)^{\frac{1}{1+\rho}} \right]^{1+\rho}$$

Proof.

Nous ne rappellerons pas ici la démonstration complète de l'inégalité mais quelques élément de preuve. Dans [5], $\|G(S, U)\|_\rho$ est récrit comme l'intégration d'une fonction concave selon une mesure de probabilité Q . L'application de l'inégalité classique de Jensen conduit à une borne inférieure valable quelle que soit Q . Le facteur multiplicatif quant à lui vient de l'inégalité :

$$\sum_{u \in \mathcal{U}} \frac{1}{G(\sigma, u)} = \sum_{i=1}^p \frac{1}{i} \leq 1 + \log(p)$$

Enfin, un choix particulier de Q livre le résultat. □

Une application pratique du Théorème 7 est de garantir un temps de décodage minimal pour tout espion. L'émetteur, de son côté sera intéressé par la maximisation de la borne inférieure.

Problèmes étendus

Comme annoncé en introduction, on donne maintenant accès à Charlie à une variable aléatoire V corrélée avec le message U . Dans la situation cryptographique classique, V représente le message chiffré que Charlie intercepte sur le réseau lorsqu' Alice communique avec Bob. Il s'agit alors d'une fonction de U et donc d'un élément déterministe à U fixé.

Nous généralisons légèrement en supposant V variable à U fixé. Cela permet d'ajouter des informations telles que la fréquence utilisée, l'antenne d'émission, la position de l'émetteur, du récepteur, ou toute information utile que Charlie pourrait observer. Au final, tout est intégré dans la variable aléatoire V corrélée à U et vivant dans un alphabet \mathcal{V} de taille q .

Comme mentionné dans l'article [5], Charlie choisit maintenant sa stratégie en accord avec l'observation V . Typiquement, il adapte ses essais (comprendre son ordre) en fonction du message chiffré observé qui est contenu dans V . La loi de probabilité du couple (U, V) est notée $\mathbb{P}_{U,V} = \pi$ tandis que ses marges sont $\mathbb{P}_U = \mu$ et $\mathbb{P}_V = \nu$.

La fonction de gain s'exprime maintenant selon la notation $G(S, U|V)$ qui insiste à dessein sur la connaissance de la valeur de V en utilisant le symbole $|$ pour "sachant que".

Reste qu'au final, à chaque valeur $V = v$, une stratégie originale S_v (voir Définition 19) est construite et conduit à un gain original $G(S_v, U)$ de sorte que l'on a :

$$G(S, U|V) = \sum_{v \in \mathcal{V}} G(S_v, U) \mathbb{1}_{V=v} \tag{4.1}$$

Le même article propose une généralisation de la proposition 7 que nous rapportons ici :

Théorème 8 (Borne inférieure généralisée de l'efficacité).

Pour toute stratégie, le temps de reconstruction du message est minoré selon l'inégalité :

$$\mathbb{E}_{(S,U,V) \sim \mathbb{P}_{S,U,V}} [G(S,U|V)^\rho] \geq (1 + \log(p))^{-\rho} \sum_{v \in \mathcal{V}} \left[\sum_{u \in \mathcal{U}} (\pi_{u,v})^{\frac{1}{1+\rho}} \right]^{1+\rho}$$

Proof.

Le résultat provient de la décomposition déjà remarquée dans l'équation 4.1 de la stratégie globale S en des stratégies originales à $V = v$ fixé. Pour chaque valeur v , la stratégie locale (ou originale) rentre dans le cadre de la proposition 7 de sorte qu'elle en respecte la conclusion. Le résultat global en découle immédiatement. \square

Couplage d'*indetermination* et renforcement du cryptage

Affranchissons-nous légèrement de la littérature et mesurons la performance de Charlie selon sa capacité à retrouver le message u du premier coup. C'est une mesure raisonnable d'autant plus que s'il observe une séquence de messages, Charlie peut devoir passer rapidement de l'un à l'autre, trop rapidement pour avoir plusieurs essais sur chaque.

Définition 21 (Performance one-shot).

Pour une stratégie S fixée, nous définissons la performance selon la formule ci-dessous qui mesure la probabilité de retrouver la valeur u après un essai :

$$M(S, U, V) = \mathbb{P}_{(S,U,V) \sim \mathbb{P}_{S,U,V}} (S[1] = U)$$

Remarque 16 (Performance one-shot généralisée).

Nous pourrions aisément introduire une mesure sous le nom performance " k shots" évaluant la probabilité de retrouver le message après k essais. Cette mesure peut avoir un intérêt, comme mentionné par un rapporteur anonyme de notre article à GSI, typiquement en fixant k égal au nombre de tentatives que Charlie peut faire entre deux messages d'Alice. Une astuce pour couvrir ce cas consiste à faire répéter à Alice k fois ses messages et à s'en tenir à la performance one-shot.

Nous remarquerions alors habilement que la performance " k shots" vaut 1 dès lors que $k \geq q$ pour toute stratégie raisonnable. Nous ne rentrerons pas plus dans le détail nous limitant à deux mesures : premier coup ou nombre infini d'essais.

Nous supposons comme pour construire la stratégie optimale sous la première performance que Charlie connaît la distribution $\mathbb{P}_{U,V} = \pi$. Par exemple, il peut auparavant avoir eu accès à un historique de message clairs lors d'une étape préliminaire.

Deux stratégies se distinguent immédiatement :

1. S_{max} : retourne systématiquement, à v fixé (observé par hypothèse), le message \tilde{u} le plus probable selon la marge $\mathbb{P}_{U|V=v}$
2. S_{margin} : retourne, à v fixé, la réalisation \tilde{u} d'une variable aléatoire suivant la loi $\mathbb{P}_{U|V=v}$

De la même manière que les auteurs de [69] ont prouvé que S_{max} maximise la mesure de performance G , on peut montrer qu'elle maximise également la performance one-shot.

De fait, M se récrit :

$$M(S, U, V) = \sum_{v=1}^q \nu_v \mathbb{P}_{(S_v, U) \sim \mathbb{P}_{S_v} \otimes \mathbb{P}_{U|V=v}} (S_v[1] = U) \quad (4.2)$$

$$= \sum_{u=1}^p \sum_{v=1}^q \pi_{u,v} \mathbb{P}_{S_v} (S_v[1] = u). \quad (4.3)$$

Comme \mathbb{P}_{S_v} ne dépend que de v , M est maximal sous S_{max} , quand $\mathbb{P}_{S_v} = \delta_{u_v}$ avec $u_v = \operatorname{argmax}_u \pi_{u,v}$ de sorte que l'on a :

$$M(S, U, V) \geq M(S_{max}, U, V) = \sum_{v=1}^q \pi_{u_v, v}. \quad (4.4)$$

Au final, nous notons $u_1 = \operatorname{argmax}_u \mu_u$ et remarquons que $\sum_{v=1}^q \pi_{u_v, v} \geq \mu_{u_1}$ ce qui nous conduit à la proposition 16.

Propriété 16 (Performance optimale de Charlie).

Nous supposons les marges μ et ν figées. Alors, pour n'importe quel couplage π entre les messages U et les chiffrés V , la meilleure performance one-shot que Charlie puisse accomplir a toujours lieu sous S_{max} . De surcroît, elle admet une borne inférieure μ_{u_1} qui ne fait pas intervenir π ; en résumé :

$$M(S, U, V) \geq M(S_{max}, U, V) = \sum_{v=1}^q \pi_{u_v, v} \geq \mu_{u_1}. \quad (4.5)$$

Supposons, but naturel, qu'Alice souhaite minimiser la performance one-shot de Charlie. Nous supposons également qu'elle travaille à marges fixées (μ sur \mathcal{U} et ν sur \mathcal{V}). Il s'agit en fait d'une hypothèse réaliste : l'alphabet \mathcal{U} dans lequel les messages sont composés respecte typiquement une distribution sur les lettres ; la variable V quant à elle peut devoir respecter des taux d'occupation sur les fréquences ou des durées de localisation d'Alice pour reprendre les exemples qu'elle peut encoder. Au final, Alice ne peut agir que sur le couplage entre U et V d'où l'idée de calculer la valeur de la meilleure performance sous chaque couplage canonique. On obtient en utilisant la proposition 16 et en remplaçant π par ses deux valeurs :

$$M(S_{max}, \pi^\times) = \mu_{u_1} \quad M(S_{max}, \pi^+) = \mu_{u_1}$$

Nous avons donc montré que les deux couplages étaient optimaux pour Alice : ils fournissent la pire performance maximale pour Charlie.

Si la deuxième stratégie n'est pas optimale pour la mesure de performance choisie, elle est plus difficile à contrer pour Alice. Sous S_{marge} en effet, il s'agit de réduire les collisions entre son message U et celui généré aléatoirement par l'espion en fonction de V . La performance associée est donnée par :

$$M(S_{marge}, U, V) = \sum_{u \in \mathcal{U}} \sum_{v \in \mathcal{V}} \nu_v (\pi_{u|V=v})^2$$

Nous pouvons donc encadrer la performance de S_{marge} :

$$\frac{\|\pi\|_2^2}{\min_{v \in \mathcal{V}} \nu_v} \geq M(S_{marge}, U, V) \geq \frac{\|\pi\|_2^2}{\max_{v \in \mathcal{V}} \nu_v} \quad (4.6)$$

avec

$$\|\pi\|_2 = \sqrt{\sum_{u \in \mathcal{U}} \sum_{v \in \mathcal{V}} (\pi_{u,v})^2}$$

En appliquant l'équation (2.11) il vient :

$$\sum_{u \in \mathcal{U}} \sum_{v \in \mathcal{V}} \mathbf{M}^+(\pi_{u,v}) = pq \|\pi\|_2^2 - 1$$

De fait, l'étude du problème de l'espion nous ramène d'abord à l'idée de collision (définition 18) et enfin au problème 4 associé au coût \mathbf{M}^+ et dont la solution est donnée par le couplage d'*indétermination* des marges imposées.

Formellement, l'utilisation du couplage

$$\mathbb{P}_{U,V}^+ = C^+(\mathbb{P}_U, \mathbb{P}_V) = \frac{\mu_u}{q} + \frac{\nu_v}{p} - \frac{1}{pq} = \pi_{u,v}^+$$

garantit une réduction efficace des collisions (voir section 3.5.4) et limite ainsi la performance one-shot de Charlie comme exprimé dans l'équation (4.6).

Pour combattre Charlie, supposé suivre la stratégie S_{marge} , Alice doit dispatcher les messages sur le couple (U, V) suivant l'*indétermination*. Elle peut en particulier suivre le tirage constructif énoncé dans la section 3.5.

De surcroît, la borne inférieure de la performance one-shot donnée par l'équation 4.6 est minimale sous l'*indétermination* π^+ de sorte que la borne résultante est valable quel que soit le couplage π réellement utilisé :

$$\forall \pi, M(S_{marge}, U, V) \geq \frac{\|\pi^+\|_2^2}{\max_{v \in \mathcal{V}} \nu_v} \quad (4.7)$$

Cette optimalité rejoint l'observation de la Figure 3.10 : sur toute information visible V , les messages sont distribués aussi uniformément que possible ce qui réduit la probabilité d'une collision avec l'essai de Charlie qui suit précisément la marge.

Enfin, une inégalité classique entre les normes montre que la performance one-shot de S_{marge} encadre celle de S_{max} ce qui permet d'écrire l'inégalité valable quel que soit le couplage sous-jacent de (U, V) :

$$\forall \pi, \sqrt{M(S_{marge}, U, V)} \geq M(S_{max}, U, V) \geq M(S_{marge}, U, V) \geq \frac{\|\pi^+\|_2^2}{\max_{v \in \mathcal{V}} \nu_v} \quad (4.8)$$

4.2 Problème du partitionnement de tâches

Le problème du partitionnement de tâches est formalisé par l'article [18] qui fournit également une borne inférieure sur le nombre de tâches à accomplir. Nous nous tournons

plutôt vers le travail de regroupement de [59] qui s'accompagne d'une comparaison au problème de "Guessing" et d'une extension du problème original.

Formellement, nous commençons par le problème original de partitionnement de tâches : un jeu fini \mathcal{U} de p tâches est donné ainsi qu'un entier $q \leq p$. Il s'agit alors de concevoir une partition $\mathcal{A} = (\mathcal{A}_1, \dots, \mathcal{A}_q)$ de \mathcal{U} en q classes et ce afin de minimiser le nombre de tâches à accomplir sous la contrainte suivante : si une tâche u doit être réalisée dans un sous-ensemble \mathcal{A}_i ($u \in \mathcal{A}_i$) alors il faut nécessairement lancer simultanément tout le sous-ensemble et donc réaliser $|\mathcal{A}_i|$ actions.

En pratique, une tâche $U = u$ à réaliser est choisie au hasard parmi \mathcal{U} sous la loi de probabilité $\mathbb{P}_U = \mu$ qui représente la fréquence de chacune. Comme toutes, u est affectée à une unique classe $\mathcal{A}_{i(u)}$ de la partition arbitraire. Aussi, $A(u) = |\mathcal{A}_{i(u)}|$ compte le nombre d'actions à effectuer. Au final, nous jouons sur la partition pour effectuer, en moyenne, aussi peu de tâches que possible.

Habituellement, et comme pour le problème de "guessing", la performance d'une partition \mathcal{A} est exprimée par le ρ -moment de $A(U)$, formellement :

$$\mathbb{E}_U [A(U)^\rho]$$

Ainsi et semblablement au Théorème 7, les auteurs de [59] montrent que l'on a :

$$\min_{\mathcal{A}} \mathbb{E}_{U \sim \mu} [A(U)^\rho] \geq \frac{1}{q} \left[\sum_{u \in \mathcal{U}} (\mu_u)^{\frac{1}{1+\rho}} \right]^{1+\rho} \quad (4.9)$$

exprimant ainsi un nombre minimum de tâches à effectuer valable quelle que soit la partition.

En s'inspirant du problème de l'espion, ils étendent le problème de partitionnement. Nous introduisons maintenant la version enrichie qu'ils proposent. Dans celle-ci, nous ne souhaitons plus réduire le nombre de tâches à effectuer sous la contrainte de lancer le sous-ensemble complet de la tâche choisie u . À la place, le but est de réduire le nombre de tâches effectuées avant u .

De fait, dans la version originale, dès que u est tiré, une règle arbitraire impose de lancer tout le sous-ensemble $\mathcal{A}_{i(u)}$ et donc de réaliser $A(u)$ actions. Dans la nouvelle version, les tâches sont réalisées séquentiellement parmi $\mathcal{A}_{i(u)}$ selon une stratégie globale S qui peut être déterministe ou aléatoire. C'est comme si on activait un bureau $i(u)$ qui traitait les tâches dans un ordre aléatoire et que l'on attendait que la nôtre soit faite.

Un exemple d'application est l'obtention d'une série de signatures dans une administration. En ce cas, q représente le nombre de bureaux dédiés au tamponnage des documents. Chacun peut avoir le droit de tamponner avec plusieurs signatures ; c'est classique lorsque des assistants sont présents dans le processus. dans ce cas, la partition encode l'affectation des tampons à faire aux différents bureaux.

Quand un bureau $V = v$ se voit affecter un document, celui qui l'y a déposé attend que la signature ait lieu. Le bureau de son côté, suit sa propre stratégie S_v pour signer les documents affectés (comprendre que l'ordre est arbitraire, et c'est souvent le cas lorsque les documents transitent par diverses boîtes postales entre l'émetteur et le signataire). S_v peut par exemple consister à suivre toujours le même ordre dans les tâches ou bien à changer quotidiennement au gré des préférences : stratégie fixe ou aléatoire.

Sous une stratégie globale S qui consolide celles de chaque bureau, employé ou machine S_v , $\forall 1 \leq v \leq q$ et pour une tâche u à réaliser, la performance d'une partition \mathcal{A} se mesure en utilisant :

$$N_{S,\mathcal{A}}(u) = \sum_{j=1}^{|\mathcal{A}_{i(u)}|} j \mathbb{1}_{S_{i(u)}[j]=u}$$

qui compte le nombre de tâches effectuées avant la tâche u (u incluse). Une borne inférieure est donnée dans l'article [59].

Supposons maintenant que les clefs $1 \leq v \leq q$ représentent des bureaux ou des équipes se voyant affecter une proportion ν_v des tâches, celles-ci toujours distribuées selon μ . Une application classique étant l'attribution par un manager de tâches aux fréquences connues à des équipes disposant de capacités de traitement variables.

On suppose également que chaque équipe utilise la stratégie S_{marge} pour ordonner les tâches, cela signifiant qu'elle s'y applique en suivant la fréquence habituellement observée selon sa marge (donc son bureau) ; dans le cas de signatures, chacune s'occupe à chaque pas de temps d'un type de signature parmi ceux qu'elle gère en suivant les proportions habituelles.

Remarque 17 (Distribution estimée).

Dans les deux applications précédentes, l'espion comme le partitionnement, nous jouons avec des probabilités. En réalité on envoie un nombre fini et entier de lettres tout comme on réalise un nombre entier et fini de tâches, nommons-le n .

De surcroît, et comme mentionné, \mathcal{U} et \mathcal{V} sont finis. Enfin suivant le principe des matrices de contingence dont la définition est rappelée en Définition 12, pour tout choix de $u \in \mathcal{U}$ et de $v \in \mathcal{V}$, un nombre entier $n_{u,v}$ de tâches u est associé au bureau v tout comme un nombre de lettres u est envoyé en générant l'observation de v par Charlie.

Pour convertir $n_{u,v}$ en loi de probabilité, nous divisons par n :

$$\pi_{u,v} = \frac{n_{u,v}}{n} \tag{4.10}$$

réciroquement, étant donné une loi de probabilité π , nous tirerons n messages selon π et la distribution estimée se rapprochera d'autant plus de π que n augmente.

Pour conclure, nous notons que l'ajout d'une contrainte de forme sur π , à savoir de respecter la structure rationnelle de l'équation 4.10, met fin à notre capacité à résoudre les deux problèmes (Problème 5 comme Problème 6) sous une forme close. De plus, les deux couplages indétermination et indépendance résolvent le problème relâché (avec des valeurs dans \mathbb{R}). Aussi, ils constituent une solution plus efficace donc encore plus utile pour une application que la forme non relâchée ; c'est cette solution qui est ensuite approchée par les tirages.

Maintenant, nous pouvons récrire le problème de partitionnement sous la forme du problème de l'espion :

- $V = v$, auparavant un bureau, représente maintenant l'information à laquelle a accès l'espion ;
- $U = u$, auparavant une tâche à réaliser, représente maintenant un message envoyé ;
- $S = \sigma$, auparavant l'ordre dans lequel les tâches sont réalisées au sein d'un bureau, représente maintenant l'ordre suivi par Charlie pour ses propositions.

Sous ce formalisme, on mesure la probabilité $M(S, U, V)$ d'exécuter la tâche u en premier comme une extension de la performance one-shot exprimée dans la Définition 21 pour obtenir :

$$\forall \pi, M(S, U, V) \geq \frac{\|\pi\|_2^2}{\max_{1 \leq v \leq q} \nu_v} \geq \frac{\|\pi^+\|_2^2}{\max_{1 \leq v \leq q} \nu_v} \quad (4.11)$$

La borne inférieure est valable quelle que soit la partition des tâches parmi les équipes : rien ne peut justifier une performance pire (sous le prisme de la performance one-shot) pour réaliser la tâche élue.

Remarque 18 (Séparation des masses).

Pour finir, nous remarquons que permettre à V d'être aléatoire conditionnellement à U est une généralisation de la partition identique à ce que fut l'introduction du problème de Monge-Kantorovitch par rapport au problème de Monge : nous permettons l'allocation d'une masse à plusieurs bureaux.

Dans le problème initial de partitionnement, nous utilisons une partition \mathcal{A} et non une seconde variable V , ne permettant donc aucune rupture de masse ; aussi, $\mathbb{P}_V = \nu$ n'est pas proprement défini. Extrayons le de la partition \mathcal{A} en affectant à chaque bureau la somme des probabilités des tâches qui lui sont affectées. Formellement, nous définissons $V(\mathcal{A})$ comme la variable aléatoire de loi de probabilité :

$$\nu_v^{\mathcal{A}} = \mathbb{P}_{V(\mathcal{A})}(V(\mathcal{A}) = v) = \sum_{u \in \mathcal{A}_v} \mu_u;$$

ainsi que la loi du couple :

$$\pi_{u,v}^{\mathcal{A}} = \mathbb{P}_{U, V(\mathcal{A})}(U = u, V(\mathcal{A}) = v) = \mu_u \mathbb{1}_{u \in \mathcal{A}_v};$$

ce qui nous permet d'affirmer que la probabilité de satisfaire la tâche élue du premier coup dans un problème de partitionnement accepte la borne inférieure :

$$M(S, U, \mathcal{A}) = M(S, U, V(\mathcal{A})) \geq \frac{\|\pi^{\mathcal{A}}\|_2^2}{\max_{1 \leq v \leq q} \nu_v^{\mathcal{A}}} \geq \frac{\|C^+(\mu, \nu^{\mathcal{A}})\|_2^2}{\max_{1 \leq v \leq q} \nu_v^{\mathcal{A}}}.$$

De fait, un problème de partitionnement se révèle être un couplage de U et $V(\mathcal{A})$ (avec $V(\mathcal{A})|U$ déterministe) et nous savons qu'aucun couplage ne conduit à une pire performance one-shot que $C^+(U, V(\mathcal{A}))$. Une application directe est qu'aucune affectation des tâches ne saurait justifier une performance pire.

L'efficacité de l'*indétermination* dans le problème de l'espion comme dans celui du partitionnement provient directement de sa capacité à réduire les collisions. Ou bien il limite les capacités de l'espion ou bien il fournit la pire affectation possible des tâches livrant ainsi les bornes inférieures afférentes.

4.3 Clustering de graphes

4.3.1 Introduction

Lorsque le papier [41] est publié, la communauté scientifique autour de la détection de communautés dans des graphes est encore balbutiante. Le problème de trouver des

regroupements dans des graphes est justifié par la remarque initiale que, quelle que soit l'application, les connexions entre les sommets ne sauraient être aléatoires. Ainsi, dans la représentation des liens sociaux, un triangle a plus de chance d'être une clique puisque, comme exprimé dans l'article, "deux de nos amis ont une chance supérieure d'être également amis, *a minima* de se connaître, que deux personnes tirées au hasard sur la planète et ce précisément parce qu'ils nous connaissent tous les deux". En élargissant cette remarque, le problème est de regrouper les individus dans un graphe en fonction de leurs connexions, celles-ci ne représentent pas toujours un lien de connaissance mais la remarque reste valable. On souhaite donc réunir les individus, les sommets, similaires. Bien, comment définissons-nous la similarité et d'ailleurs, en la supposant quantifiée, où poser le seuil de regroupement ?

La première méthode, consiste à associer un poids $w_{i,j}$ à chaque couple de sommets (i, j) qui exprime leur ressemblance dans le graphe. Typiquement, le nombre de chemins entre eux dans celui-ci, avec une décroissance exponentielle suivant la longueur. Une fois w défini, les sommets sont réunis hiérarchiquement suivant w décroissant et jusqu'à un seuil arbitraire.

Dans l'article [41] qui nous intéresse, ils se démarquent de cette approche et se focalisent sur les nœuds entre communautés. Ainsi, ils définissent une mesure dite de "betweenness" associée au nombre de plus courts chemins passant par un sommet.

Deux années passèrent avant que les mêmes auteurs Girvan et Newman dont le nom devait rester associé au problème même du clustering (*i.e.* détection de communautés), proposent une méthode pour évaluer la qualité globale d'un clustering : la modularité ([76]). L'idée rejoint l'exemple initial du triangle d'amitié : une comparaison entre le lien effectif et la probabilité normale, générale, d'être lié.

Ainsi, ils comparent le nombre d'arêtes dans une communauté avec le nombre attendu. Ce nombre attendu est quelque peu arbitraire mais fait consensus, il s'agit du nombre espéré si les arêtes étaient distribuées au hasard sans regard pour les communautés d'appartenance. Avec les notations historiques, la modularité s'écrit :

$$Q = \sum_{i=1}^k (e_{i,i} - a_i^2)$$

où k désigne le nombre de communautés, $e_{i,j}$ la proportion des arêtes du graphe qui lient les communautés i et j , ($e_{i,i}$ comptant donc les arêtes dans la communauté i) et a_i la somme des arêtes connectés à ladite communauté. De fait, si les arêtes sont distribuées proportionnellement sans regard pour la communauté, le nombre espéré au sein de la communauté i est a_i^2 . Nous comptons donc les arêtes internes et les comparons à une situation d'équilibre décrite par l'indépendance ; plus cet écart est élevé, c'est-à-dire plus Q est grande et meilleur est le clustering. Cette relation entre Q est la qualité du clustering est illustrée dans de nombreuses applications aux graphes usuels, les mêmes que ceux de la section 4.3.5.

Cette mesure globale de la performance, fût-elle arbitraire, apporta à la communauté scientifique un objectif quantifiable pour se comparer. Arrivant de concert avec l'explosion des graphes en ligne, elle participa à l'engouement illustré par le foisonnement des sujets que nous mentionnons ci-après.

Il faut malgré tout remarquer que de nombreux problèmes subsistent. Déjà, trouver le clustering optimal sous Q est un problème *NP-complexe* ce qui justifie l'existence

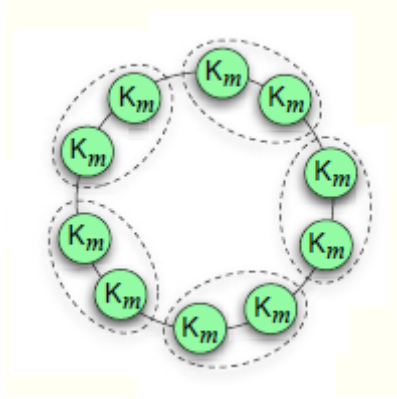


FIGURE 4.1 – Illustration de la limite de résolution

d'une liste d'heuristiques [85], l'utilisation de méthodes spectrales [34] ou encore des optimisation des structures de données [96]. La plupart de ces méthodes sont résumées dans le livre de Fortunato [36].

De surcroît, comme déjà mentionné, Q est aussi simple qu'arbitraire et n'est pas exempte de défauts. Ainsi un problème dit de limite de résolution est mentionné dès l'article [37] de 2007 et du même Fortunato ouvrant la brèche à un domaine de recherche autour de l'amélioration de la modularité. Deux écoles s'affrontent alors, les bâtisseurs et les révolutionnaires. Ou bien les auteurs s'appliquent à réparer la modularité comme dans [83] ou encore avec la "modularity density" récemment introduite [23], ou bien ils changent complètement le critère en utilisant par exemple un coefficient de clustering [75].

Enfin, la définition elle-même d'un clustering est sujette à discussion. Les communautés doivent-elles réaliser une partition ou peuvent-elles avoir une intersection non vide ? Quelques articles choisissent la deuxième option ce qui les fait définir des communautés dites "fuzzy", disons floues (voir [43]).

En résumé, au vu de la littérature, trois sujets émergent : des algorithmes pour optimiser un critère, la définition du critère le plus pertinent, l'extension du problème lui-même.

L'utilisation d'un critère consensuel comme la modularité facilite la comparaison objective (si l'on peut dire) mais s'accompagne de défauts. Il est même probable que plus un critère est utilisé et plus on lui découvre des limites. Comme déjà mentionné, un problème typique est celui de la limite de résolution [37] illustré dans la figure 4.1.

Avec K_m une clique de taille m (typique exemple de communauté puisque tout est connecté), l'article montre que si leur nombre augmente, Q finit par être maximale lorsque deux cliques sont réunies ; d'où le nom de "limite de résolution" puisque la modularité originelle ne sait distinguer des cliques trop petites par rapport à la taille globale du graphe.

Évidemment, la communauté scientifique s'empessa de corriger la définition de la modularité pour dépasser le problème. Ainsi, dans [83] les auteurs changent légèrement

la formule en introduisant un paramètre λ passant de

$$Q(C) = \sum_{c \in C} \left(\frac{a_{in}^c}{2M} - \frac{a_{tot}^c}{2M} \right)^2$$

à

$$Q(C, \lambda) = \sum_{c \in C} \left(\frac{a_{in}^c}{2M} - \lambda \frac{a_{tot}^c}{2M} \right)^2$$

avec a le poids reliant les sommets, qu'il soit inhérent au graphe (poids sur les arêtes) ou rajouté pour les besoins du clustering (nombre de chemins pondérés par exemple) et C les classes dont on évalue la pertinence

L'idée sous-jacente est que plus λ est élevé et plus Q_λ pénalise les classes ayant de nombreux sommets : cela revient essentiellement à ajuster la limite de résolution et, au final, déplace le problème.

De fait, quelques années plus tard, dans [60], ils expérimentent la modularité $Q(\cdot, \lambda)$ sur plusieurs graphes et observent deux biais caractéristiques. D'abord une tendance à fusionner les petites communautés et ensuite à séparer les plus grosses. De surcroît, comme on pouvait l'imaginer, ils constatent que λ introduit un paramètre difficile à tuner.

Plus récemment, dans [23], une nouvelle métrique appelée "modularity density" a été définie pour dépasser ces difficultés ; sa formule est bien plus compliquée et présente un premier inconvénient : elle est peu canonique. Nous la rapportons ici. Rappelons-nous que Q est définie par :

$$Q(C) = \sum_{c \in C} \left[\frac{a_c^{in}}{a} - \left(\frac{2a_c^{in} + 2a_c^{out}}{2a} \right)^2 \right]$$

Ils ajoutent une pénalisation sur la séparation des communautés (SP pour "split penalty") en comptant la proportion d'arêtes entre les communautés :

$$SP = \sum_{c_i \in C} \sum_{c_j \in C, c_i \neq c_j} \frac{a_{c_i, c_j}}{2a}$$

qu'ils intègrent dans la modularité :

$$Q_s = Q - SP$$

Puis, notant que Q_s s'accompagne de biais similaires à ceux de Q , ils normalisent en utilisant $d_{c_i} = \frac{2a_{c_i}^{in}}{|c_i||c_i|-1}$ et $d_{c_i, c_j} = \frac{2a_{c_i, c_j}}{|c_i||c_j|}$ qui représentent, pour chaque sous-ensemble, la proportion d'arêtes présente parmi toutes celles possibles (cas d'une clique). Appliquant ces arêtes normalisées dans Q_s , ils aboutissent à l'expression :

$$Q_{ds}(C) = \sum_{c \in C} \left[\frac{a_c^{in}}{a} d_c - \left(\frac{2a_c^{in} + 2a_c^{out}}{2a} d_c \right)^2 \right] - \sum_{c_i \in C} \sum_{c_j \in C, c_i \neq c_j} \frac{a_{c_i, c_j}}{2a} d_{c_i, c_j}$$

Cette nouvelle fonction a été testée dans [15] où ils remarquent un avantage important de Q_{ds} par rapport à Q : étant normalisée, elle est comparable entre deux graphes différents, quand Q compare seulement deux clustering C sur le même graphe. Pour autant, ils montrent également des limites notamment lors de l'application à des arbres.

Si la fonction objectif que nous venons de définir apparaît comme un raffinement de la modularité originelle, certains auteurs décident de complètement changer d'approche. Notamment motivés par le problème de limite de résolution, l'article [75] propose ainsi de reconstruire une fonction objectif. À tout sommet i , est associé un coefficient de clustering dans le graphe G défini par :

$$CC(i, G) = \frac{\sum_{j,k \in \mathcal{N}(i)} a_{j,k}}{d_i(d_i - 1)}$$

et conduisant à un coefficient global CC :

$$CC(G) = \frac{1}{n} \sum_{i \in V} CC(i, G)$$

Pour mesurer la pertinence d'une partition en se basant sur CC , les auteurs restreignent son application lors de chaque sommet i au graphe induit par la communauté qui l'accueille : $CC(i, G) = CC(i, G|_{C_i})$. On calcule ainsi en restreignant les arêtes à celles dans le même cluster que i . L'utilisation du CC est justifiée par sa capacité à mesurer la cohérence locale d'un sommet i par rapport à ses voisins en estimant les interactions qu'il a avec ceux-ci. En effet, nous ne faisons que calculer la proportion d'arêtes présentes entre i et ses voisins.

Comme déjà mentionné, un autre sujet, une fois le consensus établi sur un critère, est de construire une méthode pour optimiser celui-ci. Les solutions vivant dans l'espace fini des partitions, un optimal existe mais aucun algorithme ne le fournit à coup sûr en un temps raisonnable : le problème est NP -complexe.

L'algorithme de Girvan-Newman a été introduit de concert avec la modularité dans [76]. Il peut être appliqué à tout critère généralisant la modularité, nous rentrerons ultérieurement dans ces considérations. Essentiellement, il est basé sur une approche gloutonne puisqu'il consiste à choisir à chaque étape la fusion de communautés qui augmente le plus la modularité. L'initialisation de l'algorithme est triviale : chaque sommet est dans une communauté différente. L'étape de fusion est quant à elle répétée jusqu'à ce qu'aucune amélioration ne soit possible ou qu'un critère d'arrêt soit atteint. Le chemin de fusion emprunté (comprendre la séquence des étapes) ne garantit rien sinon une heuristique pertinente. L'algorithme est en réalité présenté dans [24] où Girvan n'apparaît pas. Comme remarqué dans [96], l'implémentation originale peut être appliquée jusqu'à 500,000 sommets. Certes les moyens techniques se sont depuis améliorés mais les graphes sont également de plus en plus gros. L'article propose déjà une implémentation astucieuse pour accélérer 7 fois l'optimisation originale tout en fournissant des clusters plus pertinents (comprendre avec une modularité supérieure). On observe bien ici la qualité heuristique des résultats, les algorithmes s'accompagnent souvent d'une partie aléatoire, nécessaire pour avancer mais peu reluisante. Plus récemment, [85] regroupe diverses heuristiques pour améliorer la modularité. Les performances progressent, la méthode reste heuristique.

Enfin, le problème lui-même de clustering n'est pas proprement défini. En cherchant une partition des sommets, on impose par exemple que chacun n'appartienne qu'à une classe ce qui n'est pas approprié pour certaines applications. Typiquement, on peut introduire des intersections entre les communautés ou des "fuzzy" communautés avec un poids

d'appartenance. C'est une généralisation des communautés sans intersection car il suffit de poser un seuil sur les poids d'appartenance pour les retrouver. L'article [43] propose une modularité adaptée à ce relâchement des contraintes.

Aujourd'hui, et au vu du foisonnement des articles et des pistes, l'actualité du sujet de la détection de communautés dans un graphe n'est plus à démontrer. L'apparition soudaine de graphes immenses sous les applications modernes : réseaux sociaux, communications téléphoniques, flux financiers est renforcée par notre capacité à les recueillir et à les stocker, faute de pouvoir les appréhender. Cette appréhension, justement, est confrontée à une masse trop importante et trop diluée d'informations. C'est ici qu'apparaît l'idée de regrouper les sommets du graphe en fonction de leur similarité : l'objet même des algorithmes de clustering.

Une première option consiste à fixer un nombre de classes et une distance pour se ramener à trouver des centres optimisant la somme des distances de chaque sommet au centre le plus proche. Cette approche est typiquement illustrée par l'algorithme de *K-means* dont l'idée remonte à [88]. Reste qu'en ayant fixé et le nombre de classes et la distance, le problème reste NP-complexe. Une seconde option est de construire un critère local c qui affecte un poids à chaque couple (i, j) de sommets ; un critère qui, naturellement, est d'autant plus élevé que lesdits sommets sont similaires. S'ensuit la création d'un critère global qui somme les valeurs locales si et seulement si le couple est réuni dans la même classe comme exposé dans le problème 11. Optimiser le critère global demeure NP-complexe d'où l'introduction d'heuristiques comme l'algorithme de Louvain, introduit dans [41] pour un critère canonique d'écart à l'indépendance et qui nous intéressera ici sous une forme plus générale.

4.3.2 Formalisation

Commençons par rappeler les définitions habituelles d'un graphe avant de préciser le problème de clustering et les heuristiques afférentes.

Définition 22 (Graphe pondéré).

Un graphe G est constitué d'un ensemble de n sommets $1 \leq i \leq n$. Entre chaque paire de sommets, un poids $a_{i,j}$ est placé. On note également : $2M = \sum_{i,j} a_{i,j}$.

Un moyen simple de générer aléatoirement un graphe consiste à utiliser la distribution de Edgard Gilbert :

Définition 23 (Gilbert).

On fixe le nombre n de sommets et l'on se donne un nombre $\epsilon \in [0, 1]$. Pour tout ensemble de deux sommets, on tire indépendamment une Bernoulli de paramètre ϵ pour déterminer si un lien existe. Le graphe ainsi obtenu est donc non dirigé et chaque poids vaut 1.

Remarque 19.

L'ajout d'un paramètre p représentant le poids maximum permet aisément de créer un graphe pondéré en tirant le poids selon une loi binomiale $B(\epsilon, p)$. De même, tirer des couples (et non des ensembles) de deux sommets permet de générer des graphes dirigés.

Avant de proposer la méthode de Louvain, nous définissons formellement le partitionnement d'un graphe.

Définition 24 (Partitionnement).

Nous notons x le partitionnement d'un graphe G . $x_{i,j}$ vaut 0 ou 1 et vaut 1 si et seulement si les deux sommets i et j sont dans la même classe pour x qui code donc une relation d'équivalence entre les sommets.

Comme mentionné, le chapitre est dédié à la recherche de classes, groupes, clusters ou cliques (peu importe le nom) au sein d'un graphe et caractérisés par la relation d'équivalence introduite dans la définition 24. Nous prenons le parti de ne pas fixer un nombre de classes et de nous restreindre à la seconde option de l'introduction. Plus précisément, un algorithme de clustering est basé sur l'optimisation du problème 11.

Problème 11 (Problème de clustering générique).

Pour un graphe G et un critère local c :

$$\max_x M(G, c, x) = \sum_{i=1}^n \sum_{j=1}^n c(G)_{i,j} x_{i,j}$$

sous les contraintes :

x est une relation d'équivalence

Comme souvent, nous sommes à la fois cernés par une finitude et une immensité : le nombre de classes d'équivalence sur n éléments est fini mais nous ne saurions les parcourir toutes ; c'est d'ailleurs précisément la contrainte que x soit une relation d'équivalence qui rend le problème si compliqué. Aussi, l'optimal de M existe mais nous ne savons le trouver. Nous pouvons l'approcher et c'est l'objet de l'algorithme de Louvain que nous détaillerons ultérieurement et qui fournit une heuristique d'approche.

Commençons par une remarque importante : il n'est jamais fait mention d'un nombre de classes puisque nous optimisons un critère. C'est le choix de ce critère qui est garant de la qualité du partitionnement trouvé, nonobstant l'obtention d'un résultat sous-optimal lors de l'optimisation de ce dernier.

Algorithme de Louvain

Cet algorithme repose sur deux étapes pour maximiser globalement $M(G, c, x)$ d'après les valeurs locales de c .

Initialement, chaque nœud du graphe est isolé dans sa propre communauté.

Ensuite, pour chaque nœud i , on évalue l'impact sur M de retirer i de sa communauté pour l'insérer dans une de ses voisins. Si un des changements est bénéfique (comprendre, augmente M), on effectue le meilleur. Le procédé est appliqué successivement (dans l'ordre de numérotation des sommets) et ce jusqu'à ce qu'aucune optimisation ne soit possible. L'atteinte de ce maximum local de M signe l'arrêt de la première étape.

Dans la seconde étape, tous le graphe est mis à jour : les sommets sont les classes finales de l'étape première. Les liens, quant à eux sont créés à partir des liens entre les sommets inter classes et ce en sommant les poids ; les liens intra classes créant des boucles.

Une fois le nouveau graphe généré, la première étape est relancée sur celui-ci.

L'algorithme s'arrête lorsque la différence sur $M(G, c, x)$ apportée en une étape est moindre qu'une valeur fixée en amont.

L'algorithme de Louvain est une heuristique pour plusieurs raisons. D'abord, il ne fournit pas un optimal; en cela il s'approche de *K-means* mais sans fixer le nombre de classes. Deuxièmement, il est dépendant de la numérotation des sommets puisqu'il les parcourt "dans l'ordre". On sait qu'un optimum global existe et ce quel que soit c , mais Louvain ne le trouvera pas systématiquement. Quoi qu'il en soit, les solutions apportées sont généralement suffisantes et proches de l'optimum. Cet aspect de performance de l'algorithme de Louvain, bien qu'intéressant en soi, ne sera pas l'objet de ce chapitre. Nous baserons ce dernier sur l'étude de la différence entre deux critères c canoniques à la lumière des chapitres passés.

4.3.3 Modularité originelle – indépendance

Les papiers de Newman-Girvan ([41] ou [76]) présentent un critère global canonique appelé modularité en sus de l'algorithme pour l'optimiser. Nous le précisons dans la définition 25 avant de l'interpréter.

Définition 25 (Modularité).

Étant donné un graphe G et une partition x , de fonction de poids a sur ses arêtes, la modularité s'exprime par :

$$M^\times(G, x) = \frac{1}{2M} \sum_{i,j} \left[a_{i,j} - \frac{a_{i,\cdot} a_{\cdot,j}}{2M} \right] x_{i,j}$$

Déjà, on remarque que cette modularité originale, rentre dans le cadre de notre fonction de coût générique exprimée dans le Problème 11 avec :

$$c(G)_{i,j} = m^\times(G)_{i,j} = \frac{a_{i,j}}{2M} - \frac{a_{i,\cdot} a_{\cdot,j}}{(2M)^2}$$

et que le gain local $m^\times(G)_{i,j}$ à mettre deux sommets dans la même classe n'est rien d'autre qu'un écart canonique à l'indépendance. De fait, avec la définition 22, on sait que $\pi_{i,j} = \frac{a_{i,j}}{2M}$ peut être vu comme une mesure de probabilité sur $\{1, \dots, n\}^2$ dont les marges seraient toutes les deux égales à $\mu_i = \frac{a_{i,\cdot}}{2M}$ de sorte que m^\times s'écrit encore :

$$m^\times(G)_{i,j} = 2M * (\pi_{i,j} - \mu_i \mu_j)$$

et s'exprime bien comme un critère de *déviaton canonique à l'indépendance*.

Une seconde remarque est que l'expression $m^\times(G)_{i,j}$ ne contient ni valeur absolue ni élévation au carré : des sommets non connectés entraîneront un gain négatif ce qui les empêchera d'être réunis dans la même classe; un comportement décent. S'ils sont connectés au contraire, l'importance du gain $m^\times(G)_{i,j}$ évolue positivement selon que i et j ont peu de connexions ($a_{i,\cdot}$ et $a_{\cdot,j}$ petits); là encore un comportement décent. Plus précisément, l'indépendance assurant un couplage aussi uniforme que possible dans le respect des marges (comme solution du problème 5), cela justifie la construction m^\times : le critère mesure une distance entre le poids observé et une distribution uniforme en accord avec le voisinage.

4.3.4 Modularité déformée – indétermination

Le problème 11 représente une extension de la modularité historique. Suivant la justification précédente du critère originel, on propose naturellement un critère $m^+(G)_{i,j}$ de *déviaton canonique à l'indétermination* ; un couplage qui, comme solution du problème 6, couple également aussi uniformément que possible dans le respect des marges. Une fois inséré dans la fonction de gain globale du problème 11, on obtient $M^+(G, x)$ qui optimise localement :

$$m^+(G)_{i,j} = a_{i,j} - \frac{a_{i,\cdot}}{n} - \frac{a_{\cdot,j}}{n} + \frac{2M}{n^2}$$

Symétriquement, m^+ , s'exprime comme un critère de *déviaton à l'indétermination*. De fait, avec les notations précédentes, m^+ s'écrit également :

$$m^+(G)_{i,j} = 2M * \left(\pi_{i,j} - \frac{\mu_i}{n} - \frac{\mu_j}{n} + \frac{1}{n^2} \right)$$

La section 2.4.3, a montré que la différence quadratique entre les deux couplages avait tendance à être faible. De plus, ils partagent un nombre important de propriétés comme indiqué dans le chapitre 3. Enfin, les deux critères respectent la même heuristique de croissance : ils sont importants lorsque le lien entre les deux sommets existe et que ces deux sommets ont un voisinage faible. Par ailleurs, Patricia Conde-Céspedes, dans sa thèse [25] a listé un nombre important de critères apparaissant comme un écart à l'indépendance ou à l'indétermination. Nous ne traitons ici que les deux canoniques, juste introduits en commençant par rappeler ses résultats expérimentaux.

4.3.5 Résumé de l'application expérimentale à des graphes variés

La construction parallèle du chapitre 2 sur les problèmes de transport ainsi que notre connaissance de la proximité des critères nous permettent d'attendre un comportement similaire de ceux-ci sur des graphes typiques. Nous aurons l'occasion de préciser cette pensée dans la section 4.3.6. Pour l'instant, nous rappelons les résultats de Patricia Conde-Céspedes et notons que notre intuition est validée.

La table 4.1 collecte le nombre de classes trouvé par Patricia Conde-Céspedes lors de l'application des deux critères à des graphes empiriques communs. Elle remarqua des résultats proches, ces mêmes résultats que nous pressentons et confirmerons. Introduisons, comme elle le fit, la liste des graphes utilisés.

- réseau social "Zachary karate club" : fréquemment utilisé dans l'analyse de réseaux sociaux, il est composé des 34 membres d'un club de karaté d'une université américaine (voir [103]).
- réseau social "American College Football" : collecte les matchs entre équipes américaines durant l'année 2000. Chaque sommet est une équipe et chaque lien est un match (voir [41]).
- réseau social "Jazz" : représente les collaborations entre les musiciens de jazz de l'année 1912 à l'année 1940. Chaque sommet est un groupe et ils sont connectés s'ils ont partagé un musicien. La donnée est extraite de la "The Red Hot Jazz Archive" (voir [42]).
- "Internet" : un sous-graphe de l'Internet (voir [48]).

TABLE 4.1 – Nombre de classes trouvé par chaque critère sur des graphes variés

	Karaté	Football	Jazz	Internet	Amazon	YouTube
N (nombre sommets)	34	115	198	69 949	334 863	1 134 890
M (somme des poids)	78	613	2 742	351 280	925 872	2 987 624
Nombre de classes sous M^\times	4	10	4	46	250	5 567
Nombre de classes sous M^+	4	10	6	39	246	13 985

- "Amazon" : extrait de Amazon.com. Chaque sommet représente un produit et ils sont connectés s'ils sont fréquemment achetés ensemble. (voir [101]).
- "Youtube" : chaque sommet est un utilisateur. Sur Youtube, les utilisateurs peuvent créer des groupes ; deux utilisateurs sont connectés dans le graphe s'ils ont rejoint le même groupe (voir [72]).

La table 4.1 se lit ainsi : le graphe "Internet" contient 69,949 sommets reliés par 351,280 liens ; si on applique l'algorithme de Louvain dessus avec le critère global M^\times on trouve en moyenne 46 communautés quand M^+ conduit à 39.

Comme anticipé, les deux critères sont, en moyenne, si proches que leur effet sur des graphes communs est semblable. La section 4.3.6 fournit une tentative d'explication théorique, *a minima* rationalise le résultat. La section 4.3.7, au contraire, relativise l'affirmation en proposant des graphes sur lesquels les deux critères ont des effets drastiquement différents.

4.3.6 Comparaison dans le cadre Gilbert

Comme déjà mentionné, résoudre le problème 11 est NP-complexe de sorte que nous n'espérons pas de forme close du nombre de classes pour un critère fixé, ni d'ailleurs la prédiction du déroulement de l'algorithme de Louvain sur un graphe de taille arbitraire pour ce même critère. Aussi nous nous en remettons à l'étude locale de chacun des deux critères pour extrapoler des différences ou des similitudes sur un comportement global.

On propose une expérience basée sur les graphes de Gilbert déjà introduits, pour remarquer des différences ou des similarités entre les valeurs de $m^\times(G)_{i,j}$ et $m^+(G)_{i,j}$. L'objectif est de trouver la distribution typique de chaque critère. Déjà, et ce afin de simplifier les observations, on remarque que seul le coût de référence varie entre les deux expressions m^+ et m^\times , nous le définissons formellement dans la définition 26. Aussi, nous générons des graphes suivant la distribution de Gilbert puis calculons et stockons le coût de référence sur des couples de sommets tirés au hasard au sein des graphes obtenus. L'expérience est spécifiée dans l'algorithme 3 tandis que les résultats sont recueillis et affichés dans la figure 4.2.

Définition 26 (Biais ou coût de référence).

Les deux coûts de référence extraits de m^\times et m^+ sont respectivement :

- $b_{i,j}^\times = \frac{a_{i,\cdot} a_{\cdot,j}}{2M}$
- $b_{i,j}^+ = \frac{a_{i,\cdot}}{n} + \frac{a_{\cdot,j}}{n} - \frac{2M}{n^2}$

La figure 4.2 montre deux distribution quasiment identiques. De fait, les courbes sont identiques pour toutes les valeurs fréquentes (disons avec plus de 200 réalisations). Ce

Algorithm 3 Fournit les distributions empiriques des deux coûts de référence

Input n

Input ϵ

$L_+ \leftarrow []$

$L_\times \leftarrow []$

for $R = 1 \dots 10000$ **do**

$G \leftarrow \text{Gilbert}(n, \epsilon)$

$(i, j) \leftarrow (\text{Uniforme}(n), \text{Uniforme}(n))$

$L_+ \leftarrow L_+ + (b_{i,j}^+(G))$

$L_\times \leftarrow L_\times + (b_{i,j}^\times(G))$

end for

return (L_+, L_\times)

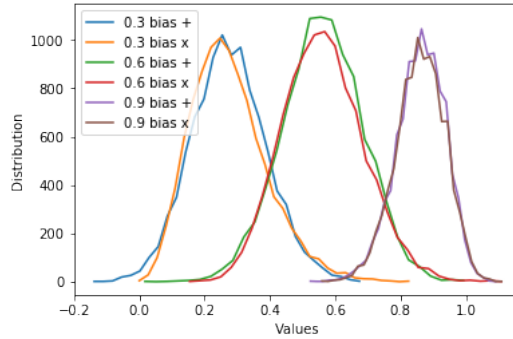


FIGURE 4.2 – Distributions empiriques des deux coûts de référence $b_{i,j}^+$ et $b_{i,j}^\times$ pour ϵ dans $[0.3, 0.6, 0.9]$

n'est pas étonnant puisque les deux optimisent des problèmes similaires et que l'on sait que l'espérance quadratique de leur différence est très faible (section 2.4.3). Nous notons également que leur moyenne est commune et égale à ϵ ; cela s'obtient aisément à partir de leurs formules. Une différence observable (*a minima* pour $\epsilon = 0.3$ et $\epsilon = 0.6$) est la tendance de b^+ aux valeurs extrêmes faibles et celle de b^\times aux valeurs extrêmes élevées.

Pour confirmer nos observations et la symétrie, nous calculons les distributions théoriques des deux biais sous Gilbert. Les valeurs de $m^\times(G)_{i,j}$ comme celles de $m^+(G)_{i,j}$ ne dépendent que des valeurs sous-jacentes de $a_{i,j}$, $a_{i,\cdot}$ et $a_{\cdot,j}$. De plus, nous pouvons facilement estimer la probabilité de la réalisation d'un événement comme indiqué dans la proposition 17.

Propriété 17 (Probabilités).

En fixant une valeur binaire b , $b \leq n_i \leq n$ et $b \leq n_j \leq n$ on obtient :

$$\begin{aligned} & \mathbb{P}(a_{i,j} = b, a_{i,\cdot} = n_i, a_{\cdot,i} = n_j) \\ &= \epsilon^b (1-\epsilon)^{1-b} \binom{n-1}{n_i-b} \epsilon^{n_i-b} (1-\epsilon)^{n-1-n_i+b} \binom{n-1}{n_j-b} \epsilon^{n_j-b} (1-\epsilon)^{n-1-n_j+b} \end{aligned}$$

Les deux valeurs $m_{i,j}^+$ et $m_{i,j}^\times$ correspondant à un groupe (b, n_i, n_j) des paramètres étant

évidentes, nous proposons la figure 4.3 qui affiche la différence entre les deux distributions pour $\epsilon = 0.3$.

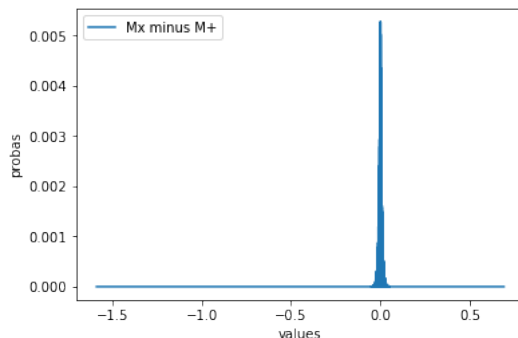


FIGURE 4.3 – distribution théorique de l'écart entre $m^\times(G)_{i,j}$ et $m^+(G)_{i,j}$

b^\times et b^+ ont des formes distinctes mais sont très proches dès que les valeurs sont probables. C'est une illustration de la section 2.4.3 : couplant deux variables à n modalités, la différence attendue moyenne est de $\frac{1}{n^2}$. La différence sur les valeurs extrêmes, elle peut être arbitrairement élevée et on l'observe sur l'empirique comme sur le théorique.

Ayant noté les différences sur les valeurs extrêmes, nous calculons celles-ci sur un graphe Gilbert générique (respectant les valeurs communes $2M = n^2\epsilon$) pour obtenir :

$$\begin{aligned} -\epsilon &\leq b^+ \leq \frac{n}{n} + \frac{n}{n} - \epsilon = 2 - \epsilon \\ 0 &\leq b^\times \leq \frac{n * n}{n * 2\epsilon} = \frac{1}{\epsilon} \end{aligned}$$

Comme attendu, depuis la figure 4.3, la différence est arbitrairement grande. Nous illustrons cette propriété au travers de la construction de graphes spécifiques.

4.3.7 Graphes spécifiques

Bien que la section 4.3.6 ait conclu à un comportement moyen similaire entre les deux critères canoniques M^+ et M^\times , encore renforcé par les expériences menées par Patricia Conde-Céspedes et résumées dans la section 4.3.5, cela ne les empêche pas de présenter des comportements spécifiques très différents ; ce sera l'objet de la présente section. Pour commencer, nous rappelons que, quel que soit le graphe initial, durant le déroulement de l'algorithme de Louvain, la mise à jour du graphe créera probablement des boucles et des poids supérieurs à 1 : les poids entre les sommets initiaux fusionnés dans une même classe par l'étape deux. Aussi, nous permettons à nos exemples de contenir les deux sans perte de généralité aucune.

En analysant les biais locaux introduits par la définition 26, on remarque que la forme produit b^\times sera petite sauf si $a_{i,\cdot}$ ET $a_{\cdot,j}$ sont grands ; de son côté, la forme additive b^+ sera petite sauf si $a_{i,\cdot}$ OU $a_{\cdot,j}$ est grand. Rappelons que le gain local m ne conduit à une fusion que s'il est suffisamment grand (*a minima* positif). En résumé :

- m^\times est pénalisé (par b^\times) si $a_{i,\cdot}$ ET $a_{\cdot,j}$ sont grands.
- m^+ est pénalisé (par b^+) dès que $a_{i,\cdot}$ OU $a_{\cdot,j}$ est grand.

Pour maximiser la forme additive aucun des deux voisinages ne peut être grand, tandis que la forme multiplicative permet d'en accepter un des deux. Utilisons cette remarque pour créer des graphes différenciant comme présenté ci-dessous.

Graphe étoilé

Le graphe spécifique que l'on présente dans la figure 4.4 contient un nœud particulier (disons le nœud 1) avec une boucle de poids 2 tandis que tous les autres sommets lui sont connectés par un poids de 1 en étoile autour.

L'intuition derrière est que, pour $m^\times(G)_{i,j}$, un sommet avec un faible poids $a_{i,j}$ peut annuler le poids élevé d'un autre a_i . grâce à la forme multiplicative, rendre ainsi $b_{i,j}^\times$ petit et donc conduire à un gain positif. Au contraire, la forme additive de $m^+(G)_{i,j}$ ne permettra pas cette correction : $b_{i,j}^+$ sera grand et donc $m_{i,j}^+$ négatif.

Au final, dans le graphe de la figure 4.4 on crée un $m^\times(G)_{i,1}$ positif pour un $m^+(G)_{i,1}$ négatif. Résultat, tout est fusionné dans une classe pour l'indépendance (\times) tandis que l'on garde le graphe tel quel avec l'indétermination (+) : un gain négatif ne saurait provoquer une fusion. Les calculs afférents :

- $m_{i,1}^+ = 1 - 1/10 - 11/10 + 20/100 = 0 \leq 0$;
- $m_{i,1}^\times = 1 - 1 * 11/20 = 0.45 > 0$;

montrent bien la variation de signe selon le critère : M^\times et M^+ produiront respectivement 1 et 10 classes.

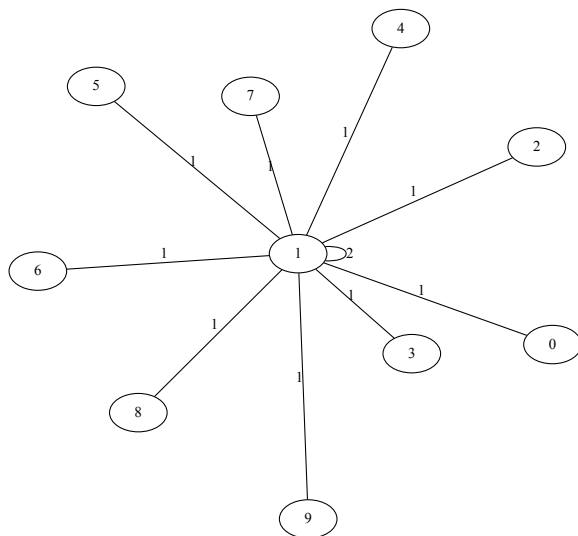


FIGURE 4.4 – Graphe étoilé fusionné par M^\times et conservé par M^+ . ($M^+ : \{0\}, \{1\}, \{2\}, \{3\}, \{4\}, \{5\}, \{6\}, \{7\}, \{8\}, \{9\}$, $M^\times : \{0, 1, 2, 3, 4, 5, 6, 7, 8, 9\}$)

Fusion de sommets non connectés

Le second graphe spécifique proposé, celui de la figure 4.5 est un cas miroir du premier : fusionné seulement par M^+ . L'intuition est que m^+ dépend de la distribution moyenne

des poids tandis que m^\times y est agnostique ; remarquer l'utilisation de m minuscule donc local.

L'exemple que nous construisons est directement inspiré de l'hypothèse 2.9 qui garantit la positivité du couplage d'indétermination, et donc du biais b^+ . Ici, le couplage est utilisé dans un cadre particulier avec une masse totale de $2M$ (individus) et n modalités sur chaque marge. Rien dans le graphe ne garantit la positivité : il faudrait essentiellement, si l'on reprend la construction de la proposition 5, empêcher les sommets d'avoir des poids trop faibles. Mais, ici, seule la formule compte et b^+ n'a pas à être une loi de probabilité.

Ignorons donc l'hypothèse 2.9. De fait, il est probable qu'un couple (i, j) (modalités (u, v) dans la formule du couplage) génère une formule d'indétermination négative. Ce qui est intéressant, c'est que, dans ce cas, ces deux sommets pourront être fusionnés sans être connectés puisque $m^+(G)_{i,j} \geq b_{i,j}^+ > 0$. Ce sera le cas pour des sommets globalement peu voire pas connectés car sinon le biais sera vite positif. En résumé, les sommets isolés (comprendre moins connectés que le reste) pourront être fusionnés et ce même sans être connectés initialement. Bien que s'agissant de sommets dégénérés, ce comportement sera probable sur un grand graphe : peu de sommets auront un b^+ négatif mais beaucoup de graphes contiendront de tels sommets. En particulier, la fin de la section 4.3.6 a montré que m^+ atteignait probablement la valeur $-\epsilon$ dans le cas de graphes de Gilbert.

Ayant émis ces remarques, nous proposons le graphe de la figure 4.5. Le nœud 0 déforme la distribution des poids afin de permettre la fusion de "petits" sommets par M^+ . Les calculs donnent pour $i \neq 0$ et $j \neq 0$:

- $m_{i,j}^+ = 0 - 1/5 - 1/5 + 14/25 = 0.16 > 0$
- $m_{0,j}^+ = 0 - 10/5 - 1/5 + 14/5 = -1.64 < 0$
- $m_{i,j}^\times = 0 - 1/14 < 0$
- $m_{0,j}^\times = 0 - 10/14 < 0$

Aussi, suivant M^+ , tous les sommets sauf 0 sont fusionnés ; le critère m^+ est sensible à la distribution globale des poids et distingue les "petits" sommets tandis que m^\times ne le peut. Au contraire, selon M^\times , aucun m^\times n'est positif et le graphe est laissé tel quel. Finalement, les partitions sont $M^+ : [\{0\}, \{1, 2, 3, 4\}]$, $M^\times : [\{0\}, \{1\}, \{2\}, \{3\}, \{4\}]$.

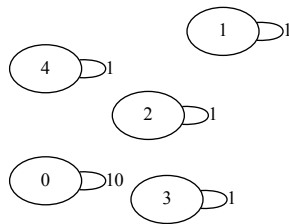


FIGURE 4.5 – Graphe dans lequel M^+ fusionne les "petits" sommets. ($M^+ : [\{0\}, \{1, 2, 3, 4\}]$, $M^\times : [\{0\}, \{1\}, \{2\}, \{3\}, \{4\}]$)

Préférence pour des sommets non connectés

Nous pouvons aller plus loin et exhiber un contexte dans lequel le critère M^+ choisira de fusionner un petit sommet avec un autre non connecté plutôt qu'un gros auquel il est connecté. Utilisons pour cela le graphe de la figure 4.6 dans lequel 0 et 1 sont connectés tandis que 2 est isolé et montrons que la meilleure option selon m^+ est de fusionner les sommets *non connectés* 1 et 2; on calcule aussi la valeur des deux critères locaux m^+ et m^\times :

- $m_{1,0}^+ = 1 - 2/3 - 6/3 + 8/9 = -7/9 < 0$
- $m_{1,2}^+ = 0 - 2/3 - 0 + 8/9 = 2/9 > 0$
- $m_{2,0}^+ = 0 - 6/3 - 0 + 8/9 = -10/9 < 0$
- $m_{1,0}^\times = 1 - 2 * 6/8 = -0.5 < 0$
- $m_{1,2}^\times = 0$
- $m_{2,0}^\times = 0$

Ainsi, en appliquant les deux critères au graphe de la figure 4.6, on génère les partitions $M^+ : [\{0\}, \{1, 2\}]$, $M^\times : [\{0\}, \{1\}, \{2\}]$.

Les classes finales sont très surprenantes pour M^+ puisque la distribution moyenne des critères entraîne un biais tellement négatif que m^+ conduit à fusionner un petit sommet à un autre petit non connecté plutôt qu'à un gros auquel il est connecté. Notons par ailleurs, que M^\times ne fusionne ni 0 ni 1 : c'est précisément causé par la boucle sur 1 qui permet au produit $a_{0,a_{\cdot,1}}$ d'être assez important.

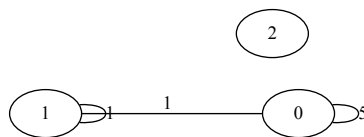


FIGURE 4.6 – Graphe conduisant M^+ à préférer fusionner des sommets non connectés. M^\times laisse le graphe tel quel. ($M^+ : [\{0\}, \{1, 2\}]$, $M^\times : [\{0\}, \{1\}, \{2\}]$)

4.3.8 Un seuil commun

Dans la section 4.3.7 précédente, nous avons fourni une liste de graphes spécifiques sur lesquels les deux critères étaient très distinguables. Tous résultaient d'une différence notée entre les deux formules m^\times et m^+ ; au final entre les deux biais b^\times et b^+ .

Au contraire, nous présentons dans cette section 4.3.8 une curiosité : une forme de graphe pour laquelle les deux critères présentent un seuil de fusion qui est, de surcroît, identique. Cet exemple présente un intérêt particulier, outre notre capacité à fixer un seuil, il fournit un entraînement pour une analyse plus générale.

On travaille sur une boucle de n classes comme celle de la figure 4.7 (pour laquelle $n = 10$) et l'on cherche un seuil sur le poids $a_{i,i}^n$ de la boucle (comme unique paramètre) au delà duquel le graphe est laissé intact par M^\times . L'intuition est que le biais b^\times sera suffisamment grand pour annihiler tous les poids de 1. Comme tous les sommets ont un

environnement identique, on sélectionne n'importe lequel des biais $b_{i,j}$: tous sont égaux. En comptant les liens, et en imposant donc $b_{i,j}^\times = b^\times \geq 1$ pour qu'aucune fusion n'ait lieu on obtient :

$$\begin{aligned} 2M &= n * a_{i,\cdot} \\ a_{i,\cdot} &= 2 + a_{i,i}^n \\ \frac{a_{i,\cdot}^2}{2M} &\geq 1 \end{aligned}$$

puis, en résolvant l'équation du second degré ($x^2 + (4-n)x + 2(2-n)$), il vient $a_{i,i}^n \geq n-2$. Dans notre exemple, $n = 10$, de sorte que le seuil est 8 ce qui explique que la figure 4.7 soit un graphe final convenable pour M^\times .

Si on s'intéresse maintenant au comportement de M^+ sur ce même graphe, on observe que l'équation du seuil est :

$$\begin{aligned} b_{i,j}^+ &\geq 1 \\ 2 * \frac{a_{i,\cdot}}{n} - \frac{2M}{n^2} &\geq 1 \\ 2 * \frac{a_{i,i}^n + 2}{n} - \frac{n(a_{i,i}^n + 2)}{n^2} &\geq 1 \\ a_{i,i}^n &\geq n - 2 \end{aligned}$$

Ainsi, et tandis que la forme du graphe a été inventée pour que m^\times ait un seuil, il apparaît que non seulement m^+ en a également un mais que les deux sont égaux. Les propriétés symétriques des deux couplages se retrouvent ici, une nouvelle fois, dans leur application au clustering.

Remarque 20 (Une curiosité?).

Comme mentionné dans la remarque 2.4.3, les deux formules sont égales dès que l'une des marges est uniforme. Or, vu la forme du graphe de la figure 4.7, tous les sommets possèdent le même environnement, de sorte qu'une distribution basée sur ce dernier ne peut être qu'uniforme; cela explique le résultat et en retire toute curiosité.

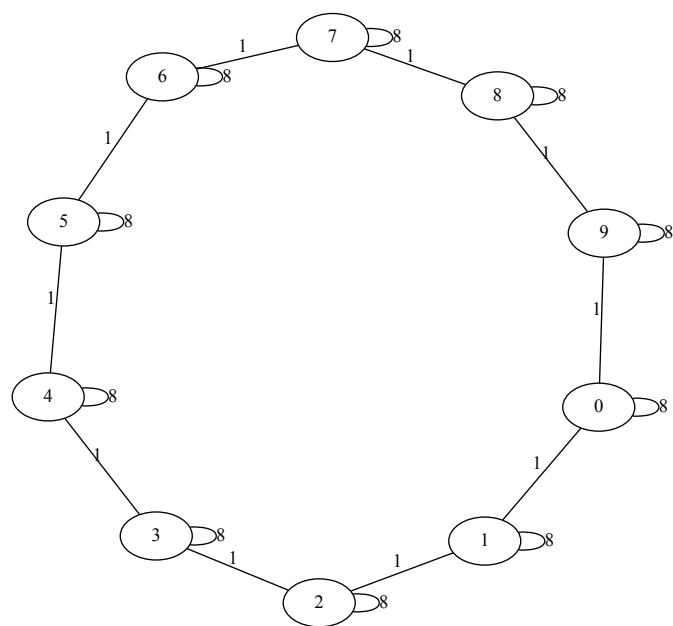


FIGURE 4.7 – Exemple d'un graphe final convenable pour M^\times comme pour M^+

Coefficients de corrélation

Tu as accru la liberté humaine
au lieu de la confisquer et tu as
ainsi imposé pour toujours à
l'être moral les affres de cette
liberté. Tu voulais être
librement aimé [...] mais ne
prévoyais-tu pas qu'il
repousserait enfin et contesterait
même ton image et ta vérité,
étant accablé sous ce fardeau
terrible : la liberté de choisir ?

Fiodor Dostoïevski, *Les Frères
Karamazov*, 1880

Les coefficients de corrélation fournissent un moyen de mesurer la dépendance entre deux variables aléatoires en résumant celle-ci par un réel souvent entre 0 et 1 calculé sur un jeu de réalisation des deux variables sur le même ensemble d'individus ; une liste importante est fournie dans [86]. Formellement, U est une première variable aléatoire sur p modalités de loi μ tandis que V est une seconde variable aléatoire sur q modalités de loi ν . La loi du couple est quant à elle notée π et la variable aléatoire correspondante W . On se donne alors n réalisations de $W : (W_1 = (U_1, V_1), \dots, W_n = (U_n, V_n))$ desquelles on veut extraire une estimation de la corrélation entre U et V . De nombreux critères dits de corrélation proposent une méthode, nous en rapporterons une liste (non exhaustive) dans ce chapitre.

Comme déjà remarqué au travers des deux définitions : définition 12 et définition 13, nous pouvons coder l'ensemble de l'information fournie par les n réalisations de deux manières différentes ; chacune avec des propriétés spécifiques.

Or, les critères de corrélations s'appliquent précisément à ladite information. Savoir comment ils se traduisent d'une écriture à l'autre est un travail important, déjà réalisé cependant par par Amel Najah Idrissi [74] qui s'appliquât à traduire la majorité des critères dans les deux formats.

Ce chapitre est l'occasion d'apporter un éclairage différent aux travaux déjà réalisés.

De fait, si nous rappellerons les différentes expressions, ce sera pour remarquer que les critères s'expriment comme un écart à l'*indétermination* ou bien à l'*indépendance*. Dûment justifié par le rôle canonique des deux fonctions de couplage relevé dans le chapitre premier, cette remarque cesse alors d'être fortuite.

Nous reviendrons également sur un critère particulier d'écart à l'*indétermination*, le Janson Vegelius, pour lequel nous proposerons des limites théoriques.

Enfin, remarquant qu'une forme commune, à savoir le produit scalaire $(X|Y)$ entre les deux matrices du codage relationnel, apparaît dans tous les critères, nous nous emploierons à l'étudier. Cette dernière partie sera l'occasion de remarquer que sa valeur sur un tirage de n variables repose en moyenne sur une permutation de n éléments. En s'appuyant sur cette décomposition nous proposerons une méthode pour approcher la solution du problème 13 sur des entiers pour laquelle aucune forme close n'existe.

5.1 Codage relationnel

Cette section regroupe une sélection de critères de corrélation qui, bien que non exhaustive, a deux mérites. D'abord elle contient les critères les plus utilisés, ensuite elle s'accompagne d'une ligne cohérente pour justifier les différences entre chacun.

Deux critères se distingueront comme nous remarquerons à la fin. Un bien connu, le χ_2 , et un second, le Janson-Vegelius dont l'historique de construction sera rappelé ultérieurement et qui apparaîtra ici comme dual du χ_2 selon le codage contingenciel.

Afin de disposer des deux encodages mentionnés en introduction, nous allons appliquer le principe 10 à plusieurs reprises. Nous reportons le lecteur aux définitions 12 et 13 pour les notations utilisées ci-après.

5.1.1 Critère de Rand

Nous commençons par rappeler ici la définition du très célèbre critère de Rand sous sa forme de contingence (disons "contingencielle").

Définition 27 (Expression de contingence du critère de Rand).

$$RAND(U, V) = \frac{2 \sum_{u=1}^p \sum_{v=1}^q n_{u,v}^2 - \sum_{u=1}^p n_{u,\cdot}^2 - \sum_{v=1}^q n_{\cdot,v}^2 + n_{\cdot,\cdot}^2}{n_{\cdot,\cdot}^2}$$

Ce critère est évalué et comparé à $1/2$ pour juger la corrélation entre les deux variables. Le critère de Rand habituel admet bien sûr une forme relationnelle que l'on obtient facilement en prenant $p = q = 2$ dans la proposition 20. Sous cette forme il apparaît comme le critère de Condorcet qui repose sur l'égalité des voix pour et contre, égalité que nous avons mentionnée dans la définition 16.

Propriété 18 (Critère de Rand-Condorcet).

Le critère de Rand multiplié par $n_{\cdot,\cdot}^2$ est le fameux critère de Condorcet.

Démonstration.

Imposer $p = q = 2$ dans l'égalité 3.14 puis remplacer les quantités $A_0, A_1, Q_{0,1}, Q_{1,0}$ par leur valeur en écriture relationnelle donne :

$$A_0 + A_1 = Q_{0,1} + Q_{1,0}$$

Et précisément, en utilisant le principe de transfert relationnel décrit dans la propriété 10 il vient :

$$\begin{aligned}
N^2 \left(RAND(U, V) - \frac{1}{2} \right) &= 2 \sum_{u=1}^p \sum_{v=1}^q n_{u,v}^2 - \sum_{u=1}^p n_{u,\cdot}^2 - \sum_{v=1}^q n_{\cdot,v}^2 + \frac{N^2}{2} \\
&= \frac{1}{2} \left(\sum_{i=1}^N \sum_{j=1}^N x_{i,j} y_{i,j} + \sum_{i=1}^N \sum_{j=1}^N \overline{x_{i,j} y_{i,j}} - \sum_{i=1}^N \sum_{j=1}^N x_{i,j} \overline{y_{i,j}} - \sum_{i=1}^N \sum_{j=1}^N \overline{x_{i,j}} y_{i,j} \right) \\
&= \frac{1}{2} (A_0 + A_1 - Q_{0,1} - Q_{1,0})
\end{aligned}$$

□

Ainsi, le numérateur de la quantité $2 * RAND(U, V) - 1$ vaut zéro en cas d'indétermination c'est-à-dire lorsque les nombres d'accords et de désaccords sont égaux. Cette interprétation du critère usuel de Rand nous permet de dériver l'interprétation 19 suivante.

Propriété 19 (Critère de Rand et indétermination).

*Le critère de Rand dans sa version affine $2 * RAND(U, V) - 1$ vaut 0 quand U et V sont dans un couplage d'indétermination.*

Remarque 21 (Rand comme écart à l'indétermination).

En utilisant la formule 3.11, on constate que l'écriture du critère $2RAND(U, V) - 1$ correspond exactement à un écart quadratique à l'indétermination dans le cas $p = q = 2$.

Un besoin de généralisation pour proposer un critère d'écart quadratique à l'indétermination dans un cadre plus général que celui de Rand amène à introduire le critère de Janson-Vegelius.

5.1.2 Critère de Janson-Vegelius

Ce critère dont nous reportons l'introduction historique à la section qui lui sera consacrée peut être vu comme la généralisation du critère de Rand en tant que les deux sont égaux lorsque $p = q = 2$. Nous allons utiliser les formules habituelles de la propriété 10 pour montrer la définition-proposition suivante :

Propriété 20 (Expression de contingence et relationnelle de Janson-Vegelius).

$$\begin{aligned}
JV(U, V) &= \frac{pq \sum_{u=1}^p \sum_{v=1}^q n_{u,v}^2 - p \sum_{u=1}^p n_{u,\cdot}^2 - q \sum_{v=1}^q n_{\cdot,v}^2 + n_{\cdot,\cdot}^2}{\sqrt{p(p-2) \sum_{u=1}^p n_{u,\cdot}^2 + n_{\cdot,\cdot}^2} \sqrt{q(q-2) \sum_{v=1}^q n_{\cdot,v}^2 + n_{\cdot,\cdot}^2}} \\
&= \frac{\sum_{i=1}^n \sum_{j=1}^n [x_{i,j} - \frac{1}{p}] [y_{i,j} - \frac{1}{q}]}{\sqrt{\sum_{i=1}^n \sum_{j=1}^n [x_{i,j} - \frac{1}{p}]^2} \sqrt{\sum_{i=1}^n \sum_{j=1}^n [y_{i,j} - \frac{1}{q}]^2}}
\end{aligned}$$

Démonstration.

On part de la seconde expression et la développe : N_1 est le numérateur de la première expression et N_2 celui de la deuxième.

$$\begin{aligned}
N_2 &= \sum [x_{i,j} - \frac{1}{p}][y_{i,j} - \frac{1}{q}] \\
&= \sum [x_{i,j}y_{i,j} - \frac{x_{i,j}}{q} - \frac{y_{i,j}}{p} + \frac{1}{pq}] \\
&= \frac{1}{pq} \sum [pqx_{i,j}y_{i,j} - px_{i,j} - qy_{i,j} + 1] \\
&= \frac{1}{pq} N_1
\end{aligned}$$

D_{1L} est la partie de gauche du dénominateur de la première expression et l'on définit symétriquement D_{1R} , D_{2L} and D_{2R} .

$$\begin{aligned}
D_{2L}^2 &= \sum [x_{i,j} - \frac{1}{p}]^2 \\
&= \sum [x_{i,j}^2 - 2\frac{x_{i,j}}{p} + \frac{1}{p^2}] \\
&= \sum [x_{i,j} - 2\frac{x_{i,j}}{p} + \frac{1}{p^2}] \\
&= \frac{1}{p^2} \sum [p^2x_{i,j} - 2px_{i,j} + 1] \\
&= \frac{p(p-2)}{p^2} (\sum x_{i,j} + N^2) \\
&= \frac{1}{p^2} D_{1L}^2
\end{aligned}$$

On a de façon similaire $D_{2R}^2 = \frac{1}{q^2} D_{1R}^2$. En combinant toutes ces égalités, on obtient l'équivalence des expressions. \square

Sous cette forme, le coefficient apparaît comme un cosinus dans l'espace relationnel. Par ailleurs, on voit qu'il généralise Rand de façon formelle (lorsque $p=q=2$, ils sont égaux) et aussi comme écart à l'indétermination d'après la formule 3.11. Cette formule montre au passage que le numérateur du Janson-Végélius est positif et donc que le JV est positif; propriété qui n'est pas évidente sous la forme relationnelle tandis que l'inégalité $JV \leq 1$ est au contraire évidente sous cette forme.

On obtient ainsi par combinaison des deux approches les propriétés suivantes :

Propriété 21 (Propriétés sur les valeurs du JV).

- $0 \leq JV(U, V) \leq 1$
- $JV(U, V) = 0$ ssi $n_{u,v} = \frac{n_{u,\cdot}}{p} + \frac{n_{\cdot,v}}{q} - \frac{n_{\cdot\cdot}}{pq}$
- $JV(U, V) = 1$ ssi $X = Y$ et donc $p = q$

Si l'on s'intéresse au numérateur du JV, on obtient aisément, en combinant l'égalité 3.11 et la proposition 20 :

Propriété 22 (Numérateur JV).

$$N_{JV} = \sum_{u=1}^p \sum_{v=1}^q \left[n_{u,v} - \frac{n_{u,\cdot}}{q} - \frac{n_{\cdot,v}}{p} + \frac{n_{\cdot,\cdot}}{pq} \right]^2 = \sum_{i=1}^n \sum_{j=1}^n \left[x_{i,j} - \frac{1}{p} \right] \left[y_{i,j} - \frac{1}{q} \right]$$

Le premier membre $\sum_{u=1}^p \sum_{v=1}^q n_{u,v}^2$ apparaît dans la majorité des critères, c'était évidemment le cas pour Rand, il s'agit du produit scalaire $(X|Y)$ que nous aurons l'occasion d'étudier en détail ultérieurement.

5.1.3 Critère du χ_2

On cherche cette fois à établir la forme relationnelle du χ_2 comme écart à l'indépendance. Celle-ci s'écrit facilement : $\frac{n_{u,\cdot} n_{\cdot,v}}{n_{\cdot,\cdot}}$ de sorte que la forme de contingence est bien celle proposée dans la proposition 23.

Propriété 23 (Ecriture du χ_2).

$$\begin{aligned} \chi_2(U, V) &= \sum_{u=1}^p \sum_{v=1}^q \frac{n_{\cdot,\cdot}}{n_{u,\cdot} n_{\cdot,v}} \left[n_{u,v} - \frac{n_{u,\cdot} n_{\cdot,v}}{n_{\cdot,\cdot}} \right]^2 \\ &= n_{\cdot,\cdot} \sqrt{(p-1)(q-1)} \frac{\sum_{i=1}^N \sum_{j=1}^N \left(\frac{x_{i,j}}{x_{i,\cdot}} - \frac{1}{N} \right) \left(\frac{y_{i,j}}{y_{\cdot,j}} - \frac{1}{N} \right)}{\sqrt{\sum_{i=1}^N \sum_{j=1}^N \left(\frac{x_{i,j}}{x_{i,\cdot}} - \frac{1}{N} \right)^2 \sum_{i=1}^N \sum_{j=1}^N \left(\frac{y_{i,j}}{y_{\cdot,j}} - \frac{1}{N} \right)^2}} \end{aligned}$$

Démonstration.

On reprend les notations de décomposition utilisées pour la preuve de la proposition 20 et l'on applique comme toujours les formules de la propriété 10.

Le premier numérateur s'écrit :

$$\begin{aligned} N_1 &= \sum_{u=1}^p \sum_{v=1}^q \frac{1}{n_{u,\cdot} n_{\cdot,v}} \left[n_{u,v} - \frac{n_{u,\cdot} n_{\cdot,v}}{n_{\cdot,\cdot}} \right]^2 \\ &= \sum \frac{1}{n_{u,\cdot} n_{\cdot,v}} \left[n_{u,v}^2 - 2 \frac{n_{u,v} n_{u,\cdot} n_{\cdot,v}}{n_{\cdot,\cdot}} + \frac{n_{u,\cdot}^2 n_{\cdot,v}^2}{n_{\cdot,\cdot}^2} \right] \\ &= \sum \left[\frac{n_{u,v}^2}{n_{u,\cdot} n_{\cdot,v}} - 2 \frac{n_{u,v}}{n_{\cdot,\cdot}} + \frac{n_{u,\cdot} n_{\cdot,v}}{n_{\cdot,\cdot}^2} \right] \\ &= \sum \left[\frac{n_{u,v}^2}{n_{u,\cdot} n_{\cdot,v}} \right] - 2 + 1 \\ &= \sum \left[\frac{n_{u,v}^2}{n_{u,\cdot} n_{\cdot,v}} \right] - 1 \end{aligned}$$

De l'autre côté nous avons :

$$\begin{aligned}
N_2 &= \sum_{i=1}^N \sum_{j=1}^N \left(\frac{x_{i,j}}{x_{i,\cdot}} - \frac{1}{N} \right) \left(\frac{y_{i,j}}{y_{\cdot,j}} - \frac{1}{N} \right) \\
&= \sum \left[\frac{x_{i,j}}{x_{i,\cdot}} \frac{y_{i,j}}{y_{\cdot,j}} - \frac{x_{i,j}}{Nx_{i,\cdot}} - \frac{y_{i,j}}{Ny_{\cdot,j}} + \frac{1}{N} \right] \\
&= \sum \left[\frac{n_{u,v}^2}{n_{u,\cdot} n_{\cdot,v}} \right] - 1
\end{aligned}$$

Et enfin le deuxième dénominateur vaut :

$$\begin{aligned}
D_{2L}^2 &= \sum_{i=1}^N \sum_{j=1}^N \left(\frac{x_{i,j}}{x_{i,\cdot}} - \frac{1}{N} \right)^2 \\
&= \sum \left[\frac{x_{i,j}^2}{x_{i,\cdot}^2} - 2 \frac{x_{i,j}}{Nx_{i,\cdot}} + \frac{1}{N^2} \right] \\
&= \sum_i \left[\frac{1}{x_{i,\cdot}} - 2 \frac{x_{i,\cdot}}{Nx_{i,\cdot}} \right] + 1 \\
&= \# \{ \text{modalités de X} \} - 2 + 1 \\
&= p - 1
\end{aligned}$$

Symétriquement, on obtient $D_{2R} = \sqrt{q-1}$ ce qui conduit au résultat. \square

Sous sa forme relationnelle, le χ_2 apparaît lui aussi comme un produit scalaire mais pondéré par une fonction du nombre de classes. On obtient par ailleurs les propriétés suivantes, évidentes en utilisant les deux formes.

Propriété 24 (Valeurs du χ_2).

- $0 \leq \frac{\chi_2(U,V)}{n_{\cdot,\cdot}} \leq \sqrt{(p-1)(q-1)}$
- $\chi_2(U,V) = 0$ ssi $n_{u,v} = \frac{n_{u,\cdot} n_{\cdot,v}}{n_{\cdot,\cdot}}$
- $\chi_2(U,V) = n_{\cdot,\cdot} \sqrt{(p-1)(q-1)}$ ssi $X = Y$

Si l'on s'intéresse au numérateur du χ_2 , on montre, moins aisément que pour le JV, l'égalité suivante.

Propriété 25 (Numérateur χ_2).

$$N_{\chi_2} = \sum_{u=1}^p \sum_{v=1}^q \left[n_{u,v} - \frac{n_{u,\cdot} n_{\cdot,v}}{n_{\cdot,\cdot}} \right]^2 = \sum_{i=1}^n \sum_{j=1}^n \left(x_{i,j} - \frac{x_{i,\cdot}}{N} - \frac{x_{\cdot,j}}{N} + \frac{x_{\cdot,\cdot}}{N^2} \right) \left(y_{i,j} - \frac{y_{i,\cdot}}{N} - \frac{y_{\cdot,j}}{N} + \frac{y_{\cdot,\cdot}}{N^2} \right)$$

Démonstration.

Nous partons avec les notations habituelles du membre de droite que nous développons à pondération par N^2 près.

On remarque que la somme de N_{2R} (partie droite de N_2) est nulle selon à la fois i et j . Ainsi, seul un facteur en (i, j) peut être conservé après sommation. Le calcul se simplifie grandement :

$$\begin{aligned}
N_2 &= \sum_{i=1}^n \sum_{j=1}^n \left(x_{i,j} - \frac{x_{i,\cdot}}{N} - \frac{x_{\cdot,j}}{N} + \frac{x_{\cdot,\cdot}}{N^2} \right) \left(y_{i,j} - \frac{y_{i,\cdot}}{N} - \frac{y_{\cdot,j}}{N} + \frac{y_{\cdot,\cdot}}{N^2} \right) \\
&= \sum_{i=1}^n \sum_{j=1}^n x_{ij} \left(y_{i,j} - \frac{y_{i,\cdot}}{N} - \frac{y_{\cdot,j}}{N} + \frac{y_{\cdot,\cdot}}{N^2} \right) \\
&= \sum_{i=1}^n \sum_{j=1}^n \left(x_{i,j} y_{i,j} - \frac{x_{i,\cdot} y_{i,\cdot}}{N} - \frac{x_{\cdot,j} y_{\cdot,j}}{N} + \frac{x_{\cdot,\cdot} y_{\cdot,\cdot}}{N^2} \right) \\
&\quad \text{Avec les formules du principe de transfert relationnel :} \\
&= \sum_{u=1}^p \sum_{v=1}^q \left[n_{u,v}^2 - 2 \frac{n_{u,v} n_{u,\cdot} n_{\cdot,v}}{n_{\cdot,\cdot}} + \frac{n_{u,\cdot}^2 n_{\cdot,v}^2}{n_{\cdot,\cdot}^2} \right]
\end{aligned}$$

on retrouve donc bien la forme de contingence du numérateur. \square

Remarque 22 (Expression symétrique inutile ?).

Vu le développement, on constate que l'on aurait pu simplifier la forme relationnelle du numérateur. Un souci de symétrie et le développement que nous pousserons dans la partie suivante justifieront cette forme particulière.

Remarque 23.

La forme étonnante des résultats sur les valeurs du χ_2 et notamment le fait qu'il fasse apparaître $\sqrt{(p-1)(q-1)}$ conduit à définir un nouveau critère pondéré : le critère de Tchuprow.

5.1.4 Critère de Tchuprow

S'agissant des mêmes calculs que ceux portant sur le critère du χ_2 , nous ne ferons que livrer les résultats concernant ce critère sous la forme de la définition-propriété suivante.

Propriété 26 (Expression de contingence et relationnelle de Tchuprow).

$$\begin{aligned}
T(U, V) &= \frac{\chi_2(U, V)}{n_{\cdot,\cdot} \sqrt{p-1} \sqrt{q-1}} \\
&= \sum_{u=1}^p \sum_{v=1}^q \frac{1}{n_{u,\cdot} n_{\cdot,v}} \frac{\left[n_{u,v} - \frac{n_{u,\cdot} n_{\cdot,v}}{n_{\cdot,\cdot}} \right]^2}{\sqrt{p-1} \sqrt{q-1}} \\
&= \frac{\sum_{i=1}^N \sum_{j=1}^N \left(\frac{x_{i,j}}{x_{i,\cdot}} - \frac{1}{N} \right) \left(\frac{y_{i,j}}{y_{\cdot,j}} - \frac{1}{N} \right)}{\sqrt{\sum_{i=1}^N \sum_{j=1}^N \left(\frac{x_{i,j}}{x_{i,\cdot}} - \frac{1}{N} \right)^2 \sum_{i=1}^N \sum_{j=1}^N \left(\frac{y_{i,j}}{y_{\cdot,j}} - \frac{1}{N} \right)^2}}
\end{aligned}$$

Introduisons maintenant un dernier critère, plus complexe que les précédents : celui de Lerman-Chah-Gini dit LCG.

5.1.5 Critère de Lerman-Chah-Gini

Nous proposons la définition-propriété suivante pour introduire ce nouveau critère.

Propriété 27 (Critère de Lerman-Chah-Gini).

$$\begin{aligned}
 LCG(U, V) &= \frac{\sum_{i=1}^N \sum_{j=1}^N \left(x_{i,j} - \frac{x_{\cdot,\cdot}}{N^2}\right) \left(y_{i,j} - \frac{y_{\cdot,\cdot}}{N^2}\right)}{\sqrt{\sum_{i=1}^N \sum_{j=1}^N \left(x_{i,j} - \frac{x_{\cdot,\cdot}}{N^2}\right)^2 \sum_{i=1}^N \sum_{j=1}^N \left(y_{i,j} - \frac{y_{\cdot,\cdot}}{N^2}\right)^2}} \\
 &= \frac{\sum_{u=1}^p \sum_{v=1}^q \left(n_{u,v}^2 - \frac{n_{u,\cdot}^2 n_{\cdot,v}^2}{N}\right)}{\sqrt{\sum_{u=1}^p n_{u,\cdot}^2 - \left[\sum_{u=1}^p \frac{n_{u,\cdot}^2}{N}\right]^2} \sqrt{\sum_{v=1}^q n_{\cdot,v}^2 - \left[\sum_{v=1}^q \frac{n_{\cdot,v}^2}{N}\right]^2}}
 \end{aligned}$$

Démonstration.

Là encore, nous utiliserons les formules présentes dans la propriété 10.

$$\begin{aligned}
 LCG(U, V) &= \frac{\sum_{i=1}^N \sum_{j=1}^N \left(x_{i,j} - \frac{x_{\cdot,\cdot}}{N^2}\right) \left(y_{i,j} - \frac{y_{\cdot,\cdot}}{N^2}\right)}{\sqrt{\sum_{i=1}^N \sum_{j=1}^N \left(x_{i,j} - \frac{x_{\cdot,\cdot}}{N^2}\right)^2 \sum_{i=1}^N \sum_{j=1}^N \left(y_{i,j} - \frac{y_{\cdot,\cdot}}{N^2}\right)^2}} \\
 &= \frac{\sum_{i=1}^N \sum_{j=1}^N \left(x_{i,j} y_{i,j} - \frac{x_{\cdot,\cdot} y_{\cdot,\cdot}}{N^2}\right)}{\sqrt{\sum_{i=1}^N \sum_{j=1}^N \left(x_{i,j}^2 - \frac{x_{\cdot,\cdot}^2}{N^2}\right) \sum_{i=1}^N \sum_{j=1}^N \left(y_{i,j}^2 - \frac{y_{\cdot,\cdot}^2}{N^2}\right)}}
 \end{aligned}$$

on utilise les formules de transfert et le fait que $x_{i,j}^2 = x_{i,j}$ pour avoir

$$\begin{aligned}
 &= \frac{\sum_{i=1}^N \sum_{j=1}^N \left(x_{i,j} y_{i,j}\right) - \frac{x_{\cdot,\cdot} y_{\cdot,\cdot}}{N^2}}{\sqrt{\sum_{i=1}^N \sum_{j=1}^N \left(x_{i,j} - \frac{x_{\cdot,\cdot}}{N^2}\right) \sum_{i=1}^N \sum_{j=1}^N \left(y_{i,j} - \frac{y_{\cdot,\cdot}}{N^2}\right)}} \\
 &= \frac{\sum_{u=1}^p \sum_{v=1}^q \left(n_{u,v}^2 - \left(\sum_{u=1}^p \frac{n_{u,\cdot}^2}{N}\right) \left(\sum_{v=1}^q \frac{n_{\cdot,v}^2}{N}\right)\right)}{\sqrt{\sum_{u=1}^p n_{u,\cdot}^2 - \left[\sum_{u=1}^p \frac{n_{u,\cdot}^2}{N}\right]^2} \sqrt{\sum_{v=1}^q n_{\cdot,v}^2 - \left[\sum_{v=1}^q \frac{n_{\cdot,v}^2}{N}\right]^2}}
 \end{aligned}$$

□

Ce critère a plusieurs propriétés intéressantes qui rassemblent les deux couplages de notre étude. D'abord, il est facile de constater qu'il s'annule pour l'indépendance. Par ailleurs, à marges fixées, le théorème 2 nous permet d'affirmer que n^+ en minimise le numérateur. Par n^+ , notation simplificatrice, il faut entendre un tirage de taille n qui

réalise parfaitement le couplage selon l'indétermination des marges. Cela suppose des propriétés très spécifiques sur les marges afin que la formule :

$$n_{u,v}^+ = \frac{n_{u,\cdot}}{q} + \frac{n_{\cdot,v}}{p} - \frac{n_{\cdot,\cdot}}{pq}$$

soit entière pour tout (u, v) . C'est certes peu probable. Dans le cas général, les marges μ_u et ν_u dans \mathbb{Q} conduiront à un π^+ dans \mathbb{Q} qui sera approché par la réalisation sans jamais être égal.

Aussi, n^+ et n^\times réalisant (lorsqu'aucune des marges n'est uniforme) deux couplages différents, on déduit du point de réalisation de ce minimum que le critère LCG peut être négatif.

Par ailleurs, ce critère s'insère dans la suite d'inégalités que nous montrerons par la suite. Revenons pour l'instant à la dualité entre le χ_2 et le JV, dualité qui justifie l'écriture complexe remarquée dans la remarque 22.

5.1.6 Dualité indépendance-indétermination

Nous insistons sur une dualité apparaissant dans les expressions précédemment fournies entre le χ_2 qui est un écart à l'indépendance et le JV qui est un écart à l'indétermination.

Nous commençons par justifier les centrages apparaissant dans les expressions relationnelles des deux critères.

Remarque 24 (Déviation à la moyenne).

Dans l'expression du JV, nous comparons $x_{i,j}$ à $\frac{1}{p}$ (voir la proposition 20) tandis que $\frac{x_{i,j}}{x_{i,\cdot}}$ est comparé à $\frac{1}{N}$ dans l'expression du χ_2 (voir la proposition 23). Cela vient des égalités présentées ici et permettant d'interpréter chaque centrage comme une moyenne :

$$\frac{1}{N^2} \sum_{i=1}^N \sum_{j=1}^N x_{i,j} = \frac{1}{N^2} \sum_{i=1}^N x_{i,\cdot} = \frac{1}{N^2} \sum_{u=1}^p n_{u,\cdot}^2$$

Aussi, s'il y a p classes, on s'attend à avoir en moyenne $\frac{N}{p}$ éléments dans chaque ce qui conduit bien à une moyenne en $\frac{1}{p}$.

De façon similaire,

$$\frac{1}{N^2} \sum_{i=1}^N \sum_{j=1}^N \frac{x_{i,j}}{x_{i,\cdot}} = \frac{1}{N^2} \sum_{i=1}^N 1 = \frac{1}{N}$$

Passons maintenant à la constatation d'une dualité entre les deux critères qui permet de faire perdurer la symétrie entre les deux couplages.

Propriété 28 (Dualité JV et χ_2).

Partant des propositions 20 et 23, on remarque une symétrie déjà relevée dans [63]. Celle-ci se résume dans le tableau suivant :

	<i>Codage de contingence</i>	<i>Codage relationnel</i>
χ_2	$\sum_{u=1}^p \sum_{v=1}^q \left[n_{u,v} - \frac{n_{u,\cdot} n_{\cdot,v}}{n_{\cdot,\cdot}} \right]^2$	$\sum_{i=1}^n \sum_{j=1}^n \left(x_{i,j} - \frac{x_{i,\cdot}}{N} - \frac{x_{\cdot,j}}{N} + \frac{x_{\cdot,\cdot}}{N^2} \right) \left(y_{i,j} - \frac{y_{i,\cdot}}{N} - \frac{y_{\cdot,j}}{N} + \frac{y_{\cdot,\cdot}}{N^2} \right)$
JV	$\left[n_{u,v} - \frac{n_{u,\cdot}}{q} - \frac{n_{\cdot,v}}{p} + \frac{n_{\cdot,\cdot}}{pq} \right]^2$	$\sum_{i=1}^n \sum_{j=1}^n \left[x_{i,j} - \frac{1}{p} \right] \left[y_{i,j} - \frac{1}{q} \right]$

Ainsi, la forme de couplage propre à l'indétermination apparaît en codage relationnel pour le χ_2 et en codage de contingence pour le JV .

5.1.7 Comparaison des valeurs de critères

Nous justifions ici l'introduction des nombreux critères présentés précédemment, en effet, outre qu'ils s'appuient pour la plupart sur un écart à l'indépendance ou à l'indétermination et continuent ainsi notre parallèle, ils entretiennent entre eux une relation d'ordre que nous souhaitons présenter dans cette partie. Les indices de Janson-Vegelius et de Lerman-Chah-Gini's sont les plus complexes mais aussi les plus cohérents parmi ceux présentés dans ce document ou dans les travaux de la thèse [74]. Une question naturelle est alors de comprendre les similitudes qu'ils peuvent présenter dans leurs valeurs malgré leurs structures différentes. Nous allons utiliser l'angle de leurs expressions relationnelles pour apporter une expression explicite de similitude.

Introduisons tout d'abord la notion de densité d'une relation.

Définition 28 (Densité d'une relation).

La densité est simplement la moyenne des valeurs présentes dans la matrice relationnelle X conformément à la définition 13.

$$\delta_X = \frac{1}{N^2} \sum_{i=1}^N \sum_{j=1}^N x_{i,j} = \frac{x_{\cdot,\cdot}}{N^2}$$

En général, $0 \leq \delta_X \leq 1$, les relations réflexives ont en plus $\frac{1}{N} \leq \delta_X \leq 1$. Nous reprenons maintenant les expressions relationnelles des deux indices qui nous intéressent ici (Propriété 20 et Propriété 27) et les réécrivons en fonction de la notion de densité tout juste introduite.

$$\begin{aligned}
JV(U, V) &= \frac{\sum_{i=1}^N \sum_{j=1}^N \left(x_{i,j} - \frac{1}{p}\right) \left(y_{i,j} - \frac{1}{q}\right)}{\sqrt{\sum_{i=1}^N \sum_{j=1}^N \left(x_{i,j} - \frac{1}{p}\right)^2 \sum_{i=1}^N \sum_{j=1}^N \left(y_{i,j} - \frac{1}{q}\right)^2}} \\
&= \frac{\sum_{i=1}^N \sum_{j=1}^N x_{i,j} y_{i,j} - N^2 \left(\frac{\delta_X}{q} + \frac{\delta_Y}{p} - \frac{1}{pq}\right)}{N^2 \sqrt{\delta_X(1 - \delta_X) + \left(\delta_X - \frac{1}{p}\right)^2} \sqrt{\delta_Y(1 - \delta_Y) + \left(\delta_Y - \frac{1}{q}\right)^2}} \\
LCG(U, V) &= \frac{N^2 \left[\sum_{i=1}^N \sum_{j=1}^N x_{i,j} y_{i,j} \right] - x_{\cdot, \cdot} y_{\cdot, \cdot}}{\sqrt{N^2 x_{\cdot, \cdot} - x_{\cdot, \cdot}^2} \sqrt{N^2 y_{\cdot, \cdot} - y_{\cdot, \cdot}^2}} \\
&= \frac{\sum_{i=1}^N \sum_{j=1}^N x_{i,j} y_{i,j} - N^2 \delta_X \delta_Y}{N^2 \sqrt{\delta_X(1 - \delta_X)} \sqrt{\delta_Y(1 - \delta_Y)}}
\end{aligned}$$

Propriété 29 (Majoration commune de JV et LCG).

Les critères Janson-Vegelius et Lerman-Chah-Gini sont tous les deux majorés par :

$$\frac{\frac{pq}{N^2} \sum_{i=1}^N \sum_{j=1}^N x_{i,j} y_{i,j} - 1}{\sqrt{p-1} \sqrt{q-1}}$$

De plus, dans le cas où les marges sont uniformes, il s'agit d'une égalité et le majorant se réécrit :

$$LCG(X, Y) = JV(X, Y) = \frac{pq\delta_{X,Y} - 1}{\sqrt{p-1} \sqrt{q-1}}$$

Où l'on a généralisé la notion de densité au couplage : la matrice représentant le couple étant le produit terme à terme des deux matrices.

Démonstration.

Nous savons que X et Y encodent des relations d'équivalence. En particulier, en notant $C_u, u = 1, \dots, p$ les classes de X :

$$\sum_{i=1}^N \sum_{j=1}^N \frac{x_{i,j}}{x_{i, \cdot} x_{\cdot, j}} = \sum_{i=1}^N \frac{1}{x_{i, \cdot}} = \sum_{u=1}^p \sum_{i \in C_u} \frac{1}{x_{i, \cdot}} = \sum_{u=1}^p \sum_{i \in C_u} \frac{1}{\#C_u} = \sum_{u=1}^p 1 = p$$

On définit la moyenne harmonique H_X des valeurs marginales $\{x_{i, \cdot}\}$ par

$$\frac{1}{H_X} = \frac{1}{N} \sum_{i=1}^N \frac{1}{x_{i, \cdot}} = \frac{p}{N}$$

Un résultat bien connu montre que la moyenne arithmétique habituelle $A_X = N\delta_X$ est toujours supérieure à la moyenne harmonique. Cette relation se traduit ici par :

$$\frac{1}{p} \leq \delta_X$$

Simplifions alors les expressions des deux critères en utilisant cette propriété et le fait que $(\delta_X - \frac{1}{p})^2$ et $(\delta_Y - \frac{1}{q})^2$ sont positifs.

$$\begin{aligned}
JV(X, Y) &= \frac{\sum_{i=1}^N \sum_{j=1}^N x_{i,j} y_{i,j} - N^2 \left(\frac{\delta_X}{q} + \frac{\delta_Y}{p} - \frac{1}{pq} \right)}{N^2 \sqrt{\delta_X(1 - \delta_X) + (\delta_X - \frac{1}{p})^2} \sqrt{\delta_Y(1 - \delta_Y) + (\delta_Y - \frac{1}{q})^2}} \\
&\leq \frac{\sum_{i=1}^N \sum_{j=1}^N x_{i,j} y_{i,j} - \frac{N^2}{pq}}{N^2 \sqrt{\delta_X(1 - \delta_X)} \sqrt{\delta_Y(1 - \delta_Y)}}
\end{aligned}$$

On combine pour la formule suivante les deux inégalités $\frac{1}{p} \leq \delta_X$ et $\frac{1}{q} \leq \delta_Y$ pour déduire $\frac{N^2}{pq} \leq N^2 \delta_X \delta_Y$.

$$\begin{aligned}
LCG(X, Y) &= \frac{\sum_{i=1}^N \sum_{j=1}^N x_{i,j} y_{i,j} - N^2 \delta_X \delta_Y}{N^2 \sqrt{\delta_X(1 - \delta_X)} \sqrt{\delta_Y(1 - \delta_Y)}} \\
&\leq \frac{\sum_{i=1}^N \sum_{j=1}^N x_{i,j} y_{i,j} - \frac{N^2}{pq}}{N^2 \sqrt{\delta_X(1 - \delta_X)} \sqrt{\delta_Y(1 - \delta_Y)}}
\end{aligned}$$

Dans le cas particulier où les marges sont uniformes c'est-à-dire où $n_{u,\cdot}$ et $n_{\cdot,v}$ valent respectivement $\frac{N}{p}$ et $\frac{N}{q}$, $\delta_X = \frac{1}{p}$ et $\delta_Y = \frac{1}{q}$, les inégalités précédentes se transforment en égalités et les deux critères sont égaux. Le dénominateur vaut alors $\frac{N^2}{pq} \sqrt{p-1} \sqrt{q-1}$ et le numérateur $\sum_{i=1}^N \sum_{j=1}^N x_{i,j} y_{i,j} - \frac{N^2}{pq}$ ce qui termine la preuve. \square

Remarque 25. *En appliquant le même processus au critère de Tschuprow défini par la propriété 26, on peut montrer qu'il prend la même valeur que les deux autres dans le cas de marges uniformes.*

Démonstration.

$$\begin{aligned}
T(X, Y) &= \frac{\sum_{i=1}^N \sum_{j=1}^N \left(\frac{x_{i,j}}{x_{i,\cdot}} - \frac{1}{N} \right) \left(\frac{y_{i,j}}{y_{\cdot,j}} - \frac{1}{N} \right)}{\sqrt{\sum_{i=1}^N \sum_{j=1}^N \left(\frac{x_{i,j}}{x_{i,\cdot}} - \frac{1}{N} \right)^2 \sum_{i=1}^N \sum_{j=1}^N \left(\frac{y_{i,j}}{y_{\cdot,j}} - \frac{1}{N} \right)^2}} \\
&= \frac{\sum_{i=1}^N \sum_{j=1}^N \left(\frac{x_{i,j} y_{i,j}}{x_{i,\cdot} y_{\cdot,j}} \right) - 1}{\sqrt{\sum_{i=1}^N \sum_{j=1}^N \left(\frac{x_{i,j}^2}{x_{i,\cdot}^2} \right) - 1} \sqrt{\sum_{i=1}^N \sum_{j=1}^N \left(\frac{y_{i,j}^2}{N y_{\cdot,j}} \right) - 1}} \\
&= \frac{\sum_{i=1}^N \sum_{j=1}^N \left(\frac{x_{i,j} y_{i,j}}{x_{i,\cdot} y_{\cdot,j}} \right) - 1}{\sqrt{p-1} \sqrt{q-1}} \\
&\quad \text{dans le cas de marges uniformes :} \\
&= \frac{\sum_{i=1}^N \sum_{j=1}^N \left(\frac{x_{i,j} y_{i,j}}{N/pN/q} \right) - 1}{\sqrt{p-1} \sqrt{q-1}}
\end{aligned}$$

Ce qui termine la preuve. □

Ainsi les critères que nous avons introduit se rejoignent dans le cas de marges uniformes (le χ_2 n'étant qu'une pondération différente du Tchuprow et Rand un cas particulier du JV). Nous rejoignons l'égalité des couplages remarquée dans la proposition 3.

Pour renforcer la proximité entre les critères, nous pouvons remarquer que les formes relationnelles s'apparentent souvent à des produits scalaires. Dans la prochaine section, nous allons généraliser et suivre les travaux développés dans [74] pour extraire une forme générique.

5.1.8 Forme générique des critères

En regardant la structure des critères précédemment introduits, on se rend compte qu'une forme générale se dessine : celle d'un cosinus dans l'espace relationnel. Pour obtenir une définition générique, il suffit de définir une notion de moyenne qui généralise celle de la définition 28.

Définition 29 (Indice de corrélation générique).

Tous les coefficients que nous avons rencontrés respectent la forme ci-dessous :

$$\rho_{\eta_X, \eta_Y, \Phi, \Psi}(X, Y) = \frac{\sum_{i=1}^N \sum_{j=1}^N (\phi(x_{i,j}) - \overline{\phi_{X\eta_X}}) (\psi(x_{i,j}) - \overline{\psi_{Y\eta_Y}})}{\sqrt{\sum_{i=1}^N \sum_{j=1}^N (\phi(x_{i,j}) - \overline{\phi_{X\eta_X}})^2} \sqrt{\sum_{i=1}^N \sum_{j=1}^N (\psi(x_{i,j}) - \overline{\psi_{Y\eta_Y}})^2}}$$

avec ϕ et ψ des fonctions positives sur les $x_{i,j}$ et les $y_{i,j}$; $\eta_{i,j}^X$ et $\eta_{i,j}^Y$ des coefficients positifs sommant à 1; $\overline{\phi_X}_{\eta_X} = \sum_{i=1}^N \sum_{j=1}^N \eta_{i,j}^X \phi(x_{i,j})$ la valeur moyenne prise par ϕ sur la matrice X

selon les coefficients dictés par η^X ; et de même $\overline{\psi_Y}_{\eta_Y} = \sum_{i=1}^N \sum_{j=1}^N \eta_{i,j}^Y \psi(y_{i,j})$.

L'utilisation du produit scalaire $(x|y) = \sum_{i,j} x_{i,j} y_{i,j}$ et de la norme associée permet de condenser l'écriture :

$$\rho_{\eta_X, \eta_Y, \Phi, \Psi}(X, Y) = \frac{(\phi_X - \overline{\phi_X}_{\eta_X} | \psi_Y - \overline{\psi_Y}_{\eta_Y})}{\|\phi_X - \overline{\phi_X}_{\eta_X}\| \|\psi_Y - \overline{\psi_Y}_{\eta_Y}\|} = \cos(\phi_X - \overline{\phi_X}_{\eta_X}, \psi_Y - \overline{\psi_Y}_{\eta_Y}) \quad (5.1)$$

Remarque 26 (Exemples).

Typiquement, on peut choisir comme fonctions $\phi = Id$ ou $\phi(x_{i,j}) = \frac{x_{i,j}}{x_{i,\cdot}}$ (quand définie) et garder $\eta = \frac{1}{n^2}$.

Remarque 27 (Critère à 1).

En général, les deux variables sous-jacentes jouant le même rôle, on choisit $\phi = \psi$ et $\eta^X = \eta^Y$. Dans ce cas, le critère résultant vaut 1 dès que $X = Y$. Ce comportement est cohérent car il s'agit de la corrélation maximale.

Donnons maintenant quelques exemples de moyennes $\overline{\phi_X}_{\eta_X}$ pour des fonctions ϕ et des pondérations η^X spécifiques en retrouvant au passage des critères déjà définis. Tous les critères précédemment introduits traitant symétriquement X et Y , on a systématiquement $\psi = \phi$; l'obtention des centrages (les $\overline{\phi_X}_{\eta_X}$ et $\overline{\psi_Y}_{\eta_Y}$) nécessitera parfois des η différents. Pour plus de simplicité, on cesse néanmoins d'introduire le η dans la notation (il reste évidemment dans le calcul!) et l'on pose $\overline{\phi_X} = \overline{\phi_X}_{\eta_X}$ et $\overline{\psi_Y} = \overline{\psi_Y}_{\eta_Y}$.

Loi uniforme \rightarrow Critère de Lerman-Chah-Gini

On choisit classiquement $\eta = \frac{1}{N^2}$ et $\phi = Id$ et l'on a donc $\overline{\phi_X} = \frac{x_{\cdot,\cdot}}{N^2}$. En reportant dans la définition 29 on obtient alors bien le coefficient défini en 27.

$$LCG(X, Y) = \rho_{\frac{1}{N^2}, \frac{1}{N^2}, Id, Id}(X, Y) = \frac{\sum_{i=1}^N \sum_{j=1}^N (x_{i,j} - \frac{x_{\cdot,\cdot}}{N^2}) (y_{i,j} - \frac{y_{\cdot,\cdot}}{N^2})}{\sqrt{\sum_{i=1}^N \sum_{j=1}^N (x_{i,j} - \frac{x_{\cdot,\cdot}}{N^2})^2} \sqrt{\sum_{i=1}^N \sum_{j=1}^N (y_{i,j} - \frac{y_{\cdot,\cdot}}{N^2})^2}}$$

Loi uniforme appliquée à $\phi(x) = \frac{x_{ij}}{x_{i,\cdot}}$ \rightarrow critère de Tchuprow

En gardant l'hypothèse précédente de densité uniforme sur le couple (i, j) mais en posant $\phi(x) = \frac{x_{ij}}{x_{i,\cdot}}$, on obtient la valeur $\overline{\phi_X}$ suivante :

$$\overline{\phi_X} = \sum_{i=1}^N \sum_{j=1}^N \frac{x_{ij}}{x_{i,\cdot}} \frac{1}{N^2} = \frac{N}{N^2} = \frac{1}{N}$$

En reportant dans la Définition 29, on obtient le critère défini en 26 :

$$T(X, Y) = \rho_{\frac{1}{N^2}, \frac{1}{N^2}, \frac{x_{ij}}{x_{i,\cdot}}, \frac{y_{ij}}{y_{i,\cdot}}}(X, Y) = \frac{\sum_{i=1}^N \sum_{j=1}^N \left(\frac{x_{i,j}}{x_{i,\cdot}} - \frac{1}{N} \right) \left(\frac{y_{i,j}}{y_{i,\cdot}} - \frac{1}{N} \right)}{\sqrt{\sum_{i=1}^N \sum_{j=1}^N \left(\frac{x_{i,j}}{x_{i,\cdot}} - \frac{1}{N} \right)^2 \sum_{i=1}^N \sum_{j=1}^N \left(\frac{y_{i,j}}{y_{i,\cdot}} - \frac{1}{N} \right)^2}}$$

Remarque 28. A un facteur près donné par la propriété 23, ce même facteur qui nous avait amené à définir le critère de Tchuprow, le résultat précédent montre que le critère du χ_2 d'écart à l'indépendance ne déroge pas à la règle et s'écrit comme un cosinus dans l'espace relationnel.

Taille inverse des clusters → Janson-Vegelius's Index

On pose ici $\phi = Id$ et l'on remarque la formule suivante issue de [67] :

$$\sum_{i=1}^N \sum_{j=1}^N \frac{1}{x_{i,\cdot} x_{\cdot,j}} = p^2 \quad (5.2)$$

de sorte que l'on définit les $\eta_{i,j}$ par $\eta_{i,j} = \frac{1}{x_{i,\cdot} x_{\cdot,j} p^2}$ pour avoir une somme à 1.

Remarque 29. Ici, les η sont différents d'une matrice à l'autre pour permettre d'obtenir les bons centrages.

Nous allons justifier la définition de ces η en l'appliquant au cas des relations d'équivalence. Pour toute paire $(i, j) \in X$ où X encode une relation d'équivalence, on considère que la probabilité pour un individu i d'appartenir à la même classe que j est donnée par le ratio entre la présence logique $x_{i,j}$ divisée par la taille totale de la classe. La normalisation quand à elle vient de la nécessité d'une somme à 1.

Calculons maintenant $\overline{\phi_X}$:

$$\begin{aligned} \overline{\phi_X} &= \sum_{i=1}^N \sum_{j=1}^N x_{i,j} \frac{1}{p^2 x_{i,\cdot} x_{\cdot,j}} \\ &= \frac{1}{p^2} \sum_{i=1}^N \sum_{j=1}^N \frac{x_{i,j}}{x_{i,\cdot} x_{\cdot,j}} \\ &= \frac{1}{p^2} \sum_{i=1}^N \sum_{j \simeq i} \frac{x_{i,j}}{x_{i,\cdot} x_{\cdot,j}} \\ &= \frac{1}{p^2} \sum_{i=1}^N \frac{1}{x_{i,\cdot}} \\ &= \frac{1}{p^2} \sum_{u=1}^p \sum_{i \in C_u} \frac{1}{\#C_u} \\ &= \frac{1}{p} \end{aligned}$$

On fait de même pour Y et l'on reporte dans la Définition 29 pour obtenir :

$$JV(X, Y) = \rho_{\frac{1}{x_i, x_{\cdot, j} p^2}, \frac{1}{y_i, y_{\cdot, j} p^2}, Id, Id}(X, Y) = \frac{\sum_{i=1}^N \sum_{j=1}^N \left(x_{i,j} - \frac{1}{p}\right) \left(y_{i,j} - \frac{1}{q}\right)}{\sqrt{\sum_{i=1}^N \sum_{j=1}^N \left(x_{i,j} - \frac{1}{p}\right)^2 \sum_{i=1}^N \sum_{j=1}^N \left(y_{i,j} - \frac{1}{q}\right)^2}}$$

C'est exactement la forme relationnelle du critère de Janson-Végélius que nous avons déjà introduit dans la définition 20.

Appliquer des définitions similaires à des relations d'ordre permet de retrouver des critères comme celui de Kendall ou de Somers à partir de la définition 29.

Remarque 30.

La formule générique introduite fait toujours apparaître au numérateur le facteur $(X|Y)$; aussi, son étude nous paraît intéressante et nous y dédierons la partie 5.3. Ce sera l'occasion d'utiliser et d'introduire des bornes classiques et plus ou moins anciennes, ce sera également l'occasion de le diagonaliser les matrices relationnelles et d'en extraire un algorithme d'optimisation d'un problème entier.

5.2 Focus sur le Janson-Végélius

Le but de cette section est d'étudier en détail un critère de corrélation. Nous choisissons le Janson-Végélius introduit dans la définition 20 pour son histoire et sa forme symétrique relationnelle caractéristique (voir sur ce sujet la thèse [74]). Nous commencerons par rappeler l'histoire de la création du critère JV avant de définir un cadre permettant de simuler un degré de corrélation connu et d'observer le comportement du critère, enfin, nous proposerons plusieurs propriétés limites.

5.2.1 Histoire

Littérature

Suivant les papiers [54] et [53] on cherche à comprendre la relation entre les réponses d'individus d'une population de taille n à deux QCM à réponse unique ayant respectivement p et q cases à cocher. Dans le papier [102], il s'agit de comprendre la relation entre deux partitions de cette même population. Le passage d'un problème à l'autre consiste naturellement à considérer que les individus sont classés dans la partition selon leurs réponses.

On dispose ainsi de p et q classes nominales qui sont attribuées à chaque individu par deux juges U et V .

- $U : i \in \{1..n\} \rightarrow U_i \in \{1.., u, ..p\}$
- $V : j \in \{1..n\} \rightarrow V_j \in \{1.., v, ..q\}$

Les deux juges associent à chaque élément une réalisation parmi les classes nominales disponibles. Comme présenté précédemment, deux visions permettent alors de regrouper l'ensemble des données fournies.

Vision	Tableau de contingence	Tableau relationnel
Taille du tableau	$p \times q$	$n \times n$
Valeur stockée dans chaque case	$n_{u,v}$	(X_i, Y_j)

Ayant formalisé le stockage des données, la question habituelle est d'en extraire une mesure de la dépendance.

Pour renforcer l'appellation "juge", nous supposons un jugement du même individu par deux juges justement. Dans ce cas, les modalités sont les différents jugements possibles ce qui conduit à poser $p = q$ et l'on espère bien que les jugements sont liés. Pour autant, on sait qu'ils seront, au moins parfois, différents. Poser $U = V$ est donc très restrictif et faux. Estimer à quel point ils sont proches mène au problème général ci-dessous.

Problème 12 (Estimer la dépendance).

Quel que soit le stockage choisi, le but est de concevoir un critère avec de bonnes propriétés donnant une estimation de la corrélation ou de l'indépendance entre les deux juges.

Pour faire face à ce problème, nous reprenons les étapes ayant mené, pour proposer une solution, à définir le critère dit de "Janson-Vegelius" que nous avons déjà rencontré.

Un premier critère pour ce problème est proposé dans [81]. Il consiste simplement à se baser sur le χ_2 , normalisé selon la formule

$$R_P = \sqrt{\frac{\chi_2}{n + \chi_2}}.$$

Des décennies plus tard, dans le papier [53], les auteurs proposent une suite de propriétés que nous reprenons ici, avec R le critère à définir et X, Y, Z des vecteurs regroupant les données sous la vision relationnelle.

1. $R(X, Y) = R(Y, X)$
2. $|R(X, Y)| \leq 1$
3. $R(X, X) = 1$
4. si $R(X, Z) = -1$ alors $R(X, Y) = -R(Y, Z)$
5. si $R(X, Z) = 1$ alors $R(X, Y) = R(Y, Z)$
6. si $R(X, V) = 1$ et $R(Y, Z) = 1$ alors $R(X, Z) = 1$
7. une matrice de corrélation basée sur R doit être définie positive

En bref, beaucoup de ces propriétés orientent la recherche vers un produit scalaire renormé. Un critère vérifiant ces propriétés est appelé un E-coefficient. On observe immédiatement que le critère R_P ne convient pas en ce que le point 3 n'est pas satisfait. Dans [92], Tschuprow introduit un nouveau critère :

$$R_T = \frac{\chi_2}{n\sqrt{(p-1)(q-1)}}$$

et ce bien avant la définition d'un E-coefficient.

Nous avons déjà justifié le bien-fondé de cette normalisation dans la remarque 23 et il est par ailleurs démontré dans [52] que R_T est bien un E-coefficient. L'année suivant la publication de ce papier, les mêmes auteurs en introduisent un nouveau sous la forme d'un produit scalaire normé (un cosinus) dans [53]. L'introduction proposée étant légèrement différente de l'approche relationnelle, nous la présentons ici avant de relier les deux visions.

Reprenant le second codage possible pour stocker les données (celle en $n \times n$), nous transformons légèrement la matrice de représentation en définissant :

Définition 30 (Transformation Q).

$$Q(U_i, U_j) = (p - 1)\mathbb{1}_{U_i=U_j} - \mathbb{1}_{U_i \neq U_j}$$

Remarque 31. En reprenant X issu de nos notations relationnelles, $Q(U_i, U_j) = (p - 1)x_{i,j} - (1 - x_{i,x}) = px_{i,j} + 1$

Évidemment, Q s'applique de manière similaire à V . On obtient, en l'appliquant, deux matrices (\tilde{X}, \tilde{Y}) selon la formule :

Définition 31 (Matrices transformées dites relationnelles).

$$\tilde{x}_{i,j} = Q(U_i, U_j) \text{ et } \tilde{y}_{i,j} = Q(V_i, V_j)$$

Finalement, le coefficient proposé dans le papier originel est défini par :

Définition 32.

$R_{JV}(U, V) = \frac{(\tilde{X}, \tilde{Y})}{\|\tilde{X}\| \times \|\tilde{X}\|}$ où $\|\cdot\|$ est la norme 2 sur les matrices et (\cdot, \cdot) le produit scalaire associé.

Ce n'est pas exactement la forme du JV que nous avons proposée dans la proposition 20 ; la section suivante montrera que les deux écritures sont équivalentes.

Réécriture

Nous avons introduit dans un chapitre précédent, en définition 13, les matrices relationnelles. Regardons comment se réécrit R_{JV} selon cette notion. Notons comme habituellement, X, Y les matrices relationnelles (canoniques) et \tilde{X}, \tilde{Y} celles du papier. On voit que $X = \frac{\tilde{X}+1}{p}$ et également $Y = \frac{\tilde{Y}+1}{p}$ de sorte que :

$$\begin{aligned} R_{JV}(U, V) &= \frac{(\tilde{X}, \tilde{Y})}{\|\tilde{X}\| \times \|\tilde{Y}\|} \\ &= \frac{(\tilde{X}/p, \tilde{Y}/q)}{\|\tilde{X}/p\| \times \|\tilde{Y}/q\|} \\ &= \frac{((\tilde{X} + 1 - 1)/p, (\tilde{Y} + 1 - 1)/q)}{\|(\tilde{X} + 1 - 1)/p\| \times \|(\tilde{Y} + 1 - 1)/q\|} \\ &= \frac{(X - 1/p, Y - 1/q)}{\|X - 1/p\| \times \|Y - 1/q\|} \end{aligned}$$

Finalement, avec les matrices relationnelles canoniques, le critère de Janson-Vegelius se réécrit :

$$R_{JV}(U, V) = \frac{(X - 1/p, Y - 1/q)}{\|X - 1/p\| \times \|Y - 1/q\|} \quad (5.3)$$

de sorte que l'on retrouve bien la définition donnée dans le présent document.

Cette réécriture nous permet d'apprécier les qualités du critère comme ayant été construit pour être un E-coefficient. Nous ne reviendrons pas sur la justification des points caractérisant un E-coefficient mais proposons plutôt d'étudier le JV en particulier en nous donnant un cadre paramétrique de corrélation.

5.2.2 Modèle des deux juges

On construit un modèle très simple de deux juges liés pour tester l'efficacité de R_{JV} à mesurer la dépendance. On se donne $\epsilon \in [0, 1]$, $\mathbf{B} \sim \text{Ber}(\epsilon)$ mesurant une corrélation et \mathbf{M} tirant une modalité selon p classes pour construire deux juges liés selon le tirage :

- $\mathbf{B} = 1 \rightarrow U$ et V même tirage de \mathbf{M}
- $\mathbf{B} = 0 \rightarrow U$ et V deux tirages indépendants de \mathbf{M}

Ce modèle peut paraître simple mais il permet d'appréhender efficacement le comportement de notre critère. Nous voulons dès lors évaluer $JV(X^n, Y^n, \epsilon)$, avec X^n, Y^n construits comme indiqué dans la définition 14 par une succession de tirages suivant le modèle tout juste introduit et ce pour différentes valeurs des paramètres ϵ et p .

Une première simulation qui apportera son lot de remarques consiste à répéter des tirages et calculer le JV sur chaque ensemble de tirages pour observer sa dépendance aux différents paramètres et dessiner une tendance. Nous affichons tout d'abord dans la figure 5.1 la variation du JV pour différentes valeurs de ϵ , p étant fixé.

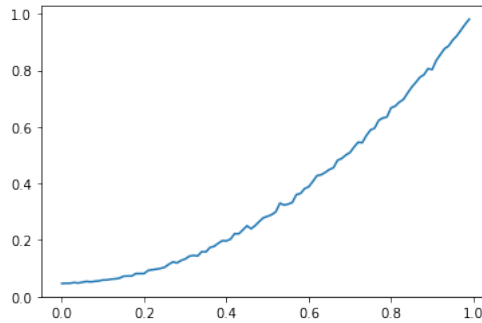


FIGURE 5.1 – Evolution du JV selon ϵ lorsque $p = 6$

On peut constater que l'évolution est proche de ϵ^2 et surtout un point intéressant : $\lim_{\epsilon \rightarrow 0} JV_{\epsilon,p} > 0$ avec les deux interprétations :

Remarque 32 (Égalité). *Si $U = V(\epsilon = 1)$, on a bien $JV = 1$*

Remarque 33 (Indépendance totale). *Si $U \neq V(\epsilon = 0)$, le JV voit sa valeur augmentée par toutes les possibilités d'égalités hasardeuses : \mathbf{M} tire les deux mêmes valeurs. On s'attendrait donc à voir JV diminuer quand p augmente.*

Pour faire suite à la seconde remarque, nous proposons de faire varier p et réalisons une succession de tirages avec les deux paramètres variables. Le résultat présenté dans la Figure 5.2 montre une tendance claire mais surprenante.

Contrairement à ce que nous supposions dans la remarque 33 on observe deux tendances :

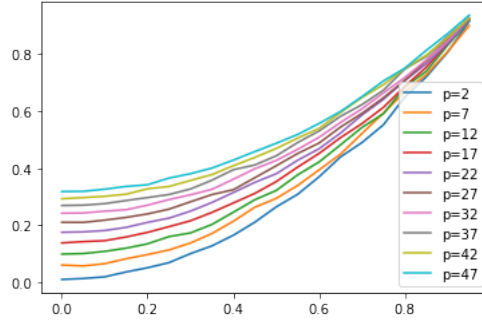


FIGURE 5.2 – Evolution du JV selon ϵ et p

1. JV augmente avec ϵ ce qui est naturel : les deux matrices dont on fait le produit scalaire selon l'équation 5.3 sont plus proches si ϵ augmente car les couples $X_{i,j} = Y_{i,j} = 1$ sont plus fréquents et les normes inchangées (cf Figure 5.3 et la proposition 32).
2. JV augmente avec p .

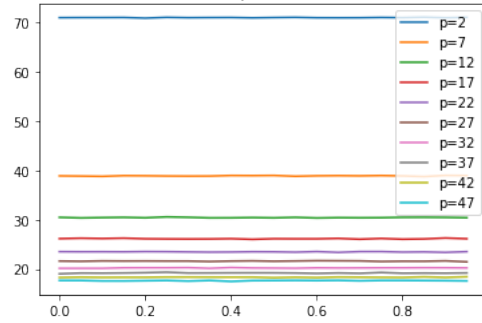


FIGURE 5.3 – Variation des normes selon ϵ et p

En détaillant la formule relationnelle de la proposition 20 on note que le résultat dépend du produit scalaire $(X|Y)$ et des normes du dénominateur : $\|X - \frac{1}{p}\|$ et $\|Y - \frac{1}{q}\|$. Or, si le numérateur, dans notre modèle dépend de la corrélation ϵ , les normes des matrices sont, elles, indépendantes de ϵ puisqu'elles résultent en moyenne des tirages de \mathbf{M} . Nous détaillerons ultérieurement l'impact de ϵ sur le numérateur dans la proposition 32.

Pour l'instant, intéressons-nous au nombre de cases remplies dans les matrices relationnelles, c'est à dire exactement à leur norme. On reprend l'expression de l'équation 5.3 :

$$X_{i,j} = \begin{cases} 1 & \text{si } i = j \\ \delta_{U_i=U_j} & \text{sinon} \end{cases}$$

et l'on remarque que $\delta_{U_i=U_j}$ vaut 1 avec probabilité $1/p$ si $i \neq j$ et avec probabilité 1 sinon. Il y a ainsi $n(n-1)$ cases à remplir et les liens entre ces cases sont complexes à traduire en probabilités : de fait les corrélations sont importantes, si $U_i = U_j$ et $U_j = U_k$ on a évidemment $U_i = U_k$. Nous aurons l'occasion de traduire ces corrélations sous la forme d'une matrice de permutation. Pour l'instant, on se place à n fixé et l'on fait croître p .

De l'expression $\mathbb{P}(U_i = U_j) = \frac{1}{p}$ on a si $p \rightarrow \infty$, $\mathbb{P}(U_i = U_j) \rightarrow 0$ dès que $i \neq j$. D'où l'on déduit les limites suivantes expliquant la croissance de JV en p .

Propriété 30 (limites $p \rightarrow \infty$).

1. $\forall n, \lim_{p \rightarrow \infty} X^n, Y^n = I_n$
2. $\forall p, \lim_{\epsilon \rightarrow 1} R_{JV}(X^n, Y^n) = 1$
3. $\forall \epsilon, \lim_{p \rightarrow \infty} R_{JV}(X^n, Y^n) = 1$

Démonstration.

1. $\lim_{p \rightarrow \infty} \mathbb{P}(U_i = U_j) = 0$ si $i \neq j$
2. pour $\epsilon = 1$, $X = Y$ et l'expression de R_{JV} est continue
3. lorsque $p \rightarrow \infty$, le point 1. et la continuité de R_{JV} donnent le résultat.

□

Cette construction paramétrique autour de deux juges corrélés fut l'occasion de prouver des propriétés du JV sur un exemple tout en pouvant les observer sur des graphiques. Notamment, nous pouvons vérifier que la diagonalisation proposée par la proposition 32 garantit des normes constantes observables dans la Figure 5.3. Observer des propriétés anticipées conforte toujours sur leur véracité.

Dans les parties suivantes, nous allons simuler puis démontrer des propriétés générales du critère en nous affranchissant du modèle des deux juges.

5.2.3 Valeur moyenne

Simulations

L'idée est ici de simuler des lois de probabilités π aléatoires sur $p \times p$ uniformément puis de calculer les valeurs du critère R_{JV} sur celles-ci afin d'observer sa distribution selon p .

Nous proposons tout d'abord la figure 5.4 qui présente la distribution du critère. Cette distribution a l'aspect d'une gaussienne restreinte à $[0, 1]$. Un élément essentiel frappe immédiatement : les valeurs se resserrent autour de 0 à mesure que p augmente. C'est une constatation, il nous reste à le montrer en théorie. Nous commencerons par le démontrer dans le cas de π^\times pour lequel la formule est simplifiée d'après la proposition 3 avant de le démontrer dans le cas général.

Pour l'instant, nous proposons d'étudier la figure 5.5 simulant justement les deux cas à traiter. Sur la courbe "Général", on tire au hasard une matrice de probabilités sur $p * p$ modalités et on calcule la valeur du R_{JV} ; l'opération est répétée 1000 fois et la moyenne est présentée. Sur la courbe "Indépendance", nous tirons au hasard une loi de probabilité sur p modalités que nous couplons ensuite avec elle-même selon une relation d'indépendance ; le même nombre de simulation est ensuite appliqué.

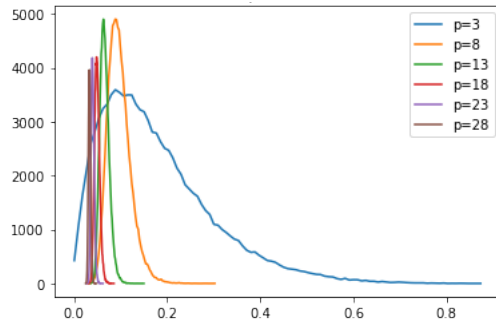


FIGURE 5.4 – Variation du JV selon des tirages de π pour différentes valeurs de p

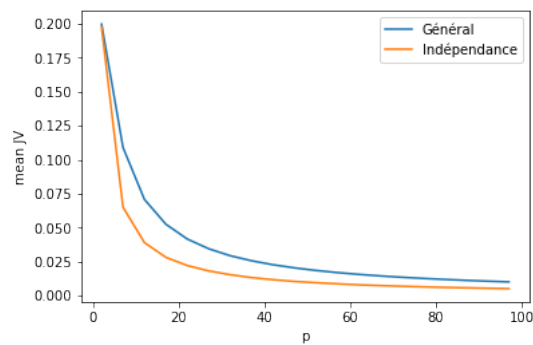


FIGURE 5.5 – Variation du JV selon des tirages de π et π^\times pour différentes valeurs de p

Théorie – cas indépendance

Cette section propose la démonstration (dans le cas du couplage indépendant de μ avec elle-même) de la propriété limite remarquée dans la section précédente et énoncé dans le théorème 9.

Théorème 9 (Limite du JV, cas indépendance).

$$\lim_{p \rightarrow \infty} \mathbb{E}_\mu [JV(\mu \otimes \mu)] = 0$$

Démonstration. On commence par utiliser une suite d'inégalités permettant de s'inté-

resser à un membre réduit de R_{JV} .

$$\begin{aligned}
JV(\mu \otimes \mu) &= \frac{p^2 \sum_{u=1}^p \left(\mu_u - \frac{1}{p}\right)^2 \sum_{u=1}^p \left(\mu_u - \frac{1}{p}\right)^2}{p(p-2)(\sum_{u=1}^p \mu_u^2) + 1} \\
&= \frac{p^2 \left(\sum_{u=1}^p \mu_u^2 - \frac{1}{p}\right)^2}{p(p-2)(\sum_{u=1}^p \mu_u^2) + 1} \\
&\leq \frac{p^2 \left(\sum_{u=1}^p \mu_u^2 - \frac{1}{p}\right)^2}{p^2(p-2)\frac{1}{p^2} + 1} \\
&\leq \frac{p^2 \left(\sum_{u=1}^p \mu_u^2 - \frac{1}{p}\right)^2}{p-1} \\
&\leq 2p \left(\sum_{u=1}^p \mu_u^2 - \frac{1}{p}\right)^2
\end{aligned}$$

Pour démontrer la convergence, nous rappelons la loi de Dirichlet déjà introduite dans la définition 7 dont nous précisons les moments pour en déduire le calcul exact de l'espérance du majorant. Cette loi est souvent présentée comme une distribution sur les distributions. C'est ainsi que nous l'utiliserons ici. En effet, on remarque que la loi uniforme sur le simplexe que nous utilisons déjà pour la démonstration de la proposition 4 n'est autre que le cas particulier $\alpha_1 = \dots = \alpha_p = 1$.

Propriété 31 (Moments de Dirichlet).

Étant donné μ_1, \dots, μ_p suivant la loi de Dirichlet de paramètre $\alpha_1, \dots, \alpha_p$, on note $\alpha_0 = \alpha..$ Alors pour tout p -uplet β_1, \dots, β_p d'entiers positifs, on a la formule (avec $\beta_0 = \beta.$) :

$$\mathbb{E} \left(\prod_{u=1}^p \mu_u^{\beta_u} \right) = \frac{\Gamma(\sum_{u=1}^p \alpha_u)}{\Gamma(\sum_{u=1}^p \alpha_u + \beta_u)} \prod_{u=1}^p \frac{\Gamma(\alpha_u + \beta_u)}{\Gamma(\alpha_u)} = \frac{\Gamma(\alpha_0)}{\Gamma(\alpha_0 + \beta_0)} \prod_{u=1}^p \frac{\Gamma(\alpha_u + \beta_u)}{\Gamma(\alpha_u)}$$

Lemme 2 (Terme tendant vers 0 et vitesse).

$$\mathbb{E} \left[\left(\sum_{u=1}^p \mu_u^2 - \frac{1}{p} \right)^2 \right] = o\left(\frac{1}{p}\right)$$

Cela suffit à montrer la convergence en espérance de R_{JV} vers 0 d'après la majoration de la première partie où nous avons isolé le terme devant tendre vers 0.

Remarque 34 (Vitesse de convergence et minimum).

Ce lemme 2 montre que la partie de gauche se rapproche de son minimum et expose de plus la vitesse de convergence. Rappelons que ce minimum est atteint uniquement dans le cas d'une marge uniforme.

Remarque 35 (contradiction avec une proposition précédente?).

On a vu dans un théorème précédent (cf proposition 4) qu'il y a très probablement au moins un élément μ_u plus petit que $\frac{1}{2p}$ mais cela n'empêche pas qu'une proportion suffisante se regroupe autour de $\frac{1}{p}$ pour que le lemme soit valide.

Démonstration.

On va essentiellement développer et utiliser la proposition 31 dans le cas $\alpha_1 = \dots = \alpha_p = 1$.

$$\mathbb{E} \left[\left(\sum_{u=1}^p \mu_u^2 - \frac{1}{p} \right)^2 \right] = \sum_{1 \leq u, v \leq p} \mathbb{E}(\mu_u^2 \mu_v^2) - \frac{2}{p} \sum_{u=1}^p \mathbb{E}(\mu_u^2) + \frac{1}{p^2}$$

Pour le cas où la puissance totale β est 4 séparée en 2 dans le premier terme ($u \neq v$) :

$$\mathbb{E}(\mu_u^2 \mu_v^2) = \frac{\Gamma(\alpha_0)}{\Gamma(\alpha_0 + \beta_0)} \frac{\Gamma(\alpha_u + 2)\Gamma(\alpha_v + 2)}{\Gamma(\alpha_u)\Gamma(\alpha_v)}; \quad (5.4)$$

$$= \frac{4}{p(p+1)(p+2)(p+3)}. \quad (5.5)$$

Pour le cas où la puissance totale β est 4 en une fois dans le premier terme ($u = v$) :

$$\mathbb{E}(\mu_u^4) = \frac{2 * 3 * 4}{p(p+1)(p+2)(p+3)}; \quad (5.6)$$

$$= \frac{24}{p(p+1)(p+2)(p+3)}. \quad (5.7)$$

Enfin, pour le cas où la puissance totale β est 2 dans le second terme

$$\mathbb{E}(\mu_u^2) = \frac{2}{p(p+1)} \quad (5.8)$$

Au final, en combinant 5.5, 5.7 et 5.8, on obtient :

$$\begin{aligned} \mathbb{E} \left[\left(\sum_{u=1}^p \mu_u^2 - \frac{1}{p} \right)^2 \right] &= \frac{4p(p-1)}{p(p+1)(p+2)(p+3)} + \frac{24p}{p(p+1)(p+2)(p+3)} - \frac{2 * 2p}{p(p+1)} + \frac{1}{p^2} \\ &= \frac{4(p-1)}{(p+1)(p+2)(p+3)} + \frac{24}{(p+1)(p+2)(p+3)} - \frac{4p}{(p+1)} + \frac{1}{p^2} \\ &= \frac{4p(p+5) - 4(p+2)(p+3)}{p(p+1)(p+2)(p+3)} + \frac{1}{p^2} \\ &= \frac{-24}{p(p+1)(p+2)(p+3)} + \frac{1}{p^2} \end{aligned}$$

Ce qui permet de conclure que même multiplié par p^r avec $0 < r < 2$, le terme évalué tend toujours vers 0, en détail :

$$\mathbb{E}_\mu [JV(\mu \otimes \mu)] = \mathcal{O}\left(\frac{1}{p}\right)$$

On termine ainsi la preuve du lemme 2 et du Théorème 9. □

□

Théorie – cas général

Ayant démontré que R_{JV} tend vers 0 dans le cas du couplage de μ aléatoire avec elle-même selon l'indépendance, nous voulons montrer le théorème suivant plus général.

Théorème 10 (Limite du JV).

$\lim_{pq \rightarrow \infty} \mathbb{E}_\pi [JV(\pi)] = 0$ où π est une loi de probabilité tirée uniformément dans le simplexe de dimension $p * q$.

Démonstration. On montre de manière similaire au cas indépendance l'inégalité

$$JV(\pi) \leq \frac{4}{\sqrt{pq}} \left(pq \sum_{u,v} (\pi_{u,v}^2) - p \sum_u (\pi_{u,\cdot}^2) - q \sum_v (\pi_{\cdot,v}^2) + 1 \right)$$

Calculons donc l'espérance du majorant en utilisant la Proposition 31 mais avec cette fois une loi de Dirichlet sur le simplexe de taille pq . On considère ainsi $(\pi_{u,v})_{1 \leq u \leq p, 1 \leq v \leq q}$ comme un vecteur de taille pq .

Pour le premier terme

$$\mathbb{E}(\pi_{u,v}^2) = \frac{2}{pq(pq+1)} \quad (5.9)$$

Pour le second terme

$$\mathbb{E}(\pi_{u,\cdot}^2) = \mathbb{E} \left(\sum_{v,v'} \pi_{u,v} \pi_{u,v'} \right) \quad (5.10)$$

$$= (q^2 - q) \frac{1}{pq(pq+1)} + q \frac{2}{pq(pq+1)} \quad (5.11)$$

$$= \frac{q+1}{p(pq+1)} \quad (5.12)$$

En combinant 5.9 et 5.12 et en jouant de symétrie pour le troisième terme, on obtient :

$$\begin{aligned} & \mathbb{E} \left(pq \sum_{u,v} (\pi_{u,v}^2) - p \sum_u (\pi_{u,\cdot}^2) - q \sum_v (\pi_{\cdot,v}^2) + 1 \right) \\ &= p^2 q^2 \frac{2}{pq(pq+1)} - p^2 \frac{q+1}{p(pq+1)} - q^2 \frac{p+1}{q(pq+1)} + 1 \\ &= \frac{2pq}{pq+1} - \frac{p(q+1)}{pq+1} - \frac{q(p+1)}{pq+1} + 1 \\ &= \frac{-p-q}{pq+1} + 1 \end{aligned}$$

De sorte que l'on a :

$$\mathbb{E}_\pi [JV(\pi)] = \mathcal{O} \left(\frac{1}{\sqrt{pq}} \right)$$

□

Remarque 36. Comme R_{JV} est positif on a aussi montré que R_{JV} tendait vers 0 en probabilités.

Dans cette section, nous avons tout d'abord appliqué le critère de Janson Vegelius dans un cadre jouet afin de proposer des limites qui, bien qu'évidentes, montrent les biais dans lesquels on tombe aisément en utilisant un critère de corrélation. Ainsi, la propension du JV à se rapprocher de 0 suppose que nous l'utilisons avec parcimonie ; si dans une application, p est grand, le rapprochement de 0 ne suppose pas une indépendance. Dans une seconde partie nous avons transposé l'observation principale inscrite dans le modèle des deux juges dans le cadre général et prouvé qu'elle restait valable.

Dans le modèle des deux juges, nous avons repoussé le calcul direct de la norme $\|X\|$ en tant que la matrice X possédait des relations internes trop complexes. La section suivante sera l'occasion de revenir sur ce point en utilisant un regroupement par bloc des modalités assorti d'une permutation des individus.

5.3 Un produit scalaire commun

Que ce soit dans les critères usuels où le produit scalaire apparaît sous sa forme relationnelle ou bien dans l'optimisation du problème 6 où il apparaît sous sa forme de contingence, la quantité

$$(X|Y) = \sum_{u=1}^v \sum_{v=1}^q n_{u,v}^2 = \sum_{i=1}^n \sum_{j=1}^n x_{i,j} y_{i,j} \quad (5.13)$$

joue un rôle essentiel dans les problèmes qui nous intéressent. Estimer sa valeur ou du moins la borner permet de mieux comprendre comment ce produit se comporte et par ailleurs de créer des critères entre 0 et 1 tout simplement en divisant par la borne supérieure.

Cette section s'applique à l'analyse de ce terme commun. Une première partie regroupe des bornes de cette quantité et notamment certaines apportant une amélioration des maximums existants et apparaissant dans la littérature sur le sujet : [4], [61], [71]. Une seconde partie en détaille la composition passant d'une dépendance aux marges et à la corrélation entre celles-ci (concept abstrait et dont l'estimation justifie précisément le pan de recherche sur les critères de corrélation) à une dépendance aux marges et à une permutation de taille n . Ainsi à nombre n de tirages fixé, on résume la valeur de $(X|Y)$ à celle d'une permutation.

Remarquant alors que de multiples permutations génèrent la même valeur, on quotiente l'ensemble des permutations par celle-ci. Survient alors une corrélation positive entre la probabilité d'un groupe de permutations et la croissance de $(X|Y)$. Nous nous appuyons sur cette corrélation pour définir une heuristique de résolution du problème 6 dans le cadre entier, précisément décrit dans le problème 13.

Commençons par les résultats historiques sur les normes de notre quantité.

5.3.1 Encadrement

Problème 13.

$$\begin{aligned} \min \text{ ou } \max \quad & \sum_{u=1}^p \sum_{v=1}^q n_{u,v}^2 = (X|Y) & (5.14) \\ \text{sous les contraintes :} & \\ & \sum_{v=1}^q n_{u,v} = n_{u,\cdot} \quad \text{et} \quad \sum_{u=1}^p n_{u,v} = n_{\cdot,v} \\ & \sum_{u=1}^p \sum_{v=1}^q n_{u,v} = N \quad \text{et} \quad n_{u,v} \geq 0 \end{aligned}$$

D'après les résultats déjà démontrés dans ce document et notamment d'après le théorème 2 (adapté), on connaît la forme du minimum atteint pour $n_{u,v}^* = \left[\frac{n_{u,\cdot}}{q} + \frac{n_{\cdot,v}}{p} - \frac{N}{pq} \right]$ (à supposer que l'on reste dans le cas entier ce qui est une hypothèse très lourde).

Reste que le maximum, lui, est inconnu. Son obtention requiert l'utilisation d'algorithmes ad hoc tandis que la borne de contingence la plus fine connue est celle de Hubert et Arabia introduite dans [4] et donnée par :

$$\sum_{u,v} n_{u,v}^2 \leq \min \left[\sum_u n_{u,\cdot}^2, \sum_v n_{\cdot,v}^2 \right]$$

Celle-ci est meilleure que la borne classique de Cauchy-Schwarz (voir l'article [61] pour une preuve contingentielle). Nous proposerons ici une preuve utilisant des transformations relationnelles. Par ailleurs, l'emploi du théorème de Hoffman-Wielandt (voir [64]) permet d'enrichir l'apport de cette borne.

Pour l'instant, commençons par rappeler quelques concepts et objets.

Définition 33 (Matrices normales).

Une matrice $N \times N$ sur l'espace complexe \mathbb{C} , A est dite "normale" lorsque $AA^* = A^*A$ avec A^* sa transposée-conjuguée.

Cela revient à dire qu'il existe une matrice unitaire U telle que UAU^* soit diagonale ; les éléments de la diagonale étant précisément les valeurs propres de A .

Toute matrice Hermitienne ($A = A^*$) est normale et, parmi les matrices réelles, les matrices symétriques sont normales.

Définition 34 (Norme de Frobenius, norme de la trace).

La norme Trace d'une matrice A est définie par $\|A\|_F^2 = \text{Tr}[AA^*] = \text{Tr}[A^*A] = \sum_{i=1}^N \sum_{j=1}^N a_{i,j}^2$.

Théorème 11 (Hoffmann-Wielandt).

Si l'on se donne A et B deux matrices réelles symétriques de taille $N \times N$ avec comme valeurs propres respectives $\lambda_1 \leq \dots \leq \lambda_N$ et $\mu_1 \leq \dots \leq \mu_N$ alors on a l'inégalité suivante :

$$\|A - B\|_F^2 = \text{Tr} [(A - B)^t(A - B)] \geq \sum_{i=1}^N (\lambda_i - \mu_i)^2$$

Remarquons alors que les matrices relationnelles X et Y que nous étudions sont par définition symétriques de sorte qu'elles rentrent dans les hypothèses du théorème 11. Par ailleurs, c'est bien connu et nous reviendrons sur ce point (voir la proposition 32) : leurs valeurs propres sont les tailles des clusters regroupant les éléments selon leur modalité. Dans le cas de p classes pour X et q pour Y et si l'on suppose $n_{1,\cdot} \geq \dots \geq n_{p,\cdot}$ pour X et $n_{\cdot,1} \geq \dots \geq n_{\cdot,q}$ pour Y on obtient : $[\lambda_1 = n_{1,\cdot} \geq \dots \geq \lambda_p = n_{p,\cdot} \geq 0, \dots, 0]$ et $[\mu_1 = n_{\cdot,1} \geq \dots \geq \mu_q = n_{\cdot,q} \geq 0, \dots, 0]$

On suppose alors $p \geq q$ et l'on utilise la Propriété 10 et le Théorème 11 pour obtenir :

$$\begin{aligned} \|X - Y\|_F^2 &= \sum_{i=1}^N \sum_{j=1}^N (x_{i,j} - y_{i,j})^2 \\ &\geq \sum_{i=1}^N (\lambda_i - \mu_i) \\ &= \sum_{u=1}^q (n_{u,\cdot} - n_{\cdot,u})^2 + \sum_{u=q+1}^p n_{u,\cdot}^2. \end{aligned}$$

En développant et en utilisant le principe de transfert relationnel, on trouve :

$$\sum_{u=1}^p \sum_{v=1}^q n_{u,v}^2 = \sum_{i=1}^N \sum_{j=1}^N x_{i,j} y_{i,j} \leq \sum_{u=1}^{\min(p,q)} n_{u,\cdot} n_{\cdot,u} \quad (5.15)$$

Où $n_{\cdot,u}$ et $n_{u,\cdot}$ sont les tailles des classes des matrices d'équivalence X et Y triées dans l'ordre décroissant. Le formalisme relationnel, nous allons le voir, permet de jouer sur l'espace pour extraire les bornes qui nous intéressent, amusons-nous avec.

Borne de Cauchy-Schwarz

En utilisant la propriété 10 et la caractère binaire des valeurs de X et Y :

$$\sum_{i=1}^N \sum_{j=1}^N x_{i,j} y_{i,j} \leq \sqrt{\sum_{i=1}^p \sum_{j=1}^q x_{i,j}^2} \sqrt{\sum_{i=1}^N \sum_{j=1}^N y_{i,j}^2} = \sqrt{\sum_{i=1}^p \sum_{j=1}^q x_{i,j}} \sqrt{\sum_{i=1}^N \sum_{j=1}^N y_{i,j}} = \sqrt{\sum_{u=1}^p n_{u,\cdot}} \sqrt{\sum_{v=1}^q n_{\cdot,v}}$$

Borne de Hubert-Arabie

En utilisant la propriété 10 et une propriété de l'opérateur min :

$$\sum_{i=1}^N \sum_{j=1}^N x_{i,j} y_{i,j} = \sum_{i=1}^N \sum_{j=1}^N \min(x_{i,j}, y_{i,j}) \leq \min \left[\sum_{i=1}^N \sum_{j=1}^N x_{i,j}, \sum_{i=1}^N \sum_{j=1}^N y_{i,j} \right] = \min \left[\sum_{u=1}^p n_{u,\cdot}, \sum_{v=1}^q n_{\cdot,v} \right]$$

Borne de Hoffman-Wielandt

Comme vu précédemment (en ordonnant de manière décroissante les $n_{u,v}$) :

$$\sum_{i=1}^N \sum_{j=1}^N x_{i,j} y_{i,j} = \sum_{u=1}^p \sum_{v=1}^q n_{u,v}^2 \leq \sum_{u=1}^{\min(p,q)} n_{u,\cdot} n_{\cdot,u}$$

Borne de Marcotorchino-Messatfa

Toujours avec la propriété 10 et celle de l'opérateur min :

$$\begin{aligned} \sum_{i=1}^N \sum_{j=1}^N x_{i,j} y_{i,j} &= \sum_{i=1}^N \sum_{j=1}^N \min(x_{i,j}, y_{i,j}) \leq \sum_{i=1}^N \min[x_{i,\cdot}, y_{i,\cdot}] = \sum_{u=1}^p \sum_{v=1}^q \sum_{i \in C_{u,v}} \min[x_{i,\cdot}, y_{i,\cdot}] \\ &= \sum_{u=1}^p \sum_{v=1}^q \sum_{i \in C_{u,v}} \min(n_{u,\cdot}, n_{\cdot,v}) = \sum_{u=1}^p \sum_{v=1}^q n_{u,v} \min(n_{u,\cdot}, n_{\cdot,v}) \end{aligned}$$

On peut montrer que cette dernière borne 5.3.1 est la plus fine. On obtient en combinant ces bornes, les encadrements suivants présentés sous leur forme relationnelle et de contingence pour la quantité qui nous intéresse à savoir : $(X|Y)$.

Encadrement de contingence

En utilisant les calculs précédent et le théorème 2 on a :

$$\sum_{u=1}^p \frac{n_{u,\cdot}^2}{q} + \sum_{v=1}^q \frac{n_{\cdot,v}^2}{p} - \frac{N^2}{pq} \leq \sum_{u=1}^p \sum_{v=1}^q n_{u,v}^2 \leq \sum_{u=1}^p \sum_{v=1}^q n_{u,v} \min[n_{u,\cdot}, n_{\cdot,v}]$$

Encadrement relationnel

Avec les formules de transfert de la propriété 10 :

$$\sum_{i=1}^N \sum_{j=1}^N \left[\frac{x_{i,j}}{q} + \frac{y_{i,j}}{p} - \frac{1}{pq} \right] \leq \sum_{i=1}^N \sum_{j=1}^N x_{i,j} y_{i,j} \leq \sum_{i=1}^N \min[x_{i,\cdot}, y_{\cdot,i}] \quad (5.16)$$

Ayant proposé les encadrements classiques du produit scalaire récurrent dans les critères de corrélation, nous allons décomposer celui-ci pour mieux le cerner. De fait, les relations internes entre les cases de X et Y nous empêchent pour l'instant de facilement l'analyser. La section 5.3.3 va justement réécrire chacune des deux matrices pour lever cette difficulté et faciliter la simulation du produit scalaire. Commençons par rappeler des questions classiques de maximisation du membre principal.

5.3.2 Maximisation

Algorithme Itératif

Une question naturelle, après avoir encadré et minimisé le coefficient récurrent dans nos problèmes d'optimisation est d'estimer exactement la valeur maximale qu'il peut atteindre dans le cas de marges fixées. Il s'agit en ce cas d'optimiser les programmes suivants : $\max \sum_{u=1}^p \sum_{v=1}^q n_{u,v}^2$ ou encore $\max \sum_{i=1}^N \sum_{j=1}^N x_{i,j} x_{i,j}$ grâce à la propriété 10. Il s'agit a priori de problèmes entiers complexes que nous nous empressons donc de relâcher en acceptant des solutions réelles.

Problème 14.

$$\max \sum_{i=1}^N \sum_{j=1}^N x_{i,j} y_{i,j} \quad (5.17)$$

sous les contraintes :

$$\sum_{i=1}^N \sum_{j=1}^N x_{i,j} = \sum_{u=1}^p n_{u,\cdot}^2 = Const_1$$

$$\sum_{i=1}^N \sum_{j=1}^N y_{i,j} = \sum_{v=1}^q n_{\cdot,v}^2 = Const_2$$

X and Y deux relations d'équivalence vérifiant donc

les contraintes linéaires introduites dans la Section 3.4.3

Le problème est non linéaire et constitue un programme ayant pour solution des variables dans $\{0, 1\}$ sous contraintes linéaires ; peu de chance de parvenir à le résoudre dans le cas général. Reste que nous pouvons profiter de la séparabilité de la fonction de coût en X et Y . Nous pouvons fixer par exemple Y^1 satisfaisant les contraintes susmentionnées dans le problème 14.

Ensuite, nous optimisons $\max_X \sum_{i=1}^N \sum_{j=1}^N x_{i,j} y_{i,j}^1$ avec les mêmes contraintes sur X que celles dans le problème 14.

Ce dernier porte toujours sur des entiers $\{0, 1\}$ mais devient linéaire et permet donc l'utilisation de méthodes pour l'obtention d'une solution optimale comme expliqué dans le livre originel de J.F. Marcotorchino et P. Michaud [68] ou encore, de façon un peu plus simple, dans le livre de Thierry Brenac [16]. Ces méthodes ne marchent que si N est supposé pas trop large. Dans le cas contraire, on utilise des heuristiques adaptées comme dans [3].

Appelons maintenant X^1 la solution obtenue en X et cherchons dans une deuxième phase une solution Y^2 maximisant $\sum_{i=1}^N \sum_{j=1}^N x_{i,j}^1 y_{i,j}^2$ avec les contraintes habituelles.

Après résolution de ce nouveau problème avec des techniques identiques, on obtient Y^2 vérifiant par construction :

$$\sum_{i=1}^N \sum_{j=1}^N x_{i,j}^1 y_{i,j}^2 \geq \sum_{i=1}^N \sum_{j=1}^N x_{i,j}^1 y_{i,j}^1$$

Aussi, après m itérations la situation suivante se présente :

$$\sum_{i=1}^N \sum_{j=1}^N x_{i,j}^{m-1} y_{i,j}^m \geq \sum_{i=1}^N \sum_{j=1}^N x_{i,j}^{m-1} y_{i,j}^{m-1} \geq \dots$$

La fonction à optimiser étant discrète et bornée, la convergence du processus est garantie. Dans le cas où il se révélerait trop coûteux à itérer, on peut utiliser les heuristiques ad-hoc proposées notamment dans des papiers tels que celui de I.C. Lerman et P. Peter [61] et de H. Messatfa [71].

L'utilisation du problème 14 comme substitut au problème 13 permet ainsi, après séparation des variables d'obtenir une fonction de coût linéaire et conduit à sa résolution.

Pour autant, il nous faut remarquer que les deux problèmes ne sont pas parfaitement équivalents. En effet, si les hypothèses du problème 13 sont vérifiées, alors le problème 14 est parfaitement défini et ses hypothèses valides. Malheureusement, si l'on part du problème 14, l'obtention des variables initiales est un travail à part entière. Formellement, nous avons :

$$\begin{aligned} \sum_{i=1}^N \sum_{j=1}^N x_{i,j} &= \sum_{u=1}^p n_{u,\cdot}^2 = Const_1 \\ \sum_{i=1}^N \sum_{j=1}^N y_{i,j} &= \sum_{v=1}^q n_{\cdot,v}^2 = Const_2 \end{aligned} \quad (5.18)$$

Or, il est ainsi bien connu que $\sum_{i=1}^K a_i^2 = \sum_{i=1}^K b_i^2$ avec $\sum_{i=1}^K a_i = \sum_{i=1}^K b_i = N$ et a_i, b_i entiers peut avoir plusieurs solutions pour peu que $N \geq 9$. Par exemple, si $N = 9, K = 3$, nous avons les deux distributions suivantes $(5 - 2 - 2)$ et $(4 - 4 - 1)$ de somme 9 et de somme des carrés $(25 + 4 + 4 = 16 + 16 + 1 = 33)$; toutes les deux égales.

Cela montre que la connaissance de $\sum_{i=1}^N \sum_{j=1}^N x_{i,j} = \sum_{u=1}^p n_{u,\cdot}^2 = Const_1$ n'est pas suffisante pour déterminer avec une correspondance certaine les $n_{u,\cdot}$ et de même pour Y .

Ainsi, on constate que la résolution du problème 14 et notamment de ses contraintes ne permet pas automatiquement celle du problème 13.

En résumé, si l'on connaît exactement les marges $n_{u,\cdot}$ et $n_{\cdot,v}$, les deux problèmes se révèlent équivalents et il est plus simple de le résoudre sous sa forme relationnelle. Malheureusement, si l'on ne connaît que les quantités 5.18, l'équivalence tombe et il nous faut résoudre le problème sous sa forme de contingence.

Coefficient avec centrage

Le coefficient que nous avons étudié est rarement utilisé tel quel : la plupart du temps, il est associé à un centrage des deux côtés. Nous allons étudier l'impact que peut produire ce décalage sur la solution d'une généralisation du problème 6.

On s'intéresse au numérateur de R_{JV} comme rappelé dans la propriété 20 en changeant les constantes, il s'agit d'une variante du problème 6. On impose les contraintes de marge et la fameuse hypothèse 2.9. Formellement, le problème s'écrit :

Problème 15 (MTM avec centrage).

$$C_{\alpha,\beta}(X, Y) = (X - \frac{\alpha}{p})(Y - \frac{\beta}{q})$$

sous les contraintes :

$$\sum_v \pi_{u,v} = \mu_u$$

$$\sum_u \pi_{u,v} = \nu_v$$

$$\sum_{u,v} \pi_{u,v} = 1$$

$$\pi_{u,v} \geq 0$$

L'énoncé du problème 15 mêle volontairement les écritures relationnelle et de contingence, chacune adaptée à une partie.

Commençons néanmoins par réécrire la fonction de coût, en utilisant les formules de passage, $0 = \sum_{i,j} (X_{i,j} - \frac{\alpha}{p})(Y_{i,j} - \frac{\beta}{q})$ se réécrit :

$$\sum_{u,v} n_{u,v}^2 = \frac{\beta}{q} \sum_u n_{u,\cdot}^2 + \frac{\alpha}{p} \sum_u n_{\cdot,v}^2 - \frac{\alpha\beta}{pq} n^2$$

En divisant par n^2 :

$$\sum_{u,v} \pi_{u,v}^2 = \frac{\beta}{q} \sum_u \pi_{u,\cdot}^2 + \frac{\alpha}{p} \sum_u \pi_{\cdot,v}^2 - \frac{\alpha\beta}{pq}$$

Le but est maintenant de trouver un π minimisant cette expression et respectant les marges. Nous ne saurons le faire sans introduire un sous-ensemble paramétrique au sein duquel optimiser. On s'inspire de la forme de π^+ et on pose $\pi = \frac{\mu_u \sqrt{\alpha}}{q} + \frac{\nu_v \sqrt{\beta}}{p} - \frac{K}{pq}$ où K est une constante à fixer.

Filons les calculs pour découvrir les valeurs optimales des diverses constantes.

Si l'on injecte cette forme dans l'équation à annuler, on obtient après simplification la forme quadratique :

$$C_{\alpha,\beta}(A, B)(\pi) = K^2 - 2K(\sqrt{\alpha} + \sqrt{\beta}) + \alpha\beta + 2\sqrt{\alpha\beta}$$

Cette expression varie entre $+\infty$ et son minimum qui vaut $\alpha\beta - \alpha - \beta$. Ce dernier peut être négatif ce qui est très différent du cadre habituel qui aboutit à un écart quadratique à l'argmin. Par ailleurs, il est atteint pour $K = \frac{\sqrt{\alpha} + \sqrt{\beta}}{\alpha\beta + 2\sqrt{\alpha\beta}}$.

Nous utilisons maintenant une condition de marges : $\sum_{u,v} \pi_{u,v} = 1$. Cela impose $K = \sqrt{\alpha} + \sqrt{\beta} - 1$ puis, $C_{\alpha,\beta} = (\alpha - 1)(\beta - 1)$. On voit donc que la position $\alpha = \beta = 1$ est très particulière en ce que la valeur de K imposée par les marges amène justement $C_{\alpha,\beta} = 0$.

Enfin, pour que $\pi_{\alpha,\beta}$ soit éligible, on a la condition de positivité modifiée suivante :

$$H_{\alpha,\beta} >: p\sqrt{\beta} \min \mu + q\sqrt{\alpha} \min \nu \geq \sqrt{\alpha} + \sqrt{\beta} - 1 \quad (5.19)$$

qui nous permet bien de retrouver l'hypothèse 2.9 si $\alpha = \beta = 1$.

L'objet de cette section était de préciser des méthodes de maximisation du coefficient. La section suivante présente une approche sensiblement différente : nous montrons que nous pouvons diagonaliser les matrices relationnelles. Cette réécriture apporte un lot de propriétés sur ces matrices et conduit surtout à la distribution théorique du coefficient principal avec les marges comme paramètres.

5.3.3 Diagonalisation

L'objectif de cette section est de proposer une distribution pour le coefficient principal et ce dans le contexte le plus général possible. Nous commencerons par préciser la forme des matrices relationnelles X et Y afin de ramener l'étude de leur produit scalaire à celle d'une permutation σ . Par la suite, nous montrerons que $(X|Y)$ ne dépend que d'une caractéristique de σ propre à notre étude : le nombre de "sauts" que nous définissons dans la propriété 34 lorsque X a deux classes de tailles r et $n-r$. Cette quantité est généralisée dans l'équation 5.20 et nous ramène aux matrices de contingence. *In fine*, pour des marges figées, nous disposerons de la distribution du coefficient principal de nombreux critères. Ce sera l'occasion de vérifier théoriquement la concentration des critères déjà remarquée empiriquement dans [44].

Produit scalaire et permutations

Définir directement une loi de probabilité sur X, Y deux matrices d'un codage relationnel est complexe car de nombreuses dépendances existent entre des cases très éloignées. Ainsi si

$$X_{i_1, i_2} = X_{i_2, i_3} = \dots = X_{i_{r-1}, i_r} = 1$$

alors

$$X_{i_r, i_{r+1}} = 1 \implies X_{i_1, i_{r+1}} = 1$$

Aussi, on utilise une réécriture plus élégante pour définir une loi de probabilités dessus.

Propriété 32 (Diagonalisation de X).

Il existe une matrice de permutation P_σ et un entier r_x égal au nombre de modalités visibles de X tels que

$$X = P_\sigma \text{Diag}(U_{n_1}, \dots, U_{n_{r_x}}) = P_\sigma D_X$$

et de même pour Y .

Avec U_i le bloc unitaire carré de taille i .

Démonstration.

Il suffit de remarquer qu'une matrice relationnelle se rapporte toujours à un tirage, quitte à utiliser l'algorithme de retour Algorithme 2. Aussi, en notant $n_u = n_{\cdot u}$, le nombre d'éléments dans chaque classe u et en réordonnant les éléments pour placer d'abord ceux dans la classe 1 puis ceux dans la classe 2 et ainsi de suite on obtient la matrice D_X . La permutation, quant à elle, vient précisément de cette réorganisation des individus. \square

S'ensuit immédiatement la formule suivante pour notre coefficient.

Propriété 33 (Expression diagonalisée du coefficient principal).

En reprenant les notations précédentes, pour une certaine matrice de permutation P ,

$$\langle X, Y \rangle = (D_X | P D_Y)$$

et par ailleurs

$$\|X\|^2 = \sum_{r=1}^{r_x} n_r^2$$

Démonstration.

Appelons P_X et P_Y deux matrices de permutation fournissant X, Y sous la forme précédente. Dans ce cas, on obtient :

$$\begin{aligned} (X|Y) &= (P_X D_X | P_Y D_Y) \\ &= (D_X | P_X^T P_Y D_Y) \end{aligned}$$

ainsi que :

$$\begin{aligned} \|X\|^2 &= \|P_X D_X\|^2 \\ &= \|D_X\|^2 \\ &= \sum_{r=1}^{r_x} n_r^2 \end{aligned}$$

Ce qui permet de conclure. □

Cette nouvelle formulation nous permet de démontrer plus facilement des propriétés sur le produit scalaire mentionné ; comme toujours, les critères sont basés sur le cosinus qu'il engendre en divisant par le produit des deux normes mais seul le coefficient principal sera sensible à la corrélation. On remarque d'ailleurs :

Remarque 37. *Estimer la dépendance entre U et V peut se faire en étudiant la proximité entre P et l'identité.*

Principe d'évaluation du coefficient principal

Nous souhaitons vérifier la remarque précédente, c'est-à-dire estimer l'impact de la matrice P sur le degré de corrélation entre les deux variables. Nous nous attendons à ce que la corrélation soit élevée lorsque $(X|Y)$ est proche de son maximal.

Afin de simplifier, nous prenons la même matrice $\Delta = \text{Diag}(U_{n_1}, \dots, U_{n_r})$ pour X et Y . La matrice P est elle tirée suivant une permutation $\sigma \in \mathfrak{S}_n$ associée. On a ainsi d'après la proposition 33 :

$$(X|Y) = (\Delta | P \Delta)$$

Nous calculons quelques paramètres pour juger de l'impact sur le coefficient redondant dans nos critères. Cette sous-section énonce simplement un principe général, sans démonstration ni rigueur tandis que les sections suivantes s'emploieront à le confirmer.

Il est évident que le produit scalaire est borné par $||\Delta||$ ce qui correspond à $\sigma = Id$; le point intéressant est de déterminer la perte engendrée par σ lorsque celle-ci s'écarte justement de l'identité.

Commençons par supposer que σ ne diffère de Id que sur un unique élément i qu'elle transfère depuis un des carrés composant Δ vers un autre. La matrice relationnelle correspondante aura à la place du carré plein, un carré avec des 0 dans la ligne et la colonne i sauf à l'emplacement (i, i) (de fait si nous reprenons l'équation 3.12, nous aurons toujours i en relation avec lui-même).

Nous prenons un exemple pour illustrer l'évolution engendrée, Δ n'a que deux carrés unitaires de tailles 4 et 3. L'application de σ inverse seulement les éléments 3 et 6.

$$(Id) \begin{bmatrix} 1 & 1 & 1 & 1 & 0 & 0 & 0 \\ 1 & 1 & 1 & 1 & 0 & 0 & 0 \\ 1 & 1 & 1 & 1 & 0 & 0 & 0 \\ 1 & 1 & 1 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 1 & 1 \\ 0 & 0 & 0 & 0 & 1 & 1 & 1 \\ 0 & 0 & 0 & 0 & 1 & 1 & 1 \end{bmatrix} \rightarrow (\sigma) \begin{bmatrix} 1 & 1 & 0 & 1 & 0 & 1 & 0 \\ 1 & 1 & 0 & 1 & 0 & 1 & 0 \\ 0 & 0 & 1 & 0 & 1 & 0 & 1 \\ 1 & 1 & 0 & 1 & 0 & 1 & 0 \\ 0 & 0 & 1 & 0 & 1 & 0 & 1 \\ 1 & 1 & 0 & 1 & 0 & 1 & 0 \\ 0 & 0 & 1 & 0 & 1 & 0 & 1 \end{bmatrix}$$

Ainsi, suite à l'action de σ , l'élément 3 change d'une classe de 4 éléments à une autre de 3 éléments auparavant, il n'a ainsi plus que 3 cases à 1 sur sa ligne ou sa colonne. C'est l'inverse pour l'élément 6. Ne sont coloriées que les cases qui rapportent des "points" lors du produit scalaire. Au final, σ fait donc perdre 10 points, chaque carré perd deux fois sa longueur moins 1. L'exemple suivant nous montrera que nous ne pouvons malheureusement pas généraliser si facilement cette perte.

Nous supposons désormais qu'une nouvelle permutation est appliquée à la matrice déjà transformée et échange cette fois les éléments 1 et 5; la permutation composée sera nommée σ' et la matrice résultante sera :

$$(\sigma') \begin{bmatrix} 1 & 0 & 1 & 0 & 0 & 0 & 1 \\ 0 & 1 & 0 & 1 & 1 & 1 & 0 \\ 1 & 0 & 1 & 0 & 0 & 0 & 1 \\ 0 & 1 & 0 & 1 & 1 & 1 & 0 \\ 0 & 1 & 0 & 1 & 1 & 1 & 0 \\ 0 & 1 & 0 & 1 & 1 & 1 & 0 \\ 1 & 0 & 1 & 0 & 0 & 0 & 1 \end{bmatrix}$$

Cette nouvelle matrice est intéressante à plusieurs titres. D'abord elle montre bien l'évolution rapide que peuvent engendrer les permutations et la complexité même avec des hypothèses fortes (les deux matrices Δ sont identiques). Ensuite, et comme annoncé, on voit que la perte en réitérant un échange est cette fois très différente, nous ne perdons que 2 points supplémentaires. Cela est dû à l'apparition de regroupements : (1, 3) comme

(5, 6) partent d'une même classe et se retrouvent dans une autre identique après un double échange ; ils ne l'étaient pas après un seul échange.

Nous allons généraliser cette idée à un σ quelconque dans la suite de cette section. Les idées essentielles se manifestant dès qu'il y a deux classes, nous nous restreignons d'abord à un Δ contenant deux blocs unitaires. Nous nous appliquerons ensuite à transposer ces idées dans un cadre général.

Calcul de la distribution du produit scalaire - cas de deux classes

On se place dans un cas très particulier pour commencer. Les matrices X et Y viennent de tirages produisant seulement deux classes et sont de plus réparties de la même manière : r éléments dans la première classe et $n - r$ dans la seconde pour chaque matrice. En bref, elles partagent leur matrice diagonale. On calcule le produit scalaire entre ces deux matrices de taille $n * n$ en remarquant comme précédemment que les deux s'écrivent :

$$\begin{aligned} X &= P_X \cdot \text{Diag}(U_r, U_{n-r}) \\ Y &= P_Y \cdot \text{Diag}(U_r, U_{n-r}) \end{aligned}$$

à deux matrices de permutation près P_X et P_Y (et c'est là tout le point), chacune est diagonale par blocs. On suppose déjà

$$X = \text{Diag}(U_r, U_{n-r})$$

ce qui revient à passer la permutation de X dans celle de Y si l'on reprend la formule de la proposition 33.

Maintenant, nous calculons quelques éléments sur les permutations. Suivant les exemples de la sous-section précédente, le nerf de la guerre est de savoir combien de 1 vont rester dans chacun des carrés proposés par X , donc parmi les r^2 et $(n - r)^2$ initiaux.

Pour ce faire, nous proposons de décomposer le problème : une proposition calcule la probabilité d'obtenir un certain nombre de sauts quand la suivante donne la valeur résultant d'un tel événement.

Dans le cas de deux classes, il est aisé de définir un "saut" comme étant, pour une permutation, le passage d'un élément d'une classe à l'autre. On regroupe alors intuitivement, les permutations suivant le nombre de "sauts" qui leur est associé sur les r premiers éléments. S'agissant d'échanges, il est évident que le même nombre de sauts a lieu sur les $n - r$ autres éléments. Comptons les permutations avec un nombre de sauts fixé (parmi les r premiers éléments).

Propriété 34 (Nombre de permutations produisant $r - k$ sauts).

On pose $A_k = \{\sigma \in \Sigma_n \text{ t.q. } |\{\sigma_1 \dots \sigma_r\} \cap \{1 \dots r\}| = k\}$, l'ensemble des permutations à $r - k$ sauts. Alors :

$$|A_k| = \frac{r!^2 (n - r)!^2}{k! (r - k)!^2 (n - 2r + k)!}$$

Démonstration. Nous partons de la formule

$$|A_k| = k! \binom{r}{k}^2 * (r - k)! \binom{n - r}{r - k} * (n - r)!$$

que nous lisons pour expliquer chacun des trois éléments comme trois étapes de construction :

1. On choisit d'abord les k éléments qui restent dans $\{1\dots r\}$, leurs positions d'arrivée dans ce même ensemble et leur ordre : $\binom{r}{k} \binom{r}{k} k!$
2. On choisit ensuite les $r - k$ éléments qui viennent de $\{r + 1\dots n\}$ dans $\{1\dots r\}$, leurs positions d'arrivée est fixée par les $r - k$ places restantes, reste seulement leur ordre dans celles-ci : $\binom{n - r}{r - k} (r - k)!$
3. On choisit enfin l'ordre des $n - r$ éléments restants qui sont choisis par défaut tout comme leur place d'arrivée : $(n - r)!$.

Il ne reste alors qu'à développer pour trouver la formule compacte de la proposition 34. □

Maintenant, on remarque que la valeur du produit scalaire dépend uniquement de la classe A_k dans laquelle se situe la permutation $\sigma \in S_n$. C'est une généralisation du principe de la sous-section 5.3.3 : on voit facilement que les éléments d'une même classe (comprendre les r premiers et les $n - r$ suivants) jouent le même rôle ; seul compte alors le nombre d'entre eux évincés de leur classe. Une fois choisi le nombre évincé de la première classe, tout est figé (dans ce cas très spécifique de deux classes).

On dispose ainsi de la probabilité d'avoir une valeur donnée du produit scalaire, reste à calculer chacune de celles-ci et c'est l'objet de la proposition suivante.

Propriété 35 (Valeur du produit scalaire à k sauts).

Si $\sigma \in A_k$ alors la valeur du produit scalaire associé est :

$$v_k = k^2 + 2 * (r - k)^2 + (n - 2r + k)^2$$

Démonstration.

Le développement de la forme de contingence du produit scalaire, à savoir, $\sum_{u,v} n_{u,v}^2$ donne immédiatement le résultat. □

Ainsi, on connaît, dans ce cadre spécifique, les valeurs possiblement prises par le produit scalaire et les probabilités de chacune ce qui nous permet de présenter les courbes suivantes.

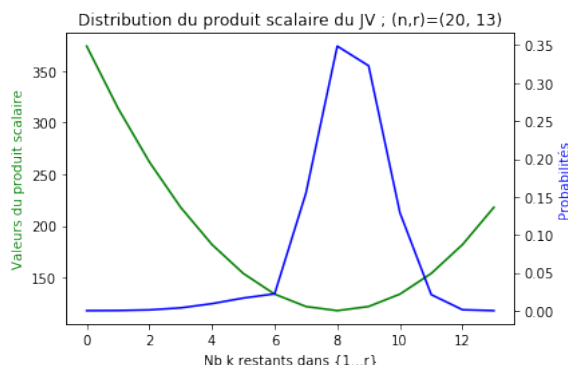


FIGURE 5.6 – Distribution du produit scalaire représentant le coefficient principal - 1

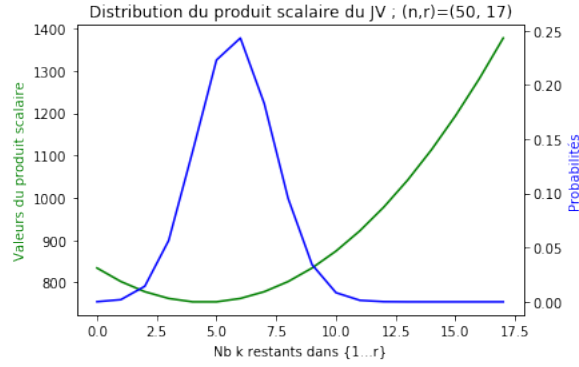


FIGURE 5.7 – Distribution du produit scalaire représentant le coefficient principal - 2

Dans ce cas très précis, on remarque déjà une très forte concentration autour de valeurs faibles. La courbe des probabilités est ainsi d'autant plus forte que les valeurs sont faibles et ce avec des probabilités très élevées pour certaines valeurs (plus de 0.3 sur la première figure).

Si nous revenons au problème initial, certes dans un cadre simplifié, nous notons donc le principe suivant :

Remarque 38 (Permutations et optimisation).

Les permutations les plus probables génèrent des valeurs de

$$(X|Y) = \sum_{u=1}^v \sum_{v=1}^q n_{u,v}^2 = \sum_{i=1}^n \sum_{j=1}^n x_{i,j} y_{i,j}$$

proches du minimum. Nous aurons l'occasion et de généraliser cette tendance dans la section 5.3.3 et de l'appliquer dans la section 5.3.4 en la reliant au problème 13.

Comme déjà annoncé, la section suivante s'emploie à transposer nos résultats dans un cadre général. Les idées sont similaires au cas de deux classes, les calculs et les notations sont eux sensiblement plus lourds.

Calcul de la distribution du produit scalaire - cas général

Nous nous plaçons maintenant dans le cadre le plus général : évaluer le produit scalaire entre deux matrices X et Y représentant les tirages de deux variables aléatoires U et V ayant respectivement p et q modalités. Après n tirages, on a, avec les notations habituelle :

$$\begin{aligned} n_u^X, & \quad \forall u \in \{1 \dots p\} \\ n_v^Y, & \quad \forall v \in \{1 \dots q\} \end{aligned}$$

qui représentent le nombre d'éléments distribués par chaque variable aléatoire dans ses différentes classes.

Le produit scalaire s'écrit ainsi :

$$v_\sigma = \left(\text{Diag}(U_{n_1^X}, \dots, U_{n_p^X}) | P_\sigma \text{Diag}(U_{n_1^Y}, \dots, U_{n_q^Y}) \right)$$

pour une permutation σ reliant les P_X et P_Y comme énoncé dans la proposition 32.

Déjà, si les tirages sont i.i.d. sous une loi π sous-jacente, lorsque n augmente, les marges statistiques se rapprochent des marges μ et ν . Dans ce cas, l'étude de v_σ se réduit à celle de σ .

De surcroît, suivant les remarques des sous-sections précédentes, la valeur v_σ dépend uniquement du nombre de 1 que l'on retrouve dans chacun des p carrés de la matrice X sous sa forme diagonale. Or, un 1 apparaît lorsque deux éléments de ce carré (la position faisant foi) sont envoyés dans la même classe par U .

Suivant ici la notation de contingence de la définition 12, on se donne ainsi un vecteur $(n_{u,v})$ donnant précisément le nombre d'éléments envoyés dans la classe u par U et dans la classe v par V . Sur ce point nous sommes en parfait accord avec la notation contingence sous-jacente mais l'originalité consiste à voir sa réalisation comme celle d'une permutation sur les n tirages. A une permutation $\sigma \in \mathfrak{S}_n$ on associe une corrélation (exprimée par les $n_{u,v}$), la position faisant foi :

$$n_{u,v} = \# \left\{ i / \sum_{r=1}^{u-1} n_{u^x} \leq i \leq \sum_{r=1}^u n_{u^x} \text{ et } \sum_{r=1}^{v-1} n_{u^y} \leq \sigma_i \leq \sum_{r=1}^v n_{u^y} \right\} \quad (5.20)$$

Comptons justement le nombre de permutations produisant un vecteur donné $(n_{u,v})$.

Propriété 36 (Nombre de permutations associées à une valeur).

Les marges n_1^X, \dots, n_p^X et n_1^Y, \dots, n_q^Y étant fixées, on se donne un vecteur $\bar{n} = (n_{u,v})_{(u,v) \in (1..p) \times (1..q)}$ qui les respecte. En ce cas, le nombre de permutations produisant \bar{n} est donné par la formule :

$$N_{\bar{n}} = \prod_{u=1}^p n_u^X! \prod_{v=1}^q n_v^Y! \prod_{u=1}^p \prod_{v=1}^q \frac{1}{n_{u,v}!}$$

Démonstration.

Commençons par une forme moins compacte mais plus lisible de la même formule avant de conclure par le calcul à une égalité.

$$N_{\bar{n}} = \prod_{u=1}^p \prod_{v=1}^q \binom{n_u^X - \sum_{r=1}^{v-1} n_{u,r}}{n_{u,v}} \binom{n_v^Y - \sum_{s=1}^{u-1} n_{s,v}}{n_{u,v}} n_{u,v}!$$

Si nous lisons la formule précédente, nous pouvons la voir comme la construction d'une permutation σ respectant \bar{n} . En effet nous choisissons les positions des éléments dans l'ordre $u = 1..p, v = 1..q$ et détaillons ci-dessous les étapes en supposant en être aux éléments $n_{u0,v0}$:

1. on a déjà enlevé $\sum_{r=1}^{v0-1} n_{u0,r}$ au départ de taille n_{u0}^X (figée par μ) et il nous faut en retirer $n_{u0,v0}$ supplémentaires d'où le premier facteur
2. on a déjà placé $\sum_{r=1}^{u0-1} n_{r,v0}$ à l'arrivée de taille n_{v0}^Y (figée par ν) et il nous faut en placer $n_{u0,v0}$ supplémentaires d'où le second facteur
3. une fois les positions (départs et arrivées) choisies, l'ordre d'arrivée donne le facteur $n_{u0,v0}!$

Reste à développer pour constater que la formule se simplifie grandement :

$$\begin{aligned}
N_{\bar{n}} &= \prod_{u=1}^p \prod_{v=1}^q \binom{n_u^X - \sum_{r=1}^{v-1} n_{u,v}}{n_{u,v}} \binom{n_v^Y - \sum_{s=1}^{u-1} n_{u,v}}{n_{u,v}} n_{u,v}! \\
&= \prod_{u=1}^p \prod_{v=1}^q \frac{(n_u^X - \sum_{r=1}^{v-1} n_{u,v})!}{n_{u,v}! (n_u^X - \sum_{r=1}^v n_{u,v})!} \frac{(n_v^Y - \sum_{s=1}^{u-1} n_{u,v})!}{n_{u,v}! (n_v^Y - \sum_{s=1}^u n_{u,v})!} n_{u,v}! \\
&= \prod_{u=1}^p \prod_{v=1}^q \frac{(n_u^X - \sum_{r=1}^{v-1} n_{u,v})!}{n_{u,v}! (n_u^X - \sum_{r=1}^v n_{u,v})!} \prod_{v=1}^q \prod_{u=1}^p \frac{(n_v^Y - \sum_{s=1}^{u-1} n_{u,v})!}{n_{u,v}! (n_v^Y - \sum_{s=1}^u n_{u,v})!} \prod_{u=1}^p \prod_{v=1}^q n_{u,v}! \\
&= \prod_{u=1}^p \prod_{v=1}^q \frac{n_u^X!}{n_{u,v}!} \prod_{v=1}^q \prod_{u=1}^p \frac{n_v^Y!}{n_{u,v}!} \prod_{u=1}^p \prod_{v=1}^q n_{u,v}! \\
&= \prod_{u=1}^p n_u^X! \prod_{v=1}^q n_v^Y! \prod_{u=1}^p \prod_{v=1}^q \frac{1}{n_{u,v}!}
\end{aligned}$$

□

Comme dans la section précédente, nous disposons maintenant de la probabilité de chaque valeur du produit scalaire, il nous reste à calculer ladite valeur associée à un vecteur \bar{n} . Cette valeur suit la formule de contingence de la fonction de coût du problème 6.

Propriété 37 (Valeur du produit scalaire).

Pour une permutation σ , associée à un vecteur de contingence $\bar{n} = (n_{u,v})_{(u,v) \in (1..p) \times (1..q)}$ la valeur de $v_\sigma = V_{\bar{n}}$ est $\sum_{u=1}^p \sum_{v=1}^q n_{u,v}^2$.

À chaque vecteur \bar{n} , nous avons donc associé une valeur et la probabilité correspondante. Il paraît aisé d'afficher la distribution.

Un dernier obstacle apparaît : la génération du vecteur \bar{n} . En effet, dans le cas de deux classes, il suffisait de figer un k tel que $1 \leq k \leq r$. C'est sensiblement plus complexe cette fois puisque le vecteur \bar{n} doit respecter toutes les marges :

$$\begin{aligned}
\sum_{v=1}^q n_{u,v} &= n_u^X, \forall u \\
\sum_{u=1}^p n_{u,v} &= n_v^Y, \forall v
\end{aligned}$$

que l'on choisit avant tout calcul de distribution.

Nous combinons l'interprétation d'une permutation σ comme réalisation d'un tirage et la quantité $N_{\bar{n}}$ associée à chacune pour générer la distribution complète. L'idée repose sur un tirage probabiliste de σ :

1. on génère $\sigma \in \mathfrak{S}_n$ aléatoirement ;
2. on construit le \bar{n}_σ correspondant selon l'équation 5.20 ;
3. on calcule le nombre de σ réalisant ce vecteur \bar{n}_σ : $N_{\bar{n}_\sigma}$ grâce à la proposition 36 ;
4. on calcule la valeur du produit scalaire associée : $V_{\bar{n}_\sigma}$ grâce à la proposition 37 ;

5. on stocke le dictionnaire des vecteurs \bar{n}_σ uniques ainsi que les valeurs $V_{\bar{n}}$ et quantités $N_{\bar{n}}$ associées ;
6. si la somme des quantités $N_{\bar{n}_\sigma}$ dans le dictionnaire vaut $n!$ on s'arrête.

Nous précisons la méthode au travers de l'algorithme 4 avant de le détourner légèrement pour l'appliquer dans un cas particulier. Il faut remarquer que l'on s'affranchit du respect des marges qui a lieu par construction via l'utilisation de 5.20, or, c'était le point essentiel. La convergence peut être longue mais elle a lieu presque sûrement et il est aisé de modifier l'algorithme pour s'arrêter une fois qu'une proportion α de la distribution a été observée ; cela changera peu sa forme si α est assez proche de 1. Par ailleurs, nous n'avons pas à parcourir l'ensemble des permutations mais seulement à en trouver une pour chaque \bar{n} possible ce qui quotiente grandement leur univers.

Algorithm 4 Calculer la distribution du produit scalaire et l'ensemble des vecteurs éligibles

```

 $(S_{\bar{n}}, LV_{\bar{n}}, LN_{\bar{n}}) \leftarrow [], [], []$  {respectivement : l'ensemble des vecteurs, les valeurs associées
et les probas associées}
while  $\sum (LN_{\bar{n}}) < n!$  do
   $\sigma \leftarrow$  permutation aléatoire de  $\{1\dots n\}$ 
  for  $(u, v) \in \{1\dots p\}, \{1\dots q\}$  do
     $\bar{n}_{u,v}$  calculé selon l'équation 5.20 {On calcule le  $\bar{n}$  associé à  $\sigma$ }
  end for
  if  $\text{not}(\bar{n} \in S_{\bar{n}})$  then
     $S_{\bar{n}}.\text{append}(\bar{n})$ 
     $LV_{\bar{n}}.\text{append}(V_{\bar{n}})$  {On utilise la proposition 37}
     $LN_{\bar{n}}.\text{append}(N_{\bar{n}}/n!)$  {On utilise la proposition 36}
  end if
end while

```

Un exemple de distribution

Nous utilisons l'algorithme 4 pour proposer un exemple de distribution, pour cela nous fixons deux marges et regardons les valeurs habituelles du coefficient principal. Les valeurs figées sont données par :

- $n = 27$
- $n^X = [5, 9, 13]$
- $n^Y = [6, 9, 6, 6]$

Ces valeurs ont été choisies pour qu'un couplage d'indétermination existe. En utilisant le résultat de l'algorithme 4, on constate que 7416 vecteurs \bar{n} sont éligibles. Parmi eux, le vecteur le moins probable (comprendre, engendré par le moins de permutations) est :

$$\bar{n} = \begin{bmatrix} 5 & 0 & 0 & 0 \\ 0 & 9 & 0 & 0 \\ 1 & 0 & 6 & 6 \end{bmatrix}$$

de probabilité $1.4941460627286446e - 10$ (il y a $1.62695614169088e + 18$ permutations engendrant ce vecteur) et la valeur du produit scalaire est 111. Le vecteur le plus probable,

au contraire, (avec une probabilité à 0.6777446540537131%) est le suivant :

$$\bar{n} = \begin{bmatrix} 1 & 2 & 1 & 1 \\ 2 & 3 & 2 & 2 \\ 3 & 4 & 3 & 3 \end{bmatrix}$$

avec une valeur associée de 52.

Nous affichons sur la figure 5.8, la distribution complète valeurs / probabilités du produit scalaire sur l'exemple choisi. La fonction de répartition est également donnée sur la figure 5.9.

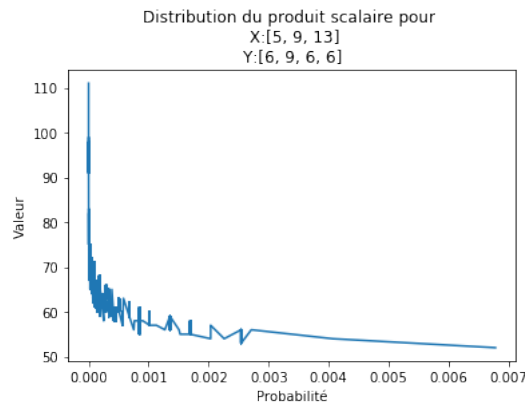


FIGURE 5.8 – Distribution du produit scalaire représentant le coefficient principal - cas général

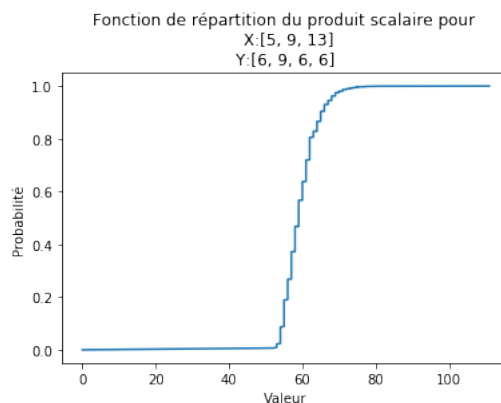


FIGURE 5.9 – Fonction de répartition du produit scalaire représentant le coefficient principal - cas général

Ainsi, la forme diagonalisée de la proposition 32 nous a amené à considérer le produit scalaire comme résultat d'une permutation puis à calculer les valeurs associées à une permutation donnée (propositions 37 et 36). L'algorithme 4 apporte alors la pièce manquante pour obtenir la distribution du coefficient principal une fois le jeu de paramètres décrivant une situation des marges figé. En bref, à marges fixées, on connaît le comportement du

coefficient principal des critères. Le reste n'étant que des centrages et des normalisations dépendant des marges, nous connaissons le comportement des critères à marges fixées.

Or, la figure 5.9 montre précisément que le produit scalaire évolue très peu : les valeurs les plus probables sont concentrées autour de 60 alors même que le minimum est à 52 et le maximum à 111. Cela signifie que la plupart des permutations réalisables conduiront à un critère de corrélation faible. Cela signifie également que les critères renverront un résultat similaire quelle que soit la permutation.

Si nous avons choisi ces marges c'est pour qu'une matrice d'indétermination soit possible. Nous remarquons ici qu'il s'agit précisément de la matrice la plus probable que nous avons déjà rapportée. On vérifie en effet aisément que la matrice respecte les marges. Quant à vérifier qu'il s'agit bien d'une matrice d'indétermination, nous laissons le lecteur choisir une des caractérisations du théorème 4.

On observe sur la figure 5.8 que (généralement) plus la valeur du coefficient diminue et plus sa probabilité est grande. Or, l'indétermination est, comme optimisation du problème 6, le couplage qui apporte la valeur la plus faible, et dispose donc (selon cette observation) de la probabilité la plus forte. Nous généralisons et exploitons cette remarque dans la section 5.3.4.

5.3.4 Minimisation

Principe

Nous partons d'un constat : le problème 6 est présenté sous une forme probabiliste essentiellement parce que, ne sachant le résoudre sous sa forme entière, nous avons relâché les hypothèses. La forme originale, appliquée par exemple aux flux migratoires, serait plutôt statistique et ne comporterait donc que des valeurs entières, comme table de contingence, devenant :

Problème 16 (Minimal Trade Model Statistique).

$$\begin{aligned} \min_{\Pi} \quad & \sum_{u,v} \Pi_{u,v}^2 \\ \text{sous les contraintes :} \quad & \\ & \sum_v \Pi_{u,v} = n_u^X, \forall u \\ & \sum_u \Pi_{u,v} = n_v^Y, \forall v \\ & \sum_{u,v} \Pi_{u,v} = n = \sum_{u=1}^p n_u^X = \sum_{v=1}^q n_v^Y \\ & \Pi_{u,v} \in \mathbb{N}, \forall (u, v) \end{aligned}$$

Résoudre ce problème est extrêmement laborieux, nous ne disposons déjà pas de formule close dans le cas réel sans ajouter l'hypothèse 2.9 et nous renforçons ici la complexité en supposant que les éléments sont entiers ! Remarquons néanmoins déjà que la fonction de coût n'est autre que notre coefficient principal : dans un cadre statistique, Π étant une table de contingence, la proposition 10 nous permet de retrouver $(X|Y)$.

On essaie donc de minimiser le coefficient principal à marges fixées, sur des valeurs entières ; en reprenant les notations de la section 5.3.3,

$$\Pi = \bar{n}.$$

Il s'agit ainsi de trouver le vecteur \bar{n} optimal et nous avons vu dans la sous-section 5.3.3 que chaque \bar{n} est associé à une permutation de n éléments. Générer des permutations, ce qui est simple, permet ainsi de produire efficacement des candidats \bar{n} . Nous disposons, de surcroît, de la probabilité de production de chaque \bar{n} grâce à la proposition 36. Ceci permet de se donner une proportion α à parcourir, de générer des permutations jusqu'à avoir parcouru cette proportion α des $\bar{n} = \Pi$ puis de retourner le meilleur Π trouvé.

Surtout, la formule énoncée par la proposition 36, outre qu'elle apporte une connaissance nécessaire pour les algorithmes 4 et 5, livre une propriété primordiale : elle est décroissante avec les $n_{u,v}$. Certes, le respect des marges empêche une décroissance totale : diminuer un $n_{u,v}$ peut selon les contraintes amener à en augmenter deux autres et globalement diminuer $N_{\bar{n}}$. Certes aussi, et c'est tout l'objet du document avec les différences de résultat suite au changement de fonction de coût entre les problèmes 5 et 6, le vecteur \bar{n} maximisant N_n sera probablement différent de Π^+ ou Π^\times .

Pour autant, une tendance subsiste, la fonction de coût menant à Π^+, Π^\times sous le respect des marges, gagne à minimiser les $n_{u,v}$ et cette tendance se retrouve dans les valeurs de $N_{\bar{n}}$: plus les $n_{u,v}$ sont faibles, plus grande et la probabilité de générer le vecteur \bar{n} correspondant. Aussi, en tirant des permutations uniformément, les valeurs faibles seront les plus probables lors de l'évaluation des fonctions de coût des problèmes 5 et 6 sur le vecteur \bar{n} .

Dit autrement, générer uniformément des permutations, aura tendance à construire des solutions proches de l'optimal pour les deux problèmes. Nous mettons en pratique cette remarque dans l'algorithme 4.

Cet algorithme présente deux propriétés intéressantes. Premièrement, ramener Π à une permutation permet de s'affranchir du respect laborieux des marges dans la construction d'un candidat ; tout est géré par l'équation 5.20. Secondement, si la distribution sur les permutation est uniforme, celle sur les Π ne l'est pas du tout, nous l'avons déjà observé dans l'exemple de sous-section 5.3.3. Or, plus que d'être non uniforme, la distribution proposée dans la figure 5.8 a une forte tendance à associer des probabilités élevées aux éléments qui fournissent un produit scalaire faible. Mieux, la matrice 5.3.3 qui est optimale dans l'exemple suivi a la probabilité maximale. Nous avons déjà conclu à la généralité de cette tendance forte probabilité / fonction de coût faible.

Remarque 39 (Défense du critère d'arrêt).

Parcourir une proportion α des permutations \mathfrak{S}_n paraît inenvisageable. Il faut se rappeler qu'il s'agit en fait d'une proportion α de l'espace quotienté par la valeur du produit scalaire, en définitive, une proportion α des dites valeurs. Or, au regard de la figure 5.9, on remarque (évidemment dans l'exemple) que certaines valeurs concentrent 25% des probabilités. Cette concentration est généralisable, comme exposé dans la proposition 36.

Remarque 40 (Arrêt sur performance).

Dans la section 5.3.4 dédiée à l'expérimentation de notre algorithme de minimisation, nous changeons légèrement le critère d'arrêt. L'optimisation probabiliste (comprendre avec

Algorithm 5 Approcher l'optimum du problème 16

Input α **Input** (n_u^X) **Input** (n_v^Y) $(S_\Pi) \leftarrow \{\}$ {Dictionnaire pour stocker les matrices Π trouvées, leur proba et leur valeur de $(X|Y)$ }**while** $\sum_k S_\Pi[k][p] < \alpha$ **do** $\sigma \leftarrow$ permutation aléatoire de $\{1\dots n\}$ **for** $(u, v) \in \{1\dots p\}, \{1\dots q\}$ **do** $\Pi_{u,v}$ calculé selon l'équation 5.20 {On calcule le Π associé à σ }**end for****if** not($\Pi \in S_\Pi$) **then** V_Π {On utilise la proposition 37 pour avoir la valeur associée} P_Π {On utilise la proposition 36 pour avoir la probabilité associée} $S_\Pi[\Pi] \leftarrow (P_\Pi, V_\Pi)$ **end if****end while** $\Pi^* \leftarrow \operatorname{argmin}\{S_\Pi, \text{key} = V_\Pi\}$

des valeurs réelles) conduit à $C^+(\mu, \nu)$ que l'on peut exprimer et pour lequel on peut calculer le produit scalaire associé : $(X|Y)_P^*$. La réponse étant dans l'espace statistique (comprendre avec des valeurs entières respectant les contraintes du problème 16), il ne s'agit certes pas de $C^+(\mu, \nu)$ mais on peut supposer que le produit scalaire minimal n'est que légèrement supérieur à $(X|Y)_P^*$.

Ainsi, par exemple avec $\beta = 1.1$, nous décidons de nous arrêter lorsque l'optimal trouvé est inférieur à $\beta(X|Y)_P^*$.

Une question reste en suspens : Π^+ a-t-elle toujours la probabilité maximale ? Cela pourrait être lié à une question sous-jacente de physique statistique, Π^+ étant la distribution de moindre coût reste que Π^+ est associée à la fonction de coût du problème 6, alors que maximiser $N_{\bar{n}}$ porte sur une fonction de coût différente. Suivant les marges, il semble donc logique que Π^+ ne réalise pas $N_{\bar{n}}^{\min}$ de la même manière que Π^+ et Π^\times ne sont pas identiques.

Expériences

Nous appliquons maintenant l'algorithme 5 dans un cadre particulier où Π^+ existe et est connu. Pour ce faire, nous partons de Π^+ .

L'objectif est d'estimer l'écart entre la solution approchée issue de notre heuristique et la valeur minimale, connue, réalisée sous Π^+ .

L'étude précise de l'heuristique est complexe, de fait comme nous allons le voir, la performance (comprendre la capacité à approcher rapidement une valeur proche de celle générée par Π^+) est difficile à prévoir. Le résultat, rapide dans la majorité des cas, peut prendre un temps infini – en théorie du moins.

Afin d'y palier, nous décidons d'approcher la valeur minimale de $(X|Y)$ liée à Π^+ à 10% près. Nous retenons ainsi le nombre de permutations générées pour atteindre un produit scalaire inférieur à 1.1 fois le minimum.

Précisément, en reprenant la construction de la partie 3.5.1, nous commençons par tirer $\Pi_{1,v}^+$ une distribution aléatoire sur 100 éléments, puis Δ une distribution également aléatoire sur 100 éléments. Ensuite nous construisons la matrice Π^+ correspondante selon ce même algorithme. Celui-ci ne permet pas de facilement contrôler la taille finale, aussi, celle-ci est aléatoire mais peu importe, nous répétons 1000 fois l'opération et rapportons la distribution des tailles dans la figure 5.10.

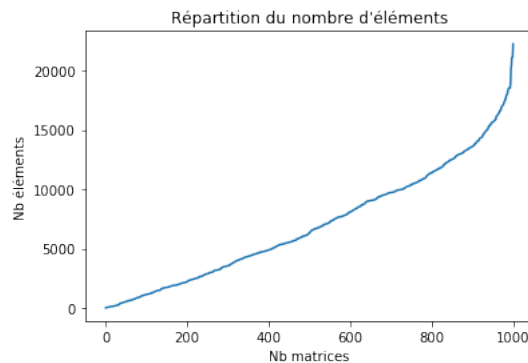


FIGURE 5.10 – Distribution du nombre d'éléments dans les matrices utilisées pour évaluer l'algorithme 5

En résumé, la taille moyenne est de 5600 éléments et les deux extrêmes 6 et 22, 241.

Nous créons ainsi 1000 matrices d'indétermination, desquelles nous extrayons les marges. Avec ces dernières, nous appliquons l'algorithme 5 légèrement modifié : nous ne nous arrêtons pas lorsqu'une partie α des valeurs associées aux permutations est parcourue mais lorsque nous trouvons une permutation qui mène à un produit scalaire majoré par 1.1 fois le minimum. Nous rappelons au passage que notre construction par la matrice d'indétermination livre justement ce minimum.

Nous souhaitons ainsi savoir combien de permutations sont tirées pour approcher l'optimal à 10% près. La figure 5.11 répond précisément à cette question en affichant la distribution du nombre de permutations nécessaire.

Nous remarquons deux choses.

D'abord, une bonne partie des cas est figé à 1000. C'est un artéfact, ainsi si après 1000 tirages, nous ne sommes pas sortis, nous cessons de chercher ; cela concerne 319 matrices et c'est un sujet à part entière de comprendre pourquoi.

Ensuite, conformément à la remarque 38 le nombre de tirages est en général très faible. Ainsi, si nous rentrons dans le détail, parmi les 681 matrices pour lesquelles nous avons approché à 10% près en moins de 1000 tirages, le nombre moyen de tirages est 13 et pour 550 matrices le premier tirage a suffit.

Comprendre pourquoi 319 matrices ont un minimum difficile à approcher est une tâche ardue. Nous regardons déjà si cela est lié au nombre d'éléments au travers de la figure 5.12.

Il n'y a pas de tendance globale, ce n'est pas parce que n augmente que le nombre de

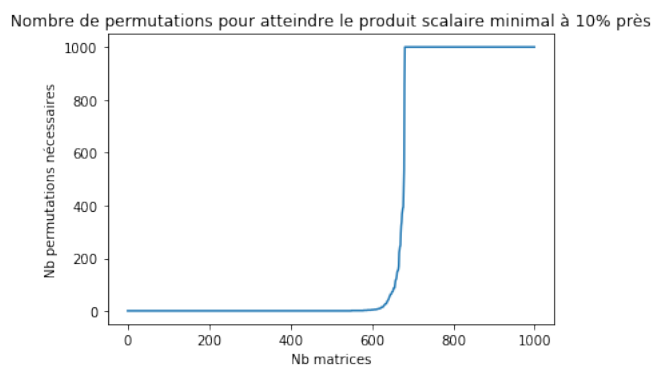


FIGURE 5.11 – Nombre de tirages avant d’approcher le minimum à 10% près

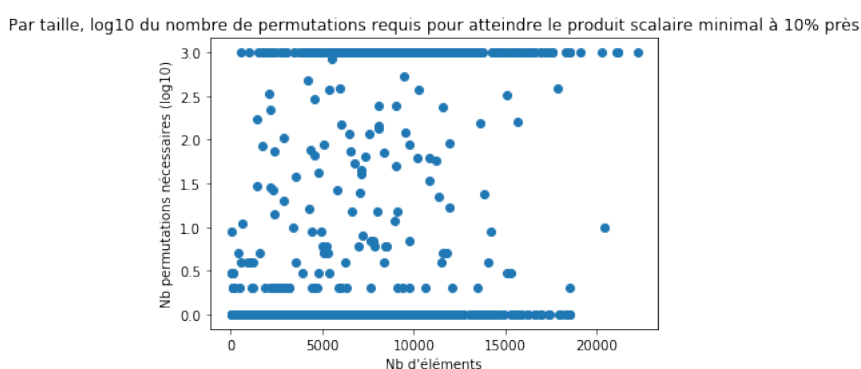


FIGURE 5.12 – Nombre de tirages avant d’approcher le minimum à 10% près en fonction du nombre d’éléments

tirages augmente et de plus, des matrices à moins de 5000 éléments sont mal approchées. Cependant, nous remarquons que les n extrêmes participent toutes à l’artéfact et donc appartiennent aux 319.

Puisqu’on n’observe pas de tendance globale, il n’y a aucune raison que, subitement, pour $n = 20,000$, l’algorithme cesse de fonctionner. Par contre, notre manière de générer les matrices d’indétermination introduit un biais. Lorsque nous tirons $\Pi_{1,v}^+$ et Δ , la matrice résultante contient d’autant plus d’éléments que $\Pi_{1,v}^+$ et Δ ont réparti les 100 individus sur beaucoup de classes. Cela suppose des petites classes et, partant, une plus grande sensibilité aux sauts dans la permutation.

Typiquement, la matrice non trouvée ayant le moins d’éléments est celle-ci :

$$\begin{bmatrix} 87 & 44 & 45 & 45 & 44 \\ 46 & 3 & 4 & 4 & 3 \\ 45 & 2 & 3 & 3 & 2 \\ 45 & 2 & 3 & 3 & 2 \\ 46 & 3 & 4 & 4 & 3 \\ 45 & 2 & 3 & 3 & 2 \end{bmatrix}$$

et les marges associées

$$n^X = (265, 60, 55, 55, 60, 55) \quad (5.21)$$

$$n^Y = (314, 56, 62, 62, 56) \quad (5.22)$$

Enfin, parmi les matrices non trouvées nous avons conservé la meilleure permutation et le facteur entre le produit scalaire résultant et l'optimal. La distribution de ce facteur est renseigné dans la figure 5.13.

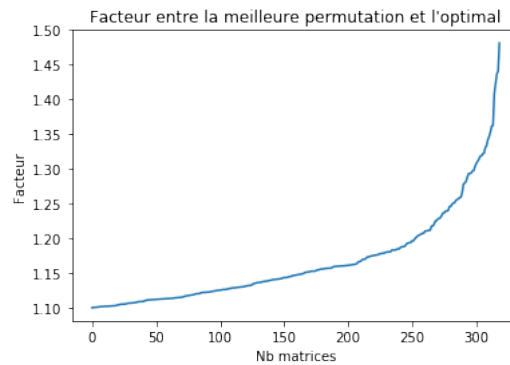


FIGURE 5.13 – Distribution du facteur lorsque celui-ci dépasse 1.1

Cette dernière figure permet de conclure que certes nous ne savons expliquer précisément pourquoi les 10% ne sont pas atteints pour 319 matrices mais qu'en tous les cas, l'algorithme fournit avec au pire 1000 tirages, un facteur 1.5 par rapport à l'optimal. De surcroît, nous savons qu'il est plus performant en général, et dans le facteur et dans le nombre de tirages.

Chapitre 6

Cas continu

T'avances comme dans des
couloirs
Tu t'arranges pour éviter les
miroirs
Mais ça continue encore et
encore
C'est que le début d'accord,
d'accord

Francis Cabrel, *Encore et
Encore*, 1985

6.1 Indétermination continue

6.1.1 Fonction de couplage cas continu

Ce chapitre étend la notion d'indétermination dans un cadre continu. La formule associée est déjà présente, notamment dans [51], mais son étude n'y apparaît pas. De surcroît, l'article mentionné fait état d'une "copule d'indétermination" dont la définition, nous allons le voir, ne peut être que locale. Transposant les marges du cas discret, nous nous donnons deux lois de probabilités μ et ν sur deux segments (respectivement $[a, a + A]$ et $[b, b + B]$) et nous supposons également que chacune dispose d'une densité que nous noterons respectivement f et g . Pour définir une loi de probabilité π sur l'espace produit, nous allons agir sur les densités (voir la définition 35) comme nous le faisons dans le cas discret (où nous sommions μ_u et ν_v).

En préparation de l'appel à des notions statistiques usuelles dans l'étude des tests (section 6.4), nous suivons et adaptons les notations introduites dans [28] et [46] ainsi que quelques lemmes et théorèmes qu'ils proposent et que nous utiliserons dans la suite. Pour commencer, nous notons $S_a = [a, a + A]$, $S_b = [b, b + B]$, $S = [a, a + A] \times [b, b + B]$ et définissons l'espace des lois de probabilités boréliennes sur chacun : $\Lambda_a = (S_a, \mathcal{B})$, $\Lambda_b = (S_b, \mathcal{B})$, $\Lambda = (S, \mathcal{B})$.

Au delà de la notion d'indétermination, nous étendons la définition de fonction de

couplage dans un espace continu.

Définition 35 (Fonction de couplage (cas continu)).

Étant donné deux lois de probabilités $\mu \in \Lambda_a$ et $\nu \in \Lambda_b$, une fonction de couplage C opère sur leurs densités (f, g) pour définir une densité $C(\mu, \nu) = C(f, g)$ agissant sur l'espace produit S . La mesure associée à cette densité respecte des propriétés similaires à celles du cas discret : il s'agit d'une loi de probabilité dont les marges sont μ et ν .

Une fonction de couplage typique est celle de d'indépendance, notée C^\times (nous étendons ici la notation discrète) qui génère une densité donnée par la formule :

$$C^\times(f, g)(x, y) = f(x)g(y), \quad \forall x \in S_a, \quad \forall y \in S_b$$

Nous souhaitons définir une fonction de couplage d'indétermination C^+ étendant la notation discrète. Comme souvent, dans le cas d'une transposition, nous pouvons l'obtenir ou bien en devinant la forme ou bien en la calculant. Nous allons suivre les deux approches en commençant par deviner.

Définition 36 (Densité de l'indétermination).

$$C^+(f, g)(x, y) = \frac{f(x)}{B} + \frac{g(y)}{A} - \frac{1}{A \times B}$$

Cette formule est une adaptation directe de la notion discrète, sans garantie ni de sa validité ni de sa construction. Pour commencer, et afin de simplifier tout calcul ultérieur, nous choisissons $a = b = 0$ et $A = B = 1$; la conversion de toute les formules se fait, nous le verrons, par application d'une transformation affine.

6.1.2 Un cas continu du problème de déviation quadratique

Intéressons-nous maintenant à la deuxième méthode pour valider notre proposition toujours sur $[0, 1] \times [0, 1]$ pour simplifier les calculs puis sur un couple de segments quelconques. On suppose également, pour justement pouvoir poser les calculs que μ et ν sont absolument continues par rapport à la mesure de Lebesgue avec pour densités respectives f et g . Ces densités vérifient évidemment la préservation de la masse $\int f = \int g = 1$ tandis que la densité π que nous cherchons respecte les deux contraintes de marges $\int_0^1 \pi(x, y) dx = \nu(y)$ et $\int_0^1 \pi(x, y) dy = \mu(x)$. Étant donné ces contraintes, le problème continu que nous devons résoudre se présente de la façon suivante sur un intervalle réduit d'abord $([0, 1]^2)$:

Problème 17 (Minimal Trade Problem continu sur un intervalle simple).

$$\begin{aligned} \min_{\pi \text{ densité sur } S} P[\pi] &= \int_0^1 \int_0^1 \pi^2(x, y) dx dy \\ \text{sous les contraintes :} & \\ &\int_0^1 \pi(x, y) dy dx = \mu(x) dx \\ &\int_0^1 \pi(x, y) dx dy = \nu(y) dy \\ &\int_0^1 \int_0^1 \pi(x, y) dx dy = 1 \end{aligned}$$

Posons enfin une dernière hypothèse similaire à celle formulée dans le cas discret (2.9), nous l'utiliserons pour conclure que la fonction trouvée est bien positive :

$$\min_x f(x) + \min_y g(y) \geq 1 \quad (6.1)$$

Calculons donc l'optimal du problème tout juste défini, le but étant d'obtenir un couplage d'indétermination continu.

Propriété 38 (Solution du MTM continu).

Une solution du problème 17 est la version continue de la solution discrète :

$$\pi^*(x, y) = \pi^+(x, y) = f(x) + g(y) - 1$$

Démonstration.

Comme, en nous inspirant du cas discret, nous avons pressenti la forme de la solution, nous pouvons utiliser cette information et développer l'inégalité

$$\int_{x=0}^1 \int_{y=0}^1 [\pi(x, y) - (f(x) + g(y) - 1)]^2 dx dy \geq 0 \quad (6.2)$$

pour obtenir successivement :

$$\begin{aligned} \int_{x=0}^1 \int_{y=0}^1 \pi^2(x, y) dx dy &\geq \int_{x=0}^1 \int_{y=0}^1 (-f^2 - g^2 - 1 - 2\pi - 2fg + 2f\pi + 2g\pi + 2f + 2g) dx dy \\ &= \int_{x=0}^1 \int_{y=0}^1 (-f^2 - g^2 - 1 - 2 - 2fg + 2f^2 + 2g^2 + 2 + 2) dx dy \\ &= \int_{x=0}^1 \int_{y=0}^1 (f^2 + g^2 - 2fg + 1) dx dy \\ &= \int_{x=0}^1 \int_{y=0}^1 (f + g - 1)^2 dx dy \end{aligned}$$

À ce stade nous avons montré l'inégalité suivante :

$$\forall \pi, P[\pi] \geq P[\pi^*] \quad (6.3)$$

Maintenant, en utilisant l'hypothèse 6.1, on constate que π^* est éligible comme fonction de densité car toujours positive. Les contraintes de marges sont par ailleurs respectées suivant des calculs similaires au cas discret. Nous pouvons donc conclure que π^* est une solution. \square

Pour passer à un cas plus général où x et y se déplacent dans des intervalles du type $[a, a + A]$ et $[b, b + B]$, on regarde l'impact sur les contraintes et la fonction de coût ; le problème se définit ainsi :

Problème 18 (Continuous Minimal Trade Problem).

$$\begin{aligned} \min_{\pi \text{ densité sur } S} P[\pi] &= \int_a^{a+A} \int_b^{b+B} \pi^2(x, y) \, dx \, dy \\ \text{sous les contraintes :} & \\ & \int_a^{a+A} \pi(x, y) \, dy = \mu(x) \, dx \\ & \int_b^{b+B} \pi(x, y) \, dx = \nu(y) \, dy \\ & \int_a^{a+A} \int_b^{b+B} \pi(x, y) \, dx \, dy = 1 \end{aligned}$$

Propriété 39 (Solution du problème 18).

Une solution du problème 18 sur un intervalle général respecte la version continue de la solution discrète :

$$\pi^*(x, y) = \pi^+(x, y) = \frac{f(x)}{B} + \frac{g(y)}{A} - \frac{1}{AB}$$

Démonstration.

Si l'on reprend les contraintes du problème 18, on constate :

- $\int_a^{a+A} \pi(x, y) \, dy = \int_0^1 \pi(a + Ax, y) \, dy = \mu(a + Ax)$ avec $x \in [0, 1]$
- $\int_b^{b+B} \pi(x, y) \, dx = \int_0^1 \pi(x, b + By) \, dx = \nu(b + By)$ avec $y \in [0, 1]$
- $\int_a^{a+A} \int_b^{b+B} \pi(x, y) \, dx \, dy = \int_0^1 \int_0^1 \frac{1}{AB} \pi(a + Ax, b + By) \, dx \, dy$

On déduit que $AB\pi(a + Ax, b + By)$ est solution du problème 17 avec comme marges $x \mapsto \mu(a + Ax)$ et $y \mapsto \nu(b + By)$.

Ces marges ont comme fonction de densité, $x \mapsto Af(a + Ax)$ et $y \mapsto Bg(b + By)$ si celles de μ et ν sont f et g . Aussi la proposition 38 montre que la densité solution respecte la forme :

$$AB\pi^*(a + Ax, b + By) = Af(a + AX) + Bg(b + BX) - 1$$

ce qui donne le résultat escompté. □

L'apparition de la formule de couplage continu est intéressante à plusieurs titres. D'abord, elle montre que C^+ s'étend au cas continu et ensuite elle introduit un couplage qui, comme tous, se rapporte à une copule que nous appellerons copule d'indétermination ; nous travaillons sur cette notion dans la section 6.2. Finalement, les deux méthodes conduisant au même résultat, nous proposons la définition ci-dessous d'un couplage d'indétermination continu :

Définition 37 (Couplage d'indétermination (continu)).

Étant donné deux variables aléatoires sur $S_1 : U$ de loi μ , de densité f et de fonction de répartition F et symétriquement V de loi définie par ν , g et G on dit que la variable aléatoire W réalise un couplage d'indétermination de U et V noté $U \oplus V$ lorsque sa densité π^+ est donnée par

$$\pi_{f,g}^+(x, y) = f(x) + g(y) - 1. \quad (6.4)$$

Pour conclure nous proposons également la fonction de répartition associée à la densité tout juste exprimée $\pi_{f,g}^+$; il s'agit juste d'une rapide intégration donnant une seconde caractérisation du couplage selon l'indétermination.

Propriété 40 (Fonction de répartition de l'indétermination (continu)).

Si $\pi_{f,g}^+(x, y) = f(x) + g(y) - 1$, avec f, g deux densités sur S_1 assurant la positivité de $\pi_{f,g}^+$ (hypothèse 6.1) et dont les fonctions de répartition sont F et G , est la densité d'une variable aléatoire dans S alors sa fonction de répartition, notée $\Pi_{F,G}^+$ est donnée par :

$$\Pi_{F,G}^+ = yF(x) + xG(y) - xy \quad (6.5)$$

6.1.3 Tirage selon l'indétermination : cas continu

Nous transposons ici la méthode de tirage sous l'indétermination que nous avons énoncée dans le cas discret (section 3.5) comme une décomposition de la densité π^+ . On note x_0 l'argument minimal de f (la condition 6.1 impose $f(x_0) > 0$ car sinon $g = 1$ et on est sous l'indépendance). Ensuite, on note $h_{x_0}(y)$ la densité de la loi conditionnelle $\mathbb{P}_{\pi_{f,g}^+}(Y|X = x_0)$ que l'on construit sur y lorsque $x = x_0$. Alors, en mimant la construction discrète d'un tirage selon l'indétermination abordé dans la proposition 15 mais dans un cadre continu on obtient le processus suivant.

1. On tire un x selon la densité f ;
2. on se donne \mathbb{U} loi uniforme sur $[0, f(x)]$ et une réalisation u de \mathbb{U} indépendante de x ;
3. si $u \leq f(x_0)$ alors on tire y selon la densité $h_{x_0}(y) = \frac{\pi_{f,g}^+(x_0, y)}{f(x_0)}$;
4. si $u \geq f(x_0)$ alors on tire y selon une loi uniforme sur $[0, 1]$.

Propriété 41 (Construction d'un tirage selon l'indétermination (cas continu)).

La construction tout juste introduite permet bien de tirer suivant le couplage de densité $C^+(f, g)$.

Démonstration.

On calcule la fonction de répartition H du couplage construit par le processus avec F et G celles associées à f et g respectivement.

$$\begin{aligned} H(x, y) &= \mathbb{P}(X \leq x, Y \leq y) \\ &= \mathbb{P}(X \leq x, Y \leq y, \mathbb{U} \leq f(x_0)) + \mathbb{P}(X \leq x, Y \leq y, \mathbb{U} \geq f(x_0)) \\ &= \mathbb{P}(Y \leq y | X \leq x, \mathbb{U} \leq f(x_0)) \mathbb{P}(\mathbb{U} \leq f(x_0), X \leq x) \\ &\quad + \mathbb{P}(Y \leq y | X \leq x, \mathbb{U} \geq f(x_0)) \mathbb{P}(\mathbb{U} \geq f(x_0), X \leq x) \\ &= \mathbb{P}(Y \leq y | \mathbb{U} \leq f(x_0)) \mathbb{P}(\mathbb{U} \leq f(x_0), X \leq x) \\ &\quad + \mathbb{P}(Y \leq y | \mathbb{U} \geq f(x_0)) \mathbb{P}(\mathbb{U} \geq f(x_0), X \leq x) \\ &= \int_0^y h_{x_0}(s) \int_0^x \left(\frac{f(x_0)}{f(r)} f(r) \right) dr ds + \int_0^y 1 \int_0^x \left(\frac{f(r) - f(x_0)}{f(r)} f(r) \right) dr ds \\ &= xf(x_0) \int_0^y (h_{x_0}(s)) ds + yF(x) - f(x_0)xy \end{aligned}$$

Maintenant, on reporte la définition de $h_{x_0}(y)$ dans la dernière équation pour obtenir

$$\begin{aligned} H(x, y) &= x f(x_0) \int_0^s \left(\frac{f(x_0) + g(s) - 1}{f(x_0)} \right) ds + y F(x) - f(x_0) x y \\ &= x \int_0^s (f(x_0) + g(s) - 1) ds + y F(x) - f(x_0) x y \\ &= x G(y) + y F(x) - x y \end{aligned}$$

Enfin, on reconnaît sous cette dernière forme la fonction de répartition $\Pi_{F,G}^+$ du couplage d'indétermination du couple (F, G) que nous avons introduit dans l'équation 6.5. \square

6.1.4 Caractérisation des marges éligibles

Dans ce qui précède, nous avons, en transposant le cas discret, supposé que l'hypothèse 6.1 était vérifiée. Tout comme dans le cas discret et la proposition 5, nous disposons d'une méthode pour adapter des marges quelconques. Elle est proposée dans la propriété 42.

Propriété 42 (Hypothèse constructive sur les marges éligibles).

Un couple (f, g) de densités respecte l'hypothèse 6.1 si et seulement s'il existe un réel $\alpha, 0 \leq \alpha \leq 1$ et un couple de densités (r, s) tel que :

$$f = (1 - \alpha)r + \alpha \quad \text{et} \quad g = \alpha s + 1 - \alpha$$

Démonstration.

Supposons dans un premier temps que f et g sont sous cette forme, et notons que nous avons $\min f \geq \alpha$ et également $\min g \geq 1 - \alpha$ de sorte que $\min f + \min g \geq 1$ et que la condition d'éligibilité 6.1 est respectée.

Maintenant, si l'on suppose la condition respectée, on définit $\alpha = \min f$ et l'on remarque que l'on a $g \geq 1 - \alpha$. Si $\alpha = 0$ ou $\alpha = 1$ alors, respectivement, G ou F est uniforme auquel cas il s'agit d'un cas dégénéré où l'indépendance et l'indétermination se confondent. Sinon, $0 < \alpha < 1$ et nous pouvons écrire :

$$f = (1 - \alpha) \frac{f - \alpha}{1 - \alpha} + \alpha$$

ainsi que :

$$g = \alpha \frac{g - (1 - \alpha)}{\alpha} + (1 - \alpha)$$

En définissant $r = \frac{f - \alpha}{1 - \alpha}$ et $s = \frac{g - (1 - \alpha)}{\alpha}$, il vient alors

$$0 \leq r, s \leq 1;$$

et

$$\int r = \int s = 1.$$

Ceci montre précisément que (r, s) est un couple de densités sur $[0, 1]$ et termine la preuve. \square

En utilisant cette dernière proposition, on peut aisément dériver de tout couple de marges un couple éligible à un couplage sous l'indétermination. Remarquons néanmoins, comme mentionné dans la preuve, que les deux valeurs $\alpha \in \{0, 1\}$ correspondent à f ou g uniforme et à une identité indépendance-indétermination. Nous devons éliminer ces deux valeurs de la plupart des calculs.

6.2 Une copule d'indétermination définissable localement

Dans ce chapitre, nous souhaitons appliquer la notion de copule, classique en probabilité, à l'indétermination afin de tenter de relier le couplage au cadre habituel de dépendance entre marges.

Commençons par rappeler la définition et les propriétés des copules avant d'introduire le très classique théorème de Sklar.

Définition 38 (Copule). *Une copule \mathbf{C} (remarquer la différence de notation avec une fonction de couplage C) est une fonction de répartition sur $[0, 1]^d$ ($d \in \mathbb{N}$) dont les marges sont uniformes. Elle est ainsi caractérisée par trois propriétés pour tout $u = (u_1, \dots, u_d)$:*

- $\mathbf{C}(u) = 0$ si un des u_i est nul ;
- $\mathbf{C}(u) = u_i$ si seule la composante u_i ne vaut pas 1 ;
- \mathbf{C} est d -croissante.

Remarque 41 (d -croissance).

La propriété de d -croissance est intimement liée à la positivité de la loi \mathbb{P} sous-jacente. Ainsi, en dimension 2, $\mathbb{P}(u_1 \leq U \leq u'_1, u_2 \leq U \leq u'_2) = \mathbf{C}(u_1, u_2) - \mathbf{C}(u_1, u'_2) - \mathbf{C}(u'_1, u_2) + \mathbf{C}(u'_1, u'_2) \geq 0$

Remarque 42 (Copule et fonction de couplage).

Les deux notions copules et fonctions de couplage sont très proches ; il en est de même de leur notation dans le présent document. Toutes deux agissent pour définir une loi sur l'espace produit, via les densités pour la fonction de couplage C et via les fonctions de répartition pour la copule \mathbf{C} . Relier les deux de manière systématique sera questionné dans les pages à venir.

Le théorème de Sklar indique que toute fonction \mathbf{C} vérifiant les conditions de la définition 38 peut être appliquée à un d -uplet de fonctions de répartition univariées (F_1, \dots, F_d) pour former une fonction de répartition dont les marges seront précisément les F_i et ce selon la formule :

$$F(x_1, \dots, x_d) = \mathbf{C}(F_1(x_1), \dots, F_d(x_d))$$

Réciproquement, pour toute fonction de répartition F , il existe une copule \mathbf{C} associée. Typiquement, si toutes les marges de F sont connues la formule est :

$$\mathbf{C}(u_1, \dots, u_d) = F(F_1^{-1}(u_1), \dots, F_d^{-1}(u_d))$$

Nous souhaitons appliquer cette définition à l'indétermination pour en extraire une *copule d'indétermination* symétrique de la *copule d'indépendance*. Nous verrons que la dépendance de la copule aux marges empêche de caractériser l'indétermination par une "copule d'indétermination". Cette affirmation sera relativisée lorsque nous montrerons que, localement, nous pouvons la définir.

6.2.1 Définition formelle

On commence par rappeler la fonction de répartition associée à la densité que nous venons d'introduire comme traduction continue du cas d'indétermination :

$$\Pi_{F,G}^+ = yF(x) + xG(y) - xy \quad (6.6)$$

Partant de cette fonction de répartition, on en déduit la copule associée dite copule d'indétermination.

Définition 39 (Copule d'indétermination).

On se donne deux variables aléatoires sur S_1 : U de loi μ , de densité f et de fonction de répartition F et symétriquement V de loi définie par ν , g et G . Alors, leur copule d'indétermination spécifique, extraite de la fonction de répartition associée à leur couplage d'indétermination est donnée par :

$$\mathbf{C}_{\mathbf{F},\mathbf{G}}^+(u, v) = v * F^{-1}(u) + u * G^{-1}(v) - F^{-1}(u) * G^{-1}(v), \quad \forall (u, v) \in S$$

A ce stade, $\mathbf{C}_{\mathbf{F},\mathbf{G}}^+$ dépend du couple de fonctions caractéristiques (F, G) , nous noterons parfois abusivement $\mathbf{C}_{\mathbf{f},\mathbf{g}}^+$ en indexant par les densités. Quoiqu'il en soit, il n'est pas encore permis de parler de copule d'indétermination générique.

Prenons un exemple : U et V possèdent pour fonctions de répartition $F(x) = x^\alpha$ et $G(y) = y^\beta$. Pour commencer, il faut assurer le respect de l'hypothèse 6.1 qui se traduit ici au niveau des densités par :

$$\begin{aligned} \min_x f + \min_y g - 1 &\geq 0 \\ \min_x \alpha * x^{\alpha-1} + \min_y \beta y^{\beta-1} &\geq 1 \end{aligned}$$

l'hypothèse 6.1 impose α et β inférieurs à 1 et

$$\alpha + \beta \geq 1$$

Sous cette dernière hypothèse, la copule associée est alors :

$$\mathbf{C}_{\alpha,\beta}^+(\mathbf{u}, \mathbf{v}) = uv^{\frac{1}{\beta}} + u^{\frac{1}{\alpha}}v - uv \quad (6.7)$$

Remarque 43 (Dépendance aux marges).

Une remarque que nous aurons loisir de détailler dans la suite de cette section : l'apparition de α et β dans la formule permet de voir immédiatement une dépendance importante de la copule aux marges qui sont couplées selon l'indétermination.

Pour l'instant, on ne saurait parler de la copule d'indétermination mais d'une copule d'indétermination parmi toutes celles vérifiant la formule 39 pour des marges données.

6.2.2 Lien entre marges et copule

Essayons de mieux comprendre le lien entre une copule d'indétermination et les marges dont elle est issue. La copule est un moyen comme un autre de réaliser un couplage. Nous en avons défini plusieurs mais isolons-en deux : un premier consiste à prendre les densités

des marges et à coupler selon la formule 38 appliquant ainsi la fonction de couplage d'indétermination, un deuxième consiste à prendre la copule dite d'indétermination donnée par la formule 39 et d'agir directement sur les fonctions de répartition appliquant ainsi la copule.

Couplons selon la formule 38 pour obtenir $\pi^+(x, y) = f(x)+g(y)-1$ et donc $\Pi^+(x, y) = xG(y) + yF(x) - xy$.

La seconde méthode de couplage comporte un problème qu'il est aisé d'isoler ici : la copule fait intervenir des fonctions F^{-1} et G^{-1} que nous avons prises quelconques pensant définir une copule d'indétermination mimant le couplage réalisé par la formule 38 mais au niveau des fonctions de répartition. On constate ici que la copule associée à un couplage d'indétermination dépend des marges que l'on veut coupler. Ce qui n'est pas le cas par exemple de la copule d'indépendance.

Dans le cas général, étant donné deux marges on a au moins trois manières de coupler :

1. définir la densité $\pi(x, y)$ vérifiant les marges (typiquement en appliquant une fonction de couplage C aux densités des marges) ;
2. définir la fonction de répartition $\Pi(x, y)$ en respectant les marges ;
3. définir la copule $\mathbf{C}(u, v)$ et l'appliquer aux fonctions de répartition des marges.

Dans le cas de l'indépendance, avec les notations habituelles, elles correspondent respectivement à $C^\times(u, v) = uv$, $\Pi^\times(x, y) = F(x)G(y)$ et $\mathbf{C}^\times(u, v) = uv$. C'est un cas spécifique ou C tout comme \mathbf{C} peuvent être définies sans référence aux marges.

Dans le cas de l'indétermination, nous n'avons pas cette chance puisque $C^+(u, v) = u+v-1$, $\Pi^+(x, y) = xG(y)+yF(x)-xy$ mais $\mathbf{C}^+(u, v) = u * F^{-1}(v) + v * G^{-1}(u) - F^{-1}(u) * G^{-1}(u)$ dépend fortement des marges. Cela est d'ailleurs bien visible sur l'exemple 6.7 qui prend des formes très différentes selon les valeurs de α et β . Ces deux paramètres n'ayant comme contraintes que d'être inférieurs à 1 et de somme supérieure à 1.

Reprenons la formule de définition des copules et essayons de comprendre ce qui rend le couplage d'indépendance spécifique (en le prenant sous sa forme densité). \mathbf{C} étant une copule, on pose $u = F(x)$ et $v = G(y)$ pour plus de facilité dans les calculs.

$$\begin{aligned} \mathbf{C}(F(x), G(y)) &= \Pi(x, y) \\ \text{en dérivant selon } x \text{ puis } y & \\ \frac{\partial \mathbf{C}(u, v)}{\partial x \partial y} &= \frac{\partial \Pi}{\partial x \partial y} \\ \frac{\partial u}{\partial x} \frac{\partial v}{\partial y} \frac{\partial \mathbf{C}(u, v)}{\partial u \partial v} &= \pi(x, y) \\ f(x)g(y) \frac{\partial \mathbf{C}(u, v)}{\partial u \partial v} &= \pi(x, y) \\ \frac{\partial \mathbf{C}(u, v)}{\partial u \partial v} &= \frac{\pi(x, y)}{f(x)g(y)} \end{aligned}$$

Ainsi le couplage d'indépendance est très particulier puisqu'il fait disparaître les marges : si $\pi = \pi^\times$, $\frac{f_g}{f_g} = 1$ est la dérivée seconde de la copule associée au couplage de F et G selon l'indépendance et cette copule se révèle indépendante de F et G justement.

Dans le cas d'un couplage selon l'indétermination, on obtient comme dérivée croisée de la copule : $\frac{f+g-1}{fg} = \frac{1}{g} + \frac{1}{f} - \frac{1}{fg}$ qui s'intègre bien pour mener à la formule 39 en suivant la formule $\frac{\partial F^{-1}(u)}{\partial u} = \frac{1}{F'(F^{-1}(u))} = \frac{1}{f(x)}$. Cette copule est bien dépendante des marges de sorte que l'on ne peut définir une copule d'indétermination générale.

Remarque 44.

Pour le calcul qui précède, nous nous sommes permis de diviser par fg . Reprenons l'hypothèse 6.1. On sait que $f + g - 1 \geq 0$ de sorte que si f s'annule, $g \geq 1$ et comme $\int_0^1 g = 1$ (on est sur le segment $[0, 1]$ car sinon, il faut diviser comme dans la solution du problème 18) on en déduit que $g = 1$ ce qui appauvrit fortement le couplage : $f + g - 1 = f$ et on retrouve l'indépendance.

6.2.3 Une copule d'indétermination locale

Nous essayons maintenant de déterminer à quel point $\mathbf{C}_{\mathbf{F},\mathbf{G}}^+$ dépend de ses marges. Pour cela, nous nous donnons quatre fonctions de densité f, g, r, s , les quatre fonctions de répartition associées et appliquons la copule d'indétermination des deux premières aux deux secondes :

$$H = H_{r,s}^{f,g}(x, y) = (\mathbf{C}_{\mathbf{F},\mathbf{G}}^+(R, S))(x, y) \tag{6.8}$$

Nous noterons cette fonction de répartition H dans la suite de cette partie. Essayons de voir si H peut représenter un couplage d'indétermination pour (R, S) et sous quelles conditions (on sait que ce n'est pas automatiquement le cas). Pour cela nous comparons H avec la fonction de répartition du couplage d'indétermination de (R, S) .

L'idée est ici de voir s'il existe des classes de couples de marges admettant la même copule d'indétermination. La proposition 43 montre que c'est localement le cas.

Propriété 43 (Une copule d'indétermination locale).

Etant donné quatre densités f, g, r, s et leur fonctions de répartition F, G, R, S , l'application de $\mathbf{C}_{\mathbf{F},\mathbf{G}}^+$ à (R, S) notée H (voir l'équation 6.8) génère un couplage d'indétermination si et seulement s'il existe un $\lambda \geq 0$ respectant

$$\frac{\max(g)}{\max(g) - 1} \geq \lambda \geq 1 - \frac{1}{\max(f)} \tag{6.9}$$

et tel que les deux équations suivantes soient satisfaites :

$$F^{-1}(x) - x = \lambda(R^{-1}(x) - x) \tag{6.10}$$

$$G^{-1}(y) - y = \frac{1}{\lambda}(S^{-1}(y) - y) \tag{6.11}$$

Démonstration.

Nous découpons la preuve en trois parties :

- montrer la relation entre (F, G) et (R, S) ;
- proposer deux encadrements de λ selon son signe ;
- éliminer le cas λ négatif ;
- vérifier que la densité h correspondante est positive.

Relation entre (F, G) et (R, S)

Calculons la dérivée seconde croisée de H en espérant tomber sur $r + s - 1$:

$$\begin{aligned} \frac{\partial^2 H(x, y)}{\partial x \partial y} &= \frac{\partial^2 (R(x) * G^{-1}(S(y)) + S(y) * F^{-1}(R(x)) - F^{-1}(R(x))G^{-1}(S(y)))}{\partial x \partial y} \\ &= \frac{r(x)s(y)}{g(G^{-1}(S(y)))} + \frac{r(x)s(y)}{f(F^{-1}(R(x)))} - \frac{r(x)s(y)}{f(F^{-1}(R(x)))g(G^{-1}(S(y)))} \end{aligned}$$

En forçant une égalité avec $r(x) + s(y) - 1$ il vient :

$$\begin{aligned} r(x) + s(y) - 1 &= \frac{r(x)s(y)}{g(G^{-1}(S(y)))} + \frac{r(x)s(y)}{f(F^{-1}(R(x)))} - \frac{r(x)s(y)}{f(F^{-1}(R(x)))g(G^{-1}(S(y)))} \quad (6.12) \\ \frac{r(x) + s(y) - 1}{r(x)s(y)} &= \frac{f(F^{-1}(R(x))) + g(G^{-1}(S(y))) - 1}{f(F^{-1}(R(x)))g(G^{-1}(S(y)))} \\ \frac{1}{r(x)} + \frac{1}{s(y)} - \frac{1}{r(x)s(y)} &= \frac{1}{f(F^{-1}(R(x)))} + \frac{1}{g(G^{-1}(S(y)))} - \frac{1}{f(F^{-1}(R(x)))g(G^{-1}(S(y)))} \\ \frac{1}{r(x)} \left(1 - \frac{1}{s(y)} \right) &= \frac{1}{f(F^{-1}(R(x)))} \left(1 - \frac{1}{g(G^{-1}(S(y)))} \right) + \frac{1}{g(G^{-1}(S(y)))} - \frac{1}{s(y)} \end{aligned}$$

Remarquons que la division par $r(x)s(y)$ nécessite certaines hypothèses que nous réglons dans la remarque 45.

Définissons quelques notations, α représente la partie "simple" et β son analogue plus "complexe" ; l'indice en x ou en y est quant à lui naturel.

$$\begin{aligned} \alpha_x &= 1 - \frac{1}{r(x)} & \alpha_y &= 1 - \frac{1}{s(y)} \\ \beta_x &= 1 - \frac{1}{f(F^{-1}(R(x)))} & \beta_y &= 1 - \frac{1}{g(G^{-1}(S(y)))} \end{aligned}$$

Alors on réécrit :

$$\begin{aligned} \frac{1}{r(x)} \alpha_y &= \frac{1}{f(F^{-1}(R(x)))} \beta_y + \alpha_y - \beta_y \\ \left(\frac{1}{r(x)} - 1 \right) \alpha_y &= \left(\frac{1}{f(F^{-1}(R(x)))} - 1 \right) \beta_y \\ \alpha_x \alpha_y &= \beta_x \beta_y \end{aligned}$$

On en déduit (voir cette fois la justification de la division dans la remarque 46) l'existence d'une constante λ telle que :

$$\frac{\alpha_y}{\beta_y} = \frac{\beta_x}{\alpha_x} = \lambda$$

Intéressons-nous aux expressions des fonctions de répartition pour mieux cerner λ . La relation $\frac{\beta_x}{\alpha_x} = \lambda$ s'écrit aussi :

$$r(x) - \frac{r(x)}{f(F^{-1}(R(x)))} = \lambda(r(x) - 1)$$

Ce qui donne en intégrant sur $[0, x]$:

$$R(x) - F^{-1}(R(x)) = \lambda(R(x) - x)$$

Comme R est une fonction de répartition dont la dérivée ne s'annule pas (hypothèse déjà formulée), R est inversible de sorte que l'on a pour tout x :

$$F^{-1} = x(1 - \lambda) + \lambda R^{-1}(x) \quad (6.13)$$

et symétriquement pour y :

$$G^{-1} = y(1 - \frac{1}{\lambda}) + \frac{1}{\lambda} S^{-1}(x) \quad (6.14)$$

On vérifie en reportant dans l'expression de la copule appliquée à (R, S) que cela suffit pour avoir $\mathbf{C}_{\mathbf{F}, \mathbf{G}}^+(R, S) = xS(y) + yR(x) - xy$.

Deux encadrements de λ

Formellement, tout concorde. Néanmoins, l'hypothèse de croissance de R rajoute des hypothèses sur λ . Supposons λ positif alors comme $F^{-1}(x) - x(1 - \lambda)$ est une fonction croissante, on a pour tout x , $\frac{1}{f(x)} \geq 1 - \lambda$ soit encore en conduisant des calculs similaires en y :

$$\frac{\max(g)}{\max(g) - 1} \geq \lambda \geq 1 - \frac{1}{\max(f)} \quad (6.15)$$

De même, si λ est supposée négative, on obtient :

$$1 - \frac{1}{\min(f)} \geq \lambda \geq \frac{\min(g)}{\min(g) - 1} \quad (6.16)$$

Ces valeurs ont un sens sauf si g vaut 1 mais dans ce cas $G = Id$ et $\mathbf{C}_{\mathbf{F}, \mathbf{G}}^+ = \mathbf{C}^\times$. Ainsi on a sans restriction $\min(f) < 1 < \max(f)$ tout comme $\min(g) < 1 < \max(g)$.

Récrivons les equations 6.13 et 6.14 :

$$F^{-1}(x) - x = \lambda(R^{-1}(x) - x) \quad (6.17)$$

$$G^{-1}(y) - y = \frac{1}{\lambda}(S^{-1}(y) - y) \quad (6.18)$$

On voit que les fonctions de répartition partageant la même copule d'indétermination sont celles dont la fonction inverse est dans l'enveloppe convexe exigible (comprendre vérifiant 6.16 ou 6.15) de F^{-1} et de l'identité.

λ positif

Enfin, il nous faut noter que l'hypothèse 6.1 garantissant la possibilité de coupler r, s selon l'indétermination n'est pas garantie. Reste que la transformation en λ est continue autour de $\lambda = 1$ de sorte que la validation de l'hypothèse sur (f, g) suffit pour λ assez proche de 1. Précisons cette remarque.

On reprend les équations 6.13 et 6.14 pour obtenir une expression de la dérivée de r et s .

$$\begin{aligned}\frac{1}{f(F^{-1}(x))} &= (1 - \lambda) + \frac{\lambda}{r(R^{-1}(x))} \\ r(R^{-1}(x)) &= \frac{1}{\frac{1}{\lambda f(F^{-1}(x))} + \frac{1}{\lambda} - 1} \\ r(\tilde{x}) &= \frac{1}{\frac{1}{\lambda f(x)} + \frac{1}{\lambda} - 1}\end{aligned}$$

Où \tilde{x} parcourt $[0, 1]$ tout comme x et de même :

$$s(\tilde{y}) = \frac{1}{\frac{\lambda}{g(y)} + \lambda - 1}$$

On souhaite donc avoir pour tout (x, y)

$$r(x) + s(y) - 1 \geq 0$$

$$r(\tilde{x}) + s(\tilde{y}) - 1 \geq 0$$

$$\frac{1}{\frac{1}{\lambda f(x)} + \frac{1}{\lambda} - 1} + \frac{1}{\frac{\lambda}{g(y)} + \lambda - 1} \geq 1$$

$r \geq 0$ d'après les inégalités 6.16 et 6.15 donc

$$\frac{\frac{\lambda}{g(y)} + \lambda - 1}{\frac{1}{\lambda f(x)} + \frac{1}{\lambda} - 1} + 1 \geq \frac{\lambda}{g(y)} + \lambda - 1$$

grâce aux mêmes inégalités

$$\frac{\lambda}{g(y)} + \lambda - 1 + \frac{1}{\lambda f(x)} + \frac{1}{\lambda} - 1 \geq \left(\frac{\lambda}{g(y)} + \lambda - 1 \right) \left(\frac{1}{\lambda f(x)} + \frac{1}{\lambda} - 1 \right)$$

qui se récrit

$$\left(\frac{1}{\lambda f} + \frac{1}{\lambda} - 2 \right) \left(\frac{\lambda}{g} + \lambda - 2 \right) \leq 1$$

$$\left(\frac{1}{f} + 1 - 2\lambda \right) \left(\frac{1}{g} + 1 - \frac{2}{\lambda} \right) \leq 1$$

Déjà, la dernière inégalité exclut définitivement le cas $\lambda \leq 0$: λ est positif et c'était l'objet de cette partie.

Transfert de la vérification de l'hypothèse 6.1

Continuons avec la résolution d'une équation du second degré. Notre dernière expression équivaut à :

$$\begin{aligned}\left(\frac{1}{\min(f)} + 1 - 2\lambda \right) \left(\frac{1}{\min(g)} + 1 - \frac{2}{\lambda} \right) &\leq 1 \\ \left(2\lambda - \frac{1}{\min(f)} - 1 \right) \left(\frac{\lambda}{\min(g)} + \lambda - 2 \right) &\geq -\lambda \\ \left(\frac{2}{m_g} + 2 \right) \lambda^2 - \lambda \left(\frac{1}{m_f m_g} + \frac{1}{m_f} + \frac{1}{m_g} + 4 \right) + \frac{2}{m_f} + 2 &\geq 0 \\ (2m_f + 2m_g)\lambda^2 - (1 + m_f + m_g + 4m_f m_g)\lambda + 2m_f + 2m_g &\geq 0\end{aligned}$$

L'équation sous-jacente n'a de solution que si l'on a :

$$2 \leq \frac{1 + m_f + m_g + 4m_fm_g}{2m_f + 2m_g}$$

Or on remarque :

$$\frac{1 + m_f + m_g + 4m_fm_g}{2m_f + 2m_g} \leq \frac{1 + 3m_f + 3m_g}{2m_f + 2m_g} = \frac{3}{2} + \frac{1}{2m_f + 2m_g} \leq 2 \quad (6.19)$$

La dernière inégalité étant obtenue grâce à la condition 6.1 valide sur (f, g) .

Remarque 45 (Division par rs).

Si r s'annule, l'hypothèse 6.1 impose $r \geq 1$ et donc $R = Id$. S'ensuit une simplification de l'équation 6.12

$$\begin{aligned} s(y) &= \frac{s(y)}{g(G^{-1}(S(y)))} + \frac{s(y)}{f(F^{-1}(y))} - \frac{s(y)}{f(F^{-1}(x))g(G^{-1}(S(y)))} \\ s(y) \left(1 - \frac{1}{g(G^{-1}(S(y)))}\right) &= \frac{1}{f(F^{-1}(x))} s(y) \left(1 - \frac{1}{g(G^{-1}(S(y)))}\right) \end{aligned}$$

Comme le cas intéressant impose f et g différents de 1, on voit que la précédente égalité ne saurait être respectée pour tout couple (x, y) .

Remarque 46 (Division par $1 - \frac{1}{f}$).

Comme on a supposé f et g différents de 1, on peut se placer sur une partie de S où aucun ne vaut 1. La division est alors permise et la preuve similaire conduit aux propriétés de la constante λ .

□

Remarque 47 (Transfert de la condition).

L'inégalité 6.19 est tout à fait remarquable : elle montre que la vérification de l'hypothèse 6.1 est automatiquement transmise de (f, g) à (r, s) dès que λ est positif tandis que la forme factorisée élimine le cas λ négatif.

6.2.4 Ecart maximal entre les deux copules

Dans cette section, nous estimons l'écart entre les deux copules canoniques : indétermination (définie localement) et indépendance. La copule d'indépendance est tout simplement :

$$\mathbf{C}^\times(u, v) = uv$$

Comme $\mathbf{C}_{\mathbf{F}, \mathbf{G}}^+$ dépend des marges, nous mesurons l'écart lorsque (F, G) est un couple éligible et ce avant de le maximiser. Sous la norme \mathbb{L}^1 , l'expression de l'écart est la suivante :

$$\begin{aligned} \Delta_1(F, G) = \|\mathbf{C}^\times - \mathbf{C}_{\mathbf{F}, \mathbf{G}}^+\|_1 &= \int_{u=0}^1 \int_{v=0}^1 |uv - vF^{-1}(u) - uG^{-1}(v) + F^{-1}(u)G^{-1}(v) - uv| \, du \, dv \\ &= \int_{u=0}^1 \int_{v=0}^1 |(F^{-1}(u) - u)(G^{-1}(v) - v)| \, du \, dv \\ &= \int_{u=0}^1 |F^{-1}(u) - u| \, du \int_{v=0}^1 |G^{-1}(v) - v| \, dv \end{aligned}$$

Nous disposons donc d'une formule close, plutôt élégante caractérisant l'écart. Le choix de la norme \mathbb{L}^1 , outre qu'il soit classique, sera motivé ultérieurement.

Nous savons déjà que la valeur est nulle dès lors qu'une des marges est uniforme mais ne connaissons pas le maximum, ni d'ailleurs, le couple de marges (F, G) qui le réalise. Les deux questions trouvent réponse dans la proposition 44.

Propriété 44.

En notant \mathcal{E}^+ l'ensemble des couples (F, G) de fonctions de répartition éligibles c'est-à-dire respectant la condition 6.1 on a :

$$\max_{(F,G) \in \mathcal{E}^+} \Delta_1(F, G) = \frac{1}{16}$$

De plus le couple densité réalisant ce maximum est donné par :

$$(f_0, g_0) = \left(u \mapsto \frac{1}{2}(1 + \delta_{u=0}), v \mapsto \frac{1}{2}(1 + \delta_{v=0}) \right)$$

Démonstration.

Travailler avec l'ensemble $(F, G) \in \mathcal{E}^+$ est assez inconfortable puisque nous émettons alors une hypothèse sur f à transmettre à F^{-1} en passant par F et de même pour G .

Avantageusement, nous utilisons la proposition constructive 42 et extrayons un α et deux densités r, s (respectivement, deux fonctions de répartition (R, S)) pour obtenir :

$$F(u) = \alpha u + (1 - \alpha)R(u);$$

ainsi que :

$$G(v) = (1 - \alpha)v + \alpha S(v).$$

Récrivons alors $\Delta_1(F, G)$:

$$\begin{aligned} \Delta_1(F, G) &= \int_{u=0}^1 |F^{-1}(u) - u| du \int_{v=0}^1 |G^{-1}(v) - v| dv \\ &= \int_{u=0}^1 |F(u) - u| f(u) du \int_{v=0}^1 |G(v) - v| g(v) dv \end{aligned}$$

À première vue, cette expression sépare F et G mais leur lien est en réalité sous-jacent, intégré dans la constante α qui les relie. Pour autant, nous pouvons maximiser chaque intégrale et reporter l'intégration de leur relation, c'est-à-dire du respect de la condition caractérisant \mathcal{E}^+ à un second temps. En menant les calculs dans la partie concernant F de Δ_1 il vient :

$$\begin{aligned} I_f &= \int_{u=0}^1 |F(u) - u| f(u) du \\ &= \int_{u=0}^1 |\alpha u + (1 - \alpha)R(u) - u| f(u) du \\ &= (1 - \alpha) \int_{u=0}^1 |R(u) - u| (\alpha + (1 - \alpha)r(u)) du \\ &= (1 - \alpha) \left[\alpha \int_{u=0}^1 |R(u) - u| du + (1 - \alpha) \int_{u=0}^1 |R^{-1}(u) - u| du \right] \end{aligned}$$

Pour chaque élément R son symétrique par rapport à la droite $y = x$ donne le même résultat de sorte que l'on peut suppose $R(u) \geq u$ en renversant toutes les parties en dessous. Immédiatement il vient alors

$$I_f \leq \frac{(1 - \alpha)}{2}. \quad (6.20)$$

Un calcul similaire sur G , conduit à l'inégalité combinée :

$$\Delta_1(F, G) \leq \frac{(1 - \alpha)\alpha}{4} \leq \frac{1}{16} \text{ atteint pour } \alpha = \frac{1}{2}$$

Nous terminons la preuve en remarquant que poser $f(u) = g(u) = \frac{1 + \delta_{u=0}}{2}$ permet de réaliser ce majorant qui devient donc le maximum. \square

Remarque 48 (Transposition dans le cas discret).

Le couple de densités (f_0, g_0) qui réalise l'écart maximal entre l'indépendance et l'indétermination est la transposition naturelle du cas discret. Ainsi, dans celui-ci, il faut maximiser le déséquilibre sur les marges, ce qui revient à définir comme probabilité $\frac{1}{2p} + \frac{1}{2}\delta_{u=0}$ avec p le nombre de modalités.

Remarque 49 (Choix de la norme L_1).

L'application du lemme de Scheffé nous permet de convertir la norme utilisée dans la propriété 44 en norme L_∞ . Formellement, nous avons également, $\forall (F, G) \in \mathcal{E}^+, \forall (U, V) \in \mathcal{B}([0, 1])^2$:

$$\left| \int_{u \in U} \int_{v \in V} \mathbf{C}_{\mathbf{F}, \mathbf{G}}^+(u, v) \, du \, dv - \int_{u \in U} \int_{v \in V} \mathbf{C}^\times(u, v) \, du \, dv \right| \leq \frac{1}{16}$$

Remarque 50 (Cas discret).

Dans la section 2.4.3 nous avons déjà démontré un résultat similaire dans le cas discret. Etant donné deux marges μ et ν prises uniformément parmi les lois de probabilité à p et q éléments respectivement, l'écart espéré en norme L^2 entre indétermination et indépendance sur ces marges tirées est inférieur à $\frac{1}{pq}$.

6.2.5 Conclusions

Au travers de la définition 39, nous avons introduit une copule d'indétermination spécifique $\mathbf{C}_{\mathbf{R}, \mathbf{G}}^+$. Nous avons également montré que sa dépendance aux marges nous empêche de caractériser l'indétermination par une copule générique. Au final, l'ensemble des copules d'indétermination est paramétrisé par les couples de marges éligibles (comprendre respectant la condition 6.1).

A priori, l'association $(F, G) \mapsto \mathbf{C}_{\mathbf{F}, \mathbf{G}}^+$ n'a aucune raison d'être injective et de fait elle ne l'est pas. Ainsi, un couple (F, G) partage la même copule d'indétermination qu'un autre couple (R, S) dès que les quatre fonctions de répartition sont liées selon les deux équations 6.10. De plus, il s'agit d'une caractérisation qui ne peut être étendue. En résumé, on demande à chaque fonction de répartition du deuxième couple d'être une combinaison linéaire de celle du premier couple et de l'identité. Sur ce segment des coefficients linéaires λ valables, toutes les copules spécifiques d'indétermination sont identiques.

6.3 Indétermination et minimisation de la vraisemblance moyenne

6.3.1 Définition de la vraisemblance moyenne

Observant un ensemble de n réalisations d'une variable W , un problème habituel et naturel est d'extraire la loi de probabilité P_π sous-jacente supposée inconnue et que nous mentionnerons parfois abusivement par sa densité π . Le problème s'accompagne souvent d'un ensemble $\Omega \subset \Lambda$ de lois de probabilité au sein duquel nous cherchons la plus vraisemblable en un sens à définir. Nous supposons dans cette section, en accord avec les notations déjà introduites que $P_\pi \in \Lambda$ s'applique à l'espace produit S .

Suivant le paradigme du maximum de vraisemblance (PMV), à chaque réalisation observée $W = w$ et pour toute loi de probabilité $P_{\pi'}$ dans Ω , on calcule la probabilité que $W' \sim P_{\pi'}$ soit proche de w c'est-à-dire appartienne à $[w, w + dw]$. Formellement :

$$\mathbb{P}_{\pi'}(W' \in [w, w + dw]) = \pi'(w) dw$$

Si nous intégrons sur les valeurs que peut prendre W on obtient la vraisemblance moyenne $\bar{L}(\pi, \pi')$ de W (ou π) sous π' :

Définition 40 (Vraisemblance moyenne).

Étant donné deux lois de probabilité P_π et $P_{\pi'}$ dont les densités sont respectivement π et π' on définit la vraisemblance moyenne entre P_π et $P_{\pi'}$ ou abusivement entre π et π' par :

$$\bar{L}(P_\pi, P_{\pi'}) = \bar{L}(\pi, \pi') = \mathbb{E}_{W \sim P_\pi}(\pi'(W)) \quad (6.21)$$

Remarque 51.

Au final, $\bar{L}(P, P')$ se réécrit $\int_S \pi(s)\pi'(s) ds$ où S est l'ensemble auquel s'appliquent P_π et $P_{\pi'}$ (pas nécessairement un espace produit). Cette formulation montre immédiatement que l'opérateur \bar{L} est symétrique.

Dans le cas discret, et comme montré dans la section 3.5.4, π^+ correspond à une minimisation des collisions (définition 18) entre deux tirages du couple de variables. Malheureusement, dans le cas continu, nous ne saurions définir une collision par une égalité sans obtenir une probabilité nulle d'apparition. Nous allons montrer que la notion de vraisemblance moyenne tout juste introduite correspond à une extension de celle de collision dans le cas continu.

Sous cette remarque que nous justifierons, on s'attend à ce que $\bar{L}(\pi, \pi)$ soit minimal quand $\pi = \pi^+$ et c'est précisément le cas puisque nous avons l'égalité suivante

$$\bar{L}(\pi, \pi) = \int_S \pi^2(w) dw \quad (6.22)$$

dans laquelle nous reconnaissons la fonction de coût du problème 17.

De surcroît dans le cas discret, le nombre de collision entre π et π^+ ne dépendait que des marges de π . Nous montrons que cette propriété s'étend au cas continu.

6.3.2 Évanouissement de la corrélation sous l'indétermination

Nous revenons à l'expression de la vraisemblance moyenne (équation 6.21) entre P_{π^+} et une loi P_h quelconque dans Λ et nous y introduisons la forme de π^+ donnée dans l'équation 36.

$$\begin{aligned}
\bar{L}(h, \pi^+) &= \mathbb{E}_{W \sim \pi^+}(h(W)) \\
&= \int_{x=0}^1 \int_{y=0}^1 \pi^+(x, y) h(x, y) dx dy \\
&= \int_{x=0}^1 \int_{y=0}^1 (f(x) + g(y) - 1) h(x, y) dx dy \\
&= \int_{x=0}^1 f(x) \partial_1 P_h(x) dx + \int_{y=0}^1 g(y) \partial_2 P_h(y) dy - 1 \tag{6.23}
\end{aligned}$$

où $\partial_1 P_h$ désigne la densité de la première marge de P_h et $\partial_2 P_h$ celle de la seconde marge.

L'égalité 6.23 montre que la vraisemblance moyenne entre h et π^+ ne dépend que des quatre marges sous-jacentes. La corrélation éventuelle au sein de P_h n'entre pas en jeu. En s'appuyant sur cette propriété nous démontrons le théorème 12

Théorème 12.

Pour deux densités données f et g sur S_1 , on note $\Omega_{f,g}$ l'ensemble des densités sur S dont le couple de marges est (f, g) . Parmi $\Omega_{f,g}$, π^+ réalise l'optimum du problème suivant :

$$\min_{\pi \in \Omega_{f,g}} \bar{L}(\pi, \pi); \tag{6.24}$$

de plus, pour tout π_0 dans $\Omega_{f,g}$ nous avons également :

$$\min_{\pi \in \Omega_{f,g}} \bar{L}(\pi, \pi) = \bar{L}(\pi^+, \pi_0). \tag{6.25}$$

Démonstration.

La première partie revient à noter à l'aide de l'équation 6.22, que le coût correspond à celui du problème 17 qui est précisément minimisé au sein de $\Omega_{f,g}$ par π^+ .

Pour la seconde partie, nous récrivons l'équation 6.23 :

$$\bar{L}(\pi^+, \pi_0) = \bar{L}(f, f) + \bar{L}(g, g) - 1$$

de sorte que l'on a :

$$\bar{L}(\pi^+, \pi_0) = \bar{L}(\pi^+, \pi^+)$$

ce qui conclut la preuve. □

6.3.3 Vraisemblance moyenne et collisions

Dans toute application statistique π n'est pas connu et le calcul de \bar{L} devra se baser sur un ensemble de n réalisations de W . Nous allons décrire dans cette partie une méthode pour approcher \bar{L} lors de données expérimentales. Par ailleurs nous allons montrer que \bar{L} correspond effectivement aux collisions dans le cas discret et en étend donc la notion.

Supposons que $s_n = W_1, \dots, W_n$ est une série de n points i.i.d. tirés sous une loi $P_\pi \in \Lambda$. On définit alors P_n , la mesure empirique associée :

$$P_n(s) = \frac{1}{n} \sum_{i=1}^n \delta_{W_i}. \quad (6.26)$$

Une question naturelle est de décider, à marges fixées (f, g) et en nous appuyant uniquement sur P_n si la densité π de W est $\pi^+ = C^+(f, g)$. Dans le cas discret, nous estimerions typiquement la probabilité de collision grâce aux réalisations et comparerions sa valeur au minimal.

$$\begin{aligned} \mathbb{P}_{C^\times(P_n, P_n)}(W'_1 = W'_2) &= \sum_{i=1}^n \frac{1}{n} P_n(W_i) \\ &= \mathbb{E}_{W' \sim P_n}(P_n(W')) \\ &= \bar{L}(P_n, P_n) \end{aligned}$$

Cette dernière égalité montre précisément que la vraisemblance moyenne étend la notion de collision de la définition 18 discrète. Malheureusement, dans le cas continu (comme P_π a une densité π), deux W_i ne seront jamais égaux entraînant donc une probabilité nulle et ce quelle que soit la corrélation intrinsèque à W . Cette remarque enlève tout intérêt à vérifier que $\bar{L}(P_n, P_n)$ est minimale puisqu'elle est tout le temps nulle!

Revenons à la propriété particulière de π^+ démontrée à la fin du théorème 12 :

$$\bar{L}(\pi^+, \pi^+) = \bar{L}(\pi^+, \pi_0) \quad (6.27)$$

et ce quelle que soit la densité π_0 sur S . Nous venons d'écrire que l'évaluation de la vraisemblance moyenne de π^+ peut se faire contre toute densité π_0 . Nous en choisissons une (ce sera π^\times dans la section 6.4) et retenons $\bar{L}(\pi_0, P_n)$ pour estimer une généralisation de la probabilité de collision.

Maintenant, si l'on suppose $P_n \rightarrow P_{\pi^+}$ en un certain sens (que nous préciserons dans la section 6.4), on s'attend à avoir :

$$\bar{L}(P_n, \pi_0) \rightarrow \bar{L}(\pi^+, \pi_0) = \bar{L}(\pi^+, \pi^+) \quad (6.28)$$

cette dernière expression étant minimale parmi $\Omega_{f,g}$ si l'on est sous π^+ .

Dans la section 6.4 nous nous emploierons à formaliser l'approche succinctement décrite notamment par l'introduction d'une topologie sur Λ pour pouvoir écrire la convergence.

6.4 Statistique de test basée sur la vraisemblance moyenne

6.4.1 Définition de t_n

Nous savons que π^\times et π^+ sont proches (voir la propriété 44) et nous attendons donc à ce que les deux valeurs $\bar{L}(\pi^\times, \pi^+) = l_0$ et $\bar{L}(\pi^\times, \pi^\times) = l_1$ le soient également. Pour autant

d'après la proposition 38, et comme $\pi^+ = C^+(f, g)$ et $\pi^\times = C^\times(f, g)$ ont les mêmes marges on sait que l'on a toujours $l_0 \leq l_1$. De surcroît, si f et g ne sont pas uniformes,

$$\eta = l_1 - l_0 > 0. \quad (6.29)$$

Nous allons utiliser cette différence stricte pour construire un test statistique d'indétermination. Pour commencer, nous définissons un ensemble autour de l'indépendance :

$$\Omega^\times = \{P_\pi \in \Omega_{f,g} / \bar{L}(\pi, \pi^\times) \geq l_1\}. \quad (6.30)$$

ainsi que deux hypothèses : H_0 correspond à $\pi = \pi^+$ tandis que son opposée est H_1 qui correspond à $\pi \in \Omega^\times$.

Remarque 52.

En utilisant la ρ -topologie que nous allons introduire autour de l'équation 6.35, nous verrons dans la proposition 45 que les deux hypothèses correspondent à des ensembles non vides et séparables de Λ .

Étant donné n observations $s_n = W_1, \dots, W_n$, on désigne par s la séquence entière et l'on définit une statistique de test t_n qui dépend uniquement des n premières coordonnées de s pour décider entre les deux hypothèses :

$$t_n(s) = \bar{L}(P_n, \pi^\times) = \int_S \pi^\times dP_n(s); \quad (6.31)$$

où P_n est défini selon l'équation 6.26.

À partir de l'observation de s , H_0 sera rejetée si $t_n(s)$ est supérieur à une valeur spécifique. Ce test est motivé par la combinaison de la section 6.3.3 qui nous permet de l'interpréter comme une estimation de la probabilité de collision entre deux tirages successifs et du théorème 12 qui nous permet (sous H_0) d'estimer $\bar{L}(\pi, \pi)$ en utilisant n'importe quel π_0 ; en particulier en utilisant π^\times .

Nous allons étudier t_n sous chacune des deux hypothèses pour calibrer le seuil. Plus spécifiquement, nous suivons la méthode employée dans la deuxième section de [45] dédiée à la pente de Bahadur. Ce travail requiert l'introduction de quelques notations que nous rapportons ci-après.

6.4.2 Quelques notions et lemmes usuels

Suivant le choix classique dans la littérature scientifique et notamment dans les deux articles précités ([28] et [46]), nous utilisons la topologie τ sur Λ ; celle-ci est définie par les voisinages de base :

$$U(P, \mathcal{P}, \epsilon) = \{Q / \forall i, |P(B_i) - Q(B_i)| < \epsilon\} \quad (6.32)$$

où \mathcal{P} parcourt toute les partitions \mathcal{B} -mesurables $\mathcal{P} = (B_1, \dots, B_k)$ de S . Sans absence de précision, les notions d'intérieur noté Ω° et d'adhérence notée $\bar{\Omega}$ d'un ensemble Ω seront dans le sens de la topologie τ .

Pour cette topologie, une séquence $(Q_n)_{n \in \mathbb{N}}$ de lois de probabilités converge vers Q si et seulement si $\lim_{n \rightarrow \infty} \int_{\mathbb{R}} f dQ_n \rightarrow \int_{\mathbb{R}} f dQ$ pour toute fonction bornée \mathcal{B} -mesurable $f : S \rightarrow \mathbb{R}$.

Nous introduisons également la divergence de Kullback-Leibler notée D_{KL} , souvent présente pour estimer la vitesse de convergence de grandes déviations, et définie par :

$$\begin{aligned} D_{KL}(P|Q) &= \int_S \log \left(\frac{dP}{dQ} \right) dP \quad \text{si } Q \ll P \\ &= \infty \quad \text{sinon} \end{aligned} \tag{6.33}$$

où l'on précise quelques conventions à savoir $\log 0 = -\infty$, $0 \cdot (\pm\infty) = 0$ et $\log(a/0) = \infty$, $\forall a > 0$.

Les propriétés de D_{KL} sont listées dans [93]. Un résultat important est sa semi-continuité inférieure :

Lemme 3 (Divergence semi-continue inférieurement).

Pour toute loi de probabilité $P \in \Lambda$, la fonction :

$$Q \in \Lambda \mapsto D_{KL}(Q, P)$$

est semi continue inférieurement au sens de la topologie τ .

Démonstration.

Il s'agit du Lemme 2.2 de [46]. □

Une extension habituelle consiste à définir D_{KL} d'un ensemble à une loi comme nous le ferions avec une distance. On se donne ainsi $\Omega \subset \Lambda$ et $P \in \Lambda$ et l'on pose :

$$D_{KL}(\Omega|P) = \inf_{Q \in \Omega} D_{KL}(Q|P) \tag{6.34}$$

avec la convention supplémentaire que la valeur est infinie si Ω est vide.

Pour la suite, il est nécessaire d'introduire une seconde topologie sur Λ . Il s'agit de la topologie ρ également classique et induite par la distance d définie comme le maximum de différence entre les deux fonctions de répartition sur l'ensemble ordonné $S = [0, 1] \times [0, 1]$:

$$d(P, Q) = \sup_{x \in S} |P(0, x) - Q(0, x)|. \tag{6.35}$$

Fortes des notions juste introduites, nous listons maintenant une liste de lemmes extraits de [46] ou [28], parfois démontrés dans le présent document lorsqu'ils sont trop spécifiques, avant de démontrer un théorème de Sanov particulier à notre problème (Théorème 13).

Le premier lemme relie les deux topologies afin d'apprécier un ouvert de l'un comme un ouvert de l'autre.

Lemme 4 (τ est plus fine que ρ).

La topologie τ est plus fine que la topologie ρ .

Démonstration.

Il s'agit du Lemme 2.1 de [46]. □

La topologie τ , nativement associée à une distance s'avère plus facile à manier et notamment pour définir un voisinage ouvert d'une loi de probabilité P comme exposé ci-dessous :

$$V_\epsilon(P) = \{Q/d(P, Q) < \epsilon\} \quad (6.36)$$

La dernière définition que nous introduirons est nécessaire pour valider une hypothèse qui conduit à la première inégalité dans le théorème 13. Il s'agit de la notion d'ensemble complètement convexe qui apparaît dans le Lemme 5.

Définition 41 (Ensemble complètement convexe).

Un ensemble de lois de probabilité $\Omega \subset \Lambda$ est complètement convexe si pour tout autre espace de probabilité (A, \mathcal{A}, μ) et tout noyau de Markov ν de (A, \mathcal{A}) vers (S, \mathcal{B}) tel que $\nu(a, \cdot) \in \Omega \forall a \in A$, la loi de probabilité $\mu\nu$ définie par :

$$\mu\nu(B) = \int_A \nu(a, B) d\mu(a), \quad B \in \mathcal{B}$$

appartient également à Ω .

Lemme 5.

Si Ω est un ensemble complètement convexe de Λ alors :

$$\overline{\lim}_{n \rightarrow \infty} \frac{1}{n} \log \mathbb{P}_\pi (P_n \in \Omega) \leq -D_{KL}(\Omega | P_\pi)$$

où P_n correspond à la mesure empirique définie dans l'équation 6.26.

Démonstration.

Il s'agit de la première partie du Théorème 1 de [28]. □

Lemme 6.

Pour toute loi de probabilité $P \in \Lambda$ le sous-espace $V_\epsilon(P)$ est complètement convexe.

Démonstration.

Pour suivre les notations habituelles, nous posons $\Omega = V_\epsilon(P)$. On se donne maintenant un espace de probabilité (A, \mathcal{A}, μ) et un noyau de Markov ν tel que $\nu(a, \cdot) \in \Omega, \forall a \in A$. Il nous faut vérifier que $\mu\nu$ définie dans la Définition 41 appartient aussi à Ω . Pour un élément $x \in S$ nous calculons :

$$\begin{aligned} |\mu\nu([0, x]) - P([0, x])| &= \left| \int_{[0, x]} \int_A d\nu(a, u) d\mu(a) - \int_{[0, x]} dP(u) \right| \\ &= \left| \int_{[0, x]} \int_A d\nu(a, u) d\mu(a) - \int_A \int_{[0, x]} dP(u) d\mu(a) \right| \\ &= \left| \int_A \left(\int_{[0, x]} d\nu(a, u) - dP(u) \right) d\mu(a) \right| \\ &\leq \int_A \left| \int_{[0, x]} d\nu(a, u) - dP(u) \right| d\mu(a) \\ &\leq \int_A \epsilon d\mu(a) \\ &= \epsilon \end{aligned}$$

L'inversion des deux intégrales est permise car toutes les parties sont positives. Au final, nous avons montré que $\mu\nu$ appartient bien à Ω qui s'avère ainsi être un ensemble complètement convexe. \square

Le dernier lemme que nous rapportons ici s'intéresse à l'inégalité complémentaire du Lemme 5.

Lemme 7.

Si $\Omega \subset \Lambda$ est tel que $D_{KL}(\Omega^\circ|P_\pi) = D_{KL}(\Omega|P_\pi)$ alors

$$-D_{KL}(\Omega|P_\pi) \leq \underline{\lim}_{n \rightarrow \infty} \frac{1}{n} \log \mathbb{P}_\pi(P_n \in \Omega)$$

Démonstration.

Ce lemme correspond à la partie B du Lemme 3.1 de [46]. \square

6.4.3 Théorème de Sanov

Nous proposons dans le théorème 13 de contrôler la vitesse de convergence d'une mesure empirique lorsqu'une hypothèse de ρ -convergence est formulée. Nous nous donnons deux mesures empiriques convergeant vers deux lois différentes P_π et $P_{\pi'}$ et montrons que nous pouvons les séparer (pour la ρ -topologie) suivant une pente exponentielle dont le facteur est essentiellement $D_{KL}(P_\pi|P_{\pi'})$.

Théorème 13 (Sanov).

Nous nous donnons deux lois de probabilités $P_\pi \in \Lambda$ et $P_{\pi'} \in \Lambda$ ainsi que deux lois empiriques extraites de n réalisations i.i.d. de chacune : P_n et P'_n . Nous supposons également que les deux lois empiriques convergent presque sûrement vers leur loi respective au sens de la ρ -topologie :

$$\lim_{n \rightarrow \infty} P_n \xrightarrow{\rho} P_\pi \quad \lim_{n \rightarrow \infty} P'_n \xrightarrow{\rho} P_{\pi'} \tag{6.37}$$

Alors, pour tout $\epsilon > 0$ et tout $\eta > 1$ l'encadrement suivant est valable :

$$\begin{aligned} \overline{\lim}_{n \rightarrow \infty} \frac{1}{n} \log \mathbb{P}_{\pi'}(P'_n \in V_\epsilon(P_n)) &\leq -D_{KL}(V_{\eta\epsilon}(P_\pi)|P_{\pi'}) \\ -D_{KL}(V_{(2-\eta)\epsilon}(P_\pi)|P_{\pi'}) &\leq \underline{\lim}_{n \rightarrow \infty} \frac{1}{n} \log \mathbb{P}_{\pi'}(P'_n \in V_\epsilon(P_n)) \end{aligned}$$

Démonstration.

Première inégalité

Nous nous donnons un $\eta > 1$ et montrons que nous avons :

$$\overline{\lim}_{n \rightarrow \infty} \frac{1}{n} \log \mathbb{P}_{\pi'}(P'_n \in V_\epsilon(P_n)) \leq -D_{KL}(V_{\eta\epsilon}(P_\pi)|P_{\pi'})$$

D'abord, nous utilisons la convergence presque sûre $P_n \xrightarrow{\rho} P_\pi$ pour obtenir un $n_\pi > 0$ tel que :

$$\forall n \geq n_\pi, d(P_n, P_\pi)_\infty \leq (1 - \frac{1}{\eta})\epsilon \leq (\eta - 1)\epsilon \tag{6.38}$$

Il s'ensuit en jouant sur l'inégalité triangulaire (sous d) :

$$V_\epsilon(P_n) \subset V_{\eta\epsilon}(P_\pi) \quad (6.39)$$

et donc :

$$\frac{1}{n} \log \mathbb{P}_{\pi'} (P'_n \in V_\epsilon(P_n)) \leq \frac{1}{n} \log \mathbb{P}_{\pi'} (P'_n \in V_{\eta\epsilon}(P_\pi))$$

En appliquant le Lemme 6 on sait (en particulier) que $V_{\eta\epsilon}(P_\pi)$ est complètement convexe. Nous vérifions donc l'hypothèse du Lemme 5 lequel nous permet de conclure :

$$\overline{\lim}_{n \rightarrow \infty} \frac{1}{n} \log \mathbb{P}_{\pi'} (P'_n \in V_\epsilon(P_n)) \leq -D_{KL}(V_{\eta\epsilon}(P_\pi) | P_{\pi'})$$

Seconde inégalité

De la même manière, en utilisant l'équation 6.38, on voit que pour $n \geq n_\pi$:

$$V_{(2-\eta)\epsilon}(P_\pi) \subset V_\epsilon(P_n) \quad (6.40)$$

ce qui implique :

$$\frac{1}{n} \log \mathbb{P}_{\pi'} (P'_n \in V_{(2-\eta)\epsilon}(P_\pi)) \leq \frac{1}{n} \log \mathbb{P}_{\pi'} (P'_n \in V_\epsilon(P_n))$$

On remarque alors que $V_{(2-\eta)\epsilon}(P_\pi)$ est un ouvert pour la topologie ρ par définition. Aussi, il s'agit également d'un ouvert pour la topologie τ en application du Lemme 4. Au final, nous appliquons le Lemme 7 pour obtenir :

$$-D_{KL}(V_{(2-\eta)\epsilon}(P_\pi) | P_{\pi'}) \leq \frac{1}{n} \log \mathbb{P}_{\pi'} (P'_n \in V_\epsilon(P_n))$$

ce qui conclut la preuve. □

Remarque 53.

Comme η est arbitraire, cela conduit à définir $U_\epsilon(P_\pi) = \bigcup_\eta V_{(2-\eta)\epsilon}(P_\pi)$. Évidemment, $U_\epsilon(P_\pi)$, comme espace ouvert inclus dans $V_\epsilon(P_\pi)$, est inclus dans $V_\epsilon(P_\pi)^\circ$. De surcroît, tout P dans $V_\epsilon(P_\pi)^\circ$ appartient en particulier à $V_\epsilon(P_\pi)$ de sorte que $d(P, P_\pi) = \alpha < \beta < \epsilon$ et $P \in V_{(2-\frac{\beta}{\epsilon})\epsilon}(P_\pi)$. Au final, $U_\epsilon(P_\pi) = V_\epsilon(P_\pi)^\circ = V_\epsilon(P_\pi)$ puisque $V_\epsilon(P_\pi)$ est déjà un ouvert de la ρ -topologie et donc de la τ -topologie (Lemme 4).

6.4.4 Analyse de la statistique de test t_n

Un premier résultat sur la statistique de test t_n touche à la séparabilité des deux hypothèses : $\{P_{\pi^+}\}$ et Ω^\times . La proposition suivante montre que les deux sous-ensembles de Λ sont séparables pour la ρ -topologie.

Propriété 45. [Séparabilité]

Dès lors que $l_1 > l_0$, il existe un $\epsilon > 0$ tel que :

$$\Omega^\times \cap V_\epsilon(P_{\pi^+}) = \emptyset \quad (6.41)$$

Démonstration.

Supposons que pour un $\epsilon > 0$, $P \in V_\epsilon(P_{\pi^+})$ alors :

$$\begin{aligned} \int_S \pi^\times dP &= \int_S \pi^\times (dP - \pi^+ + \pi^+) \\ &= l_0 + \int_S \pi^\times (dP - \pi^+) \\ &\leq l_0 + \epsilon \|\pi^\times\|_\infty \end{aligned}$$

Or, pour ϵ suffisamment petit, cette dernière quantité est strictement inférieure à l_1 ce qui permet de conclure la preuve :

$$P \in V_\epsilon(p_{\pi^+}) \implies P \notin \Omega^\times$$

□

Grâce à la ρ -topologie sur Λ nous pouvons maintenant formellement définir la convergence que nous avons supposée " $P_n \rightarrow P_\pi$ en un certain sens" dans la section précédente :

$$P_n \xrightarrow[n \rightarrow \infty]{\rho} P_\pi. \quad (6.42)$$

Dans le reste de ce chapitre, cette convergence est supposée acquise. Par ailleurs, nous supposons également que les deux densités f et g sont bornées (cela implique que π^\times et π^+ le sont aussi).

Nous allons étudier t_n sous chacune des deux hypothèses afin de calibrer le seuil. Plus spécifiquement, nous allons suivre la trame de la deuxième section de [45] dédiée à la pente de Bahadur, trame qui est également suivie dans [82] ou encore dans [17] qui l'adapte pour des données corrompues.

Dans cet article, les auteurs définissent $G_n(t) = \mathbb{P}_0(t_n < t)$ la probabilité sous H_0 d'être en dessous d'un seuil t (une bonne chose donc) avant de regarder son complémentaire lorsque le seuil t est la valeur $t_n(s)$ observée elle-même. Cette queue de distribution du test est une variable aléatoire notée L_n et définie par :

Définition 42 (Queue de distribution d'un test).

$$L_n(s) = 1 - G_n(t_n(s)) = \mathbb{P}_0(t_n \geq t_n(s)) \quad (6.43)$$

L_n représente la probabilité, sous H_0 de générer un t_n supérieur au $t_n(s)$ observé (une mauvaise chose donc, qu'il s'agit de limiter). Comme mentionné dans l'article, "plus L_n est petit et plus intenable est l'hypothèse H_0 au vu des observations". Ils étudient alors le comportement de L_n sous chacune des hypothèses. Nous adoptons cette méthode au test t_n défini précédemment.

Convergence sous H_1

On se place sous l'hypothèse $H_1 : \pi \in \Omega^\times$. Nous montrons déjà que nous disposons de la convergence presque sûre de la statistique de test.

Propriété 46 (Convergence sous H_1).

Sous $H_1 : \pi \in \Omega^\times$,

$$t_n \xrightarrow[n \rightarrow \infty]{a.s.} \bar{L}(\pi^\times, \pi) \geq l_1$$

Démonstration.

Presque sûrement, $P_n \xrightarrow[n \rightarrow \infty]{\rho} P_\pi \in \Omega^\times$.

Comme $\pi^\times : S \rightarrow \mathbb{R}$ est une fonction bornée \mathcal{B} -mesurable la convergence implique en particulier

$$\int_S \pi^\times dP_n \xrightarrow[n \rightarrow \infty]{a.s.} \int_S \pi^\times \pi = \bar{L}(\pi^\times, \pi).$$

Par ailleurs, comme $\pi \in \Omega^\times$, $\bar{L}(\pi^\times, \pi) \geq l_1$. □

Dans le cas où $\pi = \pi^\times$, la limite de la statistique de test est de la forme $\int_S \pi^2$ tout en constituant une estimation de la probabilité d'une collision. Cette propriété motive les deux remarques ci-dessous afin de mieux comprendre le sens de t_n . De fait, nous montrons dans la remarque 55 que les réalisations d'une variable aléatoire sous $\int_S \pi^2$ concentrent leurs valeurs autour du mode de π et donc à l'endroit où une "collision" est la plus probable.

Remarque 54. [Réalizations sous π^2]

π étant une densité bornée sur $S = [0, 1]^2$, nous souhaitons simuler sous P_{π^2} suivant la densité $\frac{\pi^2}{\int_S \pi^2}$. On remarque que l'on a toujours $\pi^2 \leq \|\pi\|_\infty \pi$. Aussi, en appliquant une méthode de rejet classique, on simule la variable aléatoire

$$M = \left(\frac{\pi^2(X)}{\int_S \pi^2}, \mathbb{U} \frac{\|\pi\|_\infty}{\int_S \pi^2} \pi(X) \right) \quad (6.44)$$

avec $X \sim P_\pi$ et \mathbb{U} la loi uniforme sur S .

La méthode de rejet nous permet d'affirmer que si l'on garde M uniquement lorsque $M_1 \geq M_2$ alors M_1 suivra la loi P_{π^2} : on obtient ainsi un moyen de simuler sous P_{π^2} à partir du moment où l'on sait simuler sous P_π .

De plus, dans notre application particulière, la méthode se simplifie. Ainsi :

$$\begin{aligned} \frac{\pi^2(X)}{\int_S \pi^2} &\geq \mathbb{U} \frac{\|\pi\|_\infty}{\int_S \pi^2} \pi(X) \\ \iff \pi(X) &\geq \mathbb{U} \|\pi\|_\infty \end{aligned}$$

Au final, il s'agit donc simplement de simuler indépendamment $M = (X, U) \sim P_\pi \otimes \mathbb{U}$ et de garder X si et seulement si $\pi(X) \geq U \|\pi\|_\infty$.

Remarque 55.

D'après la conclusion de la remarque 54, une loi sous P_{π^2} concentrera ses valeurs autour du mode de π . De fait, plus $\pi(X)$ est grand et plus M a des chances d'être conservé après l'étape de rejet.

Divergence sous H_0

Nous nous plaçons maintenant sous l'hypothèse complémentaire $H_0 : \pi = \pi^+$. Nous avons alors la proposition symétrique de convergence.

Propriété 47 (Convergence sous H_0).

Sous $H_0 : \pi = \pi^+$,

$$t_n \xrightarrow[n \rightarrow \infty]{a.s.} \bar{L}(\pi^\times, \pi^+) = l_0$$

Nous souhaitons estimer la probabilité asymptotique, sous H_0 d'être proche de l_1 . Un résultat de grandes déviations est énoncé ci-après et montre que l'événement est exponentiellement rare en un certain sens.

Nous commençons par une estimation de la probabilité que $t_n(s)$ avec s tiré sous $H_0 : \pi = \pi^+$ dépasse un seuil fixe t . Pour cela nous introduisons la variable aléatoire Y définie par

$$Y = \pi^\times(X) \tag{6.45}$$

avec X sous P_{π^+} .

Nous introduisons également le logarithme de la fonction génératrice des moments de Y à savoir :

$$\phi_0(t) = \log \mathbb{E}_0(\exp(tY)) = \log \left(\int_S \exp(t\pi^\times(s)) \pi^+(s) ds \right) \tag{6.46}$$

que nous supposons fini pour $|t|$ suffisamment petit ; les marges étant bornées, l'hypothèse est réalisée.

Puis, de ϕ nous extrayons sa transformée de Legendre

$$I(t) = \sup_{x \in \mathbb{R}} \{tx - \phi(t)\}. \tag{6.47}$$

Enfin, nous introduisons le sous-espace de Λ dans lequel $P_n(s)$ doit rester pour que nous ayons $t_n(s) \geq t$

$$\Omega_t = \{P \in \Lambda / L(P, \pi^\times) \geq t\}. \tag{6.48}$$

Grâce à ces objets nous pouvons maintenant écrire le théorème 14 de déviation de t_n sous H_0 .

Théorème 14 (Cramér).

Pour tout $t \geq l_0$, nous avons

$$\lim_{n \rightarrow \infty} \frac{1}{n} \log (\mathbb{P}_0(t_n \geq t)) = -I(t) = -D_{KL}(\Omega_t | P_{\pi^+}). \tag{6.49}$$

Démonstration.

Récrivons t_n en utilisant Y :

$$\begin{aligned} \mathbb{P}_0(t_n \geq t) &= \mathbb{P}_0 \left(\int_S \pi^\times(s) dP_n(s) \geq t \right) \\ &= \mathbb{P}_0 \left(\sum_{i=1}^n \pi^\times(X_i) \geq nt \right) \\ &= \mathbb{P}_0 \left(\sum_{i=1}^n Y_i \geq nt \right) \end{aligned}$$

Une application directe du théorème de Cramér's nous donne alors :

$$\lim_{n \rightarrow \infty} \frac{1}{n} \log \left(\mathbb{P}_0 \left(\sum_{i=1}^n Y_i \geq nt \right) \right) = -I(t) = -D_{KL}(\Omega_t | P_{\pi^+}) \quad (6.50)$$

ce qui conclut la preuve. \square

Le théorème précédent fournit une estimation de la rareté sous H_0 de l'événement " t_n supérieur à une valeur fixe t ". Ceci ne suffit pas. En effet, il est nécessaire de contrôler la probabilité, sous H_0 , d'avoir $t_n > t_n(s)$ lorsque s est une séquence de réalisations sous H_1 . Il s'agit en effet de l'événement : "nous ne rejetons pas H_0 alors que nous le devrions". Ce contrôle repose sur l'estimation de la pente de Bahadur (voir [7]) comme nous le verrons en conclusion.

En combinant la convergence sous H_1 de la proposition 46 et le contrôle de la déviation sous H_0 du théorème 14 nous obtenons l'estimation mentionnée comme précisé dans le théorème 15.

Théorème 15 (Pente de Bahadur).

Une séquence entière s étant tirée sous H_1 via une probabilité $P_\pi \in \Omega^\times$ nous pouvons contrôler la limite de la variable aléatoire $L_n(s)$ sous un $s' \sim H_0 : P_{\pi^+}$. De fait :

$$\lim_{n \rightarrow \infty} \frac{1}{n} \log [\mathbb{P}_0(t_n(s') \geq t_n(s))] = -I(l) = -D_{KL}(\Omega_l | P_{\pi^+}) \leq -I(l_1) = -D_{KL}(\Omega_{l_1} | P_{\pi^+}) \quad (6.51)$$

avec $l = \bar{L}(\pi^\times, \pi)$.

Démonstration.

s étant issu de P_π , nous avons presque sûrement, comme énoncé dans la proposition 46 :

$$t_n(s) \xrightarrow[n \rightarrow \infty]{a.s.} l \quad (6.52)$$

avec, comme $P_\pi \in \Omega^\times$, $l \geq l_1 > l_0$. Nous pouvons donc sélectionner un δ tel que $\frac{l-l_0}{2} > \delta > 0$. La convergence presque sûre nous fournit l'existence d'un n_0 tel que pour tout $n \geq n_0$,

$$|t_n(s) - l| \leq \delta. \quad (6.53)$$

En particulier, dès que $n \geq n_0$,

$$\mathbb{P}_0(t_n(s') \geq l + \delta) \leq \mathbb{P}_0(t_n(s') \geq t_n(s)) \leq \mathbb{P}_0(t_n(s') \leq l - \delta). \quad (6.54)$$

On applique alors le théorème 14 avec $t = l + \delta \geq l_0$ et $t = l - \delta \geq l_0$ pour obtenir

$$-I(l + \delta) \leq \lim_{n \rightarrow \infty} \frac{1}{n} \log (\mathbb{P}_0(t_n(s') \geq t_n(s))) \leq -I(l - \delta). \quad (6.55)$$

Comme δ est arbitraire, on conclut en invoquant la continuité de I .

Enfin, la borne supérieure provient de l'hypothèse H_1 elle-même : $P_\pi \in \Omega^\times$ requiert $l \geq l_1$. \square

Le théorème 15 peut être lu comme une estimation de la pente de Bahadur (voir [45]); il démontre alors que celle-ci vaut $2I(l_1)$. De plus, la propriété 45 montre que cette pente est strictement positive.

Au final, nous avons donc montré que la pente de Bahadur de notre test t_n basé sur la vraisemblance moyenne est donnée par $2I(l_1)$ et qu'elle est strictement positive. En particulier, nous en déduisons que l'événement "nous ne rejetons pas H_0 alors que nous le devrions" est exponentiellement rare en n .

Conclusion

Ce chapitre est l'occasion de résumer les apports de la thèse ainsi que les perspectives de travail qu'elle ouvre. Nous suivons logiquement l'ordre des chapitres par facilité tant pour l'écriture que pour le lecteur. Nous ajoutons par ailleurs la note (A) pour "apport" lorsque le travail mentionné est original dans la thèse et (P) pour "perspective" lorsque ce même travail ouvre une direction de recherche à creuser.

Dans le second chapitre, nous avons établi un parallèle existant entre deux fonctions de couplage ; l'indépendance et l'indétermination. Chacune est issue de l'optimisation d'un problème de transport discret avec pour fonction de coût respective l'entropie et \mathbb{L}^2 . Intéressant en soi parce qu'il regroupe les calculs de l'obtention de chaque couplage, intéressant également car il construit le problème qui nous intéresse dans la suite, ce chapitre est surtout l'occasion de justifier le choix des deux couplages en particulier. De fait en récrivant les fonctions de coût comme des divergences, les travaux d'Imre Csiszár démontrent qu'elles sont, en un sens, canoniques et, partant, que les couplages qui en résultent le sont aussi (A). Si le premier couplage est extrêmement connu, le second, bien que sous-jacent dans certains critères de corrélation, reste rarement étudié. La fin du chapitre montre que les deux sont en moyenne assez proche (A).

Le troisième chapitre s'intéresse précisément aux propriétés de l'indétermination. Toutes les marges ne pouvant être couplées selon l'indétermination, nous commençons par répondre à une question naturelle : quelle est la proportion des marges éligibles (A) ? Une étude classique est ensuite menée : nous appliquons des critères de corrélation usuels pour déterminer s'ils captent l'indétermination (A). Si le critère de Pearson distingue l'indétermination d'un couplage quelconque, celui de Spearman semble y être agnostique. Une interprétation à affiner est peut-être que l'indétermination ne se lit pas dans l'ordonnement des variables (P). Nous rappelons ensuite les propriétés de Monge vérifiées par chacun des couplages avant d'introduire l'Analyse Relationnelle Mathématique comme un codage dual des matrices de contingence. Ce nouveau codage permet d'interpréter l'indétermination comme un point d'équilibre statistique entre des votes pour et contre. Nous étendons cette égalité dans un cadre probabiliste (A). En récrivant la propriété de Monge qui le caractérise, nous proposons une décomposition du tirage d'indétermination (A). Cette décomposition nous permet de comprendre la faculté qu'a l'indétermination à minimiser les collisions, faculté que nous découvrons par ailleurs en développant la fonction de coût quadratique qu'elle optimise (A).

Le quatrième chapitre commence par appliquer l'indétermination à deux problèmes où une collision est interprétable : celui de l'espion dans lequel une collision est une réussite de ce dernier et celui du partitionnement de tâche dans lequel une collision est la réalisation d'une tâche. Nous montrons que les deux couplages y sont simultanément optimaux pour une stratégie quand l'indétermination est plus performante pour une seconde stratégie (A). Nous n'avons pas pu composer une stratégie ou un problème commun dans lequel l'indétermination est l'optimum et l'indépendance ne l'est pas (P). Une seconde partie est dédiée au partitionnement de graphe. Nous y appliquons deux critères, chacun d'écart à l'un des couplages. Si l'expérience n'est pas nouvelle, elle est éclairée par le résultat du premier chapitre. De plus, nous écartons toute recherche expérimentale et nous plaçons dans un cadre probabiliste : les graphes de Gilbert. Sous ce modèle, nous étudions le comportement général de chaque critère en quantifiant leur proximité (A). Nous terminons en exhibant des graphes pour lesquels les comportements des deux critères sont opposés (A).

Le cinquième chapitre rappelle l'écriture relationnelle et de contingence d'une liste de critères de corrélation. Nous développons une étude du critère de Janson-Végélius d'abord expérimentale puis théorique ; nous montrons notamment une propension à se rapprocher de zéro lorsque le nombre de modalités augmente qui suppose que nous ne concluons pas trop facilement à une absence de corrélation (A). Tous les critères mentionnés comportent au numérateur un produit scalaire commun sous leur écriture relationnelle. Nous utilisons une décomposition des matrices relationnelles en une matrice diagonale et une permutation et montrons que la corrélation est entièrement contenue dans la permutation. Nous étudions l'impact de cette permutation sur la valeur du produit scalaire et proposons à marges fixées la loi de probabilité de celui-ci (A). Il serait intéressant de transposer la distribution du facteur commun après centrage et normalisation pour obtenir la distribution théorique de chaque critère de corrélation (P). Finalement, en reliant ce facteur et donc la permutation à un problème de transport optimal à valeurs entières, nous proposons un algorithme probabiliste pour la résolution de celui-ci (A). L'algorithme est justifié par le lien positif entre la probabilité d'un candidat et sa performance (A). Un pan de recherche reste ouvert autour d'une étude plus poussée de la méthode à la fois théorique et expérimentale ainsi d'ailleurs qu'une comparaison des performances avec celles existantes (P).

Le sixième et dernier chapitre étend l'indétermination dans le cas continu en y transposant la notion de couplage, le problème de transport avec un coût quadratique, le calcul de la proportion des marges éligibles (A) ainsi enfin que la formule d'indétermination elle-même. Reprenant un article mentionnant une copule d'indétermination, nous montrons que celle-ci dépend des marges couplées et estimons l'ensemble exact des marges qui partagent la même copule d'indétermination (A). Nous estimons également l'écart maximal entre une copule d'indétermination et la copule d'indépendance (A). Nous définissons la vraisemblance moyenne, montrons qu'elle étend la notion de collision du cadre discret et qu'elle est minimisée par l'indétermination (A). En nous appuyant sur cette propriété, nous construisons un test statistique basé sur la vraisemblance moyenne afin de distinguer les deux formes de couplage dans le cas continu (A). Nous poursuivons une étude classique de ce test en calculant sa pente de Bahadur (A). Si la vraisemblance moyenne permet formellement d'étendre les collisions dans le cadre continu, nous ne sommes pas parvenus à interpréter sa minimisation (P). De plus, nous avons proposé un algorithme

pour simuler selon une densité au carré à partir d'une simulation selon une densité et montré une tendance à se rapprocher du mode ; cette tendance mériterait une estimation plus précise qui pourrait éclairer la notion de collision continue (P).

Minimization with respect to divergences and applications

Accepted in Geometric Science of Information, April 2021 ([13]).

Abstract : *We apply divergences to project a prior guess discrete probability law on pq elements towards a subspace defined by fixed margins constraints μ and ν on p and q elements respectively. We justify why the Kullback-Leibler and the Chi-square divergences are two canonical choices based on a 1991 work of Imre Csiszár. Besides we interpret the so called indetermination resulting from the second divergence as a construction to reduce couple matchings. Eventually, we demonstrate how both resulting probabilities arise in two information theory applications : guessing problem and task partitioning where some optimization remains to minimize a divergence projection.*

A.1 Introduction

The present document is interested in the poorly studied *indetermination* (also called *indeterminacy*) as a natural second canonical coupling, the first one being the usual independence. We extract both notions from a projection of a prior uniform guess \mathbb{U} to a subset of the set of probabilities defined by fixed margins. The problem is closely linked with contingency table adjustment exposed by Maurice Fréchet in [38] or [39] as well as by Edwards Deming and Frederick Stephan [30] or by Sir Alan G. Wilson [98] or [100]. Although we will limit ourselves to the discrete case it can be transposed in a continuous domain (see for example [90] in the same proceeding as the present paper) leading to the optimal transport theory for which an overview and recent breakthroughs are presented in [94].

When limited to the discrete case, a work of Csiszár ([29]) shows that two projection functions set apart from all the others as they are the only ones to convey "natural" properties. As we shall see, the first one leads to independence while the second one, since canonic, motivates our interest for the *indetermination* coupling it generates.

We gather here a unified construction for both couplings and show *indetermination* comes with a stunning property of couple matchings minimization that one could have expected to happen under independence. Eventually this property is applied to two concrete problems : guessing and task partitioning.

A.2 Projection using divergences

Consider two given discrete probabilities $\mu = \mu_0, \dots, \mu_p$ on p elements and $\nu = \nu_0, \dots, \nu_q$ on q elements with non zero values (it amounts to reduce p and q). We associate to them the subspace $\mathcal{L}_{\mu, \nu}$ of probability laws π on pq elements where :

$$\begin{aligned}\pi_{u,\cdot} &= \sum_{v=1}^q \pi_{u,v} = \mu_u, \forall 1 \leq u \leq p, \\ \pi_{\cdot,v} &= \sum_{u=1}^p \pi_{u,v} = \nu_v, \forall 1 \leq v \leq q,\end{aligned}$$

namely, μ and ν are the margins of π . $\mathcal{L}_{\mu, \nu}$ is thus defined by a set of linear constraints on the components of π where π is a probability law living inside the simplex S_{pq} .

As spotted in [11], we can interpret discrete transport problems related to contingency table adjustments, as projections of a prior guess, typically the uniform law $\mathbb{U} = \frac{1}{pq}, \dots, \frac{1}{pq}$, to our just defined subspace $\mathcal{L}_{\mu, \nu}$.

It happens that the projections of a prior guess that we fix equal to \mathbb{U} to a subspace \mathcal{L} of S_{pq} defined by a set of linear constraints was deeply studied in [29]. Formally, we are minimizing a divergence (see Definition 43) between \mathbb{U} and any element of \mathcal{L} as exposed in Problem 19.

Définition 43 (Divergence with functional parameter ϕ also called f-divergence).
Given, a positive function $\phi : \mathbb{R}^+ \rightarrow \mathbb{R}^+$, strictly convex at 1 with $\phi(1) = 0$, and given two discrete probabilities m and n on pq elements, we define the divergence D_ϕ as :

$$D_\phi(m|n) = \sum_{u=1}^p \sum_{v=1}^q n_{u,v} \phi \left(\frac{m_{u,v}}{n_{u,v}} \right) = \mathbb{E}_n \left[\phi \left(\frac{m_{u,v}}{n_{u,v}} \right) \right].$$

Additionally we set $0 * \phi(x) = 0$ for all $x \in \mathbb{R}^+$.

Problème 19 (Generic Projection Problem).

Given a subspace \mathcal{L} of S_{pq} defined by a set of linear constraints, the projection problem of the uniform law \mathbb{U} to \mathcal{L} using a divergence D_ϕ as defined in Definition 43, states as follows :

$$\min_{\pi \in \mathcal{L}} D_\phi(\pi | \mathbb{U}).$$

Remarque 56.

As mentioned by an anonymous reviewer, since \mathcal{L} is a m -affine subspace and D_ϕ is a f -divergence, the unicity of the optimal is stated from information geometry theory (see [78]).

A key result of the paper [29] is that, provided we add some hypotheses on the cost function we can limit ourselves to two divergences. We do not gather here those hypotheses, but one can find them obviously in the quoted paper as well as summarized

in [11]. Eventually, we obtain the two eligible divergences by setting $\phi : x \mapsto x \log(x) - x + 1$ leading to the usual Kullback-Leibler divergence :

$$D_{KL}(\pi | \mathbb{U}) = \sum_{u=1}^p \sum_{v=1}^q \pi_{u,v} \log(pq\pi_{u,v}); \quad (\text{A.1})$$

and thus to the transport problem using entropy exposed in [98] or with $\phi : x \mapsto (x - 1)^2$ leading to the so-called Pearson Chi-square divergence function :

$$D_2(\pi | \mathbb{U}) = pq \sum_{u=1}^p \sum_{v=1}^q \left(\pi_{u,v} - \frac{1}{pq} \right)^2 \quad (\text{A.2})$$

hence to the minimal transport problem exposed in [89].

Typical computations, using Karush–Kuhn–Tucker, provide for the two divergences, a closed-form expression of the projections that we summarize in the figure ?? below where π^\times is the *independence* coupling of the two margins :

$$\pi^\times(u, v) = \mu_u \nu_v; \quad (\text{A.3})$$

and π^+ is the so-called *indetermination* (or *indeterminacy*) coupling :

$$\pi^+(u, v) = \frac{\mu_u}{q} + \frac{\nu_v}{p} - \frac{1}{pq}. \quad (\text{A.4})$$

At this point we must notice that π^+ is not always defined since it may contain negative values. We add a hypothesis on the margins μ and ν to get rid of this problem :

$$p \min_u \mu_u + q \min_v \nu_v \geq 1. \quad (\text{A.5})$$

Regarding Equation A.5, a typical modification of any couple of margins to deduce eligible margins can be found in [11] while the probability to draw adapted margins if μ and ν are uniformly drawn on S_p and S_q respectively is provided in this report.

Using Mathematical Relational Analysis notations, one can show that π^+ leads to an equality between a "for" vs "against" notion which justifies the name *indetermination* historically coined in [64]. Besides, the + notation we chose in [10] comes from the Full-Monge property (see for instance the survey [19]) that the associated contingency matrix respects. Eventually, we can also extend the notion of *indeterminacy* in a continuous space and estimate an associated copula as exposed in [12]. In the present paper we shall not detail further the properties π^+ conveys and refer the interested reader to the aforementioned articles.

A.3 Mimisation of couple matchings

We insist on a property expressed at the very beginning of the *indetermination* construction. Let us say π is a probability law on pq elements. If we independently draw two occurrences of $W = (U, V) \sim \pi$, the probability of getting a matching for a couple, that is to say $U_1 = U_2$ and $V_1 = V_2$ simultaneously, is given by $\pi_{U_1, V_1} \times \pi_{U_2, V_2} = \pi_{U_1, V_1}^2$.

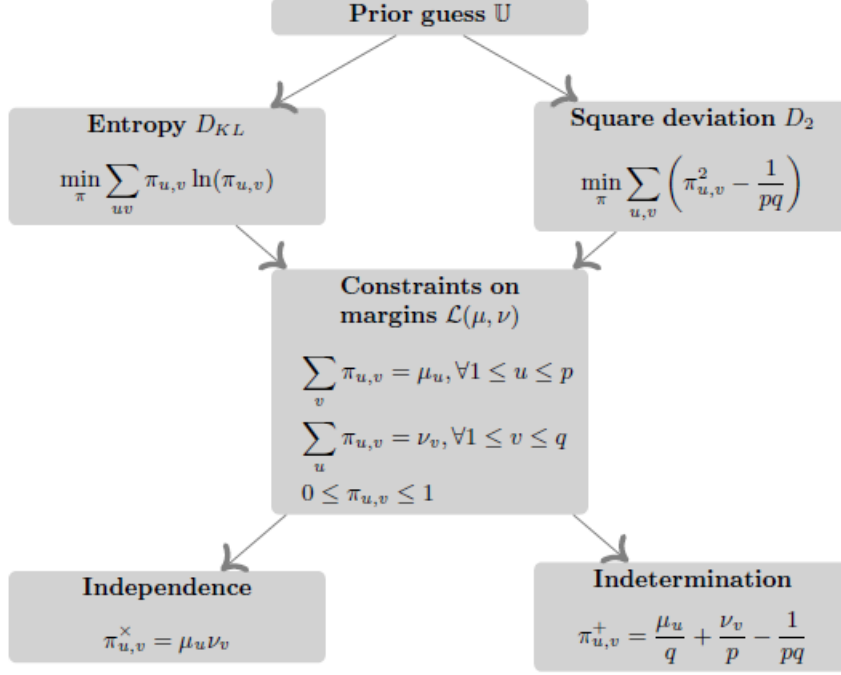


FIGURE A.1 – Summary of a symmetric construction

Besides, it turns out that

$$D_2(\pi | \mathbb{U}) = pq \sum_{u=1}^p \sum_{v=1}^q \left(\pi_{u,v} - \frac{1}{pq} \right)^2 \quad (\text{A.6})$$

$$= pq \sum_{u=1}^p \sum_{v=1}^q (\pi_{u,v})^2 - 1 \quad (\text{A.7})$$

so that the shift introduced by \mathbb{U} vanishes and minimizing $D_2(\pi | \mathbb{U})$ amounts to minimize the probability of a couple matching.

This property is illustrated by a decomposition of π^+ whose proof is available in [11] and which leads to a method for drawing under π^+ . We first set u_0 such that it corresponds to the smallest line meaning μ_{u_0} is minimal. Then, quoting $\forall u, \delta_u = \mu_u - \mu_{u_0}$, we can show that the following drawing realizes *indetermination* :

1. We draw u through μ .
2. We roll a loaded dice with the Bernoulli's skew : $I = \text{Be} \left(\frac{\delta_u}{\mu_u} \right)$.
3. If we get 0, we draw v under the distribution $(\pi_{u_0,1}^+, \dots, \pi_{u_0,q}^+)$ that we normalize to 1, else, we draw v uniformly hence under $(\frac{1}{q}, \dots, \frac{1}{q})$.

The method this coupling uses to minimize couple matchings reads as follows. If U_1 is frequent then μ_{U_1} is high, hence δ_{U_1} is high, this means that I will often draw 1 and that v will often be uniform preventing $V_1 = V_2$ as much as possible. On the opposite, if

U_1 is rare (typically $U_1 = u_0$) then $U_1 = U_2$ is already rare and I will be allowed to often draw 0 so that v will follow $(\pi_{u_0,v}^+)_v$.

Best option of course would be to stay uniform every time, this is the prior guess \mathbb{U} solution but it does not belong to $\mathcal{L}_{\mu,\nu}$. The trick with indetermination is that it hides the disequilibrium on v that is needed to respect margins inside rare lines.

A stunning result is that couple reduction is not ensured by independence but by indetermination; though the difference between both is small. One can show that in the discrete case, the expected \mathbb{L}^2 -difference is $\frac{1}{pq}$ (see [10]). Furthermore the probability of a couple matching with a variable under a second law π' only depends on its margins as stated in Proposition 48.

Propriété 48 (Matchings with another variable).

Let us suppose that π' is a second probability law on pq elements with margins μ' and ν' . We give ourselves a first couple $W = (U_1, U_2)$ drawn under π^+ and a second one $W' = (U'_1, U'_2)$ independently drawn under π' . Then the probability of couple matching only depends on the margins and is given by :

$$\mathbb{P}_{\pi^+ \times \pi'}(W = W') = \sum_{u=1}^p \left(\frac{\mu_u \mu'_u}{q} \right) + \sum_{v=1}^q \left(\frac{\nu_v \nu'_v}{p} \right) - \frac{1}{pq}. \quad (\text{A.8})$$

Démonstration. The result is due to the structure of π^+ . Formally :

$$\begin{aligned} \mathbb{P}_{\pi^+ \times \pi'}(W = W') &= \sum_{(u,v)} \pi_{u,v}^+ \pi'_{u,v} \\ &= \sum_{(u,v)} \left(\frac{\mu_u}{q} + \frac{\nu_v}{p} - \frac{1}{pq} \right) \pi'_{u,v} \\ &= \sum_{u=1}^p \left(\frac{\mu_u \mu'_u}{q} \right) + \sum_{v=1}^q \left(\frac{\nu_v \nu'_v}{p} \right) - \frac{1}{pq} \end{aligned}$$

□

In particular, the last proposition 48 shows that any $\pi' \in \mathcal{L}_{\mu,\nu}$ has the same probability of a couple matching with π^+ than π^+ itself. The next section leverages on that ability to reduce couple matching in two applications.

A.4 Applications

A.4.1 Guessing problem

In cryptography, a message u in a finite alphabet \mathcal{U} of size p is typically sent from Alice to Bob while a spy whose name is Charlie tries to decode it. A common strategy for Alice to communicate efficiently and secretly with Bob consists in encoding the message using a couple of keys (public, private) for each character or a symmetric encryption which only requires one shared key between Alice and Bob. The literature concerned with the encryption method to choose according to the situation is diverse, the most-used standard is Advanced Encryption Standard described in various articles. Possibly,

Charlie observes an encrypted message V in a second finite alphabet \mathcal{V} of size q which is a function of the message u . We choose to gather in V any additional data Charlie may observe (the channel used, the sender's location, ...). In the end U and V are correlated but not always in a deterministic link.

Related to the cryptography situation, the guessing problem quoted hereafter as Problem 20 was first introduced in the article [69] and extended in [59] to take V into account. While in cryptography Charlie tries to decode a sequence of messages, the guessing problem focuses on decoding a unique message.

Let us formalize the situation : Charlie chooses its strategy (see Definition 44) according to the value taken by the observed second variable V : he typically adapts himself to the conveyed encryption. The probability law of the couple (U, V) is quoted $\mathbb{P}_{U,V} = \pi$ while its margins are $\mathbb{P}_U = \mu$ and $\mathbb{P}_V = \nu$ with no zero values.

Problème 20 (Original Guessing Problem or Spy Problem).

When Alice sends a message $U = u$ to Bob, the spy Charlie observes a variable $V = v$ and must find out the value u of the realization. He has access to a sequence of formatted questions for any guess \tilde{u} he may have : "Does u equal \tilde{u} ?" for which the binary answer is limited to "yes/no".

Définition 44 (Strategy).

For any value v , a strategy $\sigma_v = S|v$ of Charlie is defined by an order on \mathcal{U} representing the first try, the second and so on until number p . It can be deterministic or random : we quote $\mathbb{P}_{S|v}$ its probability law. The global strategy quoted σ or S gathers all the marginal strategies S_v and is a random variable dependent only on V (not U) whose probability law is quoted \mathbb{P}_S .

Besides, for a given position $i \in [1, p]$, and an order σ , $\sigma[i]$ is the element in \mathcal{U} corresponding to the i -th try.

The usual performance $G(\sigma, u)$ of a spy consists in counting the number of questions required to eventually obtain a "yes" in Problem 20 when Charlie proposed the order $S = \sigma$ and Alice generated the message $U = u$. For instance, with an alphabet $\mathcal{U} = \{a, b, c, d\}$, if the message is $u = c$ and the strategy σ of the spy consists in the order (b, c, a, d) (meaning he first proposes message b then c, \dots) we have :

$$\begin{aligned} G(\sigma, u) &= \sum_{i=1}^p i \mathbf{1}_{u=\sigma[i]} \\ &= 1 \cdot \mathbf{1}_{u=b} + 2 \cdot \mathbf{1}_{u=c} + 3 \cdot \mathbf{1}_{u=a} + 4 \cdot \mathbf{1}_{u=d} \\ &= 2 \cdot \mathbf{1}_{u=c} \\ &= 2 \end{aligned}$$

Eventually in [69], they compute the ρ -moment of G as an estimation of the efficiency of a strategy S :

$$\|G(S, U)\|_{\rho} = \left[\mathbb{E}_{(S,U,V) \sim \mathbb{P}_{S,U,V}} (G(S, U)^{\rho}) \right]. \quad (\text{A.9})$$

We notice that G actually takes as granted that Charlie has up to p trials which is not adapted in case Alice sends a sequence of messages due to a lack of time. Let us move away from the literature and measure Charlie's performance by its probability to find out

after one trial the message u Alice sent. It is a reasonable measure as, if a sequence of messages is sent, he may have to jump from one to the following after only one trial. What actually matters (as mentioned by an anonymous reviewer) is the number of attempts k Charlie can try between two messages from Alice. It could be captured with the one-shot performance M using a trick and supposing that Alice repeats each message k times.

Définition 45 (one-shot performance).

For a given strategy S (which can depend on V with a non deterministic link), we define the following performance measure as the probability to find out the value u after one trial, formally :

$$M(S, U, V) = M(S, \pi) = \mathbb{P}_{(S,U,V)}(S[1] = U).$$

We suppose as for the original optimal strategy that the spy has access to the distribution $\pi = \mathbb{P}_{U,V}$. We can imagine he previously observed the non-encrypted messages in a preliminary step.

Two strategies immediately stand out :

1. S_{max} : systematically returns at v fixed (observed by hypothesis), the u associated with the maximal probability on the margin $\mathbb{P}_{U|V=v}$;
2. S_{margin} : returns at v fixed a random realization in \mathcal{U} under the law $\mathbb{P}_{U|V=v}$.

Similarly as in see [69] where they prove S_{max} is the best strategy in case the performance measure is G , we can show it also maximizes the one-shot performance. Actually, M rewrites :

$$M(S, U, V) = \sum_{v=1}^q \nu_v \mathbb{P}_{(S_v,U) \sim \mathbb{P}_{S_v} \otimes \mathbb{P}_{U|V=v}}(S_v[1] = U) \quad (\text{A.10})$$

$$= \sum_{u=1}^p \sum_{v=1}^q \pi_{u,v} \mathbb{P}_{S_v}(S_v[1] = u). \quad (\text{A.11})$$

Since \mathbb{P}_{S_v} only depends on v , M is maximal under S_{max} , when $\mathbb{P}_{S_v} = \delta_{u_v}$ where $u_v = \operatorname{argmax}_u \pi_{u,v}$ so that :

$$M(S, U, V) \geq M(S_{max}, U, V) = \sum_{v=1}^q \pi_{u_v,v}. \quad (\text{A.12})$$

Eventually, we quote $u_1 = \operatorname{argmax}_u \mu_u$ and notice that $\sum_{v=1}^q \pi_{u_v,v} \geq \mu_{u_1}$ leading to the proposition 49.

Propriété 49 (Charlie's best performance).

We suppose that the margins μ and ν are fixed. Then, for any coupling probability π between messages U and ciphers V , the best one-shot performance Charlie can perform always happens under S_{max} . Furthermore it admits a fixed lower-bound μ_{u_1} independent on π ; to summarize :

$$M(S, U, V) \geq M(S_{max}, U, V) = \sum_{v=1}^q \pi_{u_v,v} \geq \mu_{u_1}. \quad (\text{A.13})$$

Let us suppose, commendable task if any, that Alice wants to minimize Charlie's best one-shot performance. We also suppose that the margins μ on U and ν on V are fixed. It is a common hypothesis : the alphabet \mathcal{U} in which the messages are composed typically respects a distribution on letters ; variable V on its own, if it represents frequencies for instance may have to satisfy occupation weights on each channel. Eventually, Alice can only leverage on the coupling between U and V .

Precisely, let us now compute the corresponding value for our two couplings, both are optimal (for Alice).

$$M^\times = M(S_{max}, \pi^\times) = \mu_{u_1} \qquad M^+ = M(S_{max}, \pi^+) = \mu_{u_1}$$

Regarding the second strategy S_{margin} we know it is less efficient for Charlie in term of one-shot performance. Yet, it is by far harder to cope with for the sender who cannot easily prevent random conflicts. Consequently we come back to the reduction of couple matchings (here a success for Charlie), whose "*indetermination* coupling", we know, prevents us against. Let us unfold this remark hereafter.

Replacing \mathbb{P}_{S_v} by its value under the second strategy in Equation A.11 allows us to estimate the one-shot performance of S_{margin} which is given by :

$$M(S_{margin}, U, V) = \sum_{u \in \mathcal{U}} \sum_{v \in \mathcal{V}} \nu_v (\pi_{u|V=v})^2. \tag{A.14}$$

Concerning the strategy S_{margin} we have the two bounds :

$$\frac{\|\pi\|_2^2}{\min_{v \in \mathcal{V}} \nu_v} \geq M(S_{margin}, U, V) \geq \frac{\|\pi\|_2^2}{\max_{v \in \mathcal{V}} \nu_v}, \tag{A.15}$$

with

$$\|\pi\|_2 = \sqrt{\sum_{u \in \mathcal{U}} \sum_{v \in \mathcal{V}} (\pi_{u,v})^2}.$$

Depending on ν the bounds are more or less tight. If ν_v goes to the uniform, the situation becomes similar to S_{max} : both couplings are equal and optimal.

In any case, Equation A.15 shows that studying the guessing problem brings us back to Problem 19 associated with cost D_2 whose solution is given by the *indetermination* coupling of the margins. It guarantees an efficient reduction of conflicts (see section A.3) and eventually a weak one-shot performance under S_{margin} as expressed in Equation (A.15) and an optimal one under S_{max} .

A.4.2 Task Partitioning

Task partitioning problem is originally introduced in [18] where the authors provide a lower bound on the moment of the number of tasks to perform. Let us follow the gathering work of [59] where they also coin a generalized task partitioning problem basically adapting it as a special case of the guessing problem.

Formally, we begin with the original problem of tasks partitioning : a finite set \mathcal{U} of p tasks is given together with an integer $q \leq p$. The problem consists in creating a partition $\mathcal{A} = (\mathcal{A}_1, \dots, \mathcal{A}_q)$ of \mathcal{U} in q classes to minimize the number of tasks to perform, knowing

that if one needs to perform a task $u \in \mathcal{A}_i$, it is mandatory to launch simultaneously the whole subset of tasks included within \mathcal{A}_i .

Practically, a task $U = u$ to perform is randomly drawn from \mathcal{U} under a probability distribution $\mathbb{P}_U = \mu$ representing the tasks frequencies. As any task, the task u to perform is assigned to a unique class $\mathcal{A}_{i(u)}$ of the arbitrary partition. Hence, $A(u) = |\mathcal{A}_{i(u)}|$ counts the number of tasks to perform. Precisely, one plays on the partition knowledge to perform, in average, as few tasks as possible.

Similarly to the guessing problem, the performance of a partition \mathcal{A} is estimated through the ρ -moment of the number of tasks $A(U)$ to perform :

$$\mathbb{E}_{U \sim \mathbb{P}_U} [A(U)^\rho]. \quad (\text{A.16})$$

Inspired by the general guessing problem, paper [59] extends the task partitioning problem. Let us introduce here this generalized version, in which we are no longer interested in minimizing the number of tasks to perform but rather in reducing the number of tasks before a selected (or a chosen) task u .

Indeed, in the first version, as soon as u is drawn, an arbitrary rule imposes to perform the whole subset $\mathcal{A}_{i(u)}$ leading to realize $A(u)$ tasks. In the new version, tasks are performed sequentially in $\mathcal{A}_{i(u)}$ according to a global strategy S that can be deterministic or random. Typically, tasks may consist in a flow of signatures which an administration requires while q would be the number of workers dedicated to perform those signatures on incoming documents. A worker can be given the entitlement to perform several signatures, assistants usually do. In that case, the partition encodes the assignments of tasks to workers. When a worker $V = v$ is assigned a document, the depositor waits until the signature. At the same time the worker follows his own strategy S_v to sign his assigned documents, meaning he can always follow the same order leading to a deterministic strategy or change every day leading to a random strategy.

With a global strategy S which gathers the workers' strategies S_v , $\forall 1 \leq v \leq q$ and for a task u to perform, the performance of a partition \mathcal{A} is measured using

$$N_{S,\mathcal{A}}(u) = j / S_{i(u)}[j] = u,$$

which represents the number of tasks performed before the intended task u (u included) by the worker $i[u]$ it is affected to according to partition \mathcal{A} . A lower bound is provided in the paper [59].

Let us now suppose that the keys $1 \leq v \leq q$ are associated with offices that must perform a proportion ν_v of the incoming tasks which still follow a distribution μ . It actually appears as a sensible problem where a manager would have to distribute in advance tasks among teams according to the usual observed distribution of tasks and a list of available teams with their capacities.

Besides, we suppose that each team uses the strategy S_{marge} to perform tasks, meaning they randomly perform one according to their margin theoretical distribution ; for a document signing, they randomly sign one.

From now on, we can rewrite our task partitioning problem under the form of a guessing problem :

- $V = v$, formerly corresponds to a worker, now it represents the information the spy has access to ;

- $U = u$, formerly represents a task to perform, now it represents a sent message ;
- $S = \sigma$, formerly represents the order in which tasks are performed, now it represents the order in which Charlie proposes his guesses.

Under this formalism, we are interested in measuring the probability $M(S, U, V)$ of executing u first as an extended application of the one-shot performance of Definition 45 and we have :

$$M(S, U, V) \geq \frac{\|\pi\|_2^2}{\max_{1 \leq v \leq q} \nu_v} \geq \frac{\|\pi^+\|_2^2}{\max_{1 \leq v \leq q} \nu_v}. \quad (\text{A.17})$$

This inequality provides a lower bound for any distribution of the tasks among the team, no distribution can generate a worst "one-shot probability" of satisfying the intended task. In task partitioning actually, each task u is uniquely associated to a worker $v = i(u)$ so that the random variable representing the worker is deterministic conditionally to U : $\pi_{u,v} = \delta_{v=i(u)}$.

Independence versus Indetermination : basis of two canonical clustering criteria

Submitted to Advances in Data Analysis and Classification, July 2020. Revised March 2021 ([10]).

Abstract : *This paper aims at comparing two coupling approaches as basic layers for building clustering criteria, suited for modularizing and clustering very large networks. We briefly use "optimal transport theory" as a starting point, and a way as well, to derive two canonical couplings : "statistical independence" and "logical indetermination". A symmetric list of properties is provided and notably the so called "Monge's properties", applied to contingency matrices, and justifying the \otimes versus \oplus notation. A study is proposed, highlighting "logical indetermination", because it is, by far, lesser known. Eventually we estimate the average difference between both couplings as the key explanation of their usually close results in network clustering.*

B.1 Introduction

Network clustering (or cliques partitioning of graphs) is a key topic, concerned with a very large dedicated literature. One of the reasons of this status is the recent and power use made by the GAFAM companies about very large networks resulting of modern activities dealing with : *big social networks, cellphone communications networks, high speed financial trading, large IT networks, IOT networks etc...*. This is simultaneously associated with the IT capacity afforded today to store the really huge amounts of data, those activities force us to cope with. The sudden apparition of these big networks gave rise to a renewal of the so-called *graph theoretical domain*, used in that context for different purposes, such as : discovering the latent cliques, clustering the whole network, isolating some key parts of interest within the network, etc. In other words, this massive and raw information contained inside the networks must be analyzed *per se*, and this leads obviously to mandatory techniques, among which networks clustering plays a prominent role, with a lot of practical contextual applications.

In the scientific literature, it appears that many different methods have been dedicated to graphs clustering, one can find in [36] or more recently in [31] a quite interesting

overview on this matter. Most of them use a local criterion based on the number of paths [56], the number of shortest paths [41] or a proportion of present edges [75] which is then aggregated to define a global criterion to optimize. Some methods are based on pure decomposition of a graph, as for instance in [6] where they construct a k -clustering based on a spanning tree by removing $k - 1$ edges, some other existing methods are concerned directly with spectral analysis of graph laplacians, or with mathematical relational analysis (correlation clustering). In addition to this relatively main stream list, some methods rely on the application of very specific mathematical domains as those typically addressed through Mean Field Games theory in [26] or more promising, such as the approach given in [77] where a method based on the evolving of a discrete Ricci curvature flow was proposed.

This method deserves to be briefly exposed : given two measures μ and ν in a space endorsed with a distance d , the Wasserstein's transportation distance $W(\mu, \nu)$ is the minimum total weight to move μ to ν according to d as presented in [79] or [62]. A measure $m_x^{\alpha,p}$ to capture the neighborhood of a node x is formulated in [77]. With d the shortest path in the network, the Ricci's curvature then basically expresses whether x is closer from y using d or W between $m_x^{\alpha,p}$ and $m_y^{\alpha,p}$. Eventually, they are interested in the Ricci Flow (see [47]) which solves a differential equation where the derivative is (almost) this Ricci curvature ; consequently if the neighborhood between x and y is closer than $d(x, y)$, the derivative is strictly less than one and conversely. Iteratively updating a weight to solve the Ricci Flow differential equation, before cutting links greater than a threshold, they deliver results quite similar and even better on the usual experimental networks. Nevertheless, the lack of a canonical choice for the underlying parameters (α, p and the threshold) appears as a limitation similar to the choice of a criterion in the usual Louvain Algorithm. This drawback precisely motivates the present paper. Actually, any comparison with competitive approaches would imply a self adaptive procedure to select the technical parameters.

At that stage two aspects must be differentiated : on the one hand (i) the existence of generic algorithms to optimize a *clustering criterion* as global objective function, or, on the other hand (ii), on the network clustering criteria themselves.

Going back on the first point, (i) concerned with generic algorithms, it is well known that several methods were introduced to fit this purpose and notably the famous Louvain algorithm, whose origin is quite recent [14], and which is recognized as a very good tool by the scientific community. It is originally based upon the optimization, of a global function called *modularity* initially defined in [76] and which enabled the community to compare two clustering on a common basis. In the sequel we shall quote \mathcal{M}^\times this objective function and will show it measures *deviation from statistical independence*.

To fulfill the (ii) objective, the Louvain algorithm has been naturally generalized in [21] where the authors proposed to choose a candidate criterion among a list of global criteria, different from the usual *modularity*. Actually, the modularity, due to a resolution limit first mentioned in [37], has been modified in several articles ([83], [60] or [23]) always motivated by experimental results ; we will not detail further the list of available criteria. In her thesis [25], Patricia Conde-Céspedes, proposed some experiments on usual networks, involving \mathcal{M}^\times plus some others, showing that results may vary from one criterion to another, while being still consistent and interpretable.

In this paper, we will focus, on two network clustering criteria she applied, the original

\mathcal{M}^\times and a second quoted \mathcal{M}^+ which is locally based on a deviation to another coupling function, already latent in a paper of Fréchet [38] and that we shall call *indetermination* or *logical indetermination* (notion introduced by J.F. Marcotorchino in his seminal papers [64] and [63]).

The innovation of this paper can be stated as follows :

- We rely on a work of Csiszar on divergences [29] which assesses that the costs in a projection problem is restricted to Least Square or Entropy. Leveraging on it we show that the two chosen criteria \mathcal{M}^\times and \mathcal{M}^+ precisely result from the optimization of the two corresponding canonical discrete transportation problems.
- We gather known and new properties of the so-called *indetermination*, an equilibrium already applied in the graph clustering domain but never studied *per se* in a more general context. Furthermore, the expected difference between the two canonical coupling functions is shown to be of order $\mathcal{O}(\frac{1}{n^2})$ where n counts the number of nodes.
- We validate some of these findings, by reanalyzing more systematically the behavior of those criteria on the very simple model of Gilbert's graphs. This last item illustrates small expected difference of the previous item and explains the close experimental results found in [21]. Besides it motivates the search of situations where the two criteria side significantly apart as briefly considered in the paper.

The paper is structured as follows. In section B.2 we propose a parallel discovery of two coupling functions (\otimes) and (\oplus) using discrete *optimal transport theory*. In section B.3 is mentioned a list of dual properties related to Monge's matrices and which justify the notation \oplus/\otimes that we propose. Section B.4 deeply studies *indetermination* introducing properties that, to our knowledge, deserve to be put forward with regards to the too poor coverage which is devoted to them in the literature. Finally, Section B.5 gathers a study about the behavior of the criteria based on those coupling functions on the general Gilbert's random network model, quoting a global similarity which illustrates their symmetric construction.

B.2 Parallel discovery of two dual couplings

When we want to couple two marginal laws, the most common and straightforward way to proceed, consists in assuming *independence* and keep on computations. For instance when we use a very classical and usual criterion like the χ^2 index, we are measuring nothing but a deviation to *independence* a natural coupling in for empirical experiments.

Although being the most natural, it is not, by far, the only existing available coupling method ; actually, as introduced by Sklar in [87], any copula function will lead to a coupling function behaving on two cumulative distribution functions. In this document, we link a coupling function to a given optimal transport problem. Hence, to follow a similar approach for *indetermination* coupling, we train ourselves first by extracting *independence* coupling from the optimization of a transport problem and we generalize the principle by applying the same approach to the *indetermination* case, but with a second and different transport problem.

We already introduced the term "coupling function" several times but let us define it formally, since it will be a key notion throughout the document.

Définition 46 (Coupling function).

Given $\mu = \mu_1 \dots \mu_p$ and $\nu = \nu_1 \dots \nu_q$ two discrete probabilities called marginal distributions (or simply margins), we want to define a probability function $\pi = \pi_{u,v}$ $\{1 \leq u \leq p, 1 \leq v \leq q\}$ on the product space. A way for building it up, consists in making happen a coupling function C such that $\pi = C(\mu, \nu)$, satisfying the following constraints :

- (first margin) $C(\mu, \nu)_{u,\cdot} = \sum_{v=1}^q C(\mu, \nu)_{u,v} = \mu_u, \forall 1 \leq u \leq p$
- (second margin) $C(\mu, \nu)_{\cdot,v} = \sum_{u=1}^p C(\mu, \nu)_{u,v} = \nu_v, \forall 1 \leq v \leq q$
- (positivity) $C(\mu, \nu)_{u,v} \geq 0, \forall 1 \leq u \leq p, \forall 1 \leq v \leq q$

Remarque 57.

All coupling functions (or maps) we use will satisfy : $\pi_{u,v} = C(\mu, \nu)_{u,v} = C(\mu_u, \nu_v)$; this illustrates that π value on (u, v) only depends upon the value on the corresponding margins : μ_u and ν_v .

B.2.1 Some few words about Optimal Transport

Looking at Definition 46, we observe that a coupling function behaves as a copula in the discrete domain : acting on margins it derives a probability distribution on the product space.

We can imagine a lot of coupling functions, especially if we do not limit ourselves to Remark 57. The constraints that C has to respect, lead us to cope with some difficulties. This is the reason why we shall choose a systematic approach : minimizing a cost function and observe the link to *optimal transport* definition.

The ad-hoc discrete optimal transport problem we will be dealing with, typically looks like Problem 21, given hereafter (where MKP stands for Monge-Kantorovitch-Problem).

Problème 21 (Discrete Version of MKP).

$$\begin{aligned} \min_{\pi} \quad & \sum_{u=1}^p \sum_{v=1}^q C(\pi(u, v)) \\ \text{subject to :} \quad & \sum_{v=1}^q \pi(u, v) = \mu_u ; \forall u \in \{1, \dots, p\} \\ & \sum_{u=1}^p \pi(u, v) = \nu_v ; \forall v \in \{1, \dots, q\} \\ & \pi(u, v) \geq 0 ; \forall (u, v) \in \{1, \dots, p\} \times \{1, \dots, q\} \end{aligned}$$

The choice of a cost function \mathbf{C} depends upon the applications we want to address. Typically, we expect the global assignment to be as smooth as possible, meaning close to uniform (see both examples in the sequel). A MKP problem is then essentially given by its cost function, while margins (μ, ν) may vary. This is the reason why we shall try to solve it with a model taking the fixed margins as parameters. Let us define now an optimal coupling function C associated to a given MKP problem with fixed margins given as parameters.

Définition 47 (MKP Problem Associated with Coupling function).

For a given MKP problem P , we can define a coupling function C^P by : $C^P(\mu, \nu) = \pi^*(P)$ provided that π^* exists as a unique solution of P with margins μ and ν .

Following Definition 47 we propose the solutions of two discrete optimal transport problems that we shall use in section B.5 : each implies a structured and well-defined criterion, suitable for network clustering.

B.2.2 The Alan Wilson's Entropy Model : role of independence

First introduced by Sir Alan Wilson in 1969 for "Spatial Interaction Modeling" the "Flows Entropy Model" of Alan Wilson, can be found in his various publications : originated in [98] and developed in [99]. A fundamental justification of his approach corresponds to the following contextual situation : in a theoretical system, elements of which do not maintain affinities, it is advisable to determine the distribution of $\pi(u, v)$ (normalized frequency flows), supposing $\pi \geq 0$ which maximizes the entropy of the system under certain constraints. The objective function to be minimized is based upon the *Boltzmann's or Shannon's Entropies* so that the problem should be expressed as follows :

Problème 22 (Unbalanced PSIS).

$$\min_{\pi} - \sum_{u=1}^p \sum_{v=1}^q \pi(u, v) \ln(u, v)$$

In a situation where we have a total absence of information, the minimization of Problem 22 just amounts to satisfy the constraint that the cell values distribution is effectively a probability (i.e. : the sum of positive $\pi(u, v)$ is equal to 1). The solution of this very simple "Program of Spatial Interaction System" (PSIS) is nothing but the uniform law :

$$\pi^*(u, v) = \frac{1}{pq} \tag{B.1}$$

In other words, when we ignore everything about the way the exchanges are built up, it is necessary to use Laplace's principle of "insufficient reason" and to consider that the world trade is uniformly distributed inside the system.

By using margins, let us say information about total exports (origins flows) and total imports (destination flows), degree of disorder of the system can be drastically reduced. Indeed, totals on rows and columns are no longer free, but must satisfy marginal values μ_u and ν_v , fixed by the application as expressed in Problem 23 ; solution of which is given by theorem 16.

Problème 23 (Balanced PSIS).

$$\min_{\pi} - \sum_{u=1}^p \sum_{v=1}^q \pi(u, v) \ln(\pi(u, v))$$

subject to constraints of Problem 21

Théorème 16.

The solution of Problem 23 is $\pi^\times(u, v) = \mu_u \nu_v$.

Hence the coupling function associated to Problem 23 is nothing but "independence" :

$$C^{\text{Problem 23}}(\mu, \nu)_{u,v} = C^\times(\mu, \nu)_{u,v} = (\mu \otimes \nu)_{u,v} = \mu_u \nu_v$$

We skip the proof of theorem 16 as it is similar to the one we will develop for theorem 17 which is less common.

As a conclusion, from the direct maximization of entropy, we get the solution expressed in terms of probability and remark that the associated coupling function is nothing but "independence" (expressed with a \otimes throughout the document).

B.2.3 The minimal trade model : role of indetermination

In the "Minimal Trade Model" (see [89], [64] and [63]), the cost function aims at getting a smooth breakdown of the origins-destinations $\pi(u, v) = \frac{n_{u,v}}{n_{\cdot,\cdot}}$ which explains the term "Minimal Trade". In that case the criterion is a quadratic function measuring squared deviation of the cells values from the "no information" situation (the uniform joint distribution law related to Problem 22). Obviously, in case of free margins, the solution remains the uniform law. However, adding usual pre-conditioned constraints on margins, the least squared problem is Problem 24; solution of which is given by theorem 17.

Problème 24 (Minimal Trade Model).

$$\min_{\pi} \sum_{u,v} \left\{ \pi(u, v) - \frac{1}{pq} \right\}^2$$

subject to constraints of Problem 21

Théorème 17.

The solution of Problem 24 is $\pi^+(u, v) = \frac{\mu_u}{q} + \frac{\nu_v}{p} - \frac{1}{pq}$.

Hence the coupling function associated to Problem 24 is nothing but "indetermination" :

$$C^{\text{Problem 24}}(\mu, \nu)_{u,v} = C^+(\mu, \nu)_{u,v} = (\mu \oplus \nu)_{u,v} = \frac{\mu_u}{q} + \frac{\nu_v}{p} - \frac{1}{pq}$$

A supplementary condition, which is exogenous with regard to the previous model, can be added on the margins (which are, by the way, constant values given *a priori*), this condition (see [64]) is a simple inequality which guarantees the positivity of the frequency Matrix $\pi^*(u, v)$ we are looking for :

$$p \min_u \mu_u + q \min_v \nu_v \geq 1 \tag{B.2}$$

From now on, we shall consider that Condition B.2 applies whatever the breakdown of the μ_u and ν_v is. Notice that in the "Adjustment to Fixed Margins for Contingency Table" case, the associated values n_{uv} must be integers, and therefore returns the problem much more complex to solve, relaxation of this integrity constraint leads formally to the Problem 24.

Remarque 58 (Vanishing bias).

By developing the cost function, we obtain an interesting equality we will reuse later on :

$$\sum_{u,v} \left(\pi(u,v) - \frac{1}{pq} \right)^2 = \sum_{u,v} \pi^2(u,v) - \frac{1}{pq} \quad (\text{B.3})$$

so that the influence of the constant shift $\frac{1}{pq}$ in the squared model disappears.

Démonstration.

The proof we propose comes directly from [89] and [63]. A generalization of the canonic additive form when we relax hypothesis B.2 can be found in this report.

Using equality B.3, the Lagrangian function associated to the previous minimization model can be turned into

$$\begin{aligned} \mathbf{L}(\pi, \lambda, \omega, \theta) &= \sum_{u=1}^p \sum_{v=1}^q \pi^2(u,v) - \sum_{u=1}^p \lambda_u \left(\mu_u - \sum_{v=1}^q \pi(u,v) \right) \\ &- \sum_{v=1}^q \omega_v \left(\nu_v - \sum_{u=1}^p \pi(u,v) \right) - \theta \left(\sum_{u=1}^p \sum_{v=1}^q \pi(u,v) - 1 \right) \end{aligned}$$

Since the function to optimize is a convex one, the solution we are looking for is a minimum so that first order conditions apply and we have the following system of equations.

$$\frac{\partial \mathbf{L}(\pi, \lambda, \omega, \theta)}{\partial \pi(u,v)} = 2\pi(u,v) - \lambda_u - \omega_v - \theta = 0 \quad (\text{B.4})$$

$$\frac{\partial \mathbf{L}(\pi, \lambda, \omega, \theta)}{\partial \lambda_u} = \mu_u - \sum_{v=1}^q \pi(u,v) = 0 \quad (\text{B.5})$$

$$\frac{\partial \mathbf{L}(\pi, \lambda, \omega, \theta)}{\partial \omega_v} = \nu_v - \sum_{u=1}^p \pi(u,v) = 0 \quad (\text{B.6})$$

When supposing $\sum_v \omega_v = 0$ as Lagrange multipliers are defined within a constant near we sum B.4 on v to obtain $2\mu_u = \text{B.5} \ 2 \sum_v \pi(u,v) = q\lambda_u + q\theta$ so that

$$\lambda_u + \theta = \frac{2\mu_u}{q}, \forall u \quad (\text{B.7})$$

From B.6 we get $2\nu_v = \sum_{u=1}^p 2\pi(u,v) = \text{B.4} \ \sum_{u=1}^p \lambda_u + \omega_v + \theta = \text{B.7} \ \sum_{u=1}^p \frac{2}{q}\mu_u + \omega_v = \frac{2}{q}\mu_u + p\omega_v$ so that

$$\omega_v = \frac{2\nu_v}{p} - \frac{2}{pq}, \forall v \quad (\text{B.8})$$

Replacing into B.4 $\lambda_u + \theta$ and ω_v by their value given respectively by B.7 and B.8 we obtain :

$$\pi^*(u,v) = \frac{\mu_u}{q} + \frac{\nu_v}{p} - \frac{1}{pq}, \forall (u,v)$$

Remark, since Condition B.2 applies, the π^* expressed in the previous equation are nonnegative. We will go back to this expression, in the next sections and develop a deeper focus on it, explaining the true meaning of the term "indetermination" and some other consequences. \square

B.2.4 Expected difference between coupling

Both coupling functions are extracted from an optimal transport problem concentrating values around the uniform. Hence differences between them should be small in a certain sense. We provide in this section a measure of their proximity. We evaluate the expected value of a norm between the two couplings under uniform laws. More precisely we suppose the two margins μ and ν follow the Dirichlet's law (basically the uniformity on probability distributions). We remind here the form of that law for our application.

Définition 48 (Dirichlet's Law).

The density of a Dirichlet law \mathcal{D}_p representing a uniform law among probability law on p elements is expressed as follows :

$$f(\mu_1, \dots, \mu_p) \prod_{k=1}^p d\mu_k = \frac{1}{B(p)} \prod_{k=1}^p \mu_k^0 \prod_{k=1}^p d\mu_k = \frac{1}{B(p)} \prod_{k=1}^p d\mu_k$$

where B is the multinomial Beta function.

Having expressed a density function for μ and ν (replace p by q), we apply them two coupling functions C^+ and C^\times . As a distance, we define :

$$\Delta_p = \mathbb{E}_{(\mu, \nu) \sim \mathcal{D}_p \otimes \mathcal{D}_q} \left[\sum_{u=1}^p \sum_{v=1}^q [(\mu \otimes \nu)_{u,v} - (\mu \oplus \nu)_{u,v}]^2 \right]$$

and compute its value through the sequence :

$$\begin{aligned} \Delta_p &= \mathbb{E}_{(\mu, \nu) \sim \mathcal{D}_p \otimes \mathcal{D}_q} \left[\sum_{u=1}^p \sum_{v=1}^q \left[\left(\mu_u - \frac{1}{p} \right) \left(\nu_v - \frac{1}{q} \right) \right]^2 \right] \\ &= \mathbb{E}_{\mu \sim \mathcal{D}_p} \left[\sum_{u=1}^p \left(\mu_u - \frac{1}{p} \right)^2 \right] \mathbb{E}_{\nu \sim \mathcal{D}_q} \left[\sum_{v=1}^q \left(\nu_v - \frac{1}{q} \right)^2 \right] \\ &= pq \mathbb{E}_{\mu \sim \mathcal{D}_p} \left[\left(\mu_1 - \frac{1}{p} \right)^2 \right] \mathbb{E}_{\nu \sim \mathcal{D}_q} \left[\left(\nu_1 - \frac{1}{q} \right)^2 \right] \end{aligned}$$

Now, we notice that we need to compute the variance of \mathcal{D}_p ; as it is a known law, we use the following property :

Propriété 50 (Variance of Dirichlet law).

$$\mathbb{V}_{X \sim \mathcal{D}_p}[X] = \frac{p-1}{p^2(p+1)}$$

Proposition 50 in particular, implies that margins will concentrate their values around $\frac{1}{p}$ and $\frac{1}{q}$ respectively as soon as p or q increases respectively. As we notice that couplings equal each other when any margin is uniform, this should imply that Δ_p converges to 0 if any of the two increases. This is exactly what happens, we have the expression :

$$\Delta_p = \frac{1}{pq} \left(\frac{p-1}{p+1} \cdot \frac{q-1}{q+1} \right) \leq \frac{1}{pq}$$

B.2.5 Structural Justification based upon an axiomatic result of Imre Csiszar

Although it seems arbitrary, our restriction to these two previous coupling functions, is all but a fortuitous decision : in [29], Csiszar actually shows that, provided we verify additional intuitive properties, we must restrict ourselves to use either least square or maximum entropy as canonic "distances" between probability distributions.

Let us rewrite our transport problems in terms of the notations he uses in [29]. We notice that problems 23 and 24 aims at reducing a distance from π to the uniform law (that term actually vanishes in both), where π must satisfy constraints on its margins leading to an eligible space $L_{\mu,\nu}$ inside the simplex S_D , $D > 0$. In the first problem, the distance function is the entropy while in the second it is the norm \mathbb{L}_2 .

A general question is how to adapt a "prior guess" u^0 to verify a list of constraints. Let us say u^0 lives in S_D while the given constraints define a subspace $L \in \mathcal{L}$ (\mathcal{L} is the space of subspaces of S_D tuned by a finite list of affine constraints, see [29] for more details). To formalize it, Csiszar defines a *projection rule* Π as a function whose input is a set $L \in \mathcal{L}$ and which generates a method Π_L to project any prior guess u^0 to a vector in L :

$$\begin{aligned} \Pi : \quad \mathcal{L} &\rightarrow (S_D \rightarrow S_D) \\ L &\rightarrow \Pi_L : (u^0 \rightarrow \Pi_L(u^0) \in L) \end{aligned}$$

The article then introduces a collection of "natural" properties that we gather hereafter.

- *consistency* : if $L' \subset L$ and $\Pi_L(S_D) \subset L'$ then $\Pi_{L'} = \Pi_L$; basically, if the result of a projection to a bigger space is always inside a smaller, then the projection on the two spaces are equivalent.
- *distinctness* : if L and L' are defined by a unique constraint and they are not equal, then $\Pi_L \neq \Pi_{L'}$ (unless they both contains the initial prior guess). Typically, in \mathbb{R}^2 , minimizing $\|\cdot\|$ on two lines returns a different result as soon as they do not both contain 0.
- *continuity* : Π is continuous with regards to $L \in \mathcal{L}$; it has a continuous relation with constraints.
- *scale invariant* : $\Pi_{\lambda L}(\lambda u) = \lambda u$ for any positive λ and any $u \in S_D$.
- *local* : for any subset $J \subset \{1, \dots, D\}$, $(\Pi_L)_J = (\Pi_{L'})_J$ as soon as $L_J = L'_J$ where L_J means we only keep constraints dealing with coordinates in J and $(\Pi_L)_J$ is the restriction of the resulting vector of Π_L to the J coordinates. This property indicates that the results of Π on a set of coordinates, only depends on constraints applied to those coordinates.
- *transitive* : for any $L' \subset L$, $\Pi_{L'} = \Pi'_L \circ \Pi_L$. We can first project on a bigger space without affecting the result.

The main result of the paper [29] states as follows :

Théorème 18 (Two canonic projections).

Only two projection rules respect all the conditions quoted beforehand :

$$\begin{aligned}\Pi^2 & : L \rightarrow \Pi_L^2 : (u^0 \rightarrow \operatorname{argmin}_{v \in L} \|v - u^0\|_2) \\ \Pi^{KL} & : L \rightarrow \Pi_L^{KL} : \left(u^0 \rightarrow \operatorname{argmin}_{v \in L} \sum_{d=1}^D v_d \ln \left(\frac{v_d}{u_d^0} \right) \right)\end{aligned}$$

where Π^{KL} amounts to project using the "Kullback-Leibler" divergence.

To come back to our transport problem, the "prior guess" is the uniform law while the subspace $L \subset S_D$ is defined using the margin constraints forced by μ and ν . Then, provided we verify quoted properties, the two cost functions we used cover an exhaustive view. Eventually it justifies the two graph clustering criteria comparing the neighborhood to each equilibrium (Definition 55 and Equation B.12) are canonical.

B.3 Monge properties : a justification of the \oplus/\otimes notation

We introduce two classes of matrices, the first one is attributed to Gaspard Monge, from a basic idea appearing in his 1781 paper, (incidentally see [19], where a reference is given to Alan Hoffman¹ who first coined that point and consequently proposed the name : Monge's Matrices). For each of those Monge's matrices, we point out some remarkable equalities and, moreover, we link them to a corresponding coupling function of section B.2.

B.3.1 Monge property – Indetermination

To introduce Monge's properties, we follow the exhaustive work of Rainer Burkard, Bettina Klinz and Rüdiger Rudolf exposed in the 66-pages-long article [19] and begin with definition 49.

Définition 49 (Monge and Anti-Monge matrix).

A $p \times q$ real matrix $c_{u,v}$ is said to be a Monge matrix if it satisfies :

$$c_{u,v} + c_{u',v'} \leq c_{u',v} + c_{u,v'} \quad \forall 1 \leq u \leq u' \leq p, 1 \leq v \leq v' \leq q$$

and an Anti-Monge matrix if :

$$c_{u,v} + c_{u',v'} \geq c_{u',v} + c_{u,v'} \quad \forall 1 \leq u \leq u' \leq p, 1 \leq v \leq v' \leq q$$

Remarque 59 (Full-Monge matrix).

The important case for our purpose is the equality case when a matrix is both Monge and Anti-Monge, we will call this situation "Full-Monge" matrix.

$$c_{u,v} + c_{u',v'} = c_{u',v} + c_{u,v'} \quad \forall 1 \leq u \leq u' \leq p, 1 \leq v \leq v' \leq q$$

1. In 1961 Alan Hoffman (IBM Fellow and US Science Academy member) rediscovered Monges's observation see [49]. Hoffman showed that the Hitchcock–Kantorovich transportation problem can be solved by a very simple approach if its underlying cost matrix satisfies those Monge's properties

Although it is poorly studied, the last introduced equality fits perfectly our purpose. The inequalities on the contrary, are common and can be met in diverse situations such as cumulative distribution functions, or copula theory.

Remarque 60 (Adjacent cells).

A straightforward but important derived property is the local adjacency cells equality : it is sufficient to satisfy the property of the remark 59 on adjacent cells, to ensure the obtainment of a "Full-Monge" matrix behavior for the global set of cells i.e. :

$$c_{u,v} + c_{u+1,v+1} = c_{u+1,v} + c_{u,v+1} \quad \forall 1 \leq u \leq p, 1 \leq v \leq q$$

Remark 60 is a key property to study Monge matrices since it gives a direct $\mathcal{O}(pq)$ algorithm to verify if a matrix is Monge.

Besides, a question emerges : which density function verifies the full Monge property ? The following Proposition 51 gives an interesting answer.

Propriété 51 (Full-Monge matrix is equivalent to "Indetermination").

A "full Monge matrix" necessarily represents an "indetermination coupling".

Démonstration.

Summing on u' and v' the equality of remark 59 we straightforwardly obtain :

$$\sum_{u'} \sum_{v'} (c_{u,v} + c_{u',v'} - c_{u',v} - c_{u,v'}) = pq c_{u,v} + c_{.,.} - q c_{.,v} - p c_{u,.} = 0$$

which rewrites :

$$c_{u,v} = \frac{c_{u,.}}{q} + \frac{c_{.,v}}{p} - \frac{c_{.,.}}{pq}$$

□

By summarizing properties we get the following Theorem 19.

Théorème 19 (Full-Monge matrices).

π representing a probability matrix, the following properties are equivalent.

1. π is a Full-Monge matrix
2. $\pi_{u,v} = \pi_{u,v}^+ = \frac{\mu_u}{q} + \frac{\nu_v}{p} - \frac{1}{pq}$
3. π optimizes problem 24 for some given margins
4. All 2×2 sub-tables $\{u, v, u', v'\}$ extracted from π have the same sum on their diagonal and anti-diagonal

Last property of Theorem 19 is illustrated on Figure B.1 and justifies the \oplus notation assigned to "indetermination". Indeed, if we take blue and red arrows we get the same resulting value : 0. Using the contingency form :

$$\begin{aligned} \text{blue arrows} & : 3 + 2 - 1 - 4 = 0 \\ \text{red arrows} & : 3 + 2 - 4 - 1 = 0 \end{aligned}$$

Equality obviously remains true for the probability form.

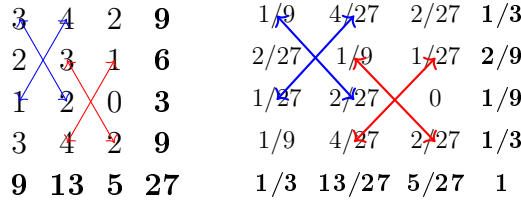


FIGURE B.1 – Example of an indetermination coupling (Statistical counting vs Probability forms)

B.3.2 Log-Monge property – Independence

We present hereafter a similar class of Matrices related, now, to independence : called Log-Monge matrices. They are built on the same principle as before through definition 50.

Définition 50 (Full-Log-Monge Matrices).

A strictly positive $p \times q$ matrix $c_{u,v}$ is "Full-Log-Monge" when :

$$\ln(c_{u,v}) + \ln(c_{u',v'}) = \ln(c_{u',v}) + \ln(c_{u,v'}) \quad \forall 1 \leq u \leq u' \leq p, 1 \leq v \leq v' \leq q$$

To immediately get the correspondence, we propose a transposition from a property to another using logarithm in Remark 7. It supposes matrices to be strictly positive (for our probability application : whole discrete space must be reached).

Remarque 61 (From Log-Monge to Monge).

We easily verify that c satisfies condition proposed in definition 50 if and only if $\ln(c)$ verifies the equivalent condition in definition 49 where logarithm is taken element-wise.

Using Remark 61, we can check that Full-Log-Monge property leads to interesting results and is linked to "independence coupling" ; without detailing their obtainment, we gather those results within Theorem 20, dual of Theorem 19.

Théorème 20 (Full-Log-Monge Matrices).

π being a strictly positive probability matrix all those properties are equivalent :

1. $\pi_{u,v}$ is Full-Log-Monge
2. $\pi_{u,v} = \pi_{u,v}^\times = \mu_u \nu_v$
3. π optimizes problem 22
4. All 2×2 sub-tables $\{u, v, u', v'\}$ extracted from π have the same product on their diagonal and anti-diagonal.

Figure B.2 illustrates "Full Log-Monge" matrices and their properties related to "independence" ; it justifies the usual \otimes notation. It is important to remark that both those matrices (in Figure B.1 and Figure B.2) optimize a problem where the unique difference is the cost functions (since the margins are strictly identical).

3	13/3	5/3	9	1/9	13/81	5/81	1/3
2	26/9	10/9	6	2/27	26/243	10/243	2/9
1	13/9	5/9	3	1/27	13/243	5/243	1/9
3	13/3	5/3	9	1/9	13/81	5/81	1/3
9	13	5	27	1/3	13/27	5/27	1

FIGURE B.2 – Example of an "independence coupling" (Contingency vs Probability forms)

B.4 Logical indetermination and Condorcet's voting equilibrium

In section B.4, our latent goal is to better understand the "indetermination coupling", that we have until now essentially introduced on a theoretical point of view. Although obtained through a similar process, "independence coupling" is straightforwardly linked to classical empirical experiences. π^+ does not share this latent simplicity and interpreting it, *per se*, is clearly a domain which deserves to be investigated. We present an attempt for helping the reader to make an accurate picture about the "indetermination" concepts.

Interest for the coupling will be reinforced as we show it corresponds to the Condorcet's majority equilibrium. Defining a "for" vs "against" notion will lead us to a formal equality interpreting "indetermination" in another space. In fact we are faced with the famous "Condorcet's voting equilibrium", which amounts to exhibit the situation where the number of opinions "for" balances exactly the number of opinions "against". The demonstration of this property requires the use of "Mathematical Relational Analysis" notations, which will be formally defined hereafter. We do not want in the context of this article to develop an exhaustive overview of this theory and its applications but pick up some results in connection with the goals we want to achieve; most of them being extracted from the following list of papers which gathers some of the most important key features about the subject : [68], [64], [71], [80], [65], [1], [2].

We also interpret the equilibrium between the "yes/for" (agreements) and the "no/against" (disagreements) as a voting "indetermination situation". This implies : since the number of votes "for" equals the number of votes "against" we are in a situation, where it is impossible to take a decision. The term : "indetermination" ("indeterminacy" or "uncertainty" should have been used as well) is a formal translation of this surprising situation. First of all, let us introduce properly Relational Analysis notations that we shall use later on.

Définition 51 (Relational Analysis notations).

Let (u_1, \dots, u_n) and (v_1, \dots, v_n) be two n probabilistic draws of $U \sim \mu$ and $V \sim \nu$. We define two associated symmetric $n \times n$ matrices X and Y by

$$X_{i,j} = \mathbf{1}_{u_i=u_j} \quad \text{and} \quad Y_{i,j} = \mathbf{1}_{v_i=v_j}, \quad \forall 1 \leq i, j \leq n$$

Basically, the two binary matrices X and Y (which correspond in fact to two binary equivalence relations based on the drawn modalities) represent agreements and disagreements of the two variables on a same draw of size n ; they are symmetric with 1 values on their diagonal. This relational coding has a lot of powerful properties, which will not be presented in this paper but which can be found in the articles we mentioned beforehand.

Definition 51 immediately provides us with an algorithm to transfer contingency representations to relational ones. The way back consists in noticing that :

$X_{i,j} = 1$ if and only if i and j share the same modality of $U \sim \mu$.

Hence we assign a modality to each class defined by the equivalence relation embedded in X : the only loss of information during this process resides in the names of modalities.

Now, we are ready to present the Theorem justifying the name "indetermination" :

Théorème 21 (π^+ and Condorcet's equilibrium).

π being a cross probability law on a set of $p \times q$ categorical variables, we shall say that π is an "indetermination coupling" on its margins, if and only if the expected number of "agreements" equals the number of "disagreements" on a 2 independent drawings of π .

Démonstration.

Let π be a probability law on $p \times q$ categorical variables ; it's defined through its values $\pi_{u,v}$, $1 \leq u \leq p$ and $1 \leq v \leq q$. U and V are random variables representing its margins. By n drawings through π , hence n samplings of (U, V) , U and V generates two partitions (equivalence relations) of the n individuals based on their modalities.

We will say that an agreement occurs when both partitions simultaneously gather or separate the individuals i and j . A disagreement occurs on the contrary when a classification regroupes i and j while the other one separates them. Formally, if X, Y encodes the n samplings as defined in Definition 51 :

- $X_{i,j}Y_{i,j} = 1$, agreement of type 11, there are pq couples of classes possible for two individuals i and j to realize this type of agreement
- $\overline{X}_{i,j}\overline{Y}_{i,j} = 1$, agreement of type 00, there are $p(p-1)q(q-1)$ couples of classes of this type
- $X_{i,j}\overline{Y}_{i,j} = 1$, disagreement of type 10, there are $pq(q-1)$ couples of classes of this type
- $\overline{X}_{i,j}Y_{i,j} = 1$, disagreement of type 01, there are $p(p-1)q$ couples of classes of this type

As quantities vary according we propose the following equality which establishes that the weighted number of agreements equals the weighted number of disagreements :

$$\frac{XY}{pq} + \frac{\overline{X}\overline{Y}}{p(p-1)q(q-1)} = \frac{X\overline{Y}}{pq(q-1)} + \frac{\overline{X}Y}{p(p-1)q} \quad (\text{B.9})$$

with the scalar product notation

$$XY = \sum_{i=1}^n \sum_{j=1}^n (X_{i,j}Y_{i,j})$$

Equality B.9 is intrinsically important. It is defined on a draw of size n and linked to a contingency indetermination. We take two draws at random independently under π :

(u_i, v_i) and (u_j, v_j) and introduce a probabilistic equality based on our 2 draws.

$$\frac{\mathbb{E}_{\pi \otimes \pi}(X_{i,j}Y_{i,j})}{pq} + \frac{\mathbb{E}_{\pi \otimes \pi}(\bar{X}_{i,j}\bar{Y}_{i,j})}{p(p-1)q(q-1)} = \frac{\mathbb{E}_{\pi \otimes \pi}(X_{i,j}\bar{Y}_{i,j})}{pq(q-1)} + \frac{\mathbb{E}_{\pi \otimes \pi}(\bar{X}_{i,j}Y_{i,j})}{p(p-1)q} \quad (\text{B.10})$$

We shall notice now that equality B.10 precisely occurs when π equals the indetermination coupling of its margins with the formula introduced in Theorem 17. Let us compute the result of two-sized independent draws under π .

- $\mathbb{E}_{\pi \otimes \pi}(X_{i,j}Y_{i,j}) = \sum_{u_i, v_i} \sum_{u_j, v_j} \pi_{u_i, v_i} \pi_{u_j, v_j} \mathbf{1}_{u_i=u_j \& v_i=v_j} = \sum_{u, v} \pi_{u, v}^2$
- $\mathbb{E}_{\pi \otimes \pi}(\bar{X}_{i,j}\bar{Y}_{i,j}) = \sum_{u_i, v_i} \sum_{u_j, v_j} \pi_{u_i, v_i} \pi_{u_j, v_j} \mathbf{1}_{u_i \neq u_j \& v_i \neq v_j} = \sum_{u, v} \pi_{u, v} (1 - \pi_{u, \cdot} - \pi_{\cdot, v} + \pi_{u, v})$
- $\mathbb{E}_{\pi \otimes \pi}(X_{i,j}\bar{Y}_{i,j}) = \sum_{u_i, v_i} \sum_{u_j, v_j} \pi_{u_i, v_i} \pi_{u_j, v_j} \mathbf{1}_{u_i=q_{u_j} \& v_i \neq v_j} = \sum_{u, v} \pi_{u, v} (\pi_{u, \cdot} - \pi_{u, v})$ (and similarly for $\mathbb{E}_{\pi \otimes \pi}(\bar{X}_{i,j}Y_{i,j})$)

Inserting into equation B.10, we get :

$$\begin{aligned} & \frac{\sum_{u, v} \pi_{u, v}^2}{pq} + \frac{\sum_{u, v} \pi_{u, v} (1 - \pi_{u, \cdot} - \pi_{\cdot, v} + \pi_{u, v})}{p(p-1)q(q-1)} \\ &= \frac{\sum_{u, v} \pi_{u, v} (\pi_{u, \cdot} - \pi_{u, v})}{pq(q-1)} + \frac{\sum_{u, v} \pi_{u, v} (\pi_{\cdot, v} - \pi_{u, v})}{p(p-1)q} \end{aligned}$$

Reducing to same denominator, we get :

$$\begin{aligned} & (p-1)(q-1) \sum_{u, v} \pi_{u, v}^2 + \sum_{u, v} \pi_{u, v} (1 - \pi_{u, \cdot} - \pi_{\cdot, v} + \pi_{u, v}) \\ &= (p-1) \sum_{u, v} \pi_{u, v} (\pi_{u, \cdot} - \pi_{u, v}) + (q-1) \sum_{u, v} \pi_{u, v} (\pi_{\cdot, v} - \pi_{u, v}) \end{aligned}$$

regrouping the similar terms yields :

$$pq \sum_{u, v} \pi_{u, v}^2 - p \sum_u \pi_{u, \cdot}^2 - q \sum_v \pi_{\cdot, v}^2 + 1 = 0$$

Making use of a classical equality similar to equation B.3, we obtain :

$$pq \sum_{u, v} (\pi_{u, v} - \pi_{u, \cdot}/q - \pi_{\cdot, v}/p + 1/pq)^2 = 0$$

Finally it holds :

$$\pi_{u, v} = \frac{\pi_{u, \cdot}}{q} + \frac{\pi_{\cdot, v}}{p} - \frac{1}{pq}$$

We have proved that $\pi = \pi^+$ if and only if the expected number of normalized agreements equals the expected number of disagreements on a 2-sized drawing. \square

In another article [11] we interpret the cost function linked with indetermination as a way to reduce couple matchings and show the Condorcet's equilibrium conveys useful applications as it "hides" the underlying distribution.

B.5 Application to network clustering

B.5.1 Introduction

We limit ourselves to Louvain Algorithm, applied as an heuristic to optimize a global objective function. In few words let us say that the global optimization is obtained iteratively by optimizing a local cost function : where two nodes are said to be *similar* if the value of a chosen criterion is high.

Conde Cespedes, in her thesis [25], gathered a large amount of network clustering criteria, coming from the scientific literature ; she took advantage of this task to give them a category label, depending upon their relationship with both "independence" or "indetermination". She compared them according to their ability to perform on various networks, and collected and stored the obtained results. Although we are in the quite same line with Patricia Conde-Céspedes, we restrict ourselves to investigate a focused study of both the canonic ones : "deviation to independence" and "deviation to indetermination" that we will reintroduce hereafter within the network theoretical context.

First let us start with some usual definitions for a graph :

Définition 52 (Weighted graph (network)).

A weighted graph G , is a graph which contains n nodes $1 \leq i \leq n$, which are linked each other through edges (i, j) linked with weights $a_{i,j}$ (representing a weighted incidence matrix). We also introduce the total weight $2M = \sum_{i,j} a_{i,j}$.

A basic way to randomly generate a network is through the Gilbert's distribution :

Définition 53 (Gilbert).

Fixing a number n of nodes and $\epsilon \in [0, 1]$, we link any set of two nodes by independently drawing through a Bernoulli law with parameter ϵ leading to a 0 – 1 weight. The obtained network is non directed and each weight is 0 or 1.

Remarque 62.

Adding a parameter p representing maximum weight, we can easily create a weighted graph by drawing a Binomial law with parameter (ϵ, p) while linking couples (instead of sets) generates directed networks.

As mentioned in section B.1, our work will be devoted to the research of classes, groupings, clusters or cliques (whatever the name) within a network. They are defined through an equivalence relation as specified in definition 54 :

Définition 54 (Graph clustering).

Let us call x , a matrix representation of a binary equivalence relation, the result of the clustering of a graph G . Then $x_{i,j}$ equals 0 or 1 and equals 1 if and only if the two nodes i and j are in the same class for x , and 0 if not.

Clustering algorithms aim at providing classes maximizing internal similarities as well as minimizing external ones. A first option is to take as input the number K of classes we are looking for, together with an associated distance (or dissimilarity index) and come up with a list of best representatives or "means" for each class. K-means algorithm whose

idea goes back to the fifties (see² [88]) typically illustrates this option. A second option, is to construct a local criterion c which assigns a weight $c_{i,j}$ to each (i, j) couple of nodes based on their similarity; the more similar they are, the higher the criterion is. We then build up a global criterion by summing up the local values $c_{i,j}$ if and only if i and j are in the same class as proposed in problem 25.

Problème 25 (Generic clustering problem).

$$\max_x \quad M(c, x) = \sum_{i=1}^n \sum_{j=1}^n c_{i,j} x_{i,j}$$

subject to :

x is an equivalence relation (see definition 51)

First let us remark that, as notably spotted in [65], [68], [80] an equivalence relation constraint can be written as :

- $x_{i,i} = 1, \forall 1 \leq i \leq n$ (reflexivity)
- $x_{i,j} = x_{j,i}, \forall 1 \leq i, j \leq n$ (symmetry)
- $x_{i,j} + x_{j,k} - x_{j,k} \leq 1, \forall 1 \leq i, j, k \leq n$ (transitivity)

Thanks to the linearity of these constraints, in addition to the linear expression of the criterion itself, the problem 25 although *a priori* NP-hard can be exactly solved (according to some conditions) through the integer relaxation of a good existing 0-1 linear programming code (see [68]). But in the context of networks clustering, the size n of the problem (here the number of nodes) can be really huge (millions for social networks) and the direct solving by linear programming, even specially tuned, is no longer possible; therefore, the use of robust heuristics becomes mandatory. As mentioned beforehand, Louvain Algorithm (see [41] or [76]) belongs to this set of methods : it does not systematically provide us with an exact optimal result but just a quite good approximate one. We concentrate on the analysis of two canonic costs for which historical experiments are reported and explained at the light of the previous sections.

Original Modularity – "Independence"

The original and famous Newman-Girvan's presentation of a global criterion for networks clustering, see [41] or [76], has been introduced in the Louvain algorithm together with a global cost called "Modularity" defined by :

Définition 55 (Modularity).

Given a partition $x_{i,j}$ and a graph G with weighted function a on its edges, the global modularity returns to :

$$M^\times(G, x) = \frac{1}{2M} \sum_{i,j} \left[a_{i,j} - \frac{a_{i,\cdot} a_{\cdot,j}}{2M} \right] x_{i,j} \quad (\text{B.11})$$

Let us first remark that the original modularity M^\times is nothing but our generic global cost function defined though Problem 25 with :

$$c_{i,j} = m^\times(G)_{i,j} = \frac{a_{i,j}}{2M} - \frac{a_{i,\cdot} a_{\cdot,j}}{(2M)^2}$$

2. factually this is the method of S. Lloyd(1957) rewritten by E.W. Forgy (1965) which corresponds to the oldest version of the K-means really used

and that the local gain $m^\times(G)_{i,j}$ to put two nodes in the same class is the local deviation to independence. Indeed, using definition 52, with $\pi_{i,j} = \frac{a_{i,j}}{2M}$ as a probability measure on $\{1 \dots n\}^2$ and margins $\mu_i = \frac{a_{i,\cdot}}{2M}$, m^\times rewrites :

$$m^\times(G)_{i,j} = 2M (\pi_{i,j} - \mu_i \mu_j)$$

and does express itself as a canonic *deviation to independence* criterion.

A second remark is that as $m^\times(G)_{i,j}$ expression does not contain absolute value or square elevation then non connected nodes will lead to negative weights preventing them from being allocated to the same class. If they are connected the importance of $m^\times(G)_{i,j}$ evolves positively as i and j have less edges ($a_{i,\cdot}$ and $a_{\cdot,j}$ small) ; here again this implies an appropriate behavior. More precisely, since independence ensures a coupling as uniform as possible with fixed margins (this is a solution of problem 22), m^\times appears as a fair construction. The criterion basically measures a distance between the observed linkage weight and an expected flat weight given by the average neighborhood.

Extended Modularity – "Indetermination"

We suggest an expression $m^+(G)_{i,j}$ which represents a deviation to indetermination. It will be used as a local cost function in Problem 25 leading to a slightly different global formula $M^+(G, x)$ to optimize locally :

$$m^+(G)_{i,j} = a_{i,j} - \frac{a_{i,\cdot}}{n} - \frac{a_{\cdot,j}}{n} + \frac{2M}{n^2}$$

Symmetrically as m^\times , it rewrites as a canonic *deviation to indetermination* :

$$m^+(G)_{i,j} = 2M * \left(\pi_{i,j} - \frac{\mu_i}{n} - \frac{\mu_j}{n} + \frac{1}{n^2} \right)$$

The global criterion being :

$$M^+(G, x) = \sum_{i,j} \left[a_{i,j} - \frac{a_{i,\cdot}}{n} - \frac{a_{\cdot,j}}{n} + \frac{2M}{n^2} \right] x_{i,j} \quad (\text{B.12})$$

We have seen that both couplings share a lot of properties as shown in section B.3 and section B.4. In the same way, Patricia Conde-Céspedes noticed that a lot of statistical criteria measuring variables correlation are based either on a "distance to independence", or on a "distance to indetermination" (see [25]). According to these remarks, our canonical *deviation to indetermination* criterion M^+ deserves to have the same types of use as those dedicated to the Newman Girvan's M^\times .

B.5.2 Summary of an application to various networks

Now the two criteria are properly introduced with a theoretical basis on their canonical structure, we gather in table B.1 the number of classes found by Patricia Conde-Céspedes, who applied both on the same empirical networks. She got similar results, as expected beforehand on a bench of experimental graphs she properly defines in [25] and for which we gather her results in table B.1. This table can be read as follows : for example,

the "Internet" network contains 69,949 nodes with 351,280 edges ; if we apply Louvain algorithm on, with the global criteria M^\times we usually find 46 communities, while M^+ leads to 39. As anticipated in section B.2.4 criteria are (in average) very close (see for instance the Amazon case); consequently their resulting effect on various networks is quite similar. Section B.5.3 of the present paper provides the reader with an explanation of the assertion Patricia Conde-Céspedes experimented.

TABLE B.1 – Number of classes found by each criteria on various networks

	Karate	Football	Jazz	Internet	Amazon	YouTube
N (nb nodes)	34	115	198	69 949	334 863	1 134 890
M (sum of weights)	78	613	2 742	351 280	925 872	2 987 624
Number of classes for criteria M^\times	4	10	4	46	250	5 567
Number of classes for criteria M^+	4	10	6	39	246	13 985

B.5.3 Gilbert Experimental Tests

As already mentioned, solving problem 25 is NP-hard so that we cannot expect precise results, neither about the number of classes for a given criterion, nor about their composition. Nevertheless, we can compare directly local values of the criterion to extrapolate a common or a distinct global run when iteratively optimized.

We propose a comparative try based on Gilbert's networks to spot differences or similarity between $m^\times(G)_{i,j}$ and $m^+(G)_{i,j}$ values. The aim is to observe the distribution of both criteria on a typical network. First, to simplify observations and as only the reference cost (*i.e.* one equilibrium of section B.2) varies between m^+ and m^\times , we only keep it by subtracting $a_{i,j}$; it is formally defined in definition 56. Then, we generate 1,000,000 networks randomly, compute each criterion on a random pairs of nodes and store the reference cost ; the results are gathered within figure B.3.

Définition 56 (Bias or reference cost).

The two bias derived from m^\times and m^+ are respectively :

$$b_{i,j}^\times = \frac{a_{i,\cdot} \cdot a_{\cdot,j}}{2M} \quad \text{and} \quad b_{i,j}^+ = \frac{a_{i,\cdot}}{n} + \frac{a_{\cdot,j}}{n} - \frac{2M}{n^2}$$

On figure B.3 we observe that the distributions of both biases are similar for any values of ϵ . Indeed, the curves are identical on their core values (those with a number of realizations upon 20,000). It illustrates their common origin which amounts to flatten a distribution (section B.2) and leads to a small expected difference (section B.2.4). Their common mean is equal to ϵ as it can be easily derived from the formulas. A difference nevertheless remains on extreme values particularly visible for $\epsilon = 0.3$. Let us now compute theoretically both distributions under Gilbert's networks to confirm their symmetry.

Propriété 52 (Probability values).

Let b be a binary value, $b \leq n_i \leq n$ and $b \leq n_j \leq n$; let us compute the following

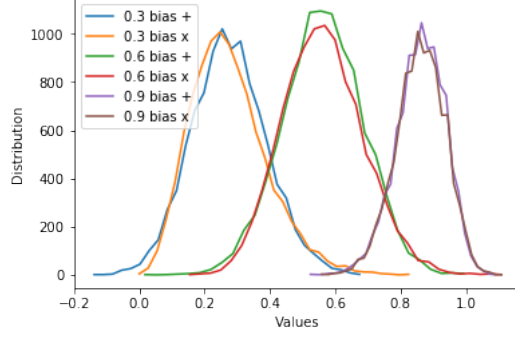


FIGURE B.3 – Empirical distribution of the two reference costs $b_{i,j}^+$ and $b_{i,j}^x$ for ϵ in $[0.3, 0.6, 0.9]$; X-axis gives the values of the bias, Y-axis gives the corresponding number of realizations

probability :

$$\begin{aligned} & \mathbb{P}(a_{i,j} = b, a_{i,\cdot} = n_i, a_{\cdot,i} = n_j) \\ &= \epsilon^b (1 - \epsilon)^{1-b} \binom{n-1}{n_i-b} \epsilon^{n_i-b} (1 - \epsilon)^{n-1-n_i+b} \binom{n-1}{n_j-b} \epsilon^{n_j-b} (1 - \epsilon)^{n-1-n_j+b} \end{aligned}$$

The corresponding value $m_{i,j}^+$ and $m_{i,j}^x$ associated to a group (b, n_i, n_j) of the parameters being evident, we propose figure B.4 which represents the difference between theoretical distributions of both criteria with $\epsilon = 0.3$. b^x and b^+ have distinct forms but their proximity on highly probable values, given on Figure B.4, illustrates section B.2.4 : if we couple two variables with n margins, expected difference is less than $\frac{1}{n^2}$.

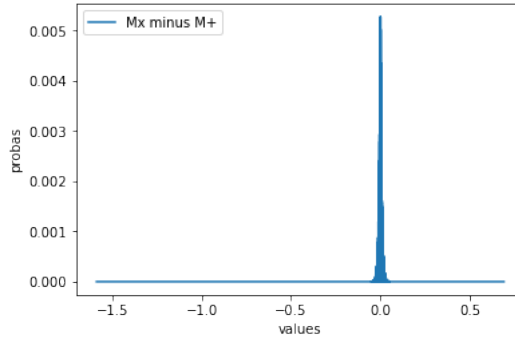


FIGURE B.4 – Theoretical distribution of the difference $m^x(G)_{i,j} - m^+(G)_{i,j}$ (same as $b_{i,j}^+ - b_{i,j}^x$) on generated graphs

Extreme values, on the contrary may differ drastically. Though it seems the opposite to Figure B.3 as m^+ comes with higher values than m^x , it's consistent because of the *minus* sign in the formula linking m with b . Having noticed that b^+ and b^x differ on their extreme values, we compute them on a general Gilbert network (respecting the common value of $2M = n^2\epsilon$), and obtain the bounds :

$$-\epsilon \leq b^+ \leq \frac{n}{n} + \frac{n}{n} - \epsilon = 2 - \epsilon \quad \text{and} \quad 0 \leq b^x \leq \frac{n \times n}{n^2\epsilon} = \frac{1}{\epsilon} \quad (\text{B.13})$$

As already expected with figure B.4 the difference between extreme values is arbitrarily high. Eventually it shows that in average, both canonical criteria will share a similar behavior hence shall be applied indifferently when the network's neighborhood weight a is close to uniform. Precisely, in any real application the chosen criterion distinguishes at each iteration a link among the neighborhood unexpected when compared to the equilibrium (either independence or indetermination). The difference between the two equilibria is so small that if one distinguishes a link, the other will so that we expect the resulting classes to be comparable if not identical. Indeed, we know we can build, following equations B.13, networks on which \mathcal{M}^+ and \mathcal{M}^\times deliver completely different results.

B.6 Conclusions

First, we followed the historical line and introduced two basis from Discrete Optimal Transport Theory : *independence* and *indetermination*. As recalled, the first one is the most intuitive and frequently used in mathematical articles as well as experimented in real life. The second notion appeared more surprising, poorly studied in the statistical literature but more commonly used by people working on Mathematical Relational Analysis Voting Theory and Analysis of Variance. Together, they cover the only two canonic projection costs as quoted in [29].

To illustrate the usefulness of the parallel construction, we turned to applications and completed the track followed by Patricia Conde-Céspedes in her thesis [25]. She gathered a list of networks clustering criteria and classified them according to their deviation to one of the mentioned coupling functions. Section B.5 reports a further analyze of the two canonical criteria. It gathers results about the general similarity of their application on various networks as well as their extreme values to set one another apart.

In each section, from optimal transport to networks, we insisted on the parallel between both notions together with their differences. As quoted beforehand, they appear as the two unique canonic structural solutions. Generally, the differences between them needs to be scanned up, either to coin a macro criteria, or to chose wisely between one or another depending on the structure of the network. In any case, the traditional use of *independence* at the expense of *indetermination* needs to be further investigated and explained.

Acknowledgements : We thank the editor and two anonymous referees for their valuable comments.

Logical indetermination coupling : a method to minimize drawing matches and its applications

Submitted to Information and Inference, December 2020 ([11]).

Abstract : *While justifying that independence is a canonic coupling, the authors show the existence of a second equilibrium to reduce the information conveyed from the margins to the joined distribution : the so-called indetermination. They use this null information property to apply indetermination to graph clustering. Furthermore, they break down a drawing under indetermination to emphasis it is the best construction to reduce couple matchings, meaning, the expected number of equal couples drawn in a row. Using this property, they notice that indetermination appears in two problems (Guessing and Task Partitioning) where couple matchings reduction is a key objective.*

C.1 Introduction

In a precedent paper [10], we highlighted a list of structural analogies between two discrete couplings namely *independence* and *indetermination* together with an application to graph clustering that we will recall hereafter in section C.4.

A discrete coupling is a function C operating on two discrete marginal laws $\mu = \mu_1 \dots \mu_p$ and $\nu = \nu_1 \dots \nu_q$ and which defines a probability law π on the product space :

$$\pi_{u,v} = C(\mu_u, \nu_v), \quad \forall 1 \leq u \leq p, 1 \leq v \leq q$$

We respectively quote both those above mentioned couplings C^\times (*independence*) and C^+ (*indetermination*); this last notion has been initially introduced by J.-F. Marcotorchino in his seminal papers [64] and [63]) while their formula will be reintroduced and rediscussed later on in section (C.2).

Their usefulness arises in statistical applications : namely, most of our usual statistical deviation criteria for contingency analysis are expressed in terms of deviations from one of the two couplings ([25] gathers a classification of them, deviation to *independence* or deviation to *indetermination*). The most famous example for *independence* is the χ^2 , widely

used in practice which computes nothing but a deviation to the *independence* coupling of the induced margins. Symmetrically the Janson-Vegelius coefficient, initially introduced in [52] as a contingency association index, measures also a deviation no longer to *independence* but rather to *indetermination*; we shall detail this point in subsection C.4.1.

Furthermore, purely theoretical considerations lead to consider *independence* and *indetermination* as the only two possible "natural" equilibria : this fact being based upon a work of Csizar [29], a summarized version of which is expressed in [10]; we will briefly explain, in this present document, the rationale behind.

While *independence* is commonly used and studied in the scientific literature, *indetermination* appears as a lesser known coupling, whose properties have been rarely presented in an explicit way. After introducing both coupling functions in section C.2, the next section, section C.3 of this paper, is precisely dedicated to the properties implied by *indetermination*. Notably, we show that it consists in reducing couple matching occurrences : drawing two couples under *indetermination*, the probability to have both couples equal will be particularly small. In addition, we decompose an *indetermination* coupling into an arbitrary margin and a uniform one leading to a constructive drawing.

Section C.4 provides several information problems where the logical *indetermination* coupling appears. First, since it is an equilibrium similar to independence, we coin a new local graph criteria as the Girvan-Newman modularity ([76]) but based on the *indetermination* coupling. Besides, leveraging on the constructive drawing as well as on the reducing "couple matchings" property, we show that *indetermination* naturally occurs in solving the so called "guessing problem" ([69]) as well as the "task partitioning problem" ([18]). The guessing problem appears in cryptography, when a spy screening a communication session tries to determine which message was sent, using to do so, some partial information. The task partitioning problem occurs in manufacturing process when we want to optimize the way to assign tasks to production teams, or to machines in job-shop scheduling.

Eventually, having deeply studied the discrete version of *indetermination*, section C.5 extends the *indetermination* notion towards a continuous version, providing straightforwardly associated density and cumulative distribution functions.

C.2 From discrete transport problem to coupling function

When we want to couple two marginal laws, the most common and straightforward way to proceed, consists in assuming *independence* and keep on computations. It is so well integrated in our mindset, that it naturally appears in real life applications, as soon as we want to build fast models up. In our scientific work, the approach is quite the same : when we use a very classical and usual criterion like the χ^2 index, we are measuring nothing but a deviation to *independence*.

Thinking about how we first introduced *independence*, we immediately suggest empirical experiments : let us say if we roll a dice twice, how should we derive the resulting probabilities from a unique dice ? Most of us will naturally apply *independence* coupling : it really relies on empirical experiments.

Although being the most natural, it is not, by far, the only existing available coupling

method; actually, as introduced by Sklar in [87], any copula function will lead to a coupling function behaving on two cumulative distribution functions. In [10] we defined a notion called "coupling function" resulting from the resolution of a fixed transport problem. Throughout the first part of this precedent paper we considered discrete optimal transport problem : how can we efficiently move a pile of sand into a corresponding hole to fill? Using Kantorovitch's theorem formulation, how can we respect given margins (in terms of masses or quantities deposited in the "origins – destinations" schema) while minimizing a cost function splitting optimally those quantities on the resulting coupling.

Formally, through modern notations, two probability measures $\mu = \mu_1 \dots \mu_p$ and $\nu = \nu_1 \dots \nu_q$ represent the initial constraints (respectively masses at origins and destinations). Typically, for instance in national economy planning : μ expresses the quantity produced by a sector and ν the consumed quantity by an industry. Then, given a cost function M which associates a positive value to any element in $[0, 1]$, the discrete "Monge Kantorovitch Problem" (MKP) can be expressed as follows :

Problème 26 (Discrete Version of MKP).

$$\begin{aligned} \min_{\pi} \quad & \sum_{u=1}^p \sum_{v=1}^q \mathbf{M}(\pi_{u,v}) \\ \text{subject to :} \quad & \sum_{v=1}^q \pi_{u,v} = \mu_u; \forall u \in \{1, \dots, p\} \\ & \sum_{u=1}^p \pi_{u,v} = \nu_v; \forall v \in \{1, \dots, q\} \\ & \pi_{u,v} \geq 0; \forall (u, v) \in \{1, \dots, p\} \times \{1, \dots, q\} \end{aligned}$$

The choice of a cost function \mathbf{M} depends upon the applications we want to address. For instance, we can force the result π to concentrate as little information as possible; this means, we shall constraint it to be as close as possible to the uniform law, referring to the product space given the margins. Other choice : we can, as well, minimize the entropy of π given the margins. Both those cases are usual approaches. They expect the global assignment to be as smooth as possible.

A MKP problem is essentially given by its cost function M , while margins (μ, ν) may vary. As such the solution to Problem 26 is written as $C^P(\mu, \nu)$

Définition 57 (MKP Problem Associated with Coupling function).

For a given MKP problem P , we define a coupling function C^P by : $C^P(\mu, \nu) = \pi^(P)$ provided that π^* exists as a unique solution of P with margins μ and ν .*

The existence of a solution depends on the cost function \mathbf{M} as well as on the margins. It has been discussed in numerous papers and we will not further discuss this point here; neither did we in [10].

In paper [10] nevertheless, we expressed Problem 26 using the projections introduced in [29]. Doing so, we were able to justify the two canonic discrete costs defined below (Equations (C.1) and (C.2)) as the only two ways leading to a structured coupling function.

$$\mathbf{M}^\times : x \mapsto x \log(pqx) \quad (\text{C.1})$$

$$\mathbf{M}^+ : x \mapsto pq \left(x - \frac{1}{pq} \right)^2 \quad (\text{C.2})$$

Under the projection formalism, let us unify the costs provided beforehand. As it will be mentioned later again, when we associate each problem to its coupling function, both share the same solution if we remove the constraints on margins, that is to say : the uniform law ($\pi_{u,v} = \frac{1}{pq}$). Actually the cost function can be formally expressed as a *divergence* (see the definition 58 below) to the uniform law (consequently to the optimal solution without constraints).

Définition 58 (Divergence with parameter).

Given, a positive function $\phi : \mathbb{R}^+ \mapsto \mathbb{R}^+$ with $\phi(1) = 0$, and given two discrete probabilities m and n on pq elements, we define the divergence D_ϕ as :

$$D_\phi(m|n) = \sum_{u=1}^p \sum_{v=1}^q n_{u,v} \phi \left(\frac{m_{u,v}}{n_{u,v}} \right) = \mathbb{E}_n \left[\phi \left(\frac{m_{u,v}}{n_{u,v}} \right) \right]$$

additionally we set $0 * \phi(x) = 0$ for all x .

Denote by \mathcal{U} the uniform distribution on $(1, \dots, p) \times (1, \dots, q)$, ie $\mathcal{U}(u, v) := 1/pq$ for all u and v . Now, for $\phi : x \mapsto x \log(x) - x + 1$ it holds :

$$\begin{aligned} D_{KL}(\pi | \mathcal{U}) &= \sum_{u=1}^p \sum_{v=1}^q \frac{1}{pq} [pq\pi_{u,v} \log(pq\pi_{u,v}) - pq\pi_{u,v} + 1] \\ &= \sum_{u=1}^p \sum_{v=1}^q \pi_{u,v} \log(pq\pi_{u,v}) \\ &= \sum_{u=1}^p \sum_{v=1}^q \mathbf{M}^\times(\pi_{u,v}) \end{aligned}$$

so that Problem 26 associated with the cost function \mathbf{M}^\times is rewritten as a minimization of D_{KL} which is the usual Kullback-Leibler divergence.

Similarly, if we pose $\phi : x \mapsto (x - 1)^2$ (the so-called Chi-square divergence function)

$$\begin{aligned} D_2(\pi | \mathcal{U}) &= \sum_{u=1}^p \sum_{v=1}^q \frac{1}{pq} (1 - pq\pi_{u,v})^2 \\ &= pq \sum_{u=1}^p \sum_{v=1}^q \left(\pi_{u,v} - \frac{1}{pq} \right)^2 \\ &= \sum_{u=1}^p \sum_{v=1}^q \mathbf{M}^+(\pi_{u,v}) \\ &= -1 + pq \sum_{u=1}^p \sum_{v=1}^q \pi_{u,v}^2 \end{aligned} \quad (\text{C.3})$$

Finally, the two costs \mathbf{M}^\times and \mathbf{M}^+ differ from each other in the way the divergence from \mathcal{U} to the uniform law π is taken into account and amounts to optimizing a divergence.

Although it seems arbitrary, the restriction to the two previous costs, is anything but a fortuitous decision : in [29], Csiszar actually shows that, provided we verify additional intuitive properties, we must restrict ourselves to use either least square or maximum entropy in Problem 26. Our transport problem aims at projecting the uniform \mathcal{U} into the eligible space of probability law respecting two margins.

A general question is how to adapt a "prior guess" u_0 to verify a list of constraints. Let us say u_0 lives in the simplex S_n while the given constraints define a subspace $L \in \mathcal{L}$ (\mathcal{L} is the space of subspaces of S_n tuned by a finite list of affine constraints, see [29] for more details). To formalize it, Csiszar defines a *projection rule* Π as a function whose input is a set $L \in \mathcal{L}$ and which generates a method Π_L to project any prior guess u_0 to a vector in L :

$$\begin{aligned} \Pi : \quad \mathcal{L} &\rightarrow (S_n \rightarrow S_n) \\ L &\rightarrow \Pi_L : (u_0 \rightarrow \Pi_L(u_0) \in L) \end{aligned}$$

The article then introduces a collection of "natural" properties that we gather hereafter.

- *consistency* : if $L' \subset L$ and $\Pi_L(S_n) \subset L'$ then $\Pi_{L'} = \Pi_L$; basically, if the result of a projection to a bigger space is always inside a smaller, then the projection on the two spaces are equivalent.
- *distinctness* : if L and L' are defined by a unique constraint and they are not equal, then $\Pi_L \neq \Pi_{L'}$ (unless they both contains the initial prior guess). Typically, in \mathbb{R}^2 , minimizing $\|\cdot\|$ on two lines returns a different result as soon as they do not both contain 0.
- *continuity* : Π is continuous with regards to $L \in \mathcal{L}$; it has a continuous relation with constraints.
- *scale invariant* : $\Pi_{\lambda L}(\lambda u) = \lambda u$ for any positive λ and any $u \in S_n$.
- *local* : for any subset $J \subset \{1, \dots, n\}$, $(\Pi_L)_J = (\Pi_{L'})_J$ as soon as $L_J = L'_J$ where L_J means we only keep constraints dealing with coordinates in J and $(\Pi_L)_J$ is the restriction of the resulting vector of Π_L to the J coordinates. This property indicates that the results of Π on a set of coordinates, only depends on constraints applied to those coordinates.
- *transitive* : for any $L' \subset L$, $\Pi_{L'} = \Pi'_{L'} \circ \Pi_L$. We can first project on a bigger space without affecting the result.

Adding a last property which guarantees that the "no interaction" solution in case we omit constraints respects a proportional behavior finally lead to restrict ourselves to the two problems we treat in this document.

C.2.1 The Alan Wilson's entropy model : role of "independence"

First introduced by Sir Alan Wilson in 1969 for "Spatial Interaction Modeling" the "Flows Entropy Model" of Alan Wilson, can be found in his various publications ; originated in [98], developed in [99], and refined in his book [100]. A fundamental justification of his approach corresponds to the following contextual situation : in a theoretical system

it is advisable to determine the distribution $\pi_{u,v}$ (normalized frequency flows), supposing $\pi \geq 0$, which maximizes the entropy of the system under some given constraints.

Without any marginal constraints, the optimal solution is nothing but the uniform law $\pi_{u,v} = \frac{1}{pq}, \forall 1 \leq u \leq p, \forall 1 \leq v \leq q$. By using margins, that is to say information about total exports (origins flows) and total imports (destination flows), degree of disorder of the system is drastically reduced. We are led to the following Problem 27; solution of which is given by Theorem 22.

Problème 27 (Balanced PSIS).

$$\begin{aligned} \min_{\pi} \quad & \sum_{u=1}^p \sum_{v=1}^q \pi_{u,v} \log(pq\pi_{u,v}) = D_{KL}(\pi | \mathcal{U}) \\ \text{subject to :} \quad & \\ & \sum_{v=1}^q \pi_{u,v} = \mu_u, \quad \forall 1 \leq u \leq p \\ & \sum_{u=1}^p \pi_{u,v} = \nu_v, \quad \forall 1 \leq v \leq q \\ & 0 \leq \pi_{u,v} \leq 1, \quad \forall 1 \leq u \leq p, 1 \leq v \leq q \end{aligned}$$

Théorème 22.

The solution of Problem 27 is $\pi^\times(u, v) = \mu_u \nu_v, \forall 1 \leq u \leq p, 1 \leq v \leq q$.

Hence the coupling function associated to Problem 27 is nothing but independence :

$$C^{\text{Problem 27}}(\mu, \nu)_{u,v} = C^\times(\mu, \nu)_{u,v} = (\mu \otimes \nu)_{u,v} = \mu_u \nu_v$$

As a conclusion, from the direct maximization of entropy, we get the solution expressed in terms of probability and remark that the associated coupling function is nothing but *independence* (denoted by \otimes in the sequel). We also note that the degree of disorder is not total : flows possess an intensity which is proportional to the weights of the partners in the world trade exchanges matrix in case of an economic application.

C.2.2 The minimal trade model : role of *indetermination*

In the "Minimal Trade Model" (see [89], [64] and [63]), we still impose the objective function to respect the marginal distributions and positivity constraints; but we change the structure of the cost function. In that case the criterion is a quadratic function measuring squared deviation of the cells values from the "no information" situation (the uniform joint distribution law). As expected, in case of free margins, the solution remains the uniform law \mathcal{U} . But, adding usual pre-conditioned constraints on margins leads to the least squares problem is Problem 28; solution of which is given by Theorem 23.

Problème 28 (Minimal Trade Model).

$$\min_{\pi} \quad pq \sum_{u,v} \left(\pi_{u,v} - \frac{1}{pq} \right)^2 = D_2(\pi | \mathcal{U})$$

subject to :

$$\begin{aligned} \sum_{v=1}^q \pi_{u,v} &= \mu_u, \quad \forall 1 \leq u \leq p \\ \sum_{u=1}^p \pi_{u,v} &= \nu_v, \quad \forall 1 \leq v \leq q \\ 0 &\leq \pi_{u,v} \leq 1, \quad \forall 1 \leq u, v \leq q \end{aligned}$$

Théorème 23.

The solution of Problem 28 is $\pi^+(u, v) = \frac{\mu_u}{q} + \frac{\nu_v}{p} - \frac{1}{pq}$.

Hence the coupling function associated to Problem 28 is nothing but indetermination :

$$C^{\text{Problem 28}}(\mu, \nu)_{u,v} = C^+(\mu, \nu)_{u,v} = (\mu \oplus \nu)_{u,v} = \frac{\mu_u}{q} + \frac{\nu_v}{p} - \frac{1}{pq}$$

A supplementary condition, which is exogenous with regard to the previous model, must be added on the margins (which are, by the way, constant values given *a priori*), this condition (see [64]) is a the following inequality which guarantees the positivity of the frequency matrix $\pi^+(u, v) = C^+(\mu, \nu)_{u,v}$:

$$p \min_u \mu_u + q \min_v \nu_v \geq 1 \tag{C.4}$$

From now on, we shall consider that Condition (C.4) applies whatever the values of the μ_u and ν_v are.

At that stage we introduced the two coupling functions in the discrete version used in graph clustering criteria. We have proposed in [10] some few properties of the *indetermination* notion ; we now consider a constructive approach to such coupling. It appears that this construction aims at reducing some matching problem. We will present two specific applications in Section C.4.

C.3 Discrete *indetermination* and associated properties

C.3.1 Monge property

The class of matrices we define here is attributed to Gaspard Monge, from a basic idea appearing in his 1781 paper. In fact, to introduce Monge's properties, we follow the exhaustive work of Rainer Burkard, Bettina Klinz and Rüdiger Rudolf exposed in their 66-pages-long article [19]. We begin with Definition 59.

Définition 59 (Monge and Anti-Monge matrix).

A $p \times q$ real matrix $c_{u,v}$ is said to be a Monge matrix if it satisfies :

$$c_{u,v} + c_{u',v'} \leq c_{u',v} + c_{u,v'} \quad \forall 1 \leq u \leq u' \leq p, 1 \leq v \leq v' \leq q$$

and an Anti-Monge matrix if :

$$c_{u,v} + c_{u',v'} \geq c_{u',v} + c_{u,v'} \quad \forall 1 \leq u \leq u' \leq p, 1 \leq v \leq v' \leq q$$

Remarque 63 (Full-Monge matrix).

The important case for our purpose is the case of equality when a matrix is both Monge and Anti-Monge, we will call this situation "Full-Monge" matrix.

$$c_{u,v} + c_{u',v'} = c_{u',v} + c_{u,v'} \quad \forall 1 \leq u \leq u' \leq p, 1 \leq v \leq v' \leq q$$

Although it was seldomly studied (see namely [63]), this last equality fits perfectly our purpose. The inequalities on the contrary, are common and can be met in various situations such as cumulative distribution functions, or copula theory.

Remarque 64 (Adjacent cells).

A straightforward but important derived property is the local adjacency cells equality : it is sufficient to satisfy the property of the Remark 63 on adjacent cells, to ensure the obtainment of a "Full-Monge" matrix behavior for the global set of cells i.e. :

$$c_{u,v} + c_{u+1,v+1} = c_{u+1,v} + c_{u,v+1} \quad \forall 1 \leq u \leq p, 1 \leq v \leq q$$

Remark 64 is a key property to study Monge matrices since it gives a direct $\mathcal{O}(pq)$ algorithm to verify if a matrix is full-Monge.

A fact remains that a question arises : which density functions verify the Full Monge property? The following Proposition 53 gives an interesting answer : all full Monge matrices derive from the density of an *indetermination* structure.

Propriété 53 (Full-Monge matrix is equivalent to *indetermination*).

A "full Monge matrix" necessarily represents an "indetermination coupling".

Démonstration.

Summing on u' and v' the equality of Remark 63 we straightforwardly obtain :

$$\sum_{u'} \sum_{v'} (c_{u,v} + c_{u',v'} - c_{u',v} - c_{u,v'}) = pq c_{u,v} + c_{\cdot,\cdot} - q c_{\cdot,v} - p c_{u,\cdot} = 0 \rightarrow c_{u,v} = \frac{c_{u,\cdot}}{q} + \frac{c_{\cdot,v}}{p} - \frac{c_{\cdot,\cdot}}{pq}$$

□

By summarizing properties of Full-Monge Matrices we get the following Theorem 24.

Théorème 24 (Full-Monge matrices).

Let π be a probability distribution on $(1, \dots, p) \times (1, \dots, q)$ then the following properties are equivalent.

1. π is a Full-Monge matrix
2. $\pi_{u,v} = \pi_{u,v}^+ = \frac{\mu_u}{q} + \frac{\nu_v}{p} - \frac{1}{pq}$

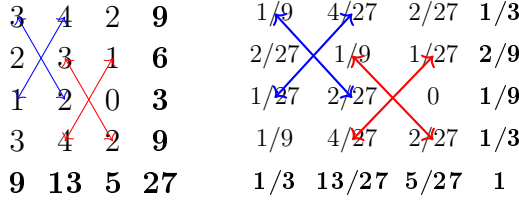


FIGURE C.1 – Example of an *indetermination* coupling (Statistical counting vs Probability forms)

3. π optimizes Problem 28 for some given margins
4. All 2×2 sub-tables $\{u, v, u', v'\}$ extracted from π have the same sum on their diagonal and anti-diagonal

Figure C.1 features the last assertion in Theorem 24 and justifies the \oplus notation assigned to *indetermination*. Indeed, if we take blue and red arrows we get the same resulting value. Using the contingency form :

$$\begin{aligned} \text{blue arrows} & : 3 + 2 - 1 - 4 = 0 \\ \text{red arrows} & : 3 + 2 - 4 - 1 = 0 \end{aligned}$$

Equality remains true for the probability form since we just have to divide the cell values by the total sum of the matrix (27 here).

C.3.2 A dependence to avoid matches

In the discrete case, Problem 28 is written as a projection of a prior guess $\mathcal{U} = \frac{1}{pq}$ (as already mentioned : the optimal solution without any initial information). As previously noticed \mathbb{L}^2 (Equation (C.1)) is the distance we project \mathcal{U} with, on the space of probability measures respecting the two margins μ and ν . Theorem 23 provides the form of the optimal solution $\pi^+ = C^+(\mu, \nu)$.

We notice that π^+ introduces a dependence between its margins, different from C^\times , the *independence* coupling.

Moreover, we notice that, unlike $\pi^+ = C^+(\mu, \nu)$, the Kullback-Leibler (or entropic) approach leads to *independence*. Indeed, Problem 27 can be formulated as a projection of \mathcal{U} on the same space of probability measures respecting the two margins μ and ν but using Kullback-Leibler. The solution of which is given by Theorem 22 : $\pi^\times = C^\times(\mu, \nu)$, generating no dependence between the margins, by definition.

To summarize, if $\mathcal{L}(\mu, \nu)$ is the space of probability measures respecting the two margins μ and ν we have :

$$\mathcal{U} \xrightarrow{KL} \mathcal{L}(\mu, \nu) = \pi^\times \quad \Bigg| \quad \mathcal{U} \xrightarrow{\mathbb{L}^2} \mathcal{L}(\mu, \nu) = \pi^+$$

Let us focus now on the *indetermination* coupling using an interpretation of the problem it solves. Indeed, removing constants, the cost function

$$\mathbf{M}^+ = pq \sum_{u,v} \left\{ \pi_{u,v} - \frac{1}{pq} \right\}^2$$

can be simply written :

$$\mathbf{M}^+ = \sum_{u=1}^p \sum_{v=1}^q \pi_{u,v}^2.$$

As we are minimizing \mathbf{M}^+ , the formulation of the problem precisely aims at reducing couple matchings. Indeed, if we independently draw two occurrences of $W = (U, V) \sim \pi$, the probability of getting a matching for a couple, that is to say $U_1 = U_2$ and $V_1 = V_2$ simultaneously, is nothing but $\pi_{U_1, V_1} \times \pi_{U_2, V_2} = \pi_{U_1, V_1}^2$.

C.3.3 Constructive definition of an *indetermination* matrix

Let us unfold hereafter the dependence introduced by the *indetermination*, associated to the \mathbb{L}^2 projection. We use the Full Monge property that characterizes *indetermination* (see Theorem 24 and which can be rewritten as :

$$\pi_{u,v}^+ - \pi_{u',v}^+ = \pi_{u,v'}^+ - \pi_{u',v'}^+$$

This last equality shows that the difference between two lines is the same whatever the column. We deduce a constructive definition of an *indetermination* law on $(1, \dots, p) \times (1, \dots, q)$, for this sake :

1. For the first line we fix an arbitrary distribution on the q columns : $(\Pi_{1,v}^+)_{1 \leq v \leq q}$.
2. For any other line u , we define $\Pi_{u,v}^+ = \Pi_{1,v}^+ + \frac{\Delta_u}{q}$, $\forall 1 \leq \dots \leq q$ where Δ_u is any real number such that $\Pi_{u,v}^+$ is always positive
3. To eventually obtain a probability matrix (summing up to 1), we set $T = \sum_{u=1}^p \sum_{v=1}^q \Pi_{u,v}^+$ and define : $\pi_{u,v}^+ = \frac{\Pi_{u,v}^+}{T}$ for any (u, v) together with $\delta_u = \frac{\Delta_u}{T}$ for any u .

There are as many *indetermination* matrices on pq elements as choices of a first line $(\pi_{1,v}^+)_{1 \leq v \leq q}$ and of increments $(\delta_1 = 0, \delta_2, \dots, \delta_p)$.

Remarque 65 (From lines to columns).

In the precedent construction, lines and columns played a different role. Obviously, any property is symmetric so that it can be easily transferred from u to v .

Remarque 66 (δ_u expresses the differences between margins).

From the precedent construction, the deduced margins are :

$$\pi_{u,\cdot}^+ = \pi_{1,\cdot}^+ + \delta_u, \forall 1 \leq u \leq p$$

where for any u , $\pi_{u,\cdot}^+ = \sum_{v=1}^q \pi_{u,v}^+$.

The division by q we introduced in the definition allows us to interpret δ_u as the difference between any margin to the first one.

We suppose the δ_u are positive ; following Remark 66 it amounts to saying that the first line corresponds to the minimal margin.

The more a δ_u is chosen high, the more probable the line u will be and the more the drawings of v on line u will be close to uniform on $1, \dots, q$. This behavior is compliant with the couple matching limiting property exposed in section C.3.2. It decreases couple matchings since when a first frequent u_1 is drawn, a conflict is probable with the second

drawn u_2 . To avoid a couple conflict the *indetermination* π^+ draws v_1 and v_2 as uniformly as possible.

A uniform margin on u returns to say $\delta = 0$. In that case, any line equals the first one and we have for any (u, v) :

$$\begin{aligned}\pi_{u,v}^+ &= \pi_{1,v}^+ \\ &= \frac{\pi_{\cdot,v}}{q} \\ &= \pi_{u,\cdot}\pi_{\cdot,v}\end{aligned}$$

meaning that we constructed *independence*. One can actually easily check that if μ is uniform and whatever ν is, $C^+(\mu, \nu) = C^\times(\mu, \nu)$, which explains the result.

On the contrary, the more both margins μ and ν differ from an uniform law and the more $\pi^+ = C^+(\mu, \nu)$ and $\pi^\times = C^\times(\mu, \nu)$ will differ ; this was already spotted in [10]. Using our new formalism, if μ is far from the uniform, the associated array δ will obviously take high values. On the corresponding lines with a high weight δ_u the influence of $\pi_{1,v}$ will disappear. As a consequence, the second variable v will be close to a uniform drawn on $1 \dots q$.

C.3.4 Drawing under *indetermination*

Given a probability law $\pi^+ = C^+(\mu, \nu)$ for two margins μ and ν we use the above constructive process of *indetermination* to propose a two steps method for drawing under π^+ . Besides, we enforce $\mu_1 \leq \mu_2 \leq \dots \leq \mu_p$. It amounts to order modalities according to their increasing probability and ends up being a renaming.

First, an unbalanced distribution $\pi_{1,v}^+$ adding up to μ_1 is extracted from π by reading its first line (we already supposed μ_1 is the smallest margin thanks to our order hypothesis). Then, we extract a list of corresponding computed $\delta_u = \mu_u - \mu_1$ to ensure that the differences between margins are respected. Eventually, a drawing is computed as follows :

1. We draw u through μ
2. We roll a loaded dice with the Bernoulli's skew : $I = \text{Be} \left(\frac{\delta_u}{\mu_u} \right)$
3. If we get 0, we draw v under the distribution $(\pi_{1,1}^+, \dots, \pi_{1,q}^+)$ else, having 1, we draw v uniformly hence under $\left(\frac{1}{q}, \dots, \frac{1}{q} \right)$.

In Proposition 54 we formally demonstrate that our new drawing realizes the *indetermination* law π^+ .

Propriété 54 (Drawing under *indetermination*).

The method we just built up does realize indetermination.

Proof.

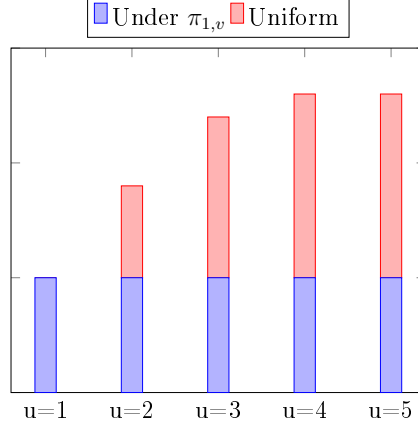


FIGURE C.2 – Distribution of v according to the margin u

We show that the resulting probabilities follow *indetermination*.

$$\begin{aligned}
\mathbb{P}(U = u, V = v) &= \mathbb{P}(V = v|U = u)\mathbb{P}(U = u) \\
&= \mathbb{P}(V = v|U = u, I = 0)\mathbb{P}(U = u)\mathbb{P}(I = 0) \\
&\quad + \mathbb{P}(V = v|U = u, I = 1)\mathbb{P}(U = u)\mathbb{P}(I = 1) \\
&= \pi_{1,v}^+ \mu_u \frac{\mu_1}{\mu_u} + \frac{1}{q} \mu_u \frac{\mu_u - \mu_1}{\mu_u} \\
&= \pi_{1,v}^+ \mu_1 + \frac{\mu_u - \mu_1}{q}
\end{aligned}$$

Now, we notice that the sum on u of those probabilities equals $\mathbb{P}(V = v)$ so that :

$$\begin{aligned}
\mathbb{P}(V = v) &= \nu_v = p\pi_{1,v}^+ \mu_1 + \frac{1 - p\mu_1}{q} \\
\text{or :} \quad \pi_{1,v}^+ \mu_1 - \frac{\mu_1}{q} &= \frac{\nu_v}{p} - \frac{1}{pq}
\end{aligned}$$

Replacing, we obtain the expected formula :

$$\mathbb{P}(U = u, V = v) = \frac{\mu_u}{q} + \frac{\nu_v}{p} - \frac{1}{pq}$$

□

The constructive method we just described creates a three steps process to realize *indetermination*; it leads to the Histogram C.2 presented just above.

This Histogram C.2 focuses on the distribution evolution. Hence, we observe that the higher μ_u is, the more uniform $V|U = u$ is. Indeed, $(\pi_{1,1}, \dots, \pi_{1,q})$ brings up a relative disequilibrium (it is arbitrary) but the complementary part is uniformly drawn. The proportion featured in Histogram C.2 is, obviously, dummy but it correctly features the evolution of the conditional probability depending on the margin. Furthermore, the sorting on μ_u , coherent with the uniform proportion, insists on the property that the important masses are at close-to-uniform.

Actually, this constructive process always behaves as a strategy to avoid "couple matchings". When having drawn a first margin u_1 with a small weight, u_2 will probably differ. Remember margins are fixed, the *indetermination* structure hides in those "small lines" any disequilibrium on $v : (\pi_{1,1}^+, \dots, \pi_{1,q}^+)$ that the margins introduced. On the contrary, having drawn a first margin u_1 with a high weight, u_2 has a high probability to be equal to u_1 ; there the close-to-uniform distribution of $V|u = u_1$ protects the couple from a matching ($v_1 = v_2$ is rare).

In the next section, dedicated to applications, we will leverage this property that *indetermination* ends up being a strategy to avoid couple matching while respecting margins.

C.4 Applications of *indetermination* structure

C.4.1 Deviation from *indetermination* : Janson Vegelius coefficient

In statistical analysis, given the values of two descriptive variables on a number n of individuals, an usual and important problem is to use a coefficient or index, measuring the correlation between the two variables.

Formally, U represents a first variable which characterizes individuals among p modalities (for instance the city where they are living, their socio-professional category, their ages, ...); a second variable V classifies them among q categories (or split them into q categories or classes).

Given realizations (U_1, \dots, U_p) and (V_1, \dots, V_q) , the categorization of n individuals, how do we measure the correlation between U and V ? Correlation typically means that the value of V is dependent on the value of U . Expressing it with "dependence" notion, we naturally define a deviation-to-independence coefficient (*i.e.* a departure from *independence* index), for instance : the χ_2 .

To do so, from the n realizations, of U we deduce an empirical margin μ counting the proportion of individuals in each modality :

$$\mu_u = \frac{\#\{i / U_i = u\}}{n}$$

similarly, an empiric margin ν out of the realizations of V and eventually an empiric margin π out of (U, V) .

The usual χ^2 can be defined as a square deviation to *independence* :

$$\chi^2(U, V) = \sum_{u=1}^p \sum_{v=1}^q \frac{(\pi_{u,v} - (\mu \otimes \nu)_{u,v})^2}{(\mu \otimes \nu)_{u,v}} \quad (\text{C.5})$$

which obviously happens to be null if and only if the empirical distribution π of the observed data is an independence coupling of the empiric margins.

Using a symmetric idea, a less used criterion, called "Janson-Vegelius Index", after the name of the inventors of this coefficient, who coined it in [52], [53] or [54] writes as a

deviation to *indetermination* :

$$JV(U, V) = \sum_{u=1}^p \sum_{v=1}^q \frac{(\pi_{u,v} - (\mu \oplus \nu)_{u,v})^2}{\sqrt{\frac{p-2}{p} (\sum_{u=1}^q \mu_u^2 + 1)} \sqrt{\frac{q-2}{q} (\sum_{v=1}^p \nu_v^2 + 1)}} \quad (\text{C.6})$$

and obviously is equal to zero if and only if the empirical π is an *indetermination* coupling of the empirical margins.

JV index, although its formulation, using contingency notations appears as non trivial, is actually just a classical cosine, or a Pearson's like correlation coefficient when rewritten in the "Mathematical Relational Analysis" Space. A list of papers which gathers some of the most important key features about the subject is [68], [64], [71], [80], [65], [66], [2].

The relational analysis space no longer encodes modalities but links between individuals. Two matrices X and Y of size $n \times n$ respectively associated to variables U and V are introduced as shown in Definition 60.

Définition 60 (Mathematical Relational Analysis notations).

Let (U_1, \dots, U_n) and (V_1, \dots, V_n) be two n probabilistic draws of $U \sim \mu$ and $V \sim \nu$ respectively. We define two associated symmetric $n \times n$ matrices X and Y by

$$\begin{aligned} X_{i,j} &= \mathbb{1}_{u_i=u_j}, \quad \forall 1 \leq i, j \leq n \\ Y_{i,j} &= \mathbb{1}_{v_i=v_j}, \quad \forall 1 \leq i, j \leq n \end{aligned}$$

Or in literal form :

- $X_{i,j} = 1$, if i and j share the same modality of variable U , $X_{i,j} = 0$ if not
- $Y_{i,j} = 1$, if i and j share the same modality of variable V , $Y_{i,j} = 0$ if not

To understand the notation, let us begin with some remarks about Definition 60. Basically, the two $\{0, 1\}$ matrices X and Y (which correspond in fact to two binary equivalence relations based on the drawn modalities) represent agreements and disagreements between the two variables on a same draw of size n ; they are symmetric with 1 values on their diagonal.

As expected, one can pass from the relational encoding to the usual contingency encoding as well as in the reciprocal way; those transfer formulas are demonstrated in the mentioned articles. Coming back to the JV index, those formulas enable us to write JV as a cosine in the relational space :

$$JV(U, V) = JV(X, Y) = \frac{\sum_{i=1}^n \sum_{j=1}^n \left(X_{i,j} - \frac{1}{p}\right) \left(Y_{i,j} - \frac{1}{q}\right)}{\sqrt{\sum_{i=1}^n \sum_{j=1}^n \left(X_{i,j} - \frac{1}{p}\right)^2 \sum_{i=1}^n \sum_{j=1}^n \left(Y_{i,j} - \frac{1}{q}\right)^2}} \quad (\text{C.7})$$

Calculations leading to Equation (C.7) from Equation (C.6) can be found in [68] or [67].

Instead to study in depth the differences between the two coefficients, $\chi^2(U, V)$ and $JV(U, V)$ we recall how they apply in clustering and refer to [10] for a more precise study.

C.4.2 Clustering problem dedicated to graphs

Clustering algorithms aim at getting a decomposition of graphs into classes, maximizing internal similarities as well as minimizing external ones. It is motivated by the fact that in any application, the connections are not random. Precisely, a triangle has a high chance of being a clique, as expressed in [41] : "two of your friends will have a greater probability of knowing one another than will two people chosen at random from the population, on account of their common acquaintance with you".

Various methods were proposed during the 20th. Due to the huge amount of papers available, we provide here an arbitrary selection while trying to highlight most of the subjects the scientific community is facing. A first option is to take as a fixed input the number K of classes we are looking for (although it is quite unrealistic for huge graphs), together with an associated distance (or dissimilarity index) and come up with a list of best representatives or "means" for each class. The output "means" tend to optimize the sum of distances from all vertices to their nearest mean. K-means algorithm whose idea goes back to the fifties (originated in [88] and rewritten by E.W. Forgy in [35]) typically illustrates this option. A second option, is to construct a local criterion w which assigns a weight $w_{i,j}$ to each (i, j) couple of vertices based on their similarity ; the more similar they are, the higher the criterion is. For instance, [56] affects the number of paths from i to j in the graph with an exponentially-decreasing factor according to their length. One then build up a global criterion W by summing up the local values $w_{i,j}$ if and only if i and j are in the same class as proposed in Problem 29.

When [41] is published, the community detection problem, though it is applied in various domains and contexts, still lacks a method to impartially evaluate and compare two results. Two years passed before the authors Girvan and Newman published a method to measure the global quality of a network clustering through the so-called modularity ([76]) : M^\times (Definition 63). The idea is to compare the number of edges in a community with the expected number in case of edges distributed without regard to the community they belong to. Having expressed a global measure for the quality of a clustering enabled the community to objectively compare their performances.

In fact, they were faced with several problems. First, finding a clustering which maximizes M^\times is a *NP*-hard problem and therefore forces the specialists to use heuristics like the original one in [24]. We can find out a list of such heuristics in [85], for instance among others : the use of spectral methods [34] as well as data structure tricks [96]. Some of those methods are gathered in the book written by Fortunato [36]. Furthermore M^\times is as plain as arbitrary since it introduces a resolution limit concern, first mentioned in [37] and which led to a research gate for improving the modularity (see in particular [25]). Either the authors decided to trick the modularity index such as in [83] or with the help of the recently introduced modularity density ([23]) or they completely changed the criteria using typically a so-called clustering coefficient [75]. Eventually, the definition itself of clustering is subject to discussion, shall communities be allowed to overlap ? some articles allow that and define fuzzy communities (see [43]).

Basically, out of the literature, three subjects stand out : algorithms to optimize a criterion, the definition of a criterion and the extension of the original problem itself.

This section both justifies the construction M^\times as a comparison to one of the two canonical equilibria which reduces the available information and introduces the second

canonical construction M^+ (see Definition C.9). Let us first start with some usual definitions for a graph :

Définition 61 (Weighted graph).

A weighted graph G contains n vertices $1 \leq i \leq n$, which are connected each other through edges (i, j) linked with weights $a_{i,j}$ (representing a weighted incidence matrix). We also introduce the total weight $2M = \sum_{i,j} a_{i,j}$.

Graph clustering is devoted to the research of classes, groupings, modules or cliques (whatever we call them) within a graph. They are defined through an equivalence relation as specified in Definition 62 :

Définition 62 (Graph clustering).

Let us call X , a matrix representation of a binary equivalence relation, the result of the clustering of a graph G . Then $X_{i,j}$ equals 0 or 1 and equals 1 if and only if the two vertices i and j are in the same class for X , and 0 if not.

Problème 29 (Generic clustering problem).

$$\begin{aligned} \max_X \quad & W(w, X) = \sum_{i=1}^n \sum_{j=1}^n w_{i,j} X_{i,j} \\ \text{subject to :} \quad & X \quad \text{is an equivalence relation} \end{aligned}$$

First let us remark that, as notably spotted in [65], [68], [80] an equivalence relation constraint can be written as :

- $X_{i,i} = 1, \forall 1 \leq i \leq n$ (reflexivity)
- $X_{i,j} = X_{j,i}, \forall 1 \leq i, j \leq n$ (symmetry)
- $X_{i,j} + X_{j,k} - X_{j,k} \leq 1, \forall 1 \leq i, j, k \leq n$ (transitivity)

Remarque 67.

Thanks to this relational formulation, appearing in the model presented just above, it is not mandatory to fix in advance the number K of clusters which is expected, this latter is got as a direct and final result of the computing process.

Modularity criterion of Newman-Girvan based upon "Independence"

The original, famous, and well known Newman-Girvan's presentation of a global criterion for graphs clustering, see [41] or [76], has been introduced inside the generic Louvain's algorithm together with a global cost called "Modularity" defined by :

Définition 63 (Modularity).

Given an equivalence relation $X_{i,j}$ (also called partition) and a graph G with weighted function $a_{i,j}$ on its edges, the global modularity N.G. criterion amounts to :

$$W^\times(G, X) = \frac{1}{2M} \sum_{i,j} \left[a_{i,j} - \frac{a_{i,\cdot} a_{\cdot,j}}{2M} \right] X_{i,j} \quad (\text{C.8})$$

Let us first remark that the original modularity W^\times is nothing but our generic global cost function defined through Problem 29 with :

$$w_{i,j} = w^\times(G)_{i,j} = \frac{a_{i,j}}{2M} - \frac{a_{i,\cdot} a_{\cdot,j}}{(2M)^2}$$

and the local gain $w^\times(G)_{i,j}$ to put two vertices in the same class is the local deviation to *independence*. Indeed, using Definition 61, we know that $\pi_{i,j} = \frac{a_{i,j}}{2M}$ can be seen as a probability measure on $\{1 \dots n\}^2$ with margins $\mu_i = \frac{a_{i,\cdot}}{2M}$ so that w^\times rewrites :

$$w^\times(G)_{i,j} = 2M (\pi_{i,j} - \mu_i \mu_j)$$

and does express itself as a canonic *deviation to independence* criterion.

Extended logical modularity based upon "*indetermination*"

Problem 29 basically represents an extension of the already introduced "Modularity criterion" towards a generic criterion based on a local input one.

We suggest an expression $w^+(G)_{i,j}$ which represents a deviation to *indetermination*. It will be used as a local cost function in Problem 29 leading to a slightly different global formula $W^+(G, X)$ to optimize locally :

$$w^+(G)_{i,j} = a_{i,j} - \frac{a_{i,\cdot}}{n} - \frac{a_{\cdot,j}}{n} + \frac{2M}{n^2}$$

Symmetrically as w^\times , it ends up being a canonic *deviation to indetermination* criterion. Indeed, keeping $\pi_{i,j} = \frac{a_{i,j}}{2M}$, w^+ rewrites :

$$w^+(G)_{i,j} = 2M * \left(\pi_{i,j} - \frac{\mu_i}{n} - \frac{\mu_j}{n} + \frac{1}{n^2} \right)$$

The global criterion being :

$$W^+(G, X) = \sum_{i,j} \left[a_{i,j} - \frac{a_{i,\cdot}}{n} - \frac{a_{\cdot,j}}{n} + \frac{2M}{n^2} \right] X_{i,j} \quad (\text{C.9})$$

It has been shown in [29] that those two costs play a specific role as deviations from canonic couplings. Inserted in the generic Louvain algorithm, they allow to get clustering solutions $X_{i,j}$ which slightly differ from each other, but not that much (see for instance [25], where some comparisons of the criteria of Definition 63 and Definition 63 are provided, in practical contexts). We will not specifically insist here on it, but rather propose two new applications of the indetermination structure.

C.4.3 Guessing or spy problem

Original problem

In cryptography, a message u in a finite alphabet \mathcal{U} of size p is typically sent from Alice to Bob while a spy whose name is Charlie tries to intercept it. A common strategy for Alice to communicate efficiently and secretly with Bob consists in encoding the message

using a couple of keys (public, private) for each character or a symmetric encryption which only requires one shared key between Alice and Bob. The literature concerned with the encryption method to choose according to the situation is diverse, the most-used standard is Advanced Encryption Standard described in various articles. Possibly, Charlie observes an encrypted message V in a second finite alphabet \mathcal{V} of size q which is a function of the message u .

Related to the cryptography situation, the guessing problem quoted hereafter as Problem 30 was first introduced in the article [69]. While in cryptography Charlie tries to decode a sequence of messages, the guessing problem focuses on decoding a unique message. Furthermore, the initial version of Problem 30 is limited due to the lack of access to any prior knowledge by the spy. A second version described in subsection C.4.3 will introduce a variable V correlated to the message, this second variable will code but will not be limited to code the encrypted message conveyed from Alice to Bob.

Though, the original version provides a collection of results that easily transpose themselves to the more realistic one. Let us formalize this simplest situation : U is a random variable which takes its values in a finite alphabet \mathcal{U} and follows the probability law $\mathbb{P}_U = \mu$. A sender "Alice" generates a sequence of independent messages under μ .

Problème 30 (Original Guessing Problem or Spy Problem).

When Alice sends a message $U = u$ to Bob, the spy Charlie must find out the value u of the realization. He has access to a sequence of formatted questions for any guess \tilde{u} he may have : "Does u equal \tilde{u} ?" for which the binary answer is limited to "yes/no".

Définition 64 (Original Strategy).

A strategy $S = \sigma$ of Charlie is defined by an order on \mathcal{U} representing the first try, the second and so on until number p . It can be deterministic or random : we quote \mathbb{P}_S its probability law.

Besides, for a given position $i \in [1, p]$, $\sigma[i]$ is the element in \mathcal{U} corresponding to the i -th try.

In [69], a measure of performance is associated to any fixed strategy σ of Charlie. It basically computes the ρ moment of G which counts the number of trials needed by Charlie to find out which message u was sent. We shall add another performance measure later on.

Définition 65 (Performance measure).

The function $G(\sigma, u)$ is defined as the number of questions required to eventually obtain a "yes" in Problem 30 when Charlie proposed the order $S = \sigma$ and Alice generated the message $U = u$. It can be a random variable even for a fixed u as soon as S is. $G(S, U)$ is a random variable and whose formal definition is :

$$G(\sigma, u) = \sum_{i=1}^p i \mathbb{1}_{\sigma[i]=u}$$

We eventually define the efficiency of a strategy S by a measure of the ρ -moment of $G(S, U)$ under the independent coupling of $S \sim P_S$ and $U \sim P_U$.

$$\|G(S, U)\|_\rho = [\mathbb{E}_{(S, U) \sim P_S \otimes P_U} (G(S, U)^\rho)]$$

The definition of $G(\sigma, u)$ precisely codes the number of trials before Charlie discovers the message u . For instance, with an alphabet $\mathcal{U} = \{a, b, c, d\}$, if the message is $u = c$ and the strategy σ of the spy consists in the order (b, c, a, d) (meaning he first proposes message b then c, \dots) we have :

$$\begin{aligned} G(\sigma, u) &= \sum_{i=1}^p i \mathbb{1}_{u=\sigma[i]} \\ &= 1 \cdot \mathbb{1}_{u=b} + 2 \cdot \mathbb{1}_{u=c} + 3 \cdot \mathbb{1}_{u=a} + 4 \cdot \mathbb{1}_{u=d} \\ &= 2 \cdot \mathbb{1}_{u=c} \\ &= 2 \end{aligned}$$

It has been proven in the same article [69] a natural result : provided $\mathbb{P}_U = \mu$ is known, the best strategy consists in proposing answers under the deterministic order σ of decreasing probabilities. That is to say we first propose the message which appears most often, then the second most probable and so on :

$$\mu_{\sigma[p]} \leq \dots \leq \mu_{\sigma[1]}$$

Besides they demonstrated a lower bound on the average number of questions which no strategy can break as it is specified in Theorem 25.

Théorème 25 (Lower bound on the efficiency).

The minimal expected number of questions to solve Problem 30 verifies the inequality :

$$\min_S \|G(S, U)\|_\rho \geq (1 + \log(p))^{-\rho} \left[\sum_{u \in \mathcal{U}} \mathbb{P}(U = u)^{\frac{1}{1+\rho}} \right]^{1+\rho}$$

Proof.

We won't gather the whole demonstration but a glimpse. In [5], $\|G(S, U)\|_\rho$ is written as the integral of some concave function with respect to some probability measure Q . By Jensen's inequality, a lower bound is obtained whatever Q is. The multiplicative factor comes from the inequality

$$\sum_{u \in \mathcal{U}} \frac{1}{G(\sigma, u)} = \sum_{i=1}^p \frac{1}{i} \leq 1 + \log(p)$$

Eventually, selecting a special value for Q leads to the result. □

A practical application of Theorem 25 is to provide a guarantee on the average time a spy will take to guess a message. The sender, on its side, is motivated by maximizing the lower bound.

Extended problems

As announced beforehand, Charlie has now access to an observed random variable V correlated with the sent message U . In the common cryptography problem it would be the encrypted message that Charlie observes when Alice sends a message, hence a

deterministic function of the message U . Here, we generalize and suppose it can also contain, for instance, the size of the message, the frequency channel used, the sender's location, the receiver, or any physical information a spy can have access to. Finally, the added information, more or less useful, is encoded into a random variable V whose values belong to a finite alphabet \mathcal{V} of size q . Obviously, V is correlated with the message U but we do not suppose their link is deterministic as it would be for an encryption.

As mentioned in the article [5], Charlie now chooses its strategy according to the value taken by the observed second variable V : he typically adapts himself to the conveyed encryption. The probability law of the couple (U, V) is quoted $\mathbb{P}_{U,V} = \pi$ while its margins are $\mathbb{P}_U = \mu$ and $\mathbb{P}_V = \nu$.

The gain function now expresses as $G(S, U|V)$: we purposely use the notation symbol "knowing V " to insist on the fact that V is known when the spy decides the strategy he uses. Eventually, for any observed value $V = v$, an original strategy S_v (see Definition 64) is built up leading to an original gain function $G(S_v, U)$ that is to say :

$$G(S, U|V) = \sum_{v \in \mathcal{V}} G(S_v, U) \mathbb{1}_{V=v}$$

The same article comes up with a generalization of Proposition 25 that we report here :

Théorème 26 (Generalized lower bound on the efficiency).

For any strategy, the average time to reconstruct the message always respects the lower bound :

$$\mathbb{E}_{(S,U,V) \sim \mathbb{P}_{S,U,V}} [G(S, U|V)^\rho] \geq (1 + \log(p))^{-\rho} \sum_{v \in \mathcal{V}} \left[\sum_{u \in \mathcal{U}} (\pi_{u,v})^{\frac{1}{1+\rho}} \right]^{1+\rho}$$

Proof.

The result is plain given that, as we already noticed, S decomposes into original strategies S_v for any fixed v . Hence, for any v , the local or original assigned strategy obeys Proposition 25 which directly leads to the result. \square

Logical *indetermination* as a lower bound

Let us move away from the literature and measure Charlie's performance by its probability to find out after one trial the message u Alice sent. It is a reasonable measure as, if a sequence of messages is sent, we may have to jump from one to the following after only one trial.

Définition 66 (one-shot performance).

For a given strategy S , we define the following performance measure as the probability to find out the value u after one trial, formally :

$$M(S, U, V) = \mathbb{P}_{(S,U,V) \sim \mathbb{P}_{S,U,V}} (S[1] = U)$$

Remarque 68 (Generalized one-shot performance).

One could easily introduce a measure whose name could be "k shots performance" evaluating the probability to guess after up to k trials. We would hence notice that if $k \geq p$ then the "k shots performance" equals 1 for any sensitive strategy. We will not detail it further here.

We suppose as for the original optimal strategy that the spy has access to the distribution $\mathbb{P}_{U,V} = \pi$. We can imagine he previously observed the non-encrypted messages in a preliminary step.

Two strategies immediately stand out :

1. S_{max} : systematically returns at v fixed (observed by hypothesis), the u associated with the maximal probability on the margin $\mathbb{P}_{U|V=v}$
2. S_{margin} : returns at v fixed a random realization of x under the law $\mathbb{P}_{U|V=v}$

While we know S_{max} is the best strategy in case the performance measure is $\|G(S, U)\|_\rho$, we have no guarantee it maximizes the one-shot performance. Furthermore, S_{margin} is by far harder to cope with for the sender who cannot easily prevent random conflicts. Consequently we come back to the reduction of couple matchings, whose "indetermination coupling", we know, prevents us against. Let us unfold this remark hereafter.

The one-shot performance of the two strategies is given by :

$$M(S_{max}, U, V) = \sum_{v \in \mathcal{V}} \nu_v \left[\max_{u \in \mathcal{U}} \pi_{u|V=v} \right] \quad (\text{C.10})$$

$$M(S_{marge}, U, V) = \sum_{u \in \mathcal{U}} \sum_{v \in \mathcal{V}} \nu_v (\pi_{u|V=v})^2 \quad (\text{C.11})$$

Let us suppose, commendable task if any, that Alice wants to minimize Charlie's one-shot performance. We also suppose that the margins μ on U and ν on V are fixed. It is a common hypothesis : the alphabet \mathcal{U} in which the messages are composed typically respects a distribution on letters ; variable V on its own, if it represents frequencies for instance may have to satisfy occupation weights on each channel.

Concerning the strategy S_{marge} we have the two bounds :

$$\frac{\|\pi\|_2^2}{\min_{v \in \mathcal{V}} \nu_v} \geq M(S_{marge}, U, V) \geq \frac{\|\pi\|_2^2}{\max_{v \in \mathcal{V}} \nu_v} \quad (\text{C.12})$$

with

$$\|\pi\|_2 = \sqrt{\sum_{u \in \mathcal{U}} \sum_{v \in \mathcal{V}} (\pi_{u,v})^2}$$

We notice using Equation (C.3) that

$$\sum_{u \in \mathcal{U}} \sum_{v \in \mathcal{V}} \mathbf{M}^+(\pi_{u,v}) = pq \|\pi\|_2^2 - 1$$

It shows that studying the guessing problem brings us back to Problem 26 associated with cost \mathbf{M}^+ whose solution is given by the *indetermination* coupling of the margins.

Formally, using the coupling

$$\mathbb{P}_{U,V}^+ = C^+(\mathbb{P}_U, \mathbb{P}_V) = \frac{\mu_u}{q} + \frac{\nu_v}{p} - \frac{1}{pq} = \pi_{u,v}^+$$

guarantees an efficient reduction of conflicts (see section C.3.2) and eventually a weak one-shot performance as expressed in Equation (C.12).

A coupling to fight Charlie who uses the strategy S_{marge} consists for Alice in dispatching the messages U among V according to the logical *indetermination*. Typically she would use the constructive drawing exposed in subsection C.3.4.

Furthermore, the lower bound of the one-shot performance given by Equation C.12 is minimal under the *indetermination* π^+ so that regardless of the coupling π actually used, we always have :

$$M(S_{marge}, U, V) \geq \frac{\|\pi^+\|_2^2}{\max_{v \in \mathcal{V}} \nu_v} \quad (\text{C.13})$$

This optimality of *indetermination* fits the observation provided on Figure C.2 : on any visible information V for Charlie, messages U are as evenly distributed as possible which reduces the probability of matching the its first guess.

Eventually, a classic norm inequality allows us to deduce that the one-shot performance of S_{marge} limits from below as well as from above the one-shot performance of S_{max} according to the inequality valid for any couple probability law on (U, V) :

$$M(S_{marge}, U, V) \leq M(S_{max}, U, V) \leq \sqrt{M(S_{marge}, U, V)}$$

C.4.4 Tasks partitioning

Task partitioning problem is originally introduced in [18] where the authors provide a lower bound on the moment of the number of tasks to perform. Let us follow the gathering work of [59] where they also coin a generalized task partitioning problem basically adapting it as a special case of the guessing problem.

Formally, we begin with the original problem of tasks partitioning : a finite set \mathcal{U} of tasks size of which is quoted p is given together with an integer $q \leq p$. The problem consists in creating a partition $\mathcal{A} = (\mathcal{A}_1, \dots, \mathcal{A}_q)$ of \mathcal{U} in q classes to minimize the number of tasks to perform, knowing that if one needs to perform a task $u \in \mathcal{A}_i$, it is mandatory to launch simultaneously the whole subset of tasks included within \mathcal{A}_i .

Practically, a task $U = u$ to perform is randomly drawn from \mathcal{U} under a probability distribution $\mathbb{P}_U = \mu$ representing the tasks frequencies. As any task, the task u to perform is assigned to a unique class $\mathcal{A}_{i(u)}$ of the arbitrary partition. Hence, $A(u) = |\mathcal{A}_{i(u)}|$ counts the number of tasks to perform. Precisely, one plays on the partition knowledge to perform, in average, as few tasks as possible.

Similarly to the guessing problem, the performance of a partition \mathcal{A} is estimated through the ρ -moment of $A(U)$, formally $\mathbb{E}_U [A(U)^\rho]$. Moreover, the authors show in [59], quite similarly as for Theorem 25 that we have :

$$\min_{\mathcal{A}} \mathbb{E}_{U \sim \mu} [A(U)^\rho] \geq \frac{1}{q} \left[\sum_{u \in \mathcal{U}} (\mu_u)^{\frac{1}{1+\rho}} \right]^{1+\rho} \quad (\text{C.14})$$

which expresses a minimum average number of tasks to perform whatever the partition is.

Inspired by the general guessing problem, they extend the task partitioning problem. Let us introduce here this generalized version, in which we are no longer interested in

minimizing the number of tasks to perform but rather in reducing the number of tasks before a selected (or a chosen) task u .

Indeed, in the first version, as soon as u is drawn, an arbitrary rule imposes to perform the whole subset $\mathcal{A}_{i(u)}$ leading to realize $A(u)$ tasks. In the new version, tasks are performed sequentially in $\mathcal{A}_{i(u)}$ according to a global strategy S that can be deterministic or random.

Typically, tasks may consists in a signatures flow which an administration requires while q would be the number of workers dedicated to perform those signatures on incoming documents. A worker can be given the entitlement to perform several signatures, assistants usually do. In that case, the partition encodes the assignments of tasks to workers. When a worker $V = v$ is assigned a document, the depositor waits until the signature. Then the worker follows his own strategy S_v to sign his assigned documents, meaning he can always follow the same order leading to a deterministic strategy or change every day leading to a random strategy.

With a global strategy S which gathers the workers' strategies S_v , $\forall 1 \leq v \leq q$ and for a task u to perform, the performance of a partition \mathcal{A} is measured using

$$N_{S,\mathcal{A}}(u)$$

which represents the number of tasks performed before the intended task u (u included). A lower bound is provided in the paper [59].

Let us now suppose the keys $1 \leq v \leq q$ are associated with offices that must perform a proportion ν_v of the incoming tasks which still follow a distribution μ . It actually appears as a sensible problem where a manager would have to distribute in advance tasks among teams according to the usual observed distribution of tasks and a list of available teams with their capacities.

Besides, we suppose each team uses the strategy S_{marge} to perform tasks, meaning they randomly perform one according to their margin theoretical distribution ; for a document signing, they randomly sign one.

Remarque 69 (Concrete estimated distribution).

In any of the previous applications, for spy as well as for tasks, we are dealing with probabilities. Actually, we send a finite and integer number n of messages and we similarly distribute a finite number n of tasks.

Moreover, \mathcal{U} and \mathcal{V} are finite. Eventually, for any $u \in \mathcal{U}$ and $v \in \mathcal{V}$, an integer number $n_{u,v}$ of tasks is associated corresponding (in the spy problem) to the number of same letters u sent using channel v .

To convert $n_{u,v}$ into a probability measure, one only needs to divide by n and define :

$$\pi_{u,v} = \frac{n_{u,v}}{n} \tag{C.15}$$

Reciprocally, given a probability measure, one will draw n messages according to π . As n increases, it will approximate the theoretical distribution better and better.

Eventually, we notice that we do not have an explicit formula to optimize any of the two transport problems (Problem 27 or Problem 28) with an additional hypothesis of π respecting the fraction-of-integers-form given in Equation (C.15). Furthermore, indetermination or independence optimizes with a relaxed hypothesis (values in \mathbb{R}) and thus provides a better solution ; to be approached by drawings.

From now on, we can rewrite our task partitioning problem under the form of a guessing problem :

- $V = v$, formerly corresponds to a worker, now it represents the information the spy has access to
- $U = u$, formerly represents a task to perform, now it represents a sent message
- $S = \sigma$, formerly represents the order in which tasks are performed, now it represents the order in which Charlie proposes his guesses

Under this formalism, we are interested in measuring the probability $M(S, U, V)$ of executing u first as an extended application of the one-shot performance of Definition 66 and we have :

$$M(S, U, V) \geq \frac{\|\pi\|_2^2}{\max_{1 \leq v \leq q} \nu_v} \geq \frac{\|\pi^+\|_2^2}{\max_{1 \leq v \leq q} \nu_v} \quad (\text{C.16})$$

This inequality provides a lower bound for any distribution of the tasks among the team, no distribution can generate a worst "one-shot probability" of satisfying the intended task.

In task partitioning actually, each u is uniquely associated to a worker $v = i(u)$ so that the random variable representing the worker is deterministic conditionally to U . Yet, it is a reducing case of the guessing problem where V is random.

Remarque 70 (Splitting mass).

Eventually, we notice that having V random conditionally on U is a generalization of task partitioning along the same lines the Monge-Kantorovith problem was an extension of the one dimension Monge problem : we allow the mass splitting possibility, since a task may be randomly assigned among several workers.

In task partitioning problem using a partition \mathcal{A} instead of V (hence allowing no mass splitting), we notice $\mathbb{P}_Y = \nu$ is not properly defined. Let us extract it from the partition \mathcal{A} by providing each worker with a probability which sums up the probabilities of the tasks he has to perform. Formally, we define $V(\mathcal{A})$ to be a random variable whose probability is :

$$\nu_v^{\mathcal{A}} = \mathbb{P}_{V(\mathcal{A})}(V(\mathcal{A}) = v) = \sum_{u \in \mathcal{A}_v} \mu_u$$

together with the couple probability :

$$\pi_{u,v}^{\mathcal{A}} = \mathbb{P}_{U, V(\mathcal{A})}(U = u, V(\mathcal{A}) = v) = \mu_u \mathbb{1}_{u \in \mathcal{A}_v}$$

It enables us to deduce that the one-shot probability of satisfying the intended task for a partitioning problem accepts as a lower bound :

$$M(S, U, \mathcal{A}) = M(S, U, V(\mathcal{A})) \geq \frac{\|\pi^{\mathcal{A}}\|_2^2}{\max_{1 \leq v \leq q} \nu_v^{\mathcal{A}}} \geq \frac{\|C^+(\mu, \nu^{\mathcal{A}})\|_2^2}{\max_{1 \leq v \leq q} \nu_v^{\mathcal{A}}}$$

Indeed, a partition problem appears as a particular coupling of U and $V(\mathcal{A})$ (where $V(\mathcal{A})|U$ is deterministic) and no coupling can be worse than $C^+(U, V(\mathcal{A}))$.

A direct application is that no office affectation should provide a worse one-shot performance...

The efficiency of *indetermination* coupling in guessing problem as well as in task partitioning directly comes from its ability to reduce couple matchings. Either it prevents the spy from discovering the message or it provides a worst strategy by preventing a task from being performed.

As mentioned in the introduction, we shall see in the next section that a continuous *indetermination* notion exists and we will construct it using both prior guess and computations.

C.5 Continuous *indetermination*

C.5.1 Coupling function

This chapter introduces a continuous *indetermination* coupling that, to the extent of our knowledge, was introduced in [63], but never studied as such. We start with two marginals probability measures μ and ν on two segments $[a, A]$, $[b, B]$ and our goal is to define a probability measure on $[a, A] \times [b, B]$ which extends *indetermination* in a continuous space. Introducing some notations, we suppose that μ and ν respectively have a density measure quoted f and g together with an associated cumulative distribution function F and G .

The coupling function notion of Definition 57 operates on the discrete weights μ_u and ν_v whose transposition appears to be $f(u) du$ and $g(v) dv$ for any $u \in [a, A]$ and $v \in [b, B]$. Using that transposition, we extend the definition of a coupling function in a continuous space

Définition 67 (Coupling functions (continuous)).

μ and ν being probability laws on segments $[a, A]$ and $[b, B]$, a coupling function C operates on their density (f, g) .

$C(f, g)$ represents a density for a probability measure on the product space whose margins precisely are μ and ν .

A typical continuous (as well as discrete) coupling function is the independance quoted C^\times (we extend here the discrete notation) which generates an eligible density under the formula :

$$C^\times(f, g)(u, v) = f(u)g(v), \quad \forall u \in [a, A], \quad \forall v \in [b, B]$$

The \otimes notation is on purpose since it corresponds to the usual continuous *independence* coupling. More precisely, the density of an *independence* coupling of the two margins μ and ν will be nothing but $C^\times(f, g)$.

We would like to define a continuous version of the *indetermination* coupling. As often, when transposing a concept, we can either obtain it through computations or using a prior guess. We will follow the two approaches. Let us first propose a prior guess.

Définition 68 (Continuous *indetermination* density).

f and g being densities of two probability measures on $[a, A]$ and $[b, B]$ respectively, C^+ operates on the couple (f, g) with the formula :

$$C^+(f, g)(u, v) = \frac{f(u)}{B-b} + \frac{g(v)}{A-a} - \frac{1}{(A-a)(B-b)}$$

The equation comes from an adaptation of the discrete one, although this adaptation seems coherent, at that stage no guarantee is given on its truthfulness neither on its construction. Let us begin with computing the margin for a fixed $u \in [a, A]$ (it is obviously symmetric for v) :

$$\begin{aligned} \int_{v=b}^B C^+(f, g)(u, v) \, du \, dv &= \int_{v=b}^B \left(\frac{f(u)}{B-b} + \frac{g(v)}{A-a} - \frac{1}{(A-a)(B-b)} \right) \, du \, dv \\ &= f(u) \, du + \frac{1}{A-a} \, du - \frac{1}{(A-a)} \, du \\ &= f(u) \, du \end{aligned}$$

It appears actually that margins are as expected. Yet, they must also be continuous densities and therefore positive, which is by the way, not always true for $C^+(f, g)$. As in the discrete case an additional hypothesis is therefore required :

$$\min_u f(u) + \min_v g(v) \geq 1 \tag{C.17}$$

this last equation being the continuous version of Hypothesis C.4.

Eventually, applied on any couple of margins whose densities respect Equation (C.17), the continuous C^+ of Definition 68 generates an eligible density which is similar to the discrete optimal coupling function which solves Problem 28. Let us unfold this remark by defining a continuous version of Problem 28.

First, to simplify any future computation, we pose $a = b = 0$ and $A = B = 1$, converting any formula will be done using a dedicated affine transformation.

We use transport problem to validate our prior guess, transposing in the continuous space the discrete Minimal Transport problem using square deviation as a cost function :

Problème 31 (Minimal Trade Problem).

$$\begin{aligned} \min_{\pi} \quad & \int_0^1 \int_0^1 \pi^2(u, v) \, dv \, du \\ \text{under constraints :} \quad & \int_{v=0}^1 \pi(u, v) \, dv \, du = \mu_u \, du \\ & \int_{u=0}^1 \pi(u, v) \, dv \, du = \nu_v \, dv \\ & \int_0^1 \int_0^1 \pi(u, v) \, du \, dv = 1 \\ & \pi \geq 0 \end{aligned}$$

We then add Inequality (C.17), which, as Hypothesis (C.4) guarantees that our prior guess function is the density of a probability law. Hence, as well as restricting ourselves to $[0, 1]$ we restrict ourselves to margins respecting the condition.

Propriété 55.

Under the ad hoc Hypothesis (C.17), the solution of Problem 31 is nothing but our prior

guess continuous coupling function of Definition 68 applied to the margins f and g , formally, the solution density function is :

$$C^+(f, g)(u, v) = c_{f,g}^+(u, v) = (f(u) + g(v) - 1)$$

Démonstration.

We use Definition 68 to efficiently solve Problem 31 by noticing that

$$\int_{u=0}^1 \int_{v=0}^1 [\pi(u, v) - (f(u) + g(v) - 1)]^2 du dv \geq 0$$

rewrites (using margins constraints) :

$$\begin{aligned} \int_{u=0}^1 \int_{v=0}^1 \pi^2(u, v) du dv &\geq \int_{u=0}^1 \int_{v=0}^1 (-f^2 - g^2 - 1 - 2\pi - 2fg + 2f\pi + 2g\pi + 2f + 2g) du dv \\ &= \int_{u=0}^1 \int_{v=0}^1 (-f^2 - g^2 - 1 - 2 - 2fg + 2f^2 + 2g^2 + 2 + 2) du dv \\ &= \int_{u=0}^1 \int_{v=0}^1 (f^2 + g^2 - 2fg + 1) du dv \\ &= \int_{u=0}^1 \int_{v=0}^1 (f + g - 1)^2 du dv \end{aligned}$$

At the end, using Hypothesis (C.17) we notice that $c_{f,g}^+(u, v) = (f(u) + g(v) - 1)$ is eligible as a density function since always positive. Margins constraints are also satisfied as computed right after the introduction of continuous C^+ . \square

Eventually, following our computations, and given two random variables on $[0, 1]$: $U \simeq \mu, (f, F)$ and $V \simeq \nu, (g, G)$ we define a third variable W on the product space $[0, 1] \times [0, 1]$. W respects the two margins probability laws μ, ν and is called *indetermination* coupling of U and V quoted $U \oplus V$. Its density is $h_{f,g}^+ = C^+(f, h)$ while its cumulative distribution function is quoted $H_{F,G}^+$ and will be computed hereafter.

To compute the cumulative distribution function associated to the just expressed density $h_{f,g}^+$ we only have to apply a quick integration of the density. It leads to a second characterization of an *indetermination* coupling.

Propriété 56 (*indetermination* cumulative distribution function).

Given two random variables U and V whose densities f, g satisfy Hypothesis (C.17), the cumulative distribution function associated to their *indetermination* coupling $U \oplus V$ is :

$$H_{F,G}^+ = vF(u) + uG(v) - uv$$

.

C.5.2 Constructive method for adapted margins

Beforehand, we assumed margins are respecting Hypothesis (C.17), we propose here a method to construct adapted margins out of any couple.

Propriété 57 (Constructive margins).

A couple (f, g) of densities fulfills Hypothesis (C.17) if and only if, it exists an $\alpha, 0 \leq \alpha \leq 1$ and a couple of densities (r, s) such that :

$$f = (1 - \alpha)r + \alpha \quad \text{and} \quad g = \alpha s + 1 - \alpha \quad (\text{C.18})$$

Démonstration.

Let us first suppose (f, g) is under this form, then we notice, $\min f \geq \alpha$ as well as $\min g \geq 1 - \alpha$, hence, $\min f + \min g \geq 1$ so that the condition is satisfied.

Now, if the condition is respected, we define $\alpha = \min f$ and have $g \geq 1 - \alpha$. If $\alpha = 0$ or $\alpha = 1$ then, respectively, G or F is uniform so that the corresponding coupling is *independence* and condition is degenerated. If not, $0 < \alpha < 1$ and we can write :

$$f = (1 - \alpha) \frac{f - \alpha}{1 - \alpha} + \alpha$$

together with :

$$g = \alpha \frac{g - (1 - \alpha)}{\alpha} + (1 - \alpha)$$

Quoting $r = \frac{f - \alpha}{1 - \alpha}$ and $s = \frac{g - (1 - \alpha)}{\alpha}$, we have $0 \leq r, s \leq 1$ and $\int r = \int s = 1$. It precisely shows that (r, s) is a couple of densities on $[0, 1]$. \square

Using the last proposition, we can easily construct a couple of margins on which an *indetermination* coupling is feasible. Though, as mentioned during the proof, if $\alpha \in \{0, 1\}$ then f or g is an uniform density for which *indetermination* and *independence* merge one with another.

C.6 Conclusion

This paper gathers a list of properties as well as applications to various problems of a canonical coupling called *indetermination*.

In section C.2 we introduced the notion of coupling function together with two particular examples : *independence* and *indetermination*. The construction of both showed they minimize the information the fixed margins constraints convey into the joined law. This no-information property encouraged us to coin graph clustering criterion out of each equilibrium since a criterion usually compare the observed graph with the expected "null graph respecting vertices degrees".

Section C.3 focuses on the study of *indetermination*. After the usual Monge property, we proposed a constructive decomposition of an *indetermination* matrix and explained how it helped to avoid couple matchings. Furthermore, we noticed this property was already expressed in the cost M^+ of the problem it solves by writing it again as the expected number of couple matchings.

Knowing it is the best construction to reduce couple matchings led us to section C.4 in which that property was used in two applications (Guessing Problem and Task Partitioning Problem) in which one can interpret a couple matching : either spy right guesses or performed tasks. We respectively demonstrated that *indetermination* could protect a sender from a spy with a specific strategy and that it appeared in the worst task partitioning method.

The presence of this second equilibrium in at least three usual problems (graph clustering, guessing problem and task partition) reinforces the usefulness of section C.3.

Finally, we extended the *indetermination* coupling in a continuous space expressing a prior guess formula as the optimal solution of a continuous version of the original Minimal Trade Model.

We repeatedly noticed in [10] that properties of *independence* usually have a symmetric transcript for *indetermination*. In the continuous case, we know we can define an *independence* copula, operating on margins cumulative distribution functions to generate the cumulative distribution function of an *independence* coupling. Yet, even if we built up a continuous *indetermination* notion in section C.5, coining an *indetermination* copula is not straightforward. We will develop the computations it requires and the applications it may have in a future article to be published.

Continuous indetermination and average likelihood minimization

Submitted to Journal of Multivariate Analysis, May 2021 ([12]).

Abstract : *The authors transpose a discrete notion of indetermination coupling in the case of continuous probabilities. They show that this coupling, expressed on densities, cannot be captured by a specific copula which acts on cumulative distribution functions without a high dependence on the margins. Furthermore, they define a notion of average likelihood which extends the discrete notion of couple matchings and demonstrate it is minimal under indetermination. Eventually, they leverage this property to build up a statistical test to distinguish indetermination and estimate its efficiency using the Bahadur's slope.*

D.1 Introduction

A key tool in statistics is the likelihood estimator which consists, given a realization $W = w$ of a random variable W to estimate the probability under a candidate law P to have $W' \sim P = w$. Typically, if P_π has a density function π it is represented by $\pi(w) dw$. In a discrete case it would represent the probability to have an exact match between W and W' . Precisely, in a precedent paper [11] we demonstrated how *indetermination* concept could reduce those matchings in the discrete case. This paper is interested in the continuous transposition of this notion. Although the definition of the continuous indetermination structure is quite simple at first glance, it is motivated by an anticipated link with the likelihood reduction as the continuous extension of discrete couple matchings and conveys, actually, some interesting properties.

We begin with the introduction of *indetermination*. Given two discrete marginal laws $\mu = \mu_1 \dots \mu_p$ and $\nu = \nu_1 \dots \nu_q$, a coupling function C constructs a probability law π on the product space where $\pi_{u,v} = C(\mu_u, \nu_v)$. Theoretical considerations based on a work of Csiszár [29], a summarized version of which is expressed in [10] lead to consider two "natural" equilibria. Shortly, Csiszár works on projection functions : given a set of constraints, a first guess has to be projected into the eligible space. Then, adding some natural properties to be respected by the projection function, it leads to consider

only two possible criteria : either least square or maximum entropy. Eventually, under some reformulations each projection function is associated to a specific coupling function, respectively : the usual independence $\pi_{u,v}^\times = C^\times(\mu, \nu)_{u,v} = \mu_u \nu_v, \forall (u, v)$ and the so-called indetermination (first quoted as such in [64]) whose formula is given by :

$$\pi_{u,v}^+ = C^+(\mu, \nu)_{u,v} = \frac{\mu_u}{q} + \frac{\nu_v}{p} - \frac{1}{pq}, \forall (u, v) \quad (\text{D.1})$$

Additional constructive properties of discrete indetermination together with interpretations and applications can be found in [11] where we develop its analysis. Some demonstrations with useful interpretations (typically justifying the name "indetermination" or "indeterminacy") require the introduction of "Mathematical Relational Analysis" which is detailed in [68], [65] or [2]. We will not present it again in this paper but we refer the interested reader to the aforementioned articles.

By construction (related to the least square cost it minimizes), π^+ is the optimal law to reduce matching in case margins are fixed. If we draw two couples $W_1 = (U_1, V_1)$ and $W_2 = (U_2, V_2)$ under π^+ the probability to have $W_1 = W_2$ is minimal. The natural transposition in the continuous case of this matching property would be possibly the motivation for a continuous indetermination.

Actually, the idea of a continuous indetermination appears first in [63] in terms of coupling, and in [51] in terms of "indetermination copula". Let us precise that the paper [51] though it conveys an interesting method to detect anomalies, lacks the theoretical foundations of some concepts it introduces. Paper [63] is more theoretical but it tackles the subject in a too short way. In consequence the continuous "indetermination copula" must be properly introduced and refers to a notion that deserves to be investigated in depth, it is precisely one goal of this article.

In this paper we show that we cannot define an "indetermination copula" *per se* but that it can be locally defined. Shortly, whilst an independence copula is easily defined without reference to the margins, an indetermination coupling uses a margin-dependent copula. A couple of margins to share an indetermination copula requires them to be close in a sense that we will define later in the paper.

Additionally, we interpret the continuous indetermination as a lower bound on the average likelihood \bar{L} , a notion that we introduce as an equivalent of discrete couple matchings. We show that based on a list of n realizations of W a random variable under indetermination, the correlation vanishes under the average likelihood estimator which becomes sensitive only to both margins. Namely, if P_n is the empiric measure associated with the n realizations, $\bar{L}(P_n, \pi)$ only depends on the margins of π whatever the underlying copula is.

Finally, we define a statistical test t_n to reject or accept the hypothesis of continuous indetermination based on its property to minimize the average likelihood. Following the usual track, detailed in [45] and dedicated to the analysis of statistical test, we compute the Bahadur's slope of t_n as an estimation of its quality.

This paper is composed of two sections, section D.2 introduces the continuous indetermination together with some properties and studies the notion of indetermination copula. Section D.3 defines a notion of average likelihood \bar{L} between two probabilities and demonstrates that, applied to probabilities in a product space, its value between π against π^+ only depends on the four associated margins. It notably implies that any π with the

same margins as π^+ has the same average likelihood as π^+ against itself. Eventually, in subsection D.3.3 we leverage on this property to build up a test to distinguish π^+ and analyse it at the light of the Bahadur's slope.

D.2 Continuous indetermination

D.2.1 Coupling function

This chapter introduces a continuous indetermination coupling that, to our knowledge, has never been studied in depth (see previous remarks). Extrapolating the discrete margins situation, we now suppose μ is a probability law on a segment $[a, A]$ and ν is a same type of probability but on a segment $[b, B]$. Let us also suppose they both have a density function, respectively f and g . To define a probability measure π we can operate on the density functions as we basically did in the discrete case (where we essentially added μ_u and ν_v).

We follow and adapt the notations introduced in [28] and [46] as well as some lemmas or theorems they cover and that we shall use in the sequel (section D.3). First, we begin by quoting $S_a = [a, A]$, $S_b = [b, B]$ and $S = [a, A] \times [b, B]$ together with the usual borelian set of probability measures on the three : Λ_a set of probability measures on (S_a, \mathcal{B}) , Λ_b on (S_b, \mathcal{B}) and Λ on (S, \mathcal{B}) .

As a first step, we extend the definition of a coupling function in a continuous space.

Définition 69 (Coupling functions (continuous case)).

$\mu \in \Lambda_a$ and $\nu \in \Lambda_b$ being probability laws, a coupling function C operates on their density (f, g) to define a density $C(\mu, \nu) = C(f, g)$ for a measure on the product space S . That measure respects some properties similar to the ones found in the discrete case : it is a probability law whose margins are μ and ν .

A typical coupling function is the independance quoted C^\times (we extend here the discrete notation) which generates an eligible density under the formula :

$$C^\times(f, g)(x, y) = f(x)g(y), \quad \forall x \in S_a, \quad \forall y \in S_b$$

We would like to define a continuous version of the indetermination coupling. As usual, when we transpose a concept, we can either obtain it through computations or through a prior guess. We will follow both approaches. Let us first propose a prior guess.

Définition 70 (Continuous indetermination density).

$$C^+(f, g)(x, y) = \frac{f(x)}{B-b} + \frac{g(y)}{A-a} - \frac{1}{(A-a)(B-b)}$$

The formula comes from an adaptation of the discrete one, no guarantee is given on its truthfulness neither on its construction. First, to simplify any future computation, we set $a = b = 0$ and $A = B = 1$, converting any formula will be done using a dedicated affine transformation.

We shall use an optimal transport problem to validate our prior guess, we transpose actually the discrete "Minimal Transport problem" using least square cost function into the continuous space :

Problème 32 (Minimal Trade Problem).

$$\begin{aligned} \min_{\pi} \quad & \int_0^1 \int_0^1 \pi^2(x, y) \, dx \, dy \\ \text{under constraints :} \quad & \int_0^1 \pi(x, y) \, dy \, dx = \mu_x \text{ (first margin)} \\ & \int_0^1 \pi(x, y) \, dx \, dy = \nu_y \text{ (second margin)} \\ & \int_0^1 \int_0^1 \pi(x, y) \, dx \, dy = 1 \text{ (mass preserving)} \\ & \pi \geq 0 \end{aligned}$$

We then add inequality [D.2](#) similar to the one used in the discrete case (see [\[11\]](#)) which ensures that our prior guess function is a probability law. As well as restricting ourselves to $[0, 1]$ we force the margins to respect the condition :

$$\min_{x \in S_1} f(x) + \min_{y \in S_1} g(y) \geq 1 \quad (\text{D.2})$$

Propriété 58.

Under the ad hoc hypothesis [D.2](#), the solution of problem [32](#) is nothing but our prior guess continuous coupling function applied to the margins f and g , formally, the associated density function is : $C^+(f, g)(x, y) = c_{f,g}^+(x, y) = (f(x) + g(y) - 1)$

Démonstration.

We use definition [70](#) to efficiently solve problem [32](#) by noticing that

$$\int_{x=0}^1 \int_{y=0}^1 [\pi(x, y) - (f(x) + g(y) - 1)]^2 \, dx \, dy \geq 0$$

Which can be rewritten (using constraints on margins) :

$$\begin{aligned} \int_{x=0}^1 \int_{y=0}^1 \pi^2(x, y) \, dx \, dy & \geq \int_{x=0}^1 \int_{y=0}^1 (-f^2 - g^2 - 1 - 2\pi - 2fg + 2f\pi + 2g\pi + 2f + 2g) \, dx \, dy \\ & = \int_{x=0}^1 \int_{y=0}^1 (-f^2 - g^2 - 1 - 2 - 2fg + 2f^2 + 2g^2 + 2 + 2) \, dx \, dy \\ & = \int_{x=0}^1 \int_{y=0}^1 (f^2 + g^2 - 2fg + 1) \, dx \, dy \\ & = \int_{x=0}^1 \int_{y=0}^1 (f + g - 1)^2 \, dx \, dy \end{aligned}$$

At this stage we have shown

$$\min_{\pi} \int_0^1 \int_0^1 \pi^2(x, y) \, dx \, dy \geq \int_{x=0}^1 \int_{y=0}^1 (C^+(f, g)(x, y))^2 \, dx \, dy$$

Eventually, using hypothesis D.2 we notice that $c_{f,g}^+(x, y) = (f(x) + g(y) - 1)$ is eligible as a density function since always positive. Margins constraints are also satisfied using the same considerations as in the discrete case. This provides the second inequality :

$$\min_{\pi} \int_0^1 \int_0^1 \pi^2(x, y) \, dx \, dy \leq \int_{x=0}^1 \int_{y=0}^1 (C^+(f, g)(x, y))^2 \, dx \, dy$$

which allows us to conclude. \square

Définition 71 (Indetermination coupling).

Given two probability laws on $S_1(S_a$ with $a = 1)$: $U \sim \mu, (f, F)$ and $V \sim \nu, (g, G)$ we say that the random variable W is an indetermination coupling of U and V , quoted $U \oplus V$ when its density π^+ is $\pi_{f,g}^+(x, y) = f(x) + g(y) - 1$

Let us compute the cumulative distribution function $\Pi_{F,G}^+$ associated to the just expressed density $\pi_{f,g}^+$; it is only a quick integration of the density leading to a second characterization of an indetermination coupling.

Propriété 59 (Cumulative distribution function for indetermination).

If $\pi_{f,g}^+(x, y) = f(x) + g(y) - 1$ is the density of a random variable, with f, g two densities on S_1 ensuring $\pi_{f,g}^+$ is positive (hypothesis D.2) and whose cumulative distribution functions are F and G respectively, then the associated cumulative distribution function, quoted $\Pi_{F,G}^+$, is given by :

$$\Pi_{F,G}^+ = yF(x) + xG(y) - xy$$

D.2.2 Constructive method for adapted margins

Beforehand, we assumed margins are respecting hypothesis D.2, we propose here a method to construct adapted margins out of any couple.

Propriété 60 (Constructive margins).

A couple (f, g) of densities fulfills hypothesis D.2 if and only if, it exists an $\alpha, 0 \leq \alpha \leq 1$ and a couple of densities (r, s) such that :

$$f = (1 - \alpha)r + \alpha \quad \text{and} \quad g = \alpha s + 1 - \alpha \quad (\text{D.3})$$

Démonstration.

Let us first suppose (f, g) are under this form, then we notice, $\min f \geq \alpha$ as well as $\min g \geq 1 - \alpha$, hence, $\min f + \min g \geq 1$ so that the condition is respected.

Now, if the condition is respected, we define $\alpha = \min f$ and have $g \geq 1 - \alpha$. If $\alpha = 0$ or $\alpha = 1$ then, respectively, g or f is uniform so that the corresponding coupling is independence and condition is degenerated. If not, $0 < \alpha < 1$ and we can write :

$$f = (1 - \alpha) \frac{f - \alpha}{1 - \alpha} + \alpha$$

together with :

$$g = \alpha \frac{g - (1 - \alpha)}{\alpha} + (1 - \alpha)$$

Quoting $r = \frac{f - \alpha}{1 - \alpha}$ and $s = \frac{g - (1 - \alpha)}{\alpha}$, we have $0 \leq r, s \leq 1$ and $\int r = \int s = 1$. It precisely shows that (r, s) is a couple of densities on S_1 . \square

Using the last proposition, we can easily build up a couple of margins on which an indetermination coupling is feasible. As mentioned during the proof, if $\alpha \in \{0, 1\}$ then f or g is a uniform density for which indetermination and independence are completely equivalent. Hence we will exclude $\alpha \in \{0, 1\}$ from most of our computations.

We repeatedly notice in the discrete case that properties of independence usually have a symmetric transcript for indetermination. In the continuous case, we know we can define an independence copula, operating on the cumulative distribution functions of the margins to generate the cumulative distribution function of an independence coupling. That property is not as common as one would expect, notably, an indetermination copula is out of existence. Let us properly explain where the difference comes from.

D.2.3 A specific indetermination copula

We first remind the definition and properties of copulas as a classic way to couple two margins together with the well-known Sklar theorem (see [87]).

Définition 72 (Copula).

A copula \mathbf{C} (in bold to distinguish from a coupling function) is a cumulative distribution function on $[0, 1]^d$ ($d \in \mathbb{N}$) whose margins are uniforms. It is defined by three properties valid for any $u = (u_1, \dots, u_d)$:

- $\mathbf{C}(u) = 0$ as soon as any u_i is null
- $\mathbf{C}(u) = u_i$ if u_i is the only component different from 1
- \mathbf{C} is d -non decreasing

A copula is used to construct a coupling law by defining it with its cumulative distributive function while a coupling function (definition 69) operates on densities ; the close notations insist on their similarity while the type of function (density or cumulative distribution function) distinguish them. One could expect any coupling function to generate a corresponding copula. The transposition is not that easy : as we shall see it may depend on margins.

The Sklar's theorem extracts and applies copulas to any probability law. More precisely, it indicates that any function C satisfying the properties stated in Definition 72 can be applied to any set of d univariate cumulative functions (F_1, \dots, F_d) to generate a multivariate cumulative function whose margins will precisely be the F_i respecting the formula :

$$F(x_1, \dots, x_d) = \mathbf{C}(F_1(x_1), \dots, F_d(x_d))$$

Reciprocally, any cumulative distribution function F corresponds to an associated copula \mathbf{C}_F . Typically, when the margins F_i of F have a closed formula, we have

$$\mathbf{C}_F(u_1, \dots, u_d) = F(F_1^{-1}(u_1), \dots, F_d^{-1}(u_d))$$

Remarque 71 (d - increasing).

The d -increasing property is closely tied to the positivity of the underlying probability law \mathbb{P} . Hence, in dimension 2, $\mathbb{P}(u_1 \leq U \leq u'_1, u_2 \leq U \leq u'_2) = \mathbf{C}(u_1, u_2) - \mathbf{C}(u_1, u'_2) - \mathbf{C}(u'_1, u_2) + \mathbf{C}(u'_1, u'_2) \geq 0$.

Extraction of the indetermination copula

We begin by applying the Sklar's theorem. We do know that when μ and ν are fixed and fulfill hypothesis D.2, we can couple them under indetermination operating on their densities ; we keep the usual notations previously introduced for their densities as well as for their cumulative distribution functions.

With indetermination, we obtain a probability on the couple space whose density is given by : $\pi_{f,g}^+ = h(x, y) = f(x) + g(y) - 1, \forall (x, y) \in S$ (remember we restricted ourselves to $S_1 = [0, 1]$).

Using the associated cumulative distribution function $\Pi_{F,G}^+$, we can express the associated copula. It is specific to the indetermination coupling (the way we generated $\pi_{f,g}^+$ hence $\Pi_{F,G}^+$) as well as to the margins F and G a priori.

Définition 73 (Specific indetermination copula).

Given two margins $U \sim \mu, (f, F)$ and $V \sim \nu, (g, G)$, their specific indetermination copula, extracted from their indetermination coupling is given by :

$$\mathbf{C}_{\mathbf{F},\mathbf{G}}^+(u, v) = v * F^{-1}(u) + u * G^{-1}(v) - F^{-1}(u) * G^{-1}(v), \forall (u, v) \in S \quad (\text{D.4})$$

Remarque 72 (Extension).

$\mathbf{C}_{\mathbf{F},\mathbf{G}}^+$ is defined by the margins, we shall insist on that later on. Yet, they are characterized by their densities, their cumulative distribution function as well as by their probability measure. Provided we can pass from one to another, we shall abusively speak of an indetermination copula for a couple of densities (f, g) ($\mathbf{C}_{\mathbf{f},\mathbf{g}}^+$) or a couple of probability measures (μ, ν) ($\mathbf{C}_{\mu,\nu}^+$).

Example

Let us introduce an example, U and V respectively follow the cumulative distribution function $F(x) = x^\alpha$ and $G(y) = y^\beta$. We first need to ensure that hypothesis D.2 is respected. Here it requires that the densities verify :

$$\begin{aligned} \min_x f + \min_y g - 1 &\geq 0 \\ \min_x \alpha * x^{\alpha-1} + \min_y \beta y^{\beta-1} &\geq 1 \end{aligned}$$

hypothesis D.2 leads to α and β less than 1 and

$$\alpha + \beta \geq 1$$

Under that last hypothesis,

$$\mathbf{C}_{\alpha,\beta}^+(u, v) = uv^{\frac{1}{\beta}} + u^{\frac{1}{\alpha}}v - u^{\frac{1}{\alpha}}v^{\frac{1}{\beta}} \quad (\text{D.5})$$

Remarque 73 (Dependence on margins).

A remark that we shall develop later : the presence of α (as well of β) inside the formula prevents it from being independent from the margins.

So far, we did not define a copula of indetermination *per se* but a specific copula of indetermination among those satisfying the formula of Definition 73 for a given couple of margins.

Dependence on margins

We unfold the link between a specific indetermination copula and its margins. The copula is a way among many to couple two laws (maybe more than two but we limit ourselves). We already defined two of them through Definition 69 which operates on densities and by Definition 72 which operates on cumulative distribution function. Replacing ourselves in the context of indetermination, it leads to formula 70, or to apply a specific indetermination copula given by formula D.4 to a pair of cumulative distribution function.

The second way for coupling exhibits a problem that we can easily isolate ; it depends on margins as functions F^{-1} and G^{-1} appear in the formula used to mimic the application on densities to cumulative distribution function. We notice that the copula associated to an indetermination coupling depends on the margins we want to couple. It is not the case (for instance) when using an independence coupling.

In general, given two margins, one can apply at least three coupling ways :

1. define a density $\pi(x, y)$ verifying the margins (typically applying a coupling function C)
2. define a cumulative distribution function $\Pi(x, y)$ verifying the margins
3. define a copula \mathbf{C} and apply it to the margins (verified by construction)

In case of independence they correspond to $C^\times(u, v) = uv$, $\Pi^\times(x, y) = F(x)G(y)$ and $\mathbf{C}^\times(u, v) = uv$. It is a particular concept where any of the three coupling methods shall be defined without reference to the associated margins.

For indetermination, it doesn't work as $C(u, v) = u + v - 1$, $\Pi^+(x, y) = xG(y) + yF(x) - xy$ but $\mathbf{C}^+(u, v) = u * F^{-1}(v) + v * G^{-1}(u) - F^{-1}(u) * G^{-1}(u)$ is highly dependent on margins. It is quite visible on the example D.5 which takes various forms when α and β vary. The only constraints of those two parameters being less than 1 and with a sum greater than 1.

Going back to independence (defined on density), we estimate why it is so specific. \mathbf{C} being a copula, we suppose $u = F(x)$ and $v = G(y)$ for more readiness. By definition, if we quote Π the cumulative distribution function and π the density :

$$\mathbf{C}(F(x), G(y)) = \Pi(x, y)$$

differentiating with respect to x and y , we get :

$$\begin{aligned} \frac{\partial \mathbf{C}(u, v)}{\partial x \partial y} &= \frac{\partial \Pi}{\partial x \partial y} \\ \frac{\partial u}{\partial x} \frac{\partial v}{\partial y} \frac{\partial \mathbf{C}(u, v)}{\partial u \partial v} &= \pi(x, y) \\ f(x)g(y) \frac{\partial \mathbf{C}(u, v)}{\partial u \partial v} &= \pi(x, y) \\ \frac{\partial \mathbf{C}(u, v)}{\partial u \partial v} &= \frac{\pi(x, y)}{f(x)g(y)} \end{aligned}$$

Hence, independence coupling is quite peculiar as margins vanish : $\frac{fg}{fg} = 1$ is the second derivative of the associated copula \mathbf{C}^\times which is precisely independent of F and G .

For an indetermination coupling, the crossed derivative of the copula \mathbf{C}^+ is : $\frac{f+g-1}{fg} = \frac{1}{g} + \frac{1}{f} - \frac{1}{fg}$ whose integration leads to formula D.4 using $\frac{\partial F^{-1}(u)}{\partial u} = \frac{1}{F'(F^{-1}(u))} = \frac{1}{f(x)}$. It is dependent on margins and one cannot define a generic indetermination copula. We shall, so far, only define a specific indetermination copula $\mathbf{C}_{\mathbf{F},\mathbf{G}}^+$

Remarque 74.

These last computations require a division by fg . Let us go back to hypothesis D.2. We know that $f + g - 1 \geq 0$ so that if there exists x such that $f(x) = 0$ then $\forall y, g(y) \geq 1$ and, as $\int_0^1 g = 1$ (we are on segment $[0, 1]$, if not we have to divide by the length as in Definition 70), we deduce $g = 1$. It leads to a poorly interesting coupling : $f + g - 1 = f$. Moreover, the computations are actually still valid since $\mathbf{C}_{\mathbf{F},\mathbf{G}}^+ = yF^{-1}(x)$ whose crossed derivative is nothing but : $\frac{1}{f}$.

D.2.4 A local indetermination copula

In this section we measure the dependence of $\mathbf{C}_{\mathbf{F},\mathbf{G}}^+$ on its margins. The idea is to check whether a given specific copula of indetermination can create indetermination coupling when applied to another couple of margins than the one which defines it. The property 61 shows that it is locally possible.

To answer, let us start with four densities f, g, r, s as well as their respective cumulative distribution functions F, G, R, S .

First step, we couple (f, g) through indetermination using the indetermination coupling function of Definition 70 and extract the associated specific copula of indetermination $\mathbf{C}_{\mathbf{F},\mathbf{G}}^+$ using Definition 73. As already explained, it can be applied to any couple of margins, hence the second step.

Second step, we apply $\mathbf{C}_{\mathbf{F},\mathbf{G}}^+$ to the margins defined by (R, S) , it leads to :

$$\Pi_{r,s}^{f,g} = \mathbf{C}_{\mathbf{F},\mathbf{G}}^+(R, S) = H \tag{D.6}$$

We shall quote H this cumulative distribution function throughout this section, it can be applied to any (x, y) in S .

Third and last step, we compare H to the coupling of (r, s) under indetermination, namely $\mathbf{C}_{\mathbf{R},\mathbf{S}}^+(R, S)$.

Since we do not expect a general equality it implies at least some requirements on (F, G, R, S) so that H represents a coupling of indetermination. Obviously, if $(F, G) = (R, S)$, the third step is trivial as both functions are equal. The natural question being : are there any other possibilities? The answer is yes, provided that the requirements expressed below are satisfied.

Propriété 61 (A local indetermination copula).

Given f, g, r, s , four densities and (R, G, R, S) the application of $\mathbf{C}_{\mathbf{F},\mathbf{G}}^+$ to (R, S) quoted H (see equation D.6) generates an indetermination coupling if and only if it exists a $\lambda \geq 0$ respecting :

$$\frac{\max(g)}{\max(g) - 1} \geq \lambda \geq 1 - \frac{1}{\max(f)} \tag{D.7}$$

such that the following equations are satisfied :

$$F^{-1}(x) - x = \lambda(R^{-1}(x) - x) \quad (\text{D.8})$$

$$G^{-1}(y) - y = \frac{1}{\lambda}(S^{-1}(y) - y) \quad (\text{D.9})$$

Remarque 75.

We notice that the cumulative distribution functions sharing the same specific indetermination copula are those whose inverse function is inside the convex eligible space (coefficient respecting equation D.7) of F^{-1} or G^{-1} and of the identity.

Démonstration.

See appendix D.5. □

Remarque 76 (Transfer of the condition).

Inequality D.42 which appears in the proof is quite remarkable : it expresses that having (f, g) respecting hypothesis D.2 automatically means that (r, s) also does, provided that λ respects equation D.7.

In the proposition 61 we used two couples of cumulative distribution functions $((F, G), (R, S))$ representing margins and expressed a condition for the specific indetermination copula of the first couple (F, G) to generate indetermination when we apply it to the second (R, S) ; formally :

$$\mathbf{C}_{\mathbf{F},\mathbf{G}}^+(R, S) = \mathbf{C}_{\mathbf{R},\mathbf{S}}^+(R, S)$$

in summary : both copula agree on one point. It does not demonstrate that they are equal.

Proposition 62 completes the previous proposition 61 and ensures the copula are the same.

Propriété 62 (Shared Indetermination Copula).

The hypotheses of proposition 61 apply if and only if $\mathbf{C}_{\mathbf{F},\mathbf{G}}^+ = \mathbf{C}_{\mathbf{R},\mathbf{S}}^+$.

Démonstration.

We begin with a simple remark, if for two couples $((f, g), (r, s))$ we have $\mathbf{C}_{\mathbf{f},\mathbf{g}}^+ = \mathbf{C}_{\mathbf{r},\mathbf{s}}^+$ then, in particular, $\mathbf{C}_{\mathbf{f},\mathbf{g}}^+(r, s) = \mathbf{C}_{\mathbf{r},\mathbf{s}}^+(r, s)$, hence, proposition 61 applies and its hypotheses are satisfied.

Now, if it exists a λ as defined in the quoted proposition linking F^{-1} to R^{-1} and G^{-1} to S^{-1} , we use it to express $\mathbf{C}_{\mathbf{f},\mathbf{g}}^+$ as a function of (r, s) :

$$\begin{aligned} \mathbf{C}_{\mathbf{f},\mathbf{g}}^+(u, v) &= vF^{-1}(u) + uG^{-1}(v) - F^{-1}(u)G^{-1}(v) \\ &= v [u + \lambda(R^{-1}(u) - u)] + uG^{-1}(v) - [u + \lambda(R^{-1}(u) - u)] G^{-1}(v) \\ &= v [u + \lambda(R^{-1}(u) - u)] - [\lambda(R^{-1}(u) - u)] G^{-1}(v) \\ &= v [u + \lambda(R^{-1}(u) - u)] - [\lambda(R^{-1}(u) - u)] \left[v + \frac{1}{\lambda}(S^{-1}(v) - v) \right] \\ &= uv - [R^{-1}(u) - u] [S^{-1}(v) - v] \\ &= \mathbf{C}_{\mathbf{r},\mathbf{s}}^+(u, v) \end{aligned}$$

It completes the proof. □

D.2.5 Maximal spread between the two copulas

Let us estimate the difference between the two copulas we just introduced. As the indetermination copula is defined locally, we compute it for any suited couple of margins (F, G) before trying to maximize it. The expression under the \mathbb{L}^1 norm amounts to :

$$\begin{aligned}\Delta_1(F, G) = \|\mathbf{C}^\times - \mathbf{C}_{\mathbf{F}, \mathbf{G}}^+\|_1 &= \int_{u=0}^1 \int_{v=0}^1 |uv - vF^{-1}(u) - uG^{-1}(v) + F^{-1}(u)G^{-1}(v)| \, du \, dv \\ &= \int_{u=0}^1 \int_{v=0}^1 |(F^{-1}(u) - u)(G^{-1}(v) - v)| \, du \, dv \\ &= \int_{u=0}^1 |F^{-1}(u) - u| \, du \int_{v=0}^1 |G^{-1}(v) - v| \, dv\end{aligned}$$

We have a closed formula of the difference between the two couplings (the common choice of the \mathbb{L}^1 norm shall be motivated later). But we already know the value is null in case any of the two margins is uniform, we do not know neither the maximum, nor for which couple it is realized. Both questions are solved within Property 63.

Propriété 63.

Quoting \mathcal{E}^+ the set of couples of cumulative distribution functions (F, G) such that their densities couple (f, g) respects condition D.2 we have :

$$\max_{(F, G) \in \mathcal{E}^+} \Delta_1(F, G) = \frac{1}{16}$$

Moreover, the couple of densities corresponding to the maximum is :

$$(f_0, g_0) = \left(u \mapsto \frac{1}{2}(1 + \delta_{u=0}), v \mapsto \frac{1}{2}(1 + \delta_{v=0}) \right)$$

Démonstration.

Working with $(F, G) \in \mathcal{E}^+$ is fairly inconvenient as it adds hypothesis on f that we have to convey to F^{-1} through F and similarly with G . To get rid of it, we use proposition 60 and extract α together with two densities r, s (respectively two cumulative distribution functions (R, S)). Finally, we obtain :

$$F(u) = \alpha u + (1 - \alpha)R(u)$$

as well as

$$G(v) = (1 - \alpha)v + \alpha S(v)$$

Let us report those functions in $\Delta_1(F, G)$:

$$\begin{aligned}\Delta_1(F, G) &= \int_{u=0}^1 |F^{-1}(u) - u| \, du \int_{v=0}^1 |G^{-1}(v) - v| \, dv \\ &= \int_{u=0}^1 |F(u) - u| f(u) \, du \int_{v=0}^1 |G(v) - v| g(v) \, dv\end{aligned}$$

The expression of Δ_1 falsely separates F and G : they are linked one to another using the constant α . Though, we can maximize each integral on its own, while the respect of condition defining \mathcal{E}^+ will appear after. We run the computations on the "F-part" of Δ_1 :

$$\begin{aligned}
I_f &= \int_{u=0}^1 |F(u) - u| f(u) \, du \\
&= \int_{u=0}^1 |\alpha u + (1 - \alpha)R(u) - u| f(u) \, du \\
&= (1 - \alpha) \int_{u=0}^1 |R(u) - u| (\alpha + (1 - \alpha)r(u)) \, du \\
&= (1 - \alpha) \left[\alpha \int_{u=0}^1 |R(u) - u| \, du + (1 - \alpha) \int_{u=0}^1 |R^{-1}(u) - u| \, du \right] \\
&\leq \frac{(1 - \alpha)}{2}
\end{aligned}$$

We immediately derive, conducting the same analysis on G :

$$\Delta_1(F, G) \leq \frac{(1 - \alpha)\alpha}{4} \leq \frac{1}{16} \text{ reached for } \alpha = \frac{1}{2}$$

Besides, it turns out that it becomes an equality if we choose (F, G) as proposed in the property, namely $f(u) = \frac{1 + \delta_{u=0}}{2}$ which finishes the proof. \square

Remarque 77 (Transposition of the discrete case).

The couple (f_0, g_0) which emphasizes most the difference between independence and indetermination appears to be the natural transposition of the discrete case. Indeed, for contingency values, each probability is $\frac{1}{2p} + \frac{1}{2}\delta_{u=0}$ where p is the number of values u .

Remarque 78 (Motivation of the L_1 -norm).

Application of the Scheffé's Lemma enables us to convert the norm we used in the above property to the L_∞ -norm. Formally, we also have, $\forall (F, G) \in \mathcal{E}^+, \forall (U, V) \in \mathcal{B}([0, 1])^2$:

$$\left| \int_{u \in U} \int_{v \in V} \mathbf{C}_{\mathbf{F}, \mathbf{G}}^+(u, v) \, du \, dv - \int_{u \in U} \int_{v \in V} \mathbf{C}^\times(u, v) \, du \, dv \right| \leq \frac{1}{16}$$

Remarque 79 (Discrete case).

In a precedent paper, we showed a similar result in the discrete case. Given two marginal laws μ and ν uniformly drawn inside the set of probability laws on p and q elements respectively, the expected L^2 norm between an indetermination and an independence coupling is less than $\frac{1}{pq}$.

D.2.6 Conclusions about an indetermination (or indeterminacy) copula

Through Definition 73, we introduced an indetermination copula $\mathbf{C}_{\mathbf{R}, \mathbf{G}}^+$ specific to a given couple of margins (F, G) . We also showed that dependence on margins prevents us

from defining a general indetermination copula : we end up with a collection of copulas parametrized by a couple of eligible margins (any respecting hypothesis D.2).

A priori, the association $(F, G) \mapsto \mathbf{C}_{\mathbf{F}, \mathbf{G}}^+$ has no reason to be injective and it is not. Precisely, a couple (F, G) defines the same indetermination copula as another (R, S) if and only if the four cumulative distribution functions respect the two equations D.8. It basically requires that each inverse cumulative distribution is a linear composition of the other cumulative distribution function and of the identity. On that segment of couples of cumulative distribution functions defined with an eligible λ , all the specific indetermination copulas are identical.

From a specific indetermination copula, we built up an indetermination copula that shall be defined and applied to a segment of couples of cumulative distribution functions and we also showed that it cannot be further extended.

D.3 Indetermination and average likelihood

Given a set of n realizations of a variable W , a common problem is to determine the underlying probability law P_π that is supposed unknown and that we shall abusively quote by its corresponding density π . It often comes with a subset $\Omega \subset \Lambda$ of probability laws among which we search for the most approaching one. Here we also suppose that $P_\pi \in \Lambda$ applies to a product space S as previously defined.

A usual method is the Maximum Likelihood Paradigm (MLP), under which for a given realization $W = w$ and for any probability law $P_{\pi'}$ in Ω , we compute the probability that $W' \sim P_{\pi'}$ belongs to $[w, w + dw]$.

$$\mathbb{P}_{\pi'}(W' \in [w, w + dw]) = \pi'(w) dw$$

If we integrate on the values W can take, we obtain the average likelihood $\bar{L}(\pi, \pi')$ of W under π' :

Définition 74 (Average likelihood).

Given two probability laws P_π and $P_{\pi'}$ whose densities are π and π' respectively we define the average likelihood between P_π and $P_{\pi'}$ or abusively between π and π' as :

$$\bar{L}(P_\pi, P_{\pi'}) = \bar{L}(\pi, \pi') = \mathbb{E}_{W \sim P_\pi}(\pi'(W)) \quad (\text{D.10})$$

Remarque 80.

Finally, $\bar{L}(P_\pi, P_{\pi'})$ rewrites $\int_S \pi(s)\pi'(s) ds$ with S the set where P_π and $P_{\pi'}$ take their values (non necessarily a product space). It immediately appears symmetric.

Now, motivated by the discrete properties of π^+ (see [11]), we link \bar{L} with a discrete notion of couple matching.

Let us go back to the discrete case, and select π among probability laws defined on a product space of pq elements with fixed margins μ and ν . Then, if $W \sim \pi$, the probability of a couple matching, that is to say the case where two independent realizations $W_1 = (U_1, V_1)$ and $W_2 = (U_2, V_2)$ are equal, is minimal when $\pi = \pi^+ = C^+(\mu, \nu)$.

Similarly, in the continuous case, we expect $\bar{L}(\pi, \pi)$ to be minimal when $\pi = \pi^+$ and this is precisely what happens since it amounts to computing the cost function of problem 32 :

$$\bar{L}(\pi, \pi) = \int_S \pi^2(w) dw \quad (\text{D.11})$$

Furthermore, in the discrete case a couple matching between π^+ and a second probability law π only depends upon the margins of π . We show this property stays true in a continuous domain in the next section.

D.3.1 Indetermination prevents correlation extraction

Going back to the definition of average likelihood given in Equation D.10, we suppose π takes the form π^+ given in 70 and compute its average likelihood with any $P_h \in \Lambda$:

$$\begin{aligned} \bar{L}(h, \pi^+) &= \mathbb{E}_{W \sim \pi^+}(h(W)) \\ &= \int_{x=0}^1 \int_{y=0}^1 \pi^+(x, y) h(x, y) dx dy \\ &= \int_{x=0}^1 \int_{y=0}^1 (f(x) + g(y) - 1) h(x, y) dx dy \\ &= \int_{x=0}^1 f(x) \partial_1 P_h(x) dx + \int_{y=0}^1 g(y) \partial_2 P_h(y) dy - 1 \end{aligned} \quad (\text{D.12})$$

where $\partial_1 P_h$ corresponds to the density of the first margin of P_h and $\partial_2 P_h$ for the second margin.

The equality of equation D.12 proves that the average likelihood between h and π^+ only depends essentially on the four underlying margins. Furthermore leveraging on this decomposition we can derive the Theorem 27

Théorème 27.

Given two densities f and g on S_1 , we quote $\Omega_{f,g}$ the set of densities on S with margins (f, g) . Among $\Omega_{f,g}$, π^+ realizes the optimum of :

$$\min_{\pi \in \Omega_{f,g}} \bar{L}(\pi, \pi); \quad (\text{D.13})$$

and, for any π_0 in $\Omega_{f,g}$ we have :

$$\min_{\pi \in \Omega_{f,g}} \bar{L}(\pi, \pi) = \bar{L}(\pi^+, \pi_0). \quad (\text{D.14})$$

Démonstration.

The first part amounts to noticing, using Equation D.11, that it corresponds to the cost of Problem 32 which is precisely minimized among $\Omega_{f,g}$ by π^+ . For the second part, we rewrite equation D.12 :

$$\bar{L}(\pi^+, \pi_0) = \bar{L}(f, f) + \bar{L}(g, g) - 1$$

so that we have :

$$\bar{L}(\pi^+, \pi_0) = \bar{L}(\pi^+, \pi^+)$$

□

D.3.2 Average likelihood as a continuous couple matching notion

In any application, π is not given and one must compute \bar{L} using a set of n realizations of W . We describe hereafter a method to approach \bar{L} ; in addition we show \bar{L} corresponds to the probability of couple matchings in the discrete case.

Suppose $s_n = W_1, \dots, W_n$ are n i.i.d. points drawn under a law $P_\pi \in \Lambda$. We quote P_n the empirical measure given by :

$$P_n(s) = \frac{1}{n} \sum_{i=1}^n \delta_{W_i} \quad (\text{D.15})$$

A natural question would be, under a hypothesis of fixed margins densities (f, g) and leveraging only on P_n , can we decide whether the underlying density π equals $\pi^+ = C^+(f, g)$? In the discrete case we could typically estimate the probability of a couple matching under P_n and check it is minimal :

$$\begin{aligned} \mathbb{P}_{C^\times(P_n, P_n)}(W'_1 = W'_2) &= \sum_{i=1}^n \frac{1}{n} P_n(W_i) \\ &= \mathbb{E}_{W' \sim P_n}(P_n(W')) \\ &= \bar{L}(P_n, P_n) \end{aligned}$$

This last equality precisely shows that the average likelihood corresponds to the discrete notion of couple matchings. Yet, in the continuous case (since P_π has a density : π), two W_i are never equal leading to a null probability for any π whatever the correlation between margins is. This property prevents us from checking that $\bar{L}(P_n, P_n)$ is minimal since it is always null.

We leverage on $\bar{L}(\pi^+, \pi^+) = \bar{L}(\pi^+, \pi_0)$ whatever $\pi_0 \in \Omega_{f,g}$ to rely on $\bar{L}(P_n, \pi_0)$ for a fixed π_0 as a way to estimate continuous couple matchings. Later on, we will choose $\pi_0 = \pi^\times = C^\times(f, g)$.

Eventually, provided $P_n \rightarrow P_{\pi^+}$ in a certain sense, we expect $\bar{L}(P_n, \pi_0)$ to converge to $\bar{L}(\pi^+, \pi_0)$ that is to say to $\bar{L}(\pi^+, \pi^+)$ which is minimal among $\Omega_{f,g}$; section D.3.3 formalizes the approach notably by introducing a topology on Λ .

D.3.3 Statistical test based on average likelihood

Definition of t_n

Since π^\times and π^+ are close in a certain sense (property 63) we expect $\bar{L}(\pi^\times, \pi^+) = l_0$ and $\bar{L}(\pi^\times, \pi^\times) = l_1$ to be close. Then, using property 58, since π^+ and π^\times have the same margins we know we always have $l_0 \leq l_1$. Eventually, provided f and g are not uniform,

$$\eta = l_1 - l_0 > 0 \quad (\text{D.16})$$

We shall take advantage on that difference to build up a statistical test of *indeterminacy*. First, we define :

$$\Omega^\times = \{P_\pi \in \Omega_{f,g} / \bar{L}(\pi, \pi^\times) \geq l_1\} \quad (\text{D.17})$$

and now we build up two hypotheses : H_0 the hypothesis $\pi = \pi^+$ while the opposite H_1 corresponds to $\pi \in \Omega^\times$.

Remarque 81.

Using the ρ -topology we shall introduce shortly, we show in Proposition 64 that both hypotheses correspond to non void and separable subsets of Λ .

Given n observations $s_n = W_1, \dots, W_n$, we quote s the whole sequence and we define a statistical test t_n depending only on the first n coordinates of s to distinguish between both hypotheses :

$$t_n(s) = \bar{L}(P_n, \pi^\times) = \int_S \pi^\times dP_n(s); \tag{D.18}$$

where P_n is defined as in Equation D.15. Observing s , H_0 will be rejected if $t_n(s)$ is greater than a specific value. The just-introduced test is motivated by the combination of subsection D.3.2 which allows us to interpret it as an estimation of continuous couple matchings and of Theorem 27 which enables us (under H_0) to estimate \bar{L} using any π_0 , in particular using π^\times . We will study t_n under the two eligible hypotheses to calibrate the threshold. Specifically, we follow the track of the second section of [45] dedicated to Bahadur's slope. It requires the introduction of some notations that we report hereafter.

Usual notations and basic lemmas

As common in the literature and notably in the two articles ([28] and [46]) cited beforehand, we apply the τ -topology on Λ which is defined by the basic neighborhoods :

$$U(P, \mathcal{P}, \epsilon) = \{Q / \forall i, |P(B_i) - Q(B_i)| < \epsilon\} \tag{D.19}$$

where $P \in \lambda$, $\epsilon > 0$ and \mathcal{P} ranges over all \mathcal{B} -measurable partitions $\mathcal{P} = (B_1, \dots, B_k)$ of S . Consequently, for any set $\Omega \subset \Lambda$, we will quote Ω° and $\bar{\Omega}$ the interior and the closure of Ω in the sense of the τ -topology.

A sequence of probability measures $(Q_n)_{n \in \mathbb{N}}$ converges to Q for this topology if and only if $\lim_{n \rightarrow \infty} \int_{\mathbb{R}} f dQ_n \rightarrow \int_{\mathbb{R}} f dQ$ for each \mathcal{B} -measurable and bounded function $f : S \rightarrow \mathbb{R}$.

We also introduce the usual Kullback-Leibler divergence quoted D_{KL} whose properties are gathered in [93] and which is defined by :

$$\begin{aligned} D_{KL}(P|Q) &= \int_S \log \left(\frac{dP}{dQ} \right) dP \text{ if } Q \ll P \\ &= \infty \text{ otherwise} \end{aligned} \tag{D.20}$$

together with the usual conventions $\log 0 = -\infty$, $0 \cdot (\pm\infty) = 0$ and $\log(a/0) = \infty$, $\forall a > 0$.

An usual result is the lower semi-continuous property of D_{KL} :

Lemme 8 (Divergence lower semi-continuous).

For any $P \in \Lambda$, the function :

$$Q \mapsto D_{KL}(Q, P)$$

is τ -lower semi-continuous.

Démonstration.

It corresponds to Lemma 2.2 of [46]. □

Furthermore, for any subset $\Omega \subset \Lambda$ and any probability law $P \in \Lambda$ we extend the definition of the divergence :

$$D_{KL}(\Omega|P) = \inf_{Q \in \Omega} D_{KL}(Q|P) \quad (\text{D.21})$$

with the convention that it equals ∞ if Ω is empty.

Eventually, we introduce a second topology for Λ , the ρ -topology which is induced by the supremum metric d defined on the ordered set $S = [0, 1] \times [0, 1]$:

$$d(P, Q) = \sup_{x \in S} |P(0, x) - Q(0, x)|. \quad (\text{D.22})$$

Enabled with those notations we list a bench of lemmas extracted from [46] or [28]. The first lemma links the two previously defined topologies and enables us to interpret the open space of one in the second.

Lemme 9 (τ is finer than ρ).

The τ -topology is finer than the ρ -topology.

Démonstration.

It corresponds to Lemma 2.1 of [46]. □

The ρ -topology is handier since it is associated to a distance d and notably to define the subset neighborhood of a probability law P as exposed below :

$$V_\epsilon(P) = \{Q/d(P, Q) < \epsilon\} \quad (\text{D.23})$$

A first result regarding t_n is related to the separability of $\{P_{\pi^+}\}$ and Ω^\times . The following proposition shows the two subsets of Λ are separable using the ρ -topology.

Propriété 64. [*Separable hypotheses*]

Supposing $l_1 > l_0$, it exists an $\epsilon > 0$ such that :

$$\Omega^\times \cap V_\epsilon(P_{\pi^+}) = \emptyset \quad (\text{D.24})$$

Démonstration.

Let us suppose for a given $\epsilon > 0$, $P \in V_\epsilon(P_{\pi^+})$ then

$$\begin{aligned} \int_S \pi^\times dP &= \int_S \pi^\times (dP - \pi^+ + \pi^+) \\ &= l_0 + \int_S \pi^\times (dP - \pi^+) \\ &\leq l_0 + \epsilon \|\pi^\times\|_\infty \end{aligned}$$

For ϵ small enough this last quantity is strictly less than l_1 which concludes the proof :

$$P \in V_\epsilon(p_{\pi^+}) \implies P \notin \Omega^\times$$

□

With the ρ -topology on Λ we can formally define the hypotheses $P_n \rightarrow P_\pi$ as quoted below :

$$P_n \xrightarrow[n \rightarrow \infty]{\rho} P_\pi. \quad (\text{D.25})$$

We suppose this convergence granted in the rest of the paper. Additionally, we suppose the two densities f and g are bounded (it implies π^\times and π^+ also are).

Analysis of t_n

To estimate the efficiency of a test, the authors of [45] define $G_n(t) = \mathbb{P}_0(t_n < t)$ together with its opposite $L_n(t) = 1 - G_n(t)$ called the tail probability of the test based on t_n . Then, they apply it for $t = t(s)$ which corresponds to the current empiric observations and define the random variable :

Définition 75 (Tail probability of a test).

$$L_n(s) = L_n(t_n(s)) = 1 - G_n(t_n(s)) = \mathbb{P}_0(t_n \geq t_n(s)) \quad (\text{D.26})$$

As quoted, in their paper, "the smaller the value of $L_n(s)$ the more untenable is the hypothesis H_0 in the light of the observations". Having said that they study the behavior of L_n under each hypothesis; we report the corresponding computations for our test t_n hereafter.

Convergence under H_1 Let us suppose $H_1 : \pi \in \Omega^\times$ then we have almost surely the convergence of our statistical test.

Propriété 65 (Convergence under H_1).

Under $H_1 : \pi \in \Omega^\times$,

$$t_n \xrightarrow[n \rightarrow \infty]{a.s.} \bar{L}(\pi^\times, \pi) \geq l_1$$

Démonstration.

Almost surely, $P_n \xrightarrow[n \rightarrow \infty]{\rho} P_\pi \in \Omega^\times$.

Since $\pi^\times : S \rightarrow \mathbb{R}$ is a \mathcal{B} -measurable and bounded function it implies in particular

$$\int_S \pi^\times dP_n \xrightarrow[n \rightarrow \infty]{a.s.} \int_S \pi^\times \pi = \bar{L}(\pi^\times, \pi).$$

Eventually, given that $\pi \in \Omega^\times$, $\bar{L}(\pi^\times, \pi) \geq l_1$. □

Remarque 82. [Realizations under π^2]

π being bounded on $S = [0, 1]^2$, we want to simulate under P_{π^2} whose density is $\frac{\pi^2}{\int_S \pi^2}$.

We notice that we always have $\pi^2 \leq \|\pi\|_\infty \pi$. Then, applying an usual reject method we simulate a couple

$$M = \left(\frac{\pi^2(X)}{\int_S \pi^2}, \mathbb{U} \frac{\|\pi\|_\infty}{\int_S \pi^2} \pi(X) \right);$$

where $X \sim P_\pi$ and \mathbb{U} is the uniform law on S .

The reject method stands that keeping M only if $M_1 \geq M_2$ conveys a M_1 under P_{π^2} : it gives a method to draw under P_{π^2} by leveraging on a method to draw under P_π .

Furthermore, in our particular application of the reject method, $M_1 \geq M_2$ can be simplified :

$$\begin{aligned} \frac{\pi^2(X)}{\int_S \pi^2} &\geq \mathbb{U} \frac{\|\pi\|_\infty}{\int_S \pi^2} \pi(X); \\ \pi(X) &\geq \mathbb{U} \|\pi\|_\infty. \end{aligned}$$

Eventually it amounts to simulate independently $M = (X, U) \sim P_\pi \otimes \mathbb{U}$ and keep X if and only if $\pi(X) \geq U \|\pi\|_\infty$.

Remarque 83.

According to the previous remark, a law under P_{π^2} will concentrate its values around the mode of π . Indeed, the higher $\pi(X)$ is the most chance we have to keep X in the reject test.

Divergence under H_0 Let us now suppose $H_0 : \pi = \pi^+$ then we similarly show :

Propriété 66 (Convergence under H_0).

Under $H_0 : \pi = \pi^+$,

$$t_n \xrightarrow[n \rightarrow \infty]{a.s.} \bar{L}(\pi^\times, \pi^+) = l_0$$

We want to estimate the asymptotic probability under H_0 to be close to l_1 . A deviation proposition is given below to validate it is exponentially rare in a certain sense.

A first result is the estimation of the probability that $t_n(s)$ with s drawn under $H_0 : \pi = \pi^+$ is larger than a fixed value t . To do so we introduce the random variables $Y = \pi^\times(X)$ where X is under the law π^+ and its logarithmic moment generating function defined by :

$$\phi_0(t) = \log \mathbb{E}_0(\exp(tY)) = \log \left(\int_S \exp(t\pi^\times(s)) \pi^+(s) \right) \quad (\text{D.27})$$

that we suppose is finite for $|t|$ small enough. Out of ϕ is coined its Legendre transform defined by :

$$I(t) = \sup_{x \in \mathbb{R}} \{tx - \phi(t)\} \quad (\text{D.28})$$

Eventually, we introduce the subset of Λ in which $P_n(s)$ must stay so that we have $t_n(s) \geq t$:

$$\Omega_t = \{P \in \Lambda \text{ s.t. } L(P, \pi^\times) \geq t\} \quad (\text{D.29})$$

We can now write a deviation theorem for t_n under H_0 as stated in Theorem 28.

Théorème 28 (Cramér).

For any $t \geq l_0$, we have

$$\lim_{n \rightarrow \infty} \frac{1}{n} \log (\mathbb{P}_0(t_n \geq t)) = -I(t) = -D_{KL}(\Omega_t | P_{\pi^+}). \quad (\text{D.30})$$

Démonstration.

We rewrite t_n using Y :

$$\begin{aligned} \mathbb{P}_0(t_n \geq t) &= \mathbb{P}_0 \left(\int_S \pi^\times(s) dP_n(s) \geq t \right) \\ &= \mathbb{P}_0 \left(\sum_{i=1}^n \pi^\times(X_i) \geq nt \right) \\ &= \mathbb{P}_0 \left(\sum_{i=1}^n Y_i \geq nt \right) \end{aligned}$$

A direct application of Cramér's theorem gives us :

$$\lim_{n \rightarrow \infty} \frac{1}{n} \log \left(\mathbb{P}_0 \left(\sum_{i=1}^n Y_i \geq nt \right) \right) = -I(t) = -D_{KL}(\Omega_t | P_{\pi^+}) \quad (\text{D.31})$$

which concludes the proof. □

The previous theorem provides an estimation of how unlikely it is to have t_n higher than a fixed value t under H_0 . More interesting is to get the probability, under H_0 , that $t_n > t_n(s)$ where s is a sequence of realizations under H_1 , namely the probability of the event "we do not reject H_0 while we should do"; this estimation relies on the Bahadur slope [7] as we shall precise later.

Combining, with s drawn under $P_\pi \in \Omega^\times$, $t_n(s) \xrightarrow[n \rightarrow \infty]{a.s.} \bar{L}(\pi^\times, \pi) \geq l_1$ and theorem 28 precisely provides this estimation as stated in theorem 29.

Théorème 29 (Bahadur's slope).

A whole given sequence s being drawn under H_1 through a probability $P_\pi \in \Omega^\times$ we estimate the limit of $L_n(s')$ where $s' \sim H_0 : P_{\pi^+}$:

$$\lim_{n \rightarrow \infty} \frac{1}{n} \log (\mathbb{P}_0(t_n(s') \geq t_n(s))) = -I(l) = -D_{KL}(\Omega_l | P_{\pi^+}) \geq -I(l_1) = -D_{KL}(\Omega_{l_1} | P_{\pi^+}) \quad (\text{D.32})$$

where $l = \bar{L}(\pi^\times, \pi)$.

Démonstration.

s being drawn under P_π , we have almost surely, as stated in proposition 65

$$t_n(s) \xrightarrow[n \rightarrow \infty]{a.s.} l. \quad (\text{D.33})$$

Furthermore, since $P_\pi \in \Omega^\times$, $l \geq l_1 > l_0$. Therefore, we select δ such that $\frac{l-l_0}{2} > \delta > 0$ and the almost sure convergence provides us with an n_0 such that for any $n \geq n_0$,

$$|t_n(s) - l| \leq \delta. \quad (\text{D.34})$$

It notably implies that for any $n \geq n_0$,

$$\mathbb{P}_0(t_n(s') \geq l + \delta) \leq \mathbb{P}_0(t_n(s') \geq t_n(s)) \leq \mathbb{P}_0(t_n(s') \leq l - \delta). \quad (\text{D.35})$$

Applying theorem 28 with $t = l + \delta \geq l_0$ and $t = l - \delta \geq l_0$ we immediately obtain

$$-I(l + \delta) \leq \lim_{n \rightarrow \infty} \frac{1}{n} \log (\mathbb{P}_0(t_n(s') \geq t_n(s))) \leq -I(l - \delta). \quad (\text{D.36})$$

Since δ is arbitrary it suffices to conclude after invoking the continuity of I .

Finally, the lower-bound comes from the hypothesis H_1 itself : $P_\pi \in \Omega^\times$ requires $l \geq l_1$. \square

The result of Theorem 29 actually consists in showing that the Bahadur slope (see [45]) is $2I(l_1)$. Furthermore, property 64 shows that it is strictly positive. Eventually we have shown that the Bahadur slope of t_n is given by $2I(l_1)$ and is strictly positive hence that the probability of the event "we do not reject H_0 while we should" is exponentially small in n .

D.4 Conclusion

In this article, we extended the discrete notion of coupling functions in a continuous version which operates on the densities of margins as a copula does on their cumulative distribution. Historical theoretical considerations on divergences conveyed a dual and exhaustive approach : either entropy or least squares, either independence or indeterminacy (called indeterminacy as well).

The transposition of the second one in the continuous case led us to consider the existence of a classic copula to capture the associated dependence. While we coined it following natural computations based on its density we proved one cannot extract the indeterminacy dependence without a deep reference to the underlying margins. Namely, although they share the same coupling function by construction, two indeterminacy laws with two respective couples of margins will (almost) never share the same indeterminacy copula. Furthermore, we computed the exact sets of couples of margins with the same indeterminacy copula.

Extending a discrete notion of couple matchings in a continuous space required the definition of the so-called average likelihood \bar{L} . Similarly to the discrete case, we demonstrated that the indeterminacy minimises the average likelihood. Using this property, we built up a test to reject or accept the indeterminacy out of the value of \bar{L} . Eventually, we followed a classic analysis of a statistic test to estimate its efficiency by the associated Bahadur slope and proving it is strictly positive.

D.5 Proof of proposition 61

Démonstration.

We split the proof in four parts :

- express the required relation between (F, G) and (R, S) (part D.5.1)
- define two bounds of λ according to its sign (part D.5.2)
- eliminate the case λ negative (part D.5.3)
- verify h (the density extracted from H) is positive (part D.5.4)

D.5.1 relation between (F, G) and (R, S)

In case an indetermination occurs, the crossed derivation of H equals $r + s - 1$. Let us compute that derivative function and compare it to the required expression.

$$\begin{aligned} \frac{\partial^2 H(x, y)}{\partial x \partial y} &= \frac{\partial^2 (R(x) * G^{-1}(S(y)) + S(y) * F^{-1}(R(x)) - F^{-1}(R(x))G^{-1}(S(y)))}{\partial x \partial y} \\ &= \frac{r(x)s(y)}{g(G^{-1}(S(y)))} + \frac{r(x)s(y)}{f(F^{-1}(R(x)))} - \frac{r(x)s(y)}{f(F^{-1}(R(x)))g(G^{-1}(S(y)))} \end{aligned}$$

Requiring an equality with $r(x) + s(y) - 1$ leads to :

$$r(x) + s(y) - 1 = \frac{r(x)s(y)}{g(G^{-1}(S(y)))} + \frac{r(x)s(y)}{f(F^{-1}(R(x)))} - \frac{r(x)s(y)}{f(F^{-1}(R(x)))g(G^{-1}(S(y)))}$$

If we divide by $r(x)s(y)$ motivated by the details we bring up in remark 84 :

$$\frac{1}{r(x)} + \frac{1}{s(y)} - \frac{1}{r(x)s(y)} = \frac{1}{f(F^{-1}(R(x)))} + \frac{1}{g(G^{-1}(S(y)))} - \frac{1}{f(F^{-1}(R(x)))g(G^{-1}(S(y)))}$$

Aggregating similar terms :

$$\frac{1}{r(x)} \left(1 - \frac{1}{s(y)}\right) = \frac{1}{f(F^{-1}(R(x)))} \left(1 - \frac{1}{g(G^{-1}(S(y)))}\right) + \frac{1}{g(G^{-1}(S(y)))} - \frac{1}{s(y)}$$

To simplify computations, we introduce some notations : α represents the "simple" part and β its "complex" analogue ; the x or y subscript is natural.

$$\begin{aligned} \alpha_x &= 1 - \frac{1}{r(x)} & \alpha_y &= 1 - \frac{1}{s(y)} \\ \beta_x &= 1 - \frac{1}{f(F^{-1}(R(x)))} & \beta_y &= 1 - \frac{1}{g(G^{-1}(S(y)))} \end{aligned}$$

It rewrites :

$$\begin{aligned} \frac{1}{r(x)}\alpha_y &= \frac{1}{f(F^{-1}(R(x)))}\beta_y + \alpha_y - \beta_y \\ \left(\frac{1}{r(x)} - 1\right)\alpha_y &= \left(\frac{1}{f(F^{-1}(R(x)))} - 1\right)\beta_y \\ \alpha_x\alpha_y &= \beta_x\beta_y \end{aligned}$$

If we divide by $\alpha_x\beta_y$ on each side to isolate terms according to their underlying variable (please refer to remark 85 for a justification of the division) we deduce the existence of a constant λ such that :

$$\frac{\alpha_y}{\beta_y} = \frac{\beta_x}{\alpha_x} = \lambda$$

As requested by the expected conclusion, we have to go back to the cumulative distribution functions. They are contained in α and β and the relation $\frac{\beta_x}{\alpha_x} = \lambda$ can be written :

$$r(x) - \frac{r(x)}{f(F^{-1}(R(x)))} = \lambda(r(x) - 1)$$

If we integrate on $[0, x]$:

$$R(x) - F^{-1}(R(x)) = \lambda(R(x) - x)$$

As R is a cumulative distribution function whose derivative (its density) never equals 0 (hypothesis already mentioned), R has an inverse function so that for all x :

$$F^{-1}(x) = x(1 - \lambda) + \lambda R^{-1}(x) \tag{D.37}$$

and symmetrically for any y :

$$G^{-1}(y) = y(1 - \frac{1}{\lambda}) + \frac{1}{\lambda} S^{-1}(y) \tag{D.38}$$

Equations D.37 et D.38 can be written :

$$\begin{aligned} F^{-1}(x) - x &= \lambda(R^{-1}(x) - x) \\ G^{-1}(y) - y &= \frac{1}{\lambda}(S^{-1}(y) - y) \end{aligned} \quad (\text{D.39})$$

We eventually check by applying $C_{f,g}^+$ to (R, S) that those relations are sufficient to formally have :

$$H = C_{f,g}^+(R, S)(x, y) = xS(y) + yR(x) - xy = C_{r,s}^+(R, S)(x, y)$$

This does not ensures H is an eligible cumulative distribution function as we shall precise it below.

D.5.2 Two bounds of λ

We defined a λ linking the four cumulative distribution functions as announced in the proposition and verified it suffices to formally have the expected equality. Yet, defining (R, S) from (F, G) using equation D.37 and equation D.38 does not always generate appropriate functions.

Indeed, the non-decreasing hypothesis on R as well as on S conveys hypothesis on λ .

Let us first suppose that λ is positive then $\lambda R^{-1}(x) = F^{-1}(x) - x(1 - \lambda)$ is a non-decreasing function as well so that for all x , $\frac{1}{f(x)} \geq 1 - \lambda$. Applying similar computations on y for g we eventually obtain :

$$\frac{\max(g)}{\max(g) - 1} \geq \lambda \geq 1 - \frac{1}{\max(f)} \quad (\text{D.40})$$

Similarly, if we suppose λ is negative, we obtain :

$$1 - \frac{1}{\min(f)} \geq \lambda \geq \frac{\min(g)}{\min(g) - 1} \quad (\text{D.41})$$

Those values exist unless g equals 1 in which case $G = Id$ and $C_{f,g}^+ = C_{f,g}^\times$. Hence, without restriction, $\min(f) < 1 < \max(f)$, $\min(g) < 1 < \max(g)$.

D.5.3 λ positive

Eventually, we notice that the hypothesis D.2 which guarantees that we can couple r, s using indetermination is not automatically transferred a priori. Then, as equations D.39 are continuous around $\lambda = 1$, the hypothesis on (f, g) should be enough as soon as λ is close to 1. Let us unfold this remark.

We go back to equations D.37 and D.38 in order to extract an expression of the densities r and s .

$$\begin{aligned} \frac{1}{f(F^{-1}(x))} &= (1 - \lambda) + \frac{\lambda}{r(R^{-1}(x))} \\ r(R^{-1}(x)) &= \frac{1}{\frac{1}{\lambda f(F^{-1}(x))} + \frac{1}{\lambda} - 1} \\ r(\tilde{x}) &= \frac{1}{\frac{1}{\lambda f(F^{-1}(R(x)))} + \frac{1}{\lambda} - 1} \end{aligned}$$

where $\tilde{x} = R^{-1}(x)$ runs over all the segment $[0, 1]$ as well as x . We similarly have :

$$s(\tilde{y}) = \frac{1}{\frac{\lambda}{g(G^{-1}(S(y)))} + \lambda - 1}$$

Hypothesis [D.2](#) requires that for any (x, y)

$$r(x) + s(y) - 1 \geq 0$$

equivalently :

$$r(\tilde{x}) + s(\tilde{y}) - 1 \geq 0$$

it precisely means :

$$\frac{1}{\frac{1}{\lambda f(x)} + \frac{1}{\lambda} - 1} + \frac{1}{\frac{\lambda}{g(y)} + \lambda - 1} \geq 1$$

Using inequalities [D.41](#) and [D.40](#) we know that $r, s \geq 0$, so, we can multiply each side by s leading to

$$\frac{\frac{\lambda}{g(y)} + \lambda - 1}{\frac{1}{\lambda f(x)} + \frac{1}{\lambda} - 1} + 1 \geq \frac{\lambda}{g(y)} + \lambda - 1.$$

Multiplying one more time on each side but by r :

$$\frac{\lambda}{g(y)} + \lambda - 1 + \frac{1}{\lambda f(x)} + \frac{1}{\lambda} - 1 \geq \left(\frac{\lambda}{g(y)} + \lambda - 1 \right) \left(\frac{1}{\lambda f(x)} + \frac{1}{\lambda} - 1 \right)$$

Which can be finally rewritten :

$$\begin{aligned} \left(\frac{1}{\lambda f} + \frac{1}{\lambda} - 2 \right) \left(\frac{\lambda}{g} + \lambda - 2 \right) &\leq 1 \\ \left(\frac{1}{f} + 1 - 2\lambda \right) \left(\frac{1}{g} + 1 - \frac{2}{\lambda} \right) &\leq 1 \end{aligned}$$

The last equality definitely excludes $\lambda \leq 0$ leading to the bounds in equation [D.40](#).

D.5.4 Transfer of hypothesis [D.2](#)

At the end of the previous section, we concluded that λ is positive. Yet, the computations began by requiring $r(x) + s(y) \geq 0$; let us resume them to obtain a second degree equation to solve. Indeed, the last line is equivalent to :

$$\left(\frac{1}{\min(f)} + 1 - 2\lambda \right) \left(\frac{1}{\min(g)} + 1 - \frac{2}{\lambda} \right) \leq 1$$

Which can be written (with obvious notations) :

$$(2m_f + 2m_g)\lambda^2 - (1 + m_f + m_g + 4m_fm_g)\lambda + 2m_f + 2m_g \geq 0$$

and that last inequality is verified if and only if :

$$\frac{1 + m_f + m_g + 4m_fm_g}{2m_f + 2m_g} \leq 2$$

Additionally we notice :

$$\frac{1 + m_f + m_g + 4m_fm_g}{2m_f + 2m_g} \leq \frac{1 + 3m_f + 3m_g}{2m_f + 2m_g} = \frac{3}{2} + \frac{1}{2m_f + 2m_g} \leq 2 \quad (\text{D.42})$$

The last inequality coming from the condition [D.2](#) valid on (f, g) which is thus transferred to (r, s) as soon as λ is suitable (meaning verifying equation [D.7](#)).

Remarque 84 (Division by rs).

If r equals 0, hypothesis [D.2](#) rewrites $r \geq 1$ hence $R = Id$. In that case, $C_{r,s}^+ = C^\times$. It can be shared with $C_{f,g}^+$ if and only if F or G equals Id but it doesn't cover any interesting case.

Remarque 85 (Division by $1 - \frac{1}{f}$).

As we supposed f and g different from 1, we can set ourselves on a interval where they never equal 1. The division is then allowed and the proof similarly conveys the expected constant λ .

□

Bibliographie

- [1] Julien Ah-Pine. *Sur des aspects algébriques et combinatoires de l'analyse relationnelle : applications en classification automatique, en théorie du choix social et en théorie des tresses*. PhD thesis, Paris 6, 2007.
- [2] Julien Ah-Pine. On aggregating binary relations using 0-1 integer linear programming. In *ISAIM*, pages 1–10, 2010.
- [3] Julien Ah-Pine and Jean-François Marcotorchino. Statistical, geometrical and logical independences between categorical variables. *Proceedings of the Applied Stochastic Models and Data Analysis ASMDA2007 Symposium*, 2007.
- [4] Phipps Arabie and Lawrence Hubert. Comparing partitions. *Proceedings of the Fourth Meeting of the International Classification Societies, Cambridge*, 1985.
- [5] Erdal Arıkan. An inequality on guessing and its application to sequential decoding. *IEEE Transactions on Information Theory*, 42, 1996.
- [6] Tetsuo Asano, Binay Bhattacharya, Mark Keil, and Frances Yao. Clustering algorithms based on minimum and maximum spanning trees. In *Proceedings of the fourth annual symposium on Computational Geometry*, pages 252–257, 1988.
- [7] Raghu R. Bahadur. Rates of convergence of estimates and test statistics. *The Annals of Mathematical Statistics*, 38(2) :303–324, 1967.
- [8] Eric Beh, Biagio Simonetti, and Luigi D’Ambra. Partitioning a non-symmetric measure of association for three-way contingency tables. *Journal of Multivariate Analysis*, 98 :1391–1411, 08 2007.
- [9] Mathias Beiglböck, Christian Léonard, and Walter Schachermayer. On the duality theory for the monge-kantorovich transport problem. *Optimal transport. Theories and applications*, pages 216–265, 2014.
- [10] Pierre Bertrand, Michel Broniatowski, and Jean-François Marcotorchino. Independence versus Indetermination : basis of two canonical clustering criteria. working paper or preprint, July 2020.
- [11] Pierre Bertrand, Michel Broniatowski, and Jean-François Marcotorchino. Logical indetermination coupling : a method to minimize drawing matches and its applications. working paper or preprint, December 2020.
- [12] Pierre Bertrand, Michel Broniatowski, and Jean-François Marcotorchino. Continuous indetermination and average likelihood minimization. working paper or preprint, May 2021.

- [13] Pierre Bertrand, Michel Broniatowski, and Jean-François Marcotorchino. Minimization with respect to divergences and applications. In *International Conference on Geometric Science of Information*, pages 818–828. Springer, 2021.
- [14] Vincent D. Blondel, Jean-Loup Guillaume, Renaud Lambiotte, and Etienne LeFebvre. Fast unfolding of communities in large networks. *Journal of statistical mechanics : theory and experiment*, 2008(10), 2008.
- [15] Federico Botta and Charo I. Del Genio. Finding network communities using modularity density. *Journal of Statistical Mechanics : Theory and Experiment*, 2016(12) :123402, 2016.
- [16] Thierry Brenac. *Contribution des méthodes de partition centrale à la mise en évidence expérimentale de catégories cognitives*. Collections de l'INRETS, 2002.
- [17] Michel Broniatowski, Jana Jurečková, Ashok K. Moses, and Emilie Miranda. Composite tests under corrupted data. *Entropy*, 21(1) :63, 2019.
- [18] Christoph Bunte and Amos Lapidoth. Encoding tasks and Rényi entropy. *IEEE Transactions on Information Theory*, 60(9) :5065–5076, Sep 2014.
- [19] Rainer E. Burkard, Bettina Klinz, and Rüdiger Rudolf. Perspectives of Monge properties in optimization. *Discrete Applied Mathematics*, 70 :95–161, 1996.
- [20] Christian Calzada. Modèles d'interaction spatiale. *Publication du Club d'Echanges sur le Transport des Marchandises, DAEI/ SES*, pages 1–47, 1999.
- [21] Romain Campigotto, Patricia Conde-Céspedes, and Jean-Loup Guillaume. A generalized and adaptive method for community detection. *arXiv preprint arXiv :1406.2518*, 2014.
- [22] Guillaume Carlier and Ivar Ekeland. Structure des villes et équilibre spatial. In *Foils presentation*. Collège de France, 2008.
- [23] Mingming Chen, Tommy Nguyen, and Boleslaw K. Szymanski. A new metric for quality of network community structure. *arXiv preprint arXiv :1507.04308*, 2015.
- [24] Aaron Clauset, Mark E. J. Newman, and Christopher Moore. Finding community structure in very large networks. *Physical review E*, 70(6) :066111, 2004.
- [25] Patricia Conde-Céspedes. *Modélisations et extensions du formalisme de l'analyse relationnelle mathématique à la modularisation des grands graphes*. PhD thesis, Paris 6, 2013.
- [26] Jean-Luc Coron. *Quelques exemples de jeux à champ moyen*. PhD thesis, Paris Sciences et Lettres (ComUE), 2017.
- [27] Thierry Cruanes, Christian Mauceri, Jean-François Marcotorchino, and Hammou Messatfa. Indexations par matrices de comparaisons par paires pour olap. *Brevet IBM*, 1997.
- [28] Imre Csiszár. Sanov property, generalized I-projection and a conditional limit theorem. *The Annals of Probability*, pages 768–793, 1984.
- [29] Imre Csiszár et al. Why least squares and maximum entropy? an axiomatic approach to inference for linear inverse problems. *The annals of statistics*, 19(4) :2032–2066, 1991.

- [30] W. Edwards Deming and Frederick F. Stephan. On the least squares adjustment of a sampled frequency table when the expected marginal totals are known. *The Annals of Mathematical Statistics*, 11 :427–444, 1940.
- [31] Patrick Doreian, Vladimir Batagelj, and Anuska Ferligoj. *Advances in Network Clustering and Blockmodeling*. John Wiley & Sons, 2020.
- [32] Sven Erlander and Neil F. Stewart. *The use of entropy maximising models*, volume 58. Utrecht, 1985.
- [33] Lawrence C. Evans. *Partial Differential Equations and Monge-Kantorovich Mass Transfer*. S. T. Yau, Editor, 1997.
- [34] Dario Fasino and Francesco Tudisco. A modularity based spectral method for simultaneous community and anti-community detection. *Linear Algebra and its Applications*, 542 :605–623, 2018.
- [35] Edward W. Forgy. Cluster analysis of multivariate data : efficiency versus interpretability of classifications. *biometrics*, 21 :768–769, 1965.
- [36] Santo Fortunato. Community detection in graphs. *Physics reports*, 486(3-5) :75–174, 2010.
- [37] Santo Fortunato and Marc Barthelemy. Resolution limit in community detection. *Proceedings of the national academy of sciences*, 104(1) :36–41, 2007.
- [38] Maurice Fréchet. Sur les tableaux de corrélations dont les marges sont données. *Annales de l'Université de Lyon, Section. A*, 14 :53–77, 1951.
- [39] Maurice Fréchet. Sur les tableaux de corrélations dont les marges et les bornes sont données. *Revue de l'Institut de Statistique*, 28 :10–32, 1960.
- [40] Bernard Fustier. Echanges commerciaux euro-méditerranéens : essai d'analyse structurale. *Revue des Sciences Economiques et de Gestion*, 3 :1–25, 2004.
- [41] Michelle Girvan and Mark E. J. Newman. Community structure in social and biological networks. *Proceedings of the National Academy of Sciences of the United States of America*, page 7821–7826, 2002.
- [42] Pablo M. Gleiser and Leon Danon. Community structure in jazz. *Advances in complex systems*, 6(04) :565–573, 2003.
- [43] Daniel Gomez, J. Tinguaro Rodríguez, Javier Yanez, and Javier Montero. A new modularity measure for fuzzy community detection problems based on overlap and grouping functions. *International Journal of Approximate Reasoning*, 74 :88–107, 2016.
- [44] Leo A. Goodman and William H. Kruskal. Measures of association for cross classifications. *Measures of association for cross classifications*, pages 2–34, 1979.
- [45] Piet Groeneboom and Jacobus Oosterhoff. Bahadur efficiency and probabilities of large deviations. *Statistica Neerlandica*, 31(1) :1–24, 1977.
- [46] Piet Groeneboom, Jacobus Oosterhoff, and Frederik H. Ruymgaart. Large deviation theorems for empirical probability measures. *The Annals of Probability*, pages 553–586, 1979.
- [47] Richard S. Hamilton et al. Three-manifolds with positive ricci curvature. *J. Differential geom*, 17(2) :255–306, 1982.

- [48] Mickaël Hoerdt and Damien Magoni. Completeness of the internet core topology collected by a fast mapping software. In *Proceedings of the 11th International Conference on Software, Telecommunications and Computer Networks*, pages 257–261, 2003.
- [49] Alan J. Hoffman. On simple linear programming problems. *Proceedings of the Seventh Symposium in Pure Mathematics of the AMS*, pages 317–327, 1963.
- [50] Myles Hollander, Douglas A. Wolfe, and Eric Chicken. *Nonparametric statistical methods*, volume 751. John Wiley & Sons, 2013.
- [51] Benoit Huyot, Yves Mabiala, and Jean-François Marcotorchino. Optimal transport, independence versus indetermination duality, impact on a new copula design. *International Conference on Geometric Science of Information*, pages 68–76, 2015.
- [52] Svante Janson and Jan Vegelius. Correlation coefficients for nominal scales. *Uppsala : Department of Statistics*, 1977.
- [53] Svante Janson and Jan Vegelius. On the applicability of truncated component analysis based on correlation coefficients for nominal scales. *Applied Psychological Measurement*, pages 135–145, 1978.
- [54] Svante Janson and Jan Vegelius. The J-index as a measure of nominal scale response agreement. *Applied Psychological Measurement*, pages 111–121, 1982.
- [55] Lenoid Kantorovich. On the translocation of masses. *Comptes Rendus (Doklady) Acad. Sci. URSS (N.S.)*, 37 :199–201, 1942.
- [56] Leo Katz. A new status index derived from sociometric analysis. *Psychometrika*, 18(1) :39–43, 1953.
- [57] Hans G. Kellerer. Duality theorems for marginal problems. *Zeitschrift für Wahrscheinlichkeits theorie und verwandte Gebiete*, pages 399–432, 1984.
- [58] Maurice G. Kendall. Rank correlation methods. *Hafner*, 1948.
- [59] M. Ashok Kumar, Albert Sunny, Ashish Thakre, and Ashisha Kumar. A unified framework for problems on guessing, source coding and task partitioning, 2019.
- [60] Andrea Lancichinetti and Santo Fortunato. Limits of modularity maximization in community detection. *Physical review E*, 84(6) :066122, 2011.
- [61] Israël-César Lerman and Philippe Peter. Structure maximale pour la somme des carrés d’une contingence aux marges fixées : une solution algorithmique programmée. *RAIRO Recherche Opérationnelle*, 22 :83–136, 1988.
- [62] John Lott and Cédric Villani. Ricci curvature for metric-measure spaces via optimal transport. *Annals of Mathematics*, pages 903–991, 2009.
- [63] Jean-François Marcotorchino and Patricia Conde-Céspedes. Optimal transport and minimal trade problem, impacts on relational metrics and applications to large graphs and networks modularity. In *International Conference on Geometric Science of Information*, pages 169–179. Springer, 2013.
- [64] Jean-François Marcotorchino. Utilisation des comparaisons par paires en statistique des contingences. *Publication du Centre Scientifique IBM de Paris et Cahiers du Séminaire Analyse des Données et Processus Stochastiques Université Libre de Bruxelles*, pages 1–57, 1984.

- [65] Jean-François Marcotorchino. Maximal association theory as a tool of research. *Classification as a tool of research*, W.Gaul and M. Schader editors, North Holland Amsterdam, 1986.
- [66] Jean-François Marcotorchino. Seriation problems :an overview. *Applied Stochastic Models and Data Analysis*, 7 :139–151, 1991.
- [67] Jean-François Marcotorchino and Najoi El Ayoubi. Paradigme logique des écritures relationnelles de quelques critères fondamentaux d’association. *Revue de Statistique Appliquée*, 39 :25–46, 1991.
- [68] Jean-François Marcotorchino and Pierre Michaud. *Optimisation en Analyse Ordinale des Données*. Ed Masson, Paris, 1979.
- [69] James L Massey. Guessing and entropy. In *Proceedings of 1994 IEEE International Symposium on Information Theory*, page 204. IEEE, 1994.
- [70] Sue McNeil and Chris Hendrickson. A note on alternative matrix entry estimation techniques. *Transportation Research*, 19B :509–519, 1985.
- [71] Hammou Messatfa. Maximal association for the sum of squares of a contingency table. *Revue RAIRO, Recherche Opérationnelle*, 24 :29–47, 1990.
- [72] Alan Mislove, Massimiliano Marcon, Krishna P. Gummadi, Peter Druschel, and Bobby Bhattacharjee. Measurement and analysis of online social networks. *Proceedings of the 5th ACM/Usenix Internet Measurement Conference (IMC’07)*, 2007.
- [73] Gaspard Monge. Mémoire sur la théorie des déblais et des remblais. *Histoire de l’Académie Royale des Sciences de Paris avec les Mémoires de Mathématique et de Physique pour la même année*, pages 666–704, 1781.
- [74] Amel Najah Idrissi. *Contribution à l’unification de critères d’association pour variables qualitatives*. PhD thesis, Paris 6, 2000.
- [75] Mariá CV Nascimento. Community detection in networks via a spectral heuristic based on the clustering coefficient. *Discrete Applied Mathematics*, 176 :89–99, 2014.
- [76] Mark E. J. Newman and Michelle Girvan. Finding and evaluating community structure in networks. *Physical review E*, 69(2) :026113, 2004.
- [77] Chien-Chun Ni, Yu-Yao Lin, Feng Luo, and Jie Gao. Community detection on networks with Ricci flow. *Scientific reports*, 9(1) :1–12, 2019.
- [78] Frank Nielsen. What is an information projection. *Notices of the AMS*, 65(3) :321–324, 2018.
- [79] Yann Ollivier. Ricci curvature of markov chains on metric spaces. *Journal of Functional Analysis*, 256(3) :810–864, 2009.
- [80] Otto Opitz and Henning Paul. Aggregation of ordinal judgements based on condorcet’s majority rule. *Data Analysis and Decision Support. Studies in Classification, Data Analysis, and Knowledge Organization*. Springer, Berlin, Heidelberg, 2005.
- [81] Karl Pearson. On the theory of contingency and its relation to association and normal correlation. *Draper’s Company Research Memoirs, Biometric Series, London : Cambridge University Press*, 1904.
- [82] Krishnan Raghavachari et al. On a theorem of bahadur on the rate of convergence of test statistics. *The Annals of Mathematical Statistics*, 41(5) :1695–1699, 1970.

- [83] Jörg Reichardt and Stefan Bornholdt. Statistical mechanics of community detection. *Physical review E*, 74(1) :016110, 2006.
- [84] Aude Rudneff-Chupin. Transport optimal et mouvements de foule. In *Présentation en Foils, Colloque GTN 2011*. Université Paris XI, Paris Sud, 2011.
- [85] Rafael Santiago and Luís C. Lamb. Efficient modularity density heuristics for large graphs. *European Journal of Operational Research*, 258(3) :844–865, 2017.
- [86] Gilbert Saporta. *Probabilités, analyse des données et statistique*. Editions Technip, 2006.
- [87] Abe Sklar. Random variables, joint distribution functions, and copulas. *Kybernetika*, 9(6) :449–460, 1973.
- [88] Hugo Steinhaus. Sur la division des corps matériels en parties. *Bulletin de l'académie polonaise des sciences, v. 4, no. 12*, page 801–804, 1957.
- [89] Eric Stemmelen. Tableaux d'échanges, description et prévision. *Cahiers du Bureau Universitaire de Recherche Opérationnelle*, 28, 1977.
- [90] Wolfgang Stummer. Optimal transport with some directed distances. In *International Conference on Geometric Science of Information*. Springer, 2021.
- [91] Yves Tillé. Methodes d'ajustement d'un tableau à des marges : propriétés des ajustements et choix des critères à optimiser. *Statistique et Analyse des Données*, 14 :53–78, 1989.
- [92] Alexander A. Tschuprow. Principles of the mathematical theory of correlation. *New York : William Hodge*, 1939.
- [93] Tim Van Erven and Peter Harremoës. Rényi divergence and kullback-leibler divergence. *IEEE Transactions on Information Theory*, 60(7) :3797–3820, 2014.
- [94] Cédric Villani. *Optimal Transport, Old and New*. Springer –Verlag, 2000.
- [95] Cédric Villani. *Topics in Optimal Transportation*, volume 58 of 10. The American Mathematical Society : Graduate Studies in Mathematics, 2003.
- [96] Ken Wakita and Toshiyuki Tsurumi. Finding community structure in mega-scale social networks. In *Proceedings of the 16th international conference on World Wide Web*, pages 1275–1276, 2007.
- [97] Frans J. Willekens. Specification and calibration of spatial interaction models a contingency- table perspective. *Tijdschrift voor Economie in Sociaal Geografie*, 74 :239–252, 1983.
- [98] Alan G. Wilson. A statistical theory of spatial distribution models. *Transportation Research*, 1 :253–269, 1967.
- [99] Alan G. Wilson. The use of entropy maximising models. *Journal of transport economics and policy*, 3 :108–126, 1969.
- [100] Alan G. Wilson. *Entropy in urban and regional modelling*, volume 1. Routledge, 2011.
- [101] Jaewon Yang and Jure Leskovec. Defining and evaluating network communities based on ground-truth. *Knowledge and Information Systems*, 42(1) :181–213, 2015.
- [102] Genane Youness and Gilbert Saporta. Sur les indices de comparaison de deux classifications. *10èmes rencontres de la Société Francophone de Classification*, 2003.

- [103] Wayne W. Zachary. An information flow model for conflict and fission in small groups. *Journal of anthropological research*, 33(4) :452–473, 1977.