Random subsequences problems : asymptotics, variance, and quantum statistics.

Clément Deslandes

INSTITUT POLYTECHNIQUE DE PARIS

NNT : 2023IPPAX130

Thèse de doctorat

Random Subsequences Problems: Asymptotics, Variance, and Quantum Statistics.

Thèse de doctorat de l'Institut Polytechnique de Paris
préparée à École polytechnique, Georgia Institute of Technology

École doctorale n°574 École doctorale de mathématiques Hadamard (EDMH)
Spécialité de doctorat : Mathématiques

Thèse présentée et soutenue à Palaiseau, le 15 décembre 2023, par

CLÉMENT DESLANDES

Composition du Jury :

Cristina Butucea Professeur, ENSAE (CREST)	Présidente
Charles Bordenave Directeur de recherche, Institut de Mathématiques de Marseille (CNRS)	Rapporteur
Jean-Christophe Breton Professeur, Université de Rennes (IRMAR)	Rapporteur
Michael Damron Professeur, Georgia Institute of Technology	Examinateur
Karim Lounici Professeur, École polytechnique (CMAP)	Directeur de thèse
Christian Houdré Professeur, Georgia Institute of Technology	Co-directeur de thèse
Pierre-Loïc Méliot Maître de conférences, Université Paris-Saclay	Examinateur

Remerciements

Je tiens à remercier Christian Houdré, qui m'a fait découvrir le monde merveilleux des sous-séquences communes et/ou croissantes. Merci infiniment pour votre grande disponibilité, votre gentillesse, et vos encouragements. Je remercie également chaleureusement Karim Lounici pour sa supervision et pour avoir rendu possible cette cotutelle internationale.

Cette thèse n'aurait pas été possible sans le soutien de ma famille et mes amis. En particulier, merci Grégoire pour les innombrables sessions de travail en commun, qui furent toujours un plaisir, et pour les discussions mathématiques stimulantes. Merci Charles pour m'avoir dépanné informatiquement. Merci à Guillaume d'avoir trouvé plus d'une fois les bons mots et de m'avoir donné confiance. Et merci à mes parents, à qui je dois tout, et qui m'ont toujours encouragé et inspiré.

Enfin, je te remercie Esther, pour ton soutien quotidien indéfectible et ton écoute inégalable, même quand il s'agit de mathématiques (parfois). Incapable d'écrire tout ce que je te dois, je te dédie les cent pages suivantes.

Contents

Introduction (condensée, en Français)

Dans ce travail, nous considérons des problèmes de mots aléatoires et leurs applications. Un mot aléatoire de longueur n est une suite finie de variables aléatoires i.i.d. à valeurs dans un ensemble fini appelé alphabet (par exemple, une suite de lancers de pièces FPPPFPPF est un mot aléatoire de longueur 8). Le point de départ est le problème suivant: étant donné deux mots aléatoires, "qu'ont-ils en commun"? Le problème d'analyser la ressemblance entre deux mots aléatoires a émergé indépendamment dans de nombreux domaines, notamment l'informatique, la biologie, la linguistique...

Malheureusement, peu de choses ont été démontrées sur ce problème fondamental de la longueur maximale d'une sous-séquence commune (notée LCS): la distribution limite, et même le comportement asymptotique de la variance, ne sont pas connus. Cependant, en modifiant légèrement le problème, il devient plus simple de trouver la distribution limite: le premier chapitre de notre travail est dédié à la limite en distribution de la longueur maximale des sous-séquences communes et croissantes. Cela signifie que l'on considère un alphabet ordonné, disons $1, \dots, m$, et les sous-séquences qui sont simplement faites d'un bloc de 1's, suivi d'un bloc de 2's, ... et ainsi de suite (la sous-séquence est croissante, mais pas strictement). Jusqu'à présent, seul le cas où les deux mots aléatoires ont leurs lettres suivant la même loi uniforme sur $1, \dots, m$ avait été traité [12], et il y avait une conjecture dans le cas d'une distribution commune aux deux mots. Nous nous plaçons dans le cas le plus général: les deux mots ont deux distributions éventuellement différentes. Dans ce cadre, nous sommes capable de donner la distribution limite, ainsi que le comportement asymptotique de l'espérance et de la variance.

Dans le chapitre deux, nous nous intéressons au problème de la variance de LCS. Déterminer si la variance est d'ordre n est un problème ouvert important (notamment car cela pourrait permettre de donner la distribution asymptotique de LCS). En introduisant des outils plus généraux, des résultats partiels pour la variance de LCS sont obtenus. Si $X_1, \dots, X_n$ sont des variables aléatoires i.i.d. et S est une fonction telle que $S(X_1, \dots, X_n)$ a une variance finie, suivant [8], posons pour $k \in \{1, \dots, n\}$

$$B_k := \mathbb{E}\frac{1}{n!}\sum_{i \in \mathfrak{S}_n} S(S^{i_1,\dots,i_{k-1}} - S^{i_1,\dots,i_k}),$$

où $\mathfrak{S}_n$ est le groupe symétrique d'ordre n et $S^{i_1,\dots,i_{k-1}}$ désigne $S(X_1, \dots, X_n)$ mais avec $X_{i_1}, \dots, X_{i_{k-1}}$ remplacés par des copies indépendantes. Il a été montré que $\operatorname{Var} S = B_1 + \cdots + B_n$ et que $(B_k)_{1 \le k \le n}$ est décroissante, nous prouvons que $(B_k)_{1 \le k \le n}$ est absolument monotone. Pour des fonctions de variables aléatoires indépendantes, diverses bornes supérieures et inférieures sont étudiées dans différents cadres. Elles sont ensuite appliquées au cas Bernoulli, Gaussien, indéfiniment divisible et à des variables aléatoires à valeurs dans un espace de Banach. Les méthodes vont du jackknife aux semi-groupes. De nouvelles applications sont présentées, permettant de retrouver et améliorer, en particulier, tous les encadrements connus de la variance de la longueur des plus longs sous-mots communs de deux mots aléatoires. Nous trouvons une nouvelle borne inférieure de la variance, d'ordre n, dans le cas d'une distribution des lettres en Bernoulli de paramètre p, avec p "petit"

mais améliorant ce qui avait déjà été fait dans ce cas. Nous donnons aussi des conditions permettant de retrouver que la variance est d'ordre n dans le cas uniforme, l'une est uniquement testée par des simulations numériques et donne un argument de plus que la variance est bien d'ordre n, l'autre est un résultat apparaissant déjà dans [28] (nous prouvons qu'il implique la linéarité asymptotique de la variance).

Dans le troisième et dernier chapitre, nous considérons la longueur maximale d'une sous-séquence croissante (notée LIS) d'un seul mot aléatoire, et le lien étonnant avec les statistiques quantiques. En effet, estimer le spectre d'une matrice de densité d'un système quantique à partir de n copies de ce système équivaut à estimer la distribution des lettres d'un mot de longueur n étant donné la forme de son tableau obtenu par l'algorithme Robinson–Schensted–Knuth (RSK). Il existe un estimateur simple du spectre: l'estimateur de Young empirique, qui renvoie simplement, si $\lambda_1, \dots, \lambda_d$ est la forme du diagramme de Young, $\lambda_1/n, \dots, \lambda_d/n$. Cet estimateur (Empirical Young Diagram, EYD) a un risque quadratique d'ordre d/n, mais existe-t-il un estimateur de risque quadratique d'ordre inférieur? C'est à cette question que nous apportons quelques éléments de réponse.

1 Les plus longues sous-séquences communes

En informatique, trouver la plus sous-séquence commune correspond à trouver le nombre minimal d'insertions et de suppressions pour passer d'un mot à l'autre. C'est pourquoi le programme Unix calcule la longueur maximale des sous-séquences communes. En biologie, un ADN est représenté par un mot aléatoire avec des lettres dans l'ensemble $\{A, T, G, C\}$. D'après la théorie de l'évolution, ce mot évolue par l'insertion de nouvelles lettres, ainsi un mot d'un ancêtre d'une espèce est une sous-séquence (un sous-mot) du mot correspondant à l'espèce. Quand deux espèces ont sous-séquence commune conséquente, on peut en déduire que cela n'est pas le fruit du hasard, et est dû au fait qu'elles ont un ancêtre commun (voir par exemple [74]). Cependant, comme [61] et [3] le notent, on doit faire attention car deux mots aléatoires ont en commun 65% de leur longueur en moyenne, ce qui peut être perçu comme contre-intuitif. Le calcul de la longueur moyenne de plus longues sous-séquences, nécessaire pour faire des statistiques en biologie, est une motivation pour l'étude des plus longues sous-séquences initiée par Chvátal et Sankoff [15] en 1975.

Soit $\mathcal{A}$ un ensemble fini, appelé alphabet, et appelons "mots" les séquences à valeurs dans $\mathcal{A}$. Par exemple, pour modéliser les brins d'ADN, on prend $\mathcal{A} = \{A, T, G, C\}$. Pour $(x_1, \dots, x_s)$, $(y_1, \dots, y_t)$ deux séquences dans $\mathcal{A}$, on note $LCS(x_1 \dots x_s; y_1 \dots y_t)$ le plus grand entier k tel qu'il existe $1 \leq i_1 < \dots < i_k \leq s$, $1 \leq j_1 < \dots < j_k \leq t$ vérifiant $a_{i_1} = b_{j_1}, \dots, a_{i_k} = b_{j_k}$, ou 0 s'il n'existe pas de tel entier.

Par exemple, $LCS(ACCGAT; GACT) = 3$ car on peut prendre $i_1 = 1, i_2 = 2, i_3 = 6$ et $j_1 = 2, j_2 = 3, j_3 = 4$, ce qui extrait le mot ACT des deux mots, mais on ne peut extraire un mot plus long. Le mot ACT est une plus longue sous-séquence, pas la seule, par exemple GAT l'est aussi. Graphiquement, cela consiste à relier un maximum de lettres identiques sans croisements (voir figure 1).

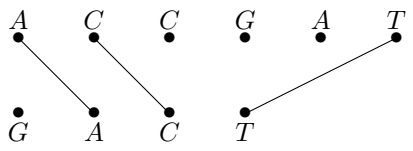


Figure 1: ACT est une sous-séquence de longueur maximale

On peut aussi voir cela comme un problème de percolation, en mettant le premier mot suivant l'axe des abscisses, le second suivant l'axe des ordonnées, le but est de trouver le chemin strictement

croissant (chaque coordonnée augmente strictement) ayant un maximal de points (représentant l'égalité des lettres). Ici, le maximum est trois (voir figure 2).

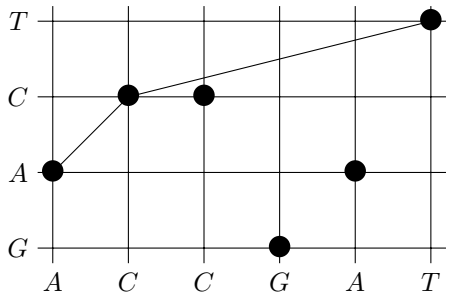


Figure 2: ACT comme un chemin strictement croissant avec un nombre maximal de points.

Cela peut être réduit à un problème de percolation sur les arêtes. Considérons le quadrillage plan avec des arêtes reliant (x, y) à $(x + 1, y + 1)$, et définissons les poids de ces arêtes comme 1 si les lettres correspondantes sont égales et zéro sinon, avec les autres arêtes (horizontales, verticales) de poids nul. On cherche un chemin de $(0, 0)$ à (s, t) de poids maximal, suivant le graphe orienté: c'est un problème de percolation de dernier passage (voir figure 3).

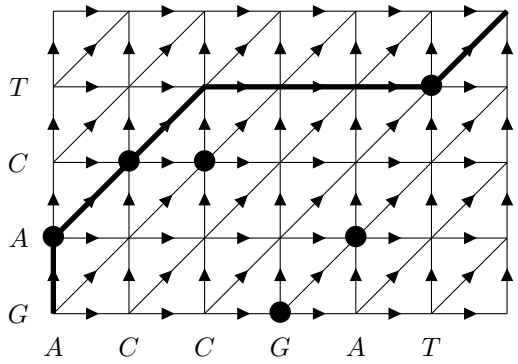


Figure 3: ACT comme un chemin de poids maximal.

On peut maintenant considérer la ressemblance entre deux mots aléatoires. Soit $(X_k)_{k\geq 1}$, $(Y_k)_{k\geq 1}$ deux suites de variables aléatoires indépendantes identiquement distribuées, à valeurs dans un alphabet fini $\mathcal{A}$. Nous nous intéressons à la variable aléatoire $LCS(X_1 \dots X_n; Y_1 \dots Y_n)$, simplement notée LC_n.

Par exemple, si $\mathcal{A} = \{0, 1\}$ et les X_k's sont des Bernoulli de paramètre $1/2$, alors LC_1 est aussi Bernoulli de paramètre $1/2$, mais $\mathbb{P}(LC_2 = 0) = 1/4$, $\mathbb{P}(LC_2 = 1) = 1/2$ et $\mathbb{P}(LC_2 = 2) = 1/4$. Il est difficile de calculer la distribution de LC_n explicitement, donc nous nous intéressons surtout à sa distribution asymptotique, en particulier pour l'espérance et la variance.

Commençons par une propriété de LC_n:

Proposition 1.1 (Chvátal et Sankoff, 1975, [15])**.** *Pour tous $m, n \geq 1$,*

$$\mathbb{E}LC_{m+n} \geq \mathbb{E}LC_m + \mathbb{E}LC_n.$$

Proof. On coupe en deux comme sur la figure 4.

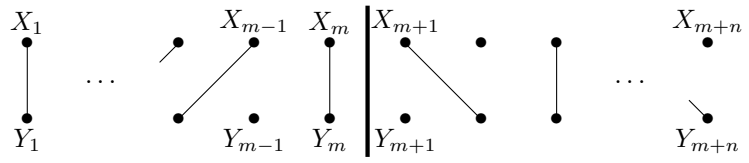


Figure 4: Couper deux mots de longueur $m+n$

On a

$$LCS(X_1 \dots X_{m+n}; Y_1 \dots Y_{m+n}) \geq LCS(X_1 \dots X_m; Y_1 \dots Y_m) + LCS(X_{m+1} \dots X_{n+m}; Y_{m+1} \dots Y_{m+n})$$

et comme $LCS(X_{m+1} \dots X_{m+n}; Y_{m+1} \dots Y_{m+n})$ a la même distribution que LC_n, le résultat suit. □

On dit que la suite $(\mathbb{E}LC_n)_{n\geq 1}$ est superadditive. Le corollaire suivant est une conséquence immédiate du lemme de Fekete (1923), et de l'inégalité $LC_n \leq n$:

Corollaire 1.2. *La suite* $(\mathbb{E}LC_n/n)_{n\geq 1}$ *converge vers* $\sup_n \mathbb{E}LC_n/n \leq 1$.

Habituellement nous notons γ la limite de $\mathbb{E}LC_n/n$, ou γ_k dans le cas d'un alphabet avec k lettres uniformément distribué. Même dans les cas les plus simples, il est difficile de déterminer γ: par exemple, il n'y a pas de valeur exacte connue de γ_2. Le meilleur encadrement est $0.788071 \leq \gamma_2 \leq 0.826280$ [52], ce qui montre l'inexactitude d'une ancienne conjecture [66], $\gamma_2 = 2/(1+\sqrt{2}) \simeq 0.828427$. Plus récemment, Tiskin [68] a montré que γ_2 est algébrique, mais sans permettre d'estimations numériques.

De nombreuses simulations numériques, par exemple [3], semblent montrer que pour un alphabet à deux lettres, γ est minimal lorsque la distribution est uniforme. Cela est intuitif: la probabilité que deux lettres coïncident est minimale quand $p = 1/2$, et nous nous attendons à avoir l'espérance de LC_n minimale quand il y a le moins (en moyenne) de paires de lettres identiques. Dans [3], les auteurs tentent de prouver cela, mais leur preuve n'est pas convaincante, et il semble que ce fait ne soit pas prouvé à ce jour.

Nous nous intéressons à présent à comment LC_n est proche de son espérance, avec des techniques de concentration classiques. On évalue ensuite la variance de LC_n à l'aide de l'inégalité d'Efron-Stein, ce qui est le point de départ du chapitre deux de notre travail.

1.1 Inégalités de concentration

On rappelle le résultat suivant (un léger raffinement par rapport à ce que donne l'application de l'inégalité d'Hoeffding):

Théorème 1.1 (McDiarmid, 1989 [53])**.** *Pour tout* $t > 0$,

$$\mathbb{P}(LC_n - \mathbb{E}LC_n \geq t) \leq \mathrm{e}^{-\frac{t^2}{n}}.$$

En appliquant la même méthode à la variable $-LC_n$, on a:

Corollaire 1.3. *Pour tout* $\epsilon > 0$,

$$\mathbb{P}\left(\left|\frac{LC_n}{n} - \frac{\mathbb{E}LC_n}{n}\right| \geq \epsilon\right) \leq 2\mathrm{e}^{-n\epsilon^2}.$$

Comme $\mathbb{E}LC_n/n$ converge vers $\gamma \in \mathbb{R}$, ce corollaire entraîne la convergence de LC_n/n vers γ. Plus précisément,

Proposition 1.4. *Presque sûrement,* $|LC_n - \mathbb{E}LC_n| = \mathcal{O}\left(\sqrt{n \log n}\right)$.

1.2 La variance

Nous avons vu que l'ordre asymptotique de l'espérance est plutôt bien compris. L'ordre de la variance l'est beaucoup moins.

Nous commençons par une borne supérieure. Le corollaire 1.3 implique $\operatorname{Var} LC_n \leq 8n$. Pour un résultat légèrement meilleur, nous utilisons le résultat suivant, une genéralisation du résultat original d'Efron-Stein [20] aux fonctions non symmétriques.

Théorème 1.2 (Steele, 1986 [65])**.** *Soient* $Z_1, \ldots, Z_n, W_1, \ldots, W_n$ *des variables aléatoires i.i.d. et* $f : \mathbb{R}^n \to \mathbb{R}$ *Borel-mesurable. Soit* $F = f(Z_1, \ldots, Z_n)$ *et, pour* $1 \leq i \leq n$,

$$F_i = f(Z_1, \ldots, Z_{i-1}, W_i, Z_{i+1}, \ldots, Z_n).$$

Alors

$$\operatorname{Var} F \leq \frac{1}{2} \sum_{i=1}^{n} \mathbb{E}((F - F_i)^2).$$

En appliquant ce résultat à LC_n:

Corollaire 1.5 (Steele, 1986 [65])**.** *Soit, pour tout* $a \in \mathcal{A}$, $p_a = \mathbb{P}(X_1 = a)$ *(les* X_i *et* Y_i *sont i.i.d. à valeurs dans* $\mathcal{A}$*). Alors*

$$\operatorname{Var} LC_n \leq n \left(1 - \sum_{a \in \mathcal{A}} p_a^2\right).$$

Nous revisitons et améliorons cette borne supérieure dans le chapitre 2. Le plus délicat est de donner une borne inférieure non triviale. Il semble que ni la divergence vers plus l'infini ni même la croissance de la variance n'aient été prouvés dans le cas uniforme binaire. Cependant, pour certaines distributions de lettres (où une lettre a une très forte probabilité d'apparaître, les autres très petites), [48] (dans le cas binaire) et [36] ont prouvés que la variance avait une borne inférieure d'ordre n, et donc était d'ordre n. Waterman [74] a fait la conjecture que la variance était toujours d'ordre n, et nous ne connaissons aucun contre-exemple à cette conjecture. Dans tous les cas, il est difficile d'évaluer l'ordre de la variance numériquement, car la variance croît lentement pour les petites valeurs de n, ce qui semble avoir conduit Chvátal et Sankoff [15] à la conjecture d'une variance d'ordre $n^{2/3}$. Mais des simulations plus récentes (Juillet 2016, [50]) sont en accord avec Waterman. Ces simulations par Monte-Carlo avec 10000 tirages, pour n allant entre 50000 à plus de 1000000, donnent l'exposant α tel que Cn^α approche au mieux la variance. L'exposant est très proche de 1 et ce pour les trois distributions de lettres testées (l'exposant trouvé est d'autant plus proche de 1 que la distribution a une entropie faible, mais dans le cas binaire uniforme, l'exposant reste plutôt proche de 1: $\alpha = 0.9086$).

Le problème de trouver l'ordre de la variance est d'autant plus intéressant qu'il a été prouvé [28] que si la variance était d'ordre n, alors LC_n, renormalisé, convergeait en distribution vers une Gaussienne. Plus précisément:

Théorème 1.3 (Houdré et Işlak, 2022)**.** *Supposons qu'il existe* $C > 0$ *tel que* $\operatorname{Var} LC_n \geq Cn$*, alors pour tout* $\eta \in (0, 1/10)$,

$$d_W\left(\frac{LC_n - \mathbb{E}LC_n}{\sqrt{\operatorname{Var} LC_n}}, \mathcal{G}\right) = \mathcal{O}\left(\frac{1}{n^\eta}\right)$$

où $\mathcal{G}$ *est une Gaussienne centrée réduite et* d_W *est la distance de Wasserstein.*

Comme mentionné dans [28], une borne inférieure plus faible suffit: Supposons qu'il existe $C > 0$ et $\beta > 9/10$ tels que $\operatorname{Var} LC_n \geq Cn^\beta$, alors pour tout $\varepsilon \in (0, \beta - 9/10)$,

$$d_W\left(\frac{LC_n - \mathbb{E}LC_n}{\sqrt{\operatorname{Var} LC_n}}, \mathcal{G}\right) = \mathcal{O}\left(\frac{1}{n^\varepsilon}\right).$$

Remarquons que la convergence vers la distribution Gaussienne pour d_W implique la convergence en probabilité et pour la distance de Kolmogorov.

Nous avons vu les premières propriétés de LC_n. On s'intéresse maintenant à une variante, la longueur des plus longues sous-suites communes et croissantes, qui est l'objet du premier chapitre de notre travail.

2 Les plus longues sous-séquences communes et croissantes

Dans cette section, notons $\mathcal{A}_m := \{1, \dots, m\}$, $m \geq 2$. Pour $(x_1, \dots, x_s)$, $(y_1, \dots, y_t)$ deux séquences prenant des valeurs dans $\mathcal{A}$, nous désignons par $LCIS(x_1 \dots x_s; y_1 \dots y_t)$ la longueur maximale d'une sous-séquence croissante (donc une sous-séquence avec un bloc de 1, suivi d'un bloc de 2,... et ainsi de suite) des deux mots. Plus formellement, nous définissons $LCIS(x_1 \dots x_s; y_1 \dots y_t)$ comme le plus grand entier k tel qu'il existe $1 \leq i_1 < \dots < i_k \leq n$ et $1 \leq j_1 < \dots < j_k \leq n$ tels que

- $\forall s \in \{1, \dots, k\}$, $x_{i_s} = y_{j_s}$,
- $x_{i_1} \leq x_{i_2} \leq \dots \leq x_{i_k}$ et $y_{j_1} \leq y_{j_2} \leq \dots \leq y_{j_k}$,

et s'il n'existe pas d'entier vérifiant ces deux conditions, nous définissons $LCIS(x_1 \dots x_s; y_1 \dots y_t) = 0$.

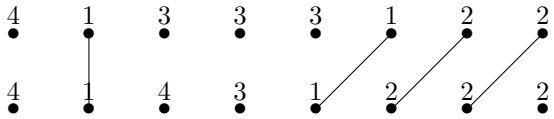


Figure 5: $(1, 1, 2, 2)$ est une sous-séquence commune croissante de longueur maximale.

Comme dans le cas du LCS, nous considérons deux suites indépendantes de variables aléatoires i.i.d. $(X_k)_{k\geq 1}$, $(Y_k)_{k\geq 1}$, de plus les Y_k peuvent avoir une distribution différente de celle des X_k. Soient $p_1^X, \dots, p_m^X$, $p_i^X > 0$, $i = 1, \dots, m$ et $p_1^Y, \dots, p_m^Y$, $p_i^Y > 0$, $i = 1, \dots, m$ leurs distributions respectives, et soit $LCI_n = LCIS(X_1 \dots X_n; Y_1, \dots Y_n)$. Ce modèle a été principalement étudié en informatique (voir par exemple [13], [60]), les principales motivations étant la généralisation des plus longues sous-séquences croissantes d'un seul mot (voir la section suivante) et les applications potentielles à la bio-informatique (par exemple [17]). La convergence en distribution de LCI_n (renormalisé) a été étudiée d'abord dans le cas binaire [31] et ensuite dans le cas uniforme à m lettres [12]. Notons qu'il est facile d'adapter à cette variante les inégalités de concentration et les bornes supérieures de variance vues dans la partie précédente.

Lorsque les lettres des deux mots suivent la même distribution uniforme, la distribution asymptotique a été trouvée (ci-dessous, $\wedge$ est l'abréviation de minimum), voir le Théorème 1.1.1 et sa généralisation conjecturée le Théorème 1.1.2. Le premier chapitre de notre travail obtient la distribution limite de LCI_n sans supposer que les X_k et Y_k $(k = 1, 2, \dots)$ ont la même distribution, et fournit également une preuve alternative du Théorème 1.1.1 ainsi qu'une preuve du Théorème conjecturé 1.1.2.

3 Les plus longues sous-séquences croissantes

On définit les sous-séquences croissantes les plus longues comme précédemment, sauf qu'il n'y a plus qu'un seul mot: Pour $x_1, \ldots, x_s \in \mathcal{A}$, notons $LIS(x_1 \ldots x_s)$ le plus grand entier k tel qu'il existe $1 \leq i_1 < \cdots < i_k \leq s$ tel que $x_{i_1} \leq x_{i_2} \leq \cdots \leq x_{i_k}$ et si aucun entier ne satisfait ces deux conditions, nous définissons $LIS(x_1 \ldots x_s) = 0$. Le LIS a d'abord été étudié lorsque les lettres sont tirées d'une permutation aléatoire. Soit π une permutation aléatoire (selon la distribution uniforme) de $\{1, \ldots, n\}$, et soit $I_n = LIS(\pi(1), \ldots, \pi(n))$. Le problème du comportement asymptotique de I_n a été introduit par Ulam [72] en 1961, et popularisé par Hammersley [26]. Des décennies plus tard, Kerov [45], Tracy et Widom [70] ont étudié le comportement asymptotique de $LI_n = LIS(X_1, \ldots, X_n)$ où $X_1, \ldots, X_n$ sont des variables aléatoires i.i.d. avec une distribution uniforme sur $\{1, \ldots, m\}$. En d'autres termes, il s'agit du problème des plus longues sous-séquences croissantes d'un mot aléatoire, plutôt que d'une permutation aléatoire. Ce problème de recherche des plus longues sous-séquences croissantes d'un mot aléatoire, ou plus précisément la généralisation telle qu'exposée ci-dessous, a un lien surprenant avec les statistiques quantiques, qui sera présenté dans la partie 4.

Commençons par la correspondance Robinson-Schensted-Knuth (RSK), qui est un outil important pour les problèmes de permutation aléatoire et de mot aléatoire. Ensuite, nous passons en revue certains résultats asymptotiques. Enfin, les théorèmes limites pour une permutation aléatoire apparaissent comme cas limites des résultats précédents (nous procédons dans cette présentation dans un ordre non chronologique).

3.1 L'algorithme de Robinson–Schensted–Knuth

Considérons un mot $w \in \{1, \ldots, m\}^n$ (il peut s'agir d'une permutation lorsque $m = n$). Schensted [63] a été le premier à relier $LIS(w)$ à la taille de la première pile lors d'un certain type de tri. Cela a ensuite été généralisé en l'algorithme de Robinson–Schensted–Knuth (RSK). Nous renvoyons à la partie 0.3.1 pour une explication de cette procédure. Cet algorithme transforme w en une paire (P, Q) de tableaux de Young (un arrangement d'entiers naturels avec des longueurs de lignes décroissantes - pour une introduction sur les tableaux de Young, voir [21]) de même forme, notée RSKshape(w). De plus, P a aussi chaque ligne croissante, et chaque colonne strictement croissante, de tels tableaux sont appelés Tableaux de Young semi-standards (SSYT). Un tableau de Young standard (SYT) est un tableau de Young dont les lignes et les colonnes sont strictement croissantes et qui contient exactement les nombres 1 à n (le nombre total de cases). Le tableau Q est un SYT. Lorsque l'entrée w est une permutation, P est également un SYT.

Une partition λ de n est une liste décroissante d'entiers non négatifs $(x_1, \ldots, x_\ell)$ tels que $x_1 + \cdots + x_\ell = n$, et s'écrit $\lambda \vdash n$. Notons $\ell(\lambda)$ la longueur de λ, le nombre d'éléments non nuls dans la liste. Pour toute partition λ, nous pouvons la compléter par un nombre arbitraire de zéros, de sorte que λ_k soit bien défini pour tout $k \geq 1$ (nul lorsque $k > \ell(\lambda)$). La forme d'un tableau de Young de taille n (ce qui signifie que le nombre total de boîtes est n) est définie comme la partition de n composée des longueurs de ses lignes.

Un diagramme de Young est un arrangement de boîtes dont les longueurs des lignes sont décroissantes (en d'autres termes, un tableau de Young avec des boîtes vides). On définit le dual d'un diagramme de Young comme sa réflexion par rapport à la droite $y = -x$.

En utilisant cette correspondance avec le dual, on peut définir pour toute partition $\lambda \vdash n$ la partition duale λ'. Sur la figure 6 par exemple, la partition est $(5, 4, 4, 2)$ et la partition duale est $(4, 4, 3, 3, 1)$.

Nous pouvons maintenant énoncer le résultat suivant:

Théorème 3.1 (Greene, 1974 [25]). *Soit $w \in \{1, \ldots, m\}^n$, soit $\lambda \vdash n$ la forme des tableaux obtenus*

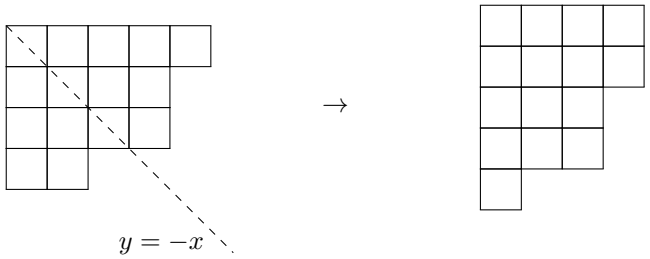


Figure 6: La correspondance entre un diagramme de Young et son dual.

par l'algorithme RSK avec le mot w, et λ' son dual. Alors, pour tout $k \leq \ell(\lambda)$, $\lambda_1 + \cdots + \lambda_k$ est égal à la somme des longueurs des k plus longues sous-séquences croissantes disjointes de w, et pour tout $\ell \leq \ell(\lambda')$, $\lambda'_1 + \cdots + \lambda'_\ell$ est égal à la somme des longueurs des plus longues ℓ sous-séquences strictement décroissantes disjointes de w.

Pour $\lambda \vdash n$, notons $\mathrm{SYT}(\lambda)$ l'ensemble des SYT de forme λ, $\mathrm{SSYT}(\lambda)$ l'ensemble des SSYT de forme λ, $\mathrm{SSYT}_m(\lambda)$ ceux dont les entrées sont dans $\{1, \ldots, m\}$, et enfin notons $f^\lambda := |\mathrm{SYT}(\lambda)|$ (cela est aussi souvent noté $\dim \lambda$). Les polynômes de Schur sont définis par

$$s_\lambda(x_1, x_2, \ldots) = \sum_{T \in \mathrm{SSYT}(\lambda)} \prod_{i=1}^{\infty} x_i^{\text{nombre d'entrées } i \text{ dans } T}.$$

Soit, pour $k \geq 1$ (et avec un léger abus de notation):

$$s_\lambda(x_1, \ldots, x_k) = s_\lambda(x_1, \ldots, x_k, 0, 0, \ldots).$$

On a, en particulier:

$$s_\lambda(x_1, \ldots, x_m) = \sum_{T \in \mathrm{SSYT}_m(\lambda)} \prod_{i=1}^{m} x_i^{\text{nombre d'entrées } i \text{ dans } T}.$$

Sans entrer dans les détails (voir la partie 0.3.1), le résultat essentiel utilisé par la suite est que si $X_1, \ldots X_n$ sont i.i.d. suivant la distribution $p_1 \ldots, p_m$, la distribution de $\mathrm{RSKshape}(X_1, \ldots, X_m)$ est explicite:

$$\mathbb{P}\left(\mathrm{RSKshape}(X_1 \ldots, X_m) = \lambda\right) = f^\lambda s_\lambda(p_1, \ldots, p_m). \tag{3.1}$$

En particulier, d'après le théorème 3.1, cela donne la distribution de LI_n.

Nous utiliserons également la formule du bialternant de Cauchy: pour tout $\lambda \vdash n$, pour tout $m \geq \ell(\lambda)$,

$$s_\lambda(x_1, \ldots, x_m) = \frac{\det\left(x_i^{\lambda_j + m - j}\right)_{1 \leq i,j \leq m}}{\Delta(x_1, \ldots, x_m)}.$$

Notons que ce polynôme est bien défini: le déterminant est alterné, il est donc divisible par $\Delta(x_1, \ldots, x_m)$. Rappelons également que lorsque $m < \ell(\lambda)$, $s_\lambda(x_1, \ldots, x_m) = 0$. Cette formule montre que les polynômes de Schur sont symétriques. En utilisant (3.1), cela implique:

Proposition 3.1. *Soit $\sigma \in \mathfrak{S}_m$, soit $X_1, \ldots, X_n$ des variables aléatoires i.i.d. avec distribution $p_1, \ldots, p_m$, soit $Y_1, \ldots, Y_n$ i.i.d. avec distribution $p_{\sigma(1)}, \ldots, p_{\sigma(m)}$. Alors,* $\mathrm{RSKshape}(X_1, \ldots, X_n)$ *et* $\mathrm{RSKshape}(Y_1, \ldots, Y_n)$ *ont la même distribution.*

La distribution de RSKshape$(X_1, \ldots, X_n)$ est appelée distribution de Schur-Weyl avec les paramètres p, n, et elle est notée $\mathrm{SW}^n(p)$. Nous écrivons également SW^n_m dans le cas particulier de la distribution uniforme $p = (1/m, \ldots, 1/m)$.

Dans la suite, et en particulier dans le chapitre 3 de notre travail, pour étudier la distribution de RSKshape$(X_1, \ldots, X_n)$ nous supposons donc, sans perte de généralité, que $p_1 \geq p_2 \geq \cdots \geq p_m$ (on dira p ordonnée).

Nous allons maintenant passer en revue les théorèmes limites, d'abord pour le modèle des mots aléatoires, puis pour le modèle des permutations aléatoires.

3.2 La forme du RSK pour les mots aléatoires

Nous passons maintenant en revue quelques résultats connus sur la distribution asymptotique de $LI_n = LIS(X_1, \ldots, X_n)$, et plus généralement, de RSKshape$(X_1, \ldots, X_n)$, pour $X_1, \ldots, X_m$ variables aléatoires i.i.d. prenant des valeurs dans $\{1, \ldots, m\}$ avec la distribution $p_1, \ldots, p_m$.

La distribution asymptotique de LI_n, et plus généralement de RSKshape, s'avère être étroitement liée aux valeurs propres de certaines matrices aléatoires: les matrices avec la distribution de l'ensemble unitaire gaussien (GUE). Nous rappelons que l'ensemble unitaire gaussien de taille m, noté GUE_m, est la distribution de probabilité des matrices hermitiennes H de taille $m \times m$ définie comme suit:

- Pour tout $i \in \{1, \ldots, m\}, H_{i,i} \sim \mathcal{N}(0, 1)$;
- Pour tout $i, j \in \{1, \ldots, m\}$ tel que $i < j$, $H_{i,j} \sim \mathcal{N}_{\mathbb{C}}(0, 1)$ (la gaussienne standard complexe, égale à $\mathcal{N}(0, 1/2) + i\mathcal{N}(0, 1/2)$);
- Ces entrées sont tirées indépendamment les unes des autres.

On peut calculer la densité de probabilité de la distribution GUE_m: pour une certaine constante de normalisation $C > 0$, la densité est $\mathrm{e}^{-\mathrm{Tr}(H^2)/2}/C$ (c'est une conséquence directe de la définition, et c'est une autre définition équivalente de GUE_m).

En utilisant les notations de [41], nous définissons également le GUE sans trace de taille m, GUE^0_m, comme la distribution de $H - (\mathrm{Tr}(H)/m)I_m$ où $H \sim \mathrm{GUE}_m$. Il s'agit de la distribution conditionnelle de GUE_m étant donné que la trace est nulle. Le premier résultat de distribution asymptotique a été obtenu par Tracy et Widom, dans le cas uniforme.

Théorème 3.2 (Tracy et Widom, 2001 [70]). *Soit $H \sim \mathrm{GUE}^0_m$ et $\mu_1(H)$ sa plus grande valeur propre. Dans le cas uniforme avec m lettres, nous avons la convergence en distribution:*

$$\frac{LI_n - n/m}{\sqrt{n/m}} \xRightarrow[n\to\infty]{} \mu_1(H).$$

Il a été conjecturé dans [70] que la convergence s'applique à l'ensemble du diagramme de Young, et prouvé plus tard par Johansson:

Théorème 3.3 (Johansson, 2001 [44]). *Soit $H \sim \mathrm{GUE}^0_m$ et $\mu_1(H) \geq \cdots \geq \mu_m(H)$ ses valeurs propres. Si $\lambda \sim \mathrm{SW}^n_m$, nous avons la convergence en distribution:*

$$\left(\frac{\lambda_1 - n/m}{\sqrt{n/m}}, \ldots, \frac{\lambda_m - n/m}{\sqrt{n/m}}\right) \xRightarrow[n\to\infty]{} (\mu_1(H), \ldots, \mu_m(H)).$$

Ce dernier résultat a été prouvé plus tôt (1994) par Kerov [45, Chap. 3, Sec. 3.4, Théorème 2]. Il a ensuite été généralisé à une distribution non uniforme. Nous suivons [41] pour les notations

suivantes. Pour toute distribution p ordonnée (c'est-à-dire avec $p_1 \geq \cdots \geq p_m$) sur $\{1, \ldots, m\}$, définissons la distribution GUE sans trace généralisée $\mathrm{GUE}^0(p)$ comme la distribution de H, où H est définie comme suit. Soit $d_1, \ldots, d_k$ les multiplicités de p, ce qui signifie $p_1 = \cdots = p_{d_1} > p_{d_1+1} = \cdots = p_{d_1+d_2} > \ldots$.

- Supposons que $H_1 \sim \mathrm{GUE}^0_{d_1}, \ldots, H_k \sim \mathrm{GUE}^0_{d_k}$ soient des matrices aléatoires indépendantes;
- Soit B la matrice $m \times m$ définie par blocs avec $H_1, \ldots, H_k$ sur sa diagonale;
- Soit $T = \sum_{i=1}^m \sqrt{p_i} B_{i,i}$;
- Enfin, pour $i, j \in \{1, \ldots, m\}$, soit $H_{i,j} = B_{i,j}$ si $i \neq j$ et $H_{i,i} = B_{i,i} - \sqrt{p_i} T$.

Comme indiqué dans [41], $\mathrm{GUE}^0(p)$ est la distribution de la somme directe d'ensembles unitaires gaussiens mutuellement indépendants $d_i \times d_i$ conditionnellement aux valeurs propres $\mu_1, \ldots \mu_d$ satisfaisant $\sqrt{p_1}\mu_1 + \cdots + \sqrt{p_m}\mu_m = 0$. Nous pouvons maintenant énoncer la généralisation du théorème précédent:

Théorème 3.4 (Its, Tracy, Widom, 2001 [42]). *Soit p une distribution ordonnée sur $\{1, \ldots, m\}$, soit $H \sim \mathrm{GUE}^0(p)$ et soit $\mu_1(H) \geq \cdots \geq \mu_m(H)$ ses valeurs propres. Si $\lambda \sim \mathrm{SW}^n(p)$, nous avons la convergence en distribution:*

$$\left(\frac{\lambda_1 - p_1 n}{\sqrt{p_1 n}}, \ldots, \frac{\lambda_m - p_m n}{\sqrt{p_m n}} \right) \xRightarrow[n \to \infty]{} (\mu_1(H), \ldots, \mu_m(H)) .$$

Remarques.

(i) En particulier, si $p_1 > \cdots > p_m$, ce qui signifie $d_1 = \cdots = d_m = 1$, la distribution limite est gaussienne multivariée.

(ii) En rappelant la Proposition 3.1, il n'y a pas de perte de généralité avec l'hypothèse que p est ordonnée.

La distribution $\mathrm{GUE}^0(p)$ n'est peut-être pas très intuitive, mais le théorème suivant permet une interprétation plus naturelle de la limite. On désigne par $p^{(i)}$, pour $1 \leq i \leq k$, la probabilité de multiplicité d_i (autrement dit, $(p_1, \ldots, p_m) = (p^{(1)}, \ldots, p^{(1)}, p^{(2)}, \ldots, p^{(2)}, \ldots, p^{(k)}, \ldots, p^{(k)})$).

Théorème 3.5 (Méliot 2012 [55], tel qu'énoncé par Wright [76]). *Soit $H \in \mathrm{GUE}^0(p)$, soit $g_1, \ldots, g_k$ des variables aléatoires gaussiennes centrées avec une covariance $\left(\mathbb{1}_{i=j} d_i - d_i d_j \sqrt{p^{(i)}} \sqrt{p^{(j)}} \right)_{1 \leq i,j \leq k}$, et pour chaque $i \in \{1, \ldots, k\}$, $H_i \sim \mathrm{GUE}^0_{d_i}$ (le vecteur g et les H_i étant indépendants). On a alors l'égalité des distributions suivantes:*

$$(\mu_1(H), \ldots, \mu_m(H)) \overset{d}{=} \left(\frac{g_1}{d_1} + \mu_1(H_1), \ldots, \frac{g_1}{d_1} + \mu_{d_1}(H_1), \frac{g_2}{d_2} + \mu_1(H_2), \ldots, \frac{g_k}{d_k} + \mu_{d_k}(H_k) \right) .$$

La distribution limite peut également être écrite sous la forme d'une fonction brownienne: cela a été fait tout d'abord dans [32] pour le LIS, puis dans [33] pour le LIS dans un cadre markovien (les lettres sont une chaîne de Markov, généralisant le cadre i.i.d.), et enfin dans toute sa généralité dans [34] pour l'ensemble du diagramme de Young toujours dans un cadre markovien (voir également [41]). L'idée principale pour obtenir de telles limites est de revoir le théorème de Greene mais avec des sous-séquences disjointes. Les mouvements browniens apparaissent alors comme des marches aléatoires renormalisées comptant le nombre d'occurrences de chaque lettre. Le principal avantage de cette approche n'est pas seulement la possibilité de généraliser à un cadre markovien, mais aussi de donner des taux de convergence non asymptotiques, comme nous

le verrons au chapitre 3. Nous rappelons un cas simple: la limite de LI_n (la longueur de la première ligne du diagramme de Young), lorsque la distribution des lettres est p. En notant d_1 la multiplicité de p_1, et $B = (B_k(t))_{1\leq k\leq d_1, t\in[0,1]}$ un mouvement brownien standard de dimension d_1, on a le résultat suivant de convergence en distribution [32, Corollaire 3.3]:

$$\frac{LI_n - p_1 n}{\sqrt{p_1 n}} \xRightarrow[n\to\infty]{} \frac{\sqrt{1-d_1p_1}-1}{d_1}\sum_{j=1}^{d_1} B_j(1) + \max_{0=t_0\leq\cdots\leq t_{d_1}=1}\left(B_j(t_j) - B_j(t_{j-1})\right).$$

Puisque la distribution limite du diagramme de Young est déjà connue (Théorème 3.4 et Théorème 3.5 ci-dessus), la fonctionnelle brownienne doit avoir la même distribution. Ceci n'est pas surprenant étant donné les liens entre certaines fonctionnelles browniennes et les valeurs propres de la GUE ([5], [24]). Plus précisément, le théorème suivant de [6], qui est une généralisation de [5], rend la connexion complète. Suivant [6], nous introduisons d'abord quelques notations. Soit $B = (B_k(t))_{1\leq k\leq M, t\in[0,1]}$ un mouvement brownien standard de dimension M. Soit $\mathcal{P}$ l'ensemble des fonctions càdlàg, non décroissantes de $[0,1]$ vers $\{1,\ldots,M\}$. Pour $\pi\in\mathcal{P}$, π peut être écrit comme $\sum_{j=1}^{M-1} j\mathbb{1}_{[t_{j-1},t_j)} + M\mathbb{1}_{[t_{M-1},t_M)}$, et notons

$$\begin{aligned}\Delta_\pi B &= \int_0^1 \mathrm{d}B_{\pi(t)}(t)\\ &= \sum_{j=1}^{m}\left(B_j(t_j) - B_j(t_{j-1})\right).\end{aligned}$$

Soit $H\sim \mathrm{GUE}_m$, et pour $1\leq k\leq M$, soit $\mu_1^k\geq\cdots\geq\mu_k^k$ les valeurs propres du principal $k\times k$ mineur de H.

Théorème 3.6 (Benaych-Georges et Houdré, 2013 [6])**.** *On a l'égalité suivante en distribution:*

$$\left(\mu_i^k\right)_{1\leq\ell\leq k\leq M} \overset{d}{=} \left(\sup\left\{\sum_{i=1}^{\ell}\Delta_{\pi_i}B; \pi_1,\ldots\pi_\ell\in\mathcal{P}, \pi_1<\cdots<\pi_\ell\leq k\right\}\right)_{1\leq\ell\leq k\leq M}.$$

Nous pouvons maintenant prouver directement que le terme de droite dans (0.3.2) a la même distribution que la limite donnée par le théorème (3.5), à savoir $\mu_1(H_1) + (g_1/d_1)$. Tout d'abord, [6, Corollaire 2] affirme que l'égalité de distribution suivante est vraie:

$$\frac{\sqrt{1-d_1p_1}-1}{d_1}\sum_{j=1}^{d_1} B_j(1) + \max_{0=t_0\leq\cdots\leq t_{d_1}=1}\left(B_j(t_j) - B_j(t_{j-1})\right) \overset{d}{=} \frac{\sqrt{1-d_1p_1}-1}{d_1}\operatorname{Tr}(H) + \mu_1(H),$$

où $H\sim\mathrm{GUE}_{d_1}$. Deuxièmement, en utilisant l'indépendance des projections d'une variable aléatoire gaussienne sur des espaces orthogonaux, $H - (\operatorname{Tr}(H)/d_1)I_{d_1}$ et $\operatorname{Tr}(H)$ sont indépendants, et

$$\begin{aligned}\frac{\sqrt{1-d_1p_1}-1}{d_1}\operatorname{Tr}(H) + \mu_1(H) &= \mu_1\left(H - \frac{\operatorname{Tr}(H)}{d_1}I_{d_1}\right) + \frac{\sqrt{1-d_1p_1}}{d_1}\operatorname{Tr}(H)\\ &= \mu_1(H^0) + \frac{\sqrt{d_1-d_1^2p_1}}{d_1}Z\end{aligned}$$

où $H' = H - (\operatorname{Tr}(H)/d_1)I_{d_1}\sim\mathrm{GUE}_{d_1}^0$ et Z est une gaussienne standard indépendante, ce qui est exactement la limite du théorème 3.5.

Notons que la limite de LCI_n, donnée au chapitre 1, est aussi une fonctionnelle brownienne, mais à ce jour, le lien avec les valeurs propres de la GUE reste inconnu.

3.3 La forme RSK pour les permutations aléatoires

Nous allons maintenant passer en revue quelques résultats connus sur la distribution asymptotique de $I_n := LIS(\pi(1), \ldots, \pi(n))$, et plus généralement, $\lambda^{(n)} := \text{RSKshape}(\pi(1), \ldots, \pi(n))$, pour π une permutation aléatoire de $\{1, \ldots, m\}$ (suivant la distribution uniforme). Ce problème, comme mentionné précédemment, est plus ancien que le modèle des mots aléatoires, et il y aurait beaucoup plus à dire, mais nous nous concentrons sur le lien avec les modèles précédents.

Nous désignons par TW la distribution de Tracy-Widom, dont la densité est définie comme la solution d'une équation de Painlevé, voir par exemple [71]. Sa moyenne est d'environ -1.771 et sa variance d'environ 0.813. Tracy et Widom ont prouvé le résultat de convergence suivant:

Théorème 3.7 (Tracy et Widom, 1994 [69]). *Soit $H_m \sim \text{GUE}_m$, et $\mu_1(H_m)$ désigne sa plus grande valeur propre. Nous avons la convergence en distribution:*

$$\frac{I_n - 2\sqrt{n}}{n^{1/6}} \Longrightarrow_{n\to\infty} TW.$$

Il s'avère que c'est aussi la distribution limite de I_n (une fois remise à l'échelle):

Théorème 3.8 (Baik, Deift et Johansson, 1999 [4]). *Nous avons:*

$$\frac{I_n - 2\sqrt{n}}{n^{1/6}} \Longrightarrow_{n\to\infty} TW.$$

Pour tout n, quand m tend vers l'infini, LI_n converge vers I_n en distribution. Ainsi, le théorème suivant est une généralisation du précédent.

Théorème 3.9 (Johansson, 2001 [44]). *Soit $m \in \mathbb{N}^{\mathbb{N}}$ tel que $(\log n)^{1/6}/m_n \xrightarrow[n\to\infty]{} 0$, alors dans le cas uniforme on a:*

$$\frac{LI_{m_n} - n/m_n - 2\sqrt{n}}{n^{1/6}\left(1 + \sqrt{n}/m_n\right)^{2/3}} \Longrightarrow_{n\to\infty} TW.$$

Comme dans le cas d'un mot aléatoire, il y a des théorèmes sur la limite de la forme des diagrammes de Young, voir la partie 0.3.3 pour plus de précisions.

4 Statistiques quantiques

Nous donnons à présent une brève introduction aux statistiques quantiques, et explorons les liens avec l'algorithme RSK. Ceci est l'objet du chapitre 3 de ce travail.

Un système quantique d-dimensionnel est un système qui se trouve dans un état mixte de d états quantiques, ce qui signifie qu'il a une probabilité p_1 d'être dans l'état $u_1 \in \mathbb{C}^d$, p_2 d'être dans l'état $u_2 \in \mathbb{C}^d$, ..., p_d d'être dans l'état $u_d \in \mathbb{C}^d$. La matrice $\rho := p_1 u_1 u_1^* + \cdots + p_d u_d u_d^*$ est positive hermitienne avec une trace égale à un, on l'appelle la matrice de densité du système. Réciproquement, pour toute matrice hermitienne positive ρ avec une trace égale à un, il existe (au moins) un système avec une matrice de densité ρ: si $(p_1, \ldots, p_d)$ sont les valeurs propres de ρ et $u_1, \ldots, u_d$ les vecteurs propres unitaires, on peut en effet considérer l'état mixte: u_1 avec une probabilité p_1, u_2 avec une probabilité p_2, ... avec une matrice de densité $p_1 u_1 u_1^* + \cdots + p_d u_d u_d^* = \rho$. Dans la suite, on appelle matrice de densité toute matrice hermitienne positive avec une trace égale à un.

Deux systèmes ayant la même matrice de densité sont physiquement indiscernables. Ainsi, on peut supposer que les vecteurs propres sont les différents états et les valeurs propres sont les

différentes probabilités de ces états. Bien sûr, il est impossible de mesurer directement ρ, mais on peut obtenir une mesure de ρ, qui est une variable aléatoire comme défini ci-dessous.

Le type le plus simple de mesure est la mesure de base: on donne une base orthonormale $v_1, \dots, v_d$ de $\mathbb{C}^d$, et le résultat de la mesure est une variable aléatoire N à valeurs dans $\{1, \dots, d\}$ avec probabilités:

$$\begin{aligned}\mathbb{P}(N = i) &= \sum_{j=1}^{d} p_j \langle u_i, v_j \rangle^2 \\ &= \mathrm{Tr}(\rho E_i^*) \text{ où } E_i := v_i v_i^* \\ &= \langle \rho, E_i \rangle \text{ avec le produit scalaire habituel sur les matrices.}\end{aligned}$$

Plus généralement, pour effectuer une mesure projective, on donne $E_1, \dots, E_d$ des projections auto-adjointes (c'est-à-dire des projections orthogonales) telles que $E_1 + \cdots + E_d = I_d$, et le résultat de la mesure est une variable aléatoire N prenant des valeurs dans $\{1, \dots, d\}$ avec probabilités:

$$\mathbb{P}(N = i) = \langle \rho, E_i \rangle.$$

Remarquons que si l'on connaît une base orthonormale de vecteurs propres (c'est-à-dire les états quantiques), alors une mesure de base avec cette base a une distribution $p_1, \dots, p_d$, donc estimer les p_i de cette manière revient à des statistiques classiques.

En statistiques quantiques, après chaque mesure, l'état s'effondre, donc il faut n copies indépendantes du système quantique pour effectuer n mesures indépendantes. Mais au lieu de faire n mesures l'une après l'autre, il est en fait préférable de considérer les n copies comme un seul système quantique avec une matrice de densité $\rho^{\otimes n}$, puis de faire une seule mesure, appelée mesure intriquée, pour estimer ρ. Parmi toutes les mesures que l'on peut effectuer sur $\rho^{\otimes n}$, il s'avère que l'une d'entre elles est optimale: une mesure projective appelée échantillonnage faible de Schur. Par "optimale", on entend optimale pour calculer n'importe quelle propriété du spectre $p_1, \dots, p_d$, au sens de [76, Theorem 2.6.3]: si nous disposons d'un algorithme pour calculer une propriété qui a un risque β, alors il existe un algorithme similaire pour le faire en utilisant uniquement l'échantillonnage faible de Schur. Par conséquent, dans ce travail, nous nous concentrons sur cette mesure de $\rho^{\otimes n}$. La famille de projecteurs auto-adjoints de cette mesure projective est donnée par le théorème de dualité Schur-Weyl, et est indexée par les partitions de n. Par conséquent, le résultat de cette mesure, l'échantillonnage faible de Schur, est une partition aléatoire $\lambda \vdash n$, et il est bien connu (voir par exemple [76]) que sa distribution est $\mathrm{SW}^n(p)$. En d'autres termes, on peut voir la mesure λ comme la forme du diagramme obtenu par l'algorithme RSK appliqué à un mot aléatoire de longueur n et de lettres tirées avec distribution p. Le problème est de trouver un bon estimateur de p, étant donné λ.

Comme vu précédemment, les distributions limites de $\lambda \sim \mathrm{SW}^n(p)$ sont bien connues, mais dans la première partie du chapitre 3, nous revisitons les taux de convergence, en raison de la nécessité de résultats d'estimation non asymptotiques. Ainsi, nous étudions quelques aspects de la convergence en distribution des diagrammes de Young associés à des mots aléatoires, obtenant des vitesses de convergence pour la distance de Kolmogorov. Puisque la longueur de la première ligne du tableau est LIS, une vitesse de convergence est donnée dans ce cas. Cela vient préciser, de manière non asymptotique, un résultat de vitesse de convergence déjà donné dans [40]. Nous donnons ensuite des résultats sur deux estimateurs du spectre, avec des simulations numériques tendant à montrer que leur risque est inférieur à celui de l'estimateur "diagramme de Young empirique" (EYD). Le premier repose sur du bootstrap (à partir l'EYD, nous voyons comment l'itérer, mais sans preuve théorique), le second est en quelque sorte un équivalent théorique du premier, où nous donnons une conjecture restante pour arriver à démontrer que son risque quadratique est inférieur. Nous prouvons ensuite une nouvelle borne pour la somme des variances d'un diagramme de Young, et enfin, nous prouvons une borne sur "l'excès" d'un diagramme de Young avec une chaîne de Markov (cela généralise un résultat déjà connu sur l'excès des diagrammes).

Introduction

This work considers some random words combinatorial problems and their applications. A random word of length n is an n-tuple of, say, i.i.d. random variables taking values in a finite set called alphabet (for example, a sequence of coin tosses $HTTTHTTH$ is a random word of length 8). The starting point of this endeavor is the following question: given two random words, "how much do they have in common"? The problem of analyzing the similarity between two random words has emerged independently in various fields, including computer science, biology, linguistics... See, for example, [62] for a description of numerous applications.

Unfortunately, too little is known on the fundamental problem of the study of the length of the Longest Common Subsequences (LCS) of two random words: the asymptotic distribution, and even the asymptotics of the variance, are unknown. However, by slightly twisting this problem, it becomes easier to find the asymptotic distribution: the first chapter of our work is dedicated to the asymptotic distribution of the length of the longest common and increasing subsequences. There we consider a totally ordered alphabet with an order, say $\{1, \ldots, m\}$, and the subsequences are simply made of a block of 1's, followed by a block of 2's, ... and so on (such a subsequence is increasing, but not strictly). In this framework, we are able to provide the asymptotic mean, variance and distribution of its maximal length.

In the second chapter, we deal with the problem of the variance of the LCS. It is an important open problem to determine whether or not the variance is linear in n. By introducing a general framework going beyond this problem, partial results in this direction are presented. Indeed, for functions of independent random variables, various upper and lower variance bounds are revisited in diverse settings. These are then specialized to the Bernoulli, Gaussian, infinitely divisible cases and to Banach space valued random variables. Frameworks and techniques vary from jackknives through semigroups and beyond. Some new applications are presented, recovering and improving, in particular, all the known estimates on the variance of the length of the longest common subsequences of two random words.

In the third and final chapter, we consider the Longest Increasing Subsequences (LIS) of one random word, and the surprising connection with quantum statistics. Indeed, estimating the spectrum of a density matrix of a quantum system with n copies of this system is equivalent to estimating the distribution of the letters of a word of size n given the Robinson–Schensted–Knuth (RSK) output shape of this word (a partition of n whose first term is the length of the LIS). Therefore, we revisit some aspects of the convergence of the cumulative shape of the RSK Young diagrams associated with random words, obtaining rates of convergence in Kolmogorov's distance. Since the length of the top row of a diagram is the length of the longest increasing subsequences of the word, a corresponding rate result follows. We then provide results on two spectrum estimators, with numerical simulations tending to prove that their risk is smaller than with the empirical Young diagram estimator. Then we bound the sum of the variances of the Young diagram, and lastly, prove a bound on the "excess" of the shape of the RSK algorithm with the help of a Markov chain.

0.1 The longest common subsequences

In computer science, finding the length of the longest common subsequences amounts to finding an edit distance corresponding to the minimum number of characters that must be deleted or inserted to go from one word to another. This is why the Unix program "diff", used to compare different versions of a file, computes the length of the longest common subsequences of two strings. In biology, a strand of DNA can be represented by a word written with the alphabet $\{A, T, G, C\}$. With the theory of evolution, this word evolves by the insertion of new letters, thus the ancestor of a species has a DNA which is a subsequence of the DNA of the latter. When two species DNA share a long common subsequence, we can infer that this is not the result of chance but because of the existence of a common ancestor (see for example [74]). However, as [61] and [3] note, one has to be careful because two random "long" strands of DNA share on average about 65% of their length, which can be perceived as counter-intuitive. The computation of this average similarity, necessary to make statistical hypotheses on the existence or not of common ancestors, is one of the motivations of the mathematical study of the similarity between two random words initiated by Chvátal and Sankoff [15] in 1975.

Let $\mathcal{A}$ be a finite set, called an alphabet, and call "words" finite sequences of variables in $\mathcal{A}$. For instance, to model DNA strands, one takes $\mathcal{A} = \{A, T, G, C\}$. For $(x_1, \ldots, x_s)$, $(y_1, \ldots, y_t)$ two sequences taking values in $\mathcal{A}$, we denote by $LCS(x_1 \ldots x_s; y_1 \ldots y_t)$ the largest integer k such that there exists $1 \leq i_1 < \cdots < i_k \leq s$, $1 \leq j_1 < \cdots < j_k \leq t$ satisfying $a_{i_1} = b_{j_1}, \ldots, a_{i_k} = b_{j_k}$, or 0 if there is no such integer.

For instance, $LCS(ACCGAT; GACT) = 3$ as one can take $i_1 = 1, i_2 = 2, i_3 = 6$ and $j_1 = 2, j_2 = 3, j_3 = 4$, which extracts the word ACT from the two words, but one cannot extract a longer word. The word ACT is an extracted word ("common ancestor") of maximum length, but it is not the only one, because one could also have extracted GAT. Graphically, finding the longest common word consists in connecting identical letters without crossings (see Figure 7).

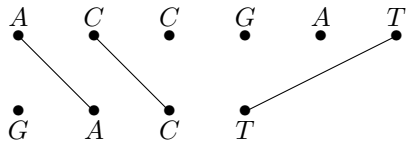


Figure 7: ACT is a subsequence of maximum length

One can also view LCS as a percolation problem (see Figure 8): if one puts the first word on the x-axis, the second one on the y-axis, and draws a dot each time the x-axis and the y-axis have the same letter, the aim is to find one strictly increasing path (along which the x-axis and the y-axis strictly increase) passing through the largest number of dots, i.e., vertices (here, the maximum is three).

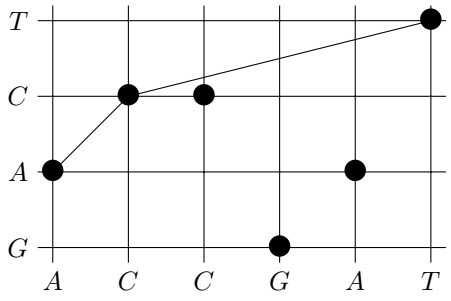


Figure 8: ACT as an increasing path containing a maximum number of dots

This can be reduced to an edge percolation problem. If one adds diagonal edges to the graph, linking (x, y) to $(x+1, y+1)$, and set their weight to one if there is a dot at (x, y), zero otherwise, with the other edges (horizontal, vertical) of zero weight, one needs to find the path from $(0, 0)$ to (s, t) of maximum weight, following the appropriately oriented edges: this is a last-passage directed percolation problem (see Figure 9).

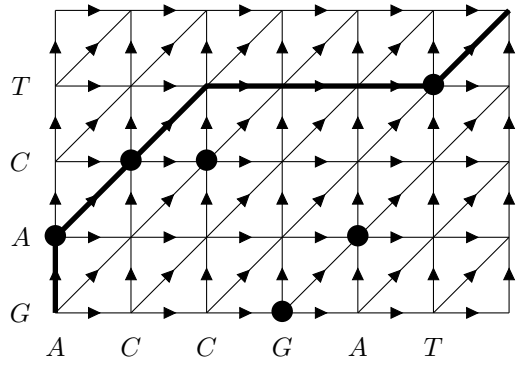


Figure 9: ACT as a path with maximum weight

We can now consider the similarity between random words. Let $(X_k)_{k\geq 1}$, $(Y_k)_{k\geq 1}$ be two sequences of independent, identically distributed random variables, with values in a finite alphabet $\mathcal{A}$. In particular, we are interested in the random variables $LCS(X_1 \dots X_n; Y_1 \dots Y_n)$, simply denoted by LC_n.

For example, if $\mathcal{A} = \{0, 1\}$ and X_k's are Bernoulli random variables with parameter $1/2$, then LC_1 is also a Bernoulli random variable with parameter $1/2$, but $\mathbb{P}(LC_2 = 0) = 1/4$, $\mathbb{P}(LC_2 = 1) = 1/2$ and $\mathbb{P}(LC_2 = 2) = 1/4$. It is difficult to compute the distribution of LC_n explicitly, so we are mainly interested in its asymptotic behavior, and in particular its expectation and variance.

Let us start with a simple property of LC_n:

Proposition 0.1.1 (Chvátal and Sankoff, 1975, [15]). *For all $m, n \geq 1$,*

$$\mathbb{E}LC_{m+n} \geq \mathbb{E}LC_m + \mathbb{E}LC_n.$$

Proof. We cut in half as in Figure 10.

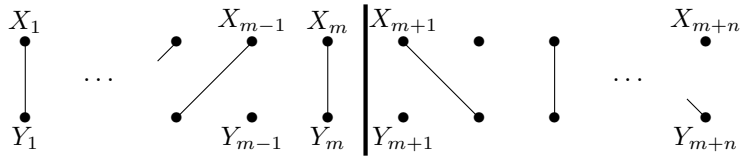


Figure 10: Cutting words of length $m + n$

We have

$$LCS(X_1 \dots X_{m+n}; Y_1 \dots Y_{m+n}) \geq LCS(X_1 \dots X_m; Y_1 \dots Y_m) + LCS(X_{m+1} \dots X_{n+m}; Y_{m+1} \dots Y_{m+n})$$

and as $LCS(X_{m+1} \dots X_{m+n}; Y_{m+1} \dots Y_{m+n})$ has the same distribution as LC_n, the result follows. □

We say that the sequence $(\mathbb{E}LC_n)_{n\geq 1}$ is super-additive. The following corollary is an immediate consequence of Fekete's lemma (1923), and the inequality $LC_n \leq n$:

Corollary 0.1.2. *The sequence* $(\mathbb{E}LC_n/n)_{n\geq 1}$ *converges to* $\sup_n \mathbb{E}LC_n/n \leq 1$.

Usually we let γ be the limit of $\mathbb{E}LC_n/n$, or γ_k in the case of an alphabet with k letters with uniform distribution. Even in the simplest cases, it is difficult to determine γ: for example, we do not have any explicit value for γ_2. The best bound is $0.788071 \leq \gamma_2 \leq 0.826280$ [52], which disproves an old conjecture of Steele [66], $\gamma_2 = 2/(1+\sqrt{2}) \simeq 0.828427$. More recently, Tiskin [68] was able to prove that γ_2 is algebraic, however, this work does not yield any numerical estimation, and is limited to the uniform, binary case.

Numerical simulations, e.g. [3], tend to show that for a two-letter alphabet the constant γ is minimal when the distribution is uniform. This is intuitive: the probability that two letters coincide being minimal when $p = 1/2$, we expect the expectation of LC_n to be minimal when there are on average the fewest number of pairs of identical letters. In [3], the authors attempt to show this fact, but their proof is not convincing, so it does not seem to have been proved, to this day.

We notice that $\gamma_k \geq 1/k$, and more generally, if the p.m.f. of the letters is $p_1, \ldots, p_k$, $\gamma \geq p_1^2 + \cdots + p_k^2$, because $LC_n \geq \sum_{j=1}^n \mathbb{1}_{X_j = Y_j}$. It is trickier to upper bound γ_k:

Theorem 0.1.3 (Chvátal and Sankoff, 1975, [15])**.** *Let, for* $0 < x < 1$, $h_k(x) = k^{x/2}x^x(1-x)^{1-x}$, *and let* y_k *be the unique real in* $(0,1)$ *such that* $h_k(y_k) = 1$. *Then* $\gamma_k \leq y_k$.

Proof. Let $c \in (0,1)$ and $g(c,n,k)$ be the number of pairs of words of length n with values in an alphabet of size k with a longest subsequence of length greater than cn, in other words, let

$$g(n,c,k) = |\{x, y \in \mathcal{A}^n;\, LCS(x;y) \geq cn\}|\,.$$

Let $m = \lceil cn \rceil$, for $x, y \in \mathcal{A}^n$, if $LCS(x;y) \geq cn$, then there exist $i_1, \ldots, i_m$ and $j_1, \ldots, j_m$ such that $X_{i_1} = Y_{j_1}, \ldots, X_{i_m} = Y_{j_m}$. We have $\binom{n}{m}^2$ possible choices for these two extractions, then k^{2n-m} choices for the letters, so $g(n,c,k) \leq \binom{n}{m}^2 k^{2n-m}$. With the help of Stirling's formula we get

$$\frac{g(n,c,k)}{k^{2n}} = \mathcal{O}\left(\frac{1}{h_k(c)^{2n}n}\right).$$

In particular, for $c = y_k$, $\mathbb{P}(LC_n \geq y_k n)$ tends to zero. Now $\mathbb{E}LC_n/n \leq y_k + \mathbb{P}(LC_n \geq y_k n)$, hence $\gamma \leq y_k$. □

We can generalize this result as follows: if the letters are drawn according to a non-trivial distribution, then $\gamma < 1$, and we have an upper bound according to the distribution. More precisely, let $p_1, \ldots, p_m$ be the p.m.f. of the letters, let $P_2 = p_1^2 + \cdots + p_m^2$, and assume $P_2 \neq 1$ (the "non-trivial" assumption), let $y \in (0,1)$ be such that $x^x(1-x)^{1-x}(1/P_2)^{x/2} = 1$, then $\gamma \leq y$. Indeed, letting $m = \lceil yn \rceil$, we have

$$\begin{aligned}\mathbb{P}(LC_n \geq yn) &\leq \binom{n}{m}^2 P_2^m \\ &= \mathcal{O}\left(\frac{P_2^m}{(y^y(1-y)^y)^{2n}n}\right) \\ &= \mathcal{O}\left(\frac{1}{n}\right),\end{aligned}$$

so we can conclude as previously that $\gamma \leq y$.

We know that $\mathbb{E}LC_n$ is super-additive, thus denoting $b_n = \mathbb{E}LC_n/n$, for all $k > 0$, $b_{kn} \geq b_n$, which may suggest that $(b_n)_{n\geq 1}$ is increasing. Computer simulations go in that direction, but it

remains an open problem to show that $(b_n)_{n\geq 1}$ is increasing [19]. Note that if a sequence $(u_n)_{n\geq 1}$ is super-additive, the sequence $(u_n/n)_{n\geq 1}$ needs not to be increasing, as the following example shows: $u_n = n - \mathbb{1}_{n \text{ is odd}}$.

We now investigate how close LC_n is to its mean, with standard concentration techniques. We then evaluate the variance of LC_n, using the Efron-Stein inequality, which is the starting point of the second chapter of our work.

0.1.1 Concentration inequalities

We recall the following theorem, often attributed to Azuma:

Theorem 0.1.4 (Hoeffding, 1963)**.** *Let $(S_k, \mathcal{F}_k)_{0\leq k\leq n}$ be a martingale. Suppose that there exist $a_k, b_k \in \mathbb{R}$ such that $a_k \leq S_k - S_{k-1} \leq b_k$ for all $1 \leq k \leq n$. Then for all $t > 0$ we have*

$$\mathbb{P}(S_n - S_0 \geq t) \leq \mathrm{e}^{-\frac{2t^2}{\sum_{k=1}^n (b_k - a_k)^2}}.$$

Corollary 0.1.5. *Let $Z_1, \ldots, Z_n$ be i.i.d. random variables, and $f : \mathbb{R}^n \mapsto \mathbb{R}$ Borel-measurable. Suppose that there exist $c_1, \ldots, c_n \in \mathbb{R}$ such that for all $1 \leq k \leq n$ and $z_1, \ldots, z_n, w_k \in \mathbb{R}$,*

$$|f(z_1, \ldots, z_{k-1}, z_k, z_{k+1}, \ldots, z_n) - f(z_1, \ldots, z_{k-1}, w_k, z_{k+1}, \ldots, z_n)| \leq c_k.$$

Then for all $t > 0$,

$$\mathbb{P}(f(Z_1, \ldots, Z_n) - \mathbb{E}(f(Z_1, \ldots, Z_n)) \geq t) \leq \mathrm{e}^{-\frac{t^2}{2\sum_{k=1}^n c_k^2}}.$$

Proof. We set, for $1 \leq k \leq n$, $\mathcal{F}_k = \sigma(Z_1, \ldots, Z_k)$ and $\mathcal{F}_0$ the trivial σ-algebra, and

$$S_k = \mathbb{E}\left(f(Z_1, \ldots, Z_n) | \mathcal{F}_k\right).$$

Clearly, $(S_k, \mathcal{F}_k)_{0\leq k\leq n}$ is a martingale, moreover, denoting by W_k a random variable with the same distribution as Z_k and independent of $Z_1, \ldots, Z_n$, $Y_{k-1} = \mathbb{E}\left(f(Z_1, \ldots, Z_{k-1}, W_k, Z_{k+1}, \ldots, Z_n) | \mathcal{F}_k\right)$ and therefore $-c_k \leq Y_k - Y_{k-1} \leq c_k$. Then, apply Theorem 0.1.4 with $a_k = -c_k, b_k = c_k$. □

Corollary 0.1.6. *For all $t > 0$,*

$$\mathbb{P}(LC_n - \mathbb{E}LC_n \geq t) \leq \mathrm{e}^{-\frac{t^2}{4n}}.$$

Proof. The variable LC_n is a function of $2n$ independent random variables:

$$LC_n = LCS(X_1 \ldots X_n; Y_1 \ldots Y_n).$$

Changing one of the variables does not change the absolute value of LC_n by more than one unit: we therefore have the hypotheses of the corollaries verified with $c_k = 1$, hence the result. □

The following result is a slight improvement of this inequality:

Theorem 0.1.7 (McDiarmid, 1989 [53])**.** *For all $t > 0$,*

$$\mathbb{P}(LC_n - \mathbb{E}LC_n \geq t) \leq \mathrm{e}^{-\frac{t^2}{n}}.$$

Proof. We use $Z_1, \ldots, Z_{2n}$ to denote the random variables $X_1, \ldots, X_n, Y_1, \ldots, Y_n$. For all $1 \leq k \leq 2n$,

$$LCS(Z_1 \ldots Z_{k-1} Z_{k+1} \ldots Z_{2n}) \leq LC_n \leq LCS(Z_1 \ldots Z_{k-1} Z_{k+1} \ldots Z_{2n}) + 1,$$

and therefore, with the previous notations,

$$\mathbb{E}(LCS(Z_1 \dots Z_{k-1} Z_{k+1} \dots Z_{2n})|\mathcal{F}_k) \leq S_k \leq \mathbb{E}(LCS(Z_1 \dots Z_{k-1} Z_{k+1} \dots Z_{2n})|\mathcal{F}_k) + 1.$$

Let $a_k = \mathbb{E}(LCS(Z_1 \dots Z_{k-1} Z_{k+1} \dots Z_{2n})|\mathcal{F}_k) - S_{k-1}$, $b_k = a_k + 1$, thus a_k and b_k are $\mathcal{F}_{k-1}$-measurable. Following the original proof of Theorem 0.1.4, when a_k, b_k are $\mathcal{F}_{k-1}$-measurable random variables, the expectation of the right-hand term $\mathrm{e}^{-\frac{2t^2}{\sum_{k=1}^n (b_k - a_k)^2}}$ is also an upper bound. This grants the result. □

By applying the same method to the variable $-LC_n$, we get the following corollary:

Corollary 0.1.8. *For all $\epsilon > 0$,*

$$\mathbb{P}\left(\left|\frac{LC_n}{n} - \frac{\mathbb{E}LC_n}{n}\right| \geq \epsilon\right) \leq 2\mathrm{e}^{-n\epsilon^2}.$$

As $\mathbb{E}LC_n/n$ converges to $\gamma \in \mathbb{R}$, the last corollary ensures the convergence in probability of LC_n/n to γ. In fact, even almost sure convergence holds true. This follows from the following result:

Proposition 0.1.9. *Almost surely, $|LC_n - \mathbb{E}LC_n| = \mathcal{O}\left(\sqrt{n \log n}\right)$.*

Proof. Thanks to Theorem 0.1.7,

$$\sum_{k=1}^{+\infty} \mathbb{P}\left(|LC_k - \mathbb{E}LC_k| \geq \sqrt{2}\sqrt{k \log k}\right) \leq \sum_{k=1}^{+\infty} \frac{2}{k^2} < +\infty.$$

Hence, by the Borel-Cantelli lemma: almost surely, from a certain rank, $|LC_k - \mathbb{E}LC_k| < \sqrt{2}\sqrt{k \log k}$, which implies the result. □

0.1.2 Rate of convergence of the expectation

The following theorem gives a useful rate of convergence, we refer to Rhee [58] for an elementary proof of it.

Theorem 0.1.10 (Alexander, 1994)**.** *There exists a universal constant $K > 0$ such that for all $n \geq 1$,*

$$\gamma - K\sqrt{\frac{\log n}{n}} \leq \frac{\mathbb{E}LC_n}{n} \leq \gamma. \tag{0.1.1}$$

In Alexander [1], it is specified that for any $K > 2 + \sqrt{2}$, the bound (0.1.1) is valid for all sufficiently large n.

With a more complete study to choose the partitions, it is possible to show that any $K > \sqrt{2}$ is also suitable for sufficiently large n (see [49] and [30] for a Markovian model). As Alexander [1] points out, even a $\sqrt{1/n}$ bound instead of $\sqrt{\log n/n}$ might be true.

0.1.3 The variance

We have seen that the asymptotics of the expectation are rather well understood, but the order of the second moment is still an open problem. Here is what is known so far, note that in Chapter 2 of our work we will present some contributions to this topic.

We start with an upper bound. First note that Corollary 0.1.8 implies $\mathrm{Var}\, LC_n \leq 8n$. To get a slightly better result, we use the following theorem, a generalization to the original Efron-Stein inequality [20] to non-symmetric functions.

Theorem 0.1.11 (Steele, 1986 [65]). *Let $Z_1, \ldots, Z_n, W_1, \ldots, W_n$ be random variables i.i.d. and $f : \mathbb{R}^n \to \mathbb{R}$ Borel-measurable. Let $F = f(Z_1, \ldots, Z_n)$ and, for $1 \leq i \leq n$,*

$$F_i = f(Z_1, \ldots, Z_{i-1}, W_i, Z_{i+1}, \ldots, Z_n).$$

Then

$$\mathrm{Var}F \leq \frac{1}{2} \sum_{i=1}^{n} \mathbb{E}((F - F_i)^2).$$

This theorem is the starting point of our investigation in Chapter 2, where we provide an alternative proof, so it is worth recalling one of the classical proofs, which is as follows.

Proof. It is a matter, just as in Corollary 0.1.5, of generalizing a result valid for a sum of n variables to any function of n variables, by writing this function as a sum. Let, for $0 \leq i \leq n$, $\mathbb{E}_i(.) = \mathbb{E}(.|Z_1, \ldots, Z_i)$ and $\mathbb{E}^i(.) = \mathbb{E}(.|Z_1, \ldots, Z_{i-1}, Z_{i+1}, \ldots, Z_n)$. For $i > 0$, let $\Delta_i = \mathbb{E}_i(F) - \mathbb{E}_{i-1}(F)$. We have $F - \mathbb{E}F = \sum_{i=1}^{n} \Delta_i$ and therefore

$$\mathrm{Var}F = \sum_{i=1}^{n} \mathbb{E}(\Delta_i^2) + 2 \sum_{j>i} \mathbb{E}(\Delta_i \Delta_j).$$

For $j > i$, $\mathbb{E}(\Delta_i \Delta_j | \mathcal{F}_i) = \Delta_i \mathbb{E}(\Delta_j | \mathcal{F}_i) = 0$, so $\mathbb{E}(\Delta_i \Delta_j) = 0$. Moreover, as $\mathbb{E}_{i-1}(F) = \mathbb{E}_i(\mathbb{E}^i(F))$, $\Delta_i = \mathbb{E}_i(F - \mathbb{E}^i(F))$, so by the conditional Jensen's inequality $\Delta_i^2 \leq \mathbb{E}_i((F - \mathbb{E}^i(F))^2)$ and $\mathbb{E}(\Delta_i^2) \leq \mathbb{E}((F - \mathbb{E}^i(F))^2) = \mathbb{E}(\mathrm{Var}^i F)$, where $\mathrm{Var}^i F = \mathbb{E}^i((F - \mathbb{E}^i(F))^2)$. So we have

$$\mathrm{Var}F \leq \mathbb{E}\left(\sum_{i=1}^{n} \mathrm{Var}^i F \right).$$

Conditionally on $Z_1, \ldots, Z_{i-1}, Z_{i+1}, \ldots, Z_n$, F and F_i are independent, so $\mathbb{E}^i((F - F_i)^2) = 2\mathbb{E}^i(F^2) - 2\mathbb{E}^i(F)^2$, so $2\mathrm{Var}^i F = \mathbb{E}^i((F - F_i)^2)$, hence $2\mathbb{E}(\mathrm{Var}^i F) = \mathbb{E}((F - F_i)^2)$, which yields the desired result. □

We now apply this result to the random variable LC_n:

Corollary 0.1.12 (Steele, 1986 [65]). *Let, for all $a \in \mathcal{A}$, $p_a = \mathbb{P}(X_1 = a)$ (recall that the X_i and Y_i are i.i.d. with values in $\mathcal{A}$). Then*

$$\mathrm{Var}\, LC_n \leq n \left(1 - \sum_{a \in \mathcal{A}} p_a^2 \right).$$

Proof. By renaming $Z_1, \ldots, Z_{2n}$ the variables $X_1, \ldots, X_n, Y_1, \ldots, Y_n$, LC_n can be written as $f(Z_1, \ldots, Z_{2n})$, and $\mathbb{E}((F - F_i)^2)$ can here be bounded above by $1 - \sum_{a \in \mathcal{A}} p_a^2$. Indeed, changing a single Z_i only changes LC_n, in absolute value, by at most 1, and there is a probability $\sum_{a \in \mathcal{A}} p_a^2$ that $Z_i = W_i$ in which case LC_n remains unchanged. □

We will revisit and improve this last bound in Chapter 2. The most challenging part of the study of the variance is to find a non-trivial lower bound. To the best of our knowledge, it is not even proved, in the uniform binary case, whether or not the variance diverges to infinity, and whether or not it increases with n. However, for some distributions of letters (where one letter has

a very high probability of appearing, the others having a very low probability), [48] (in the binary case) and [36] proved that the variance had a lower bound of order n, and therefore was of order n. Waterman [74] conjectured that the variance was always of order n, and we do not currently know of a counterexample to this conjecture. In any case, it is difficult to evaluate the order of the variance computationally, because it grows slowly for small values of n, which may have led Chvátal and Sankoff [15] to the conjecture of a variance of order $n^{2/3}$. But posterior simulations (July 2016, [50]) agree with Waterman. These Monte-Carlo simulations with 10000 draws, for n varying from 50000 to more than 1000000, give the exponent α such that Cn^α approaches the variance as closely as possible. The exponent found is very close to 1, and this for the three distributions of letters tested (the exponent is closer to 1 when the distribution has a low entropy, but the uniform binary distribution still gives an exponent quite close to 1: $\alpha = 0.9086$).

The problem of finding the order of the variance is particularly interesting because it was shown in [28] that if the variance has a linear lower bound, then LC_n, renormalized, tends to a Gaussian distribution. More precisely:

Theorem 0.1.13 (Houdré and Işlak, 2022)**.** *Suppose there is $C > 0$ such that $\operatorname{Var} LC_n \geq Cn$, then for any $\eta \in (0, 1/10)$,*

$$d_W\left(\frac{LC_n - \mathbb{E}LC_n}{\sqrt{\operatorname{Var} LC_n}}, \mathcal{G}\right) = \mathcal{O}\left(\frac{1}{n^\eta}\right)$$

where $\mathcal{G}$ is a standard normal random variable and where d_W is the Wasserstein distance.

As mentioned in [28], even a weaker bound on the variance allows to conclude: Suppose there is $C > 0$ and $\beta > 9/10$ such that $\operatorname{Var} LC_n \geq Cn^\beta$, then for any $\varepsilon \in (0, \beta - 9/10)$,

$$d_W\left(\frac{LC_n - \mathbb{E}LC_n}{\sqrt{\operatorname{Var} LC_n}}, \mathcal{G}\right) = \mathcal{O}\left(\frac{1}{n^\varepsilon}\right).$$

Note that convergence to the Gaussian distribution for d_W implies convergence in probability and for the Kolmogorov distance.

We have reviewed the basic properties of LC_n. We now turn to a variant, the longest common and increasing subsequences, which will be the main object of the first chapter of our work.

0.2 The longest common and increasing subsequences

In this section, let $\mathcal{A}_m := \{1, \dots, m\}$, $m \geq 2$. For $(x_1, \dots, x_s)$, $(y_1, \dots, y_t)$ two sequences taking values in $\mathcal{A}$, we denote by $LCIS(x_1 \dots x_s; y_1 \dots y_t)$ the maximal length of an increasing subsequence (so a subsequence with a block of 1's, followed by a block of 2's,... and so on) of both words. More formally, we let $LCIS(x_1 \dots x_s; y_1 \dots y_t)$ be the largest integer k such that there exist $1 \leq i_1 < \dots < i_k \leq n$ and $1 \leq j_1 < \dots < j_k \leq n$ such that

- $\forall s \in \{1, \dots, k\}$, $x_{i_s} = y_{j_s}$,
- $x_{i_1} \leq x_{i_2} \leq \dots \leq x_{i_k}$ and $y_{j_1} \leq y_{j_2} \leq \dots \leq y_{j_k}$,

and if no integer satisfies these two conditions, we set $LCIS(x_1 \dots x_s; y_1 \dots y_t) = 0$.

In similarity to the LCS case, we consider two independent sequences of i.i.d. random variables $(X_k)_{k\geq 1}$, $(Y_k)_{k\geq 1}$, and moreover the Y_k's may have a different distribution than the X_k's. Let $p_1^X, \dots, p_m^X$, $p_i^X > 0$, $i = 1, \dots, m$ and $p_1^Y, \dots, p_m^Y$, $p_i^Y > 0$, $i = 1, \dots, m$ be their respective p.m.f., and let $LCI_n = LCIS(X_1 \dots X_n; Y_1, \dots Y_n)$. This model has been mostly studied in computer science (see e.g. [13], [60]), the main motivations being the generalization of the longest increasing

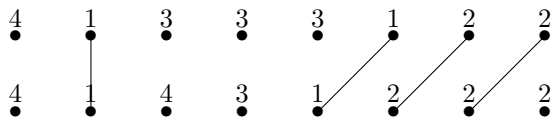


Figure 11: $(1, 1, 2, 2)$ is a common increasing subsequence of maximum length

subsequences of one single word (see the next section) and potential applications to bioinformatics (e.g. in [17]). The asymptotics of the distribution of LCI_n were studied first in the binary case [31] and then in the uniform m letters case [12]. Note that it is easy to adapt the concentration bounds and variance upper bounds seen in the previous section to this variant.

When the two words are sampled from the same uniform distribution, the asymptotic distribution has been found (below, as usual, $\wedge$ is short for minimum), see Theorem 1.1.1 and its conjectured generalization Theorem 1.1.2. The first chapter of our work obtains the limiting distribution of LCI_n, without assuming that the X_k and Y_k $(k = 1, 2, \dots)$ have the same distribution; providing also an alternative proof of Theorem 1.1.1 as well as a proof of the conjectured Theorem 1.1.2.

One may wonder if the problem of the variance is solved in this setup because, of course, the convergence in distribution does not imply the convergence of any moment. Let us prove that in fact the convergence of moments as well. Theorem 0.1.7 continues to hold in this setup (clearly, changing one variable does not change the absolute value of LCI_n by more than one unit), so we have for any $t > 0$,

$$\mathbb{P}\left(|LCI_n - \mathbb{E}LCI_n| \geq t\right) \leq 2\mathrm{e}^{-\frac{t^2}{n}},$$

hence for any $k \geq 1$,

$$\mathbb{E}\left(\frac{LCI_n - \mathbb{E}LCI_n}{\sqrt{n}}\right)^{2k} \leq 2k!,$$

so for any $k \geq 1$, the sequence $\left(\frac{LCI_n - \mathbb{E}LCI_n}{\sqrt{n}}\right)^k$ is bounded in L_2 and therefore is uniformly integrable. Denoting by L the limiting distribution of $\frac{LCI_n - \mathbb{E}LCI_n}{\sqrt{n}}$, we also have that the sequence $\left(\frac{LCI_n - \mathbb{E}LCI_n}{\sqrt{n}}\right)^k$ converges in distribution to L^k, and therefore we get that L^k is integrable and

$$\mathbb{E}\left(\frac{LCI_n - \mathbb{E}LCI_n}{\sqrt{n}}\right)^k \xrightarrow[n\to\infty]{} \mathbb{E}L^k,$$

the desired result (in particular, the problem of the variance is solved).

With this variant of the LCS problem in mind, let us recall results on the longest increasing subsequences of one single word.

0.3 The longest increasing subsequences

Define the longest increasing subsequences as previously, except that now there is only a single word: For $x_1, \dots, x_s \in \mathcal{A}$, let $LIS(x_1 \dots x_s)$ be the largest integer k such that there exist $1 \leq i_1 < \dots < i_k \leq s$ such that $x_{i_1} \leq x_{i_2} \leq \dots \leq x_{i_k}$ and if no integer satisfies these two conditions, we set $LIS(x_1 \dots x_s) = 0$. The LIS was first studied when the letters are drawn from a random permutation. Let π be a random (according to the uniform distribution) permutation of $\{1, \dots, n\}$, and let $I_n = LIS\left(\pi(1), \dots, \pi(n)\right)$. The problem of finding the asymptotics of I_n was introduced by Ulam [72] in 1961 , and made popular by Hammersley [26]. Decades later, Kerov [45], Tracy and Widom [70] investigated the asymptotics of $LI_n = LIS\left(X_1, \dots, X_n\right)$ where $X_1, \dots, X_n$ are

i.i.d. random variables with uniform distribution on $\{1, \dots, m\}$. In words, this is the problem of the longest increasing subsequences of a random word, rather than a random permutation. This problem of finding the longest increasing subsequences of a random word, or more precisely a generalization as exposed next, has a surprising connection with quantum statistics, that will be presented in Section 0.4.

Let us start this section with the Robinson–Schensted–Knuth (RSK) correspondence, which is an invaluable tool for both the random permutation and the random word problems. Then, we will review some limiting results on the random word problem. Finally, limit theorems for the random permutation problem will appear as a limiting case of the previous results (we proceed in this presentation in a non-chronological order).

0.3.1 The Robinson–Schensted–Knuth correspondence

Consider a word $w \in \{1, \dots, m\}^n$ (note that it may be permutation when $m \geq n$). Schensted [63] was the first to connect $LIS(w)$ to the size of the first pile in patience sorting. Let us recall the steps of this process. Initially, we consider an empty list L. Then, for i ranging from 1 to n, insert the letter w_i in the list L, such that the entries in L are weakly increasing (so w_i is inserted between two consecutive entries a, b so that $a \leq w_i < b$, or inserted as the last entry of the list if it is greater than all other entries). If w_i is the largest entry, return L, else, delete ("bump") the first entry strictly greater than w_i (the entry b) and return L. See Figure 12 below for an example of the patience sorting process.

1

1	3

1	2

(3 was bumped)

1	1

(2 was bumped)

1	1	1

Figure 12: Patience sorting of $w = (1, 3, 2, 1, 1)$

Schensted's theorem [63] states that the length of L, which is the result of applying patience sorting to w, is equal to $LIS(w)$. Let us denote by $|L|$ the length of L. In order to prove the theorem, one can prove, by induction on n, a stronger result: $|L| = LIS(w)$ and for any $i \in \{1, \dots, |L|\}$, LC_i is the minimal integer such that there exists a weakly increasing subsequence of w made of i letters smaller or equal to LC_i. In our previous example, this means that L gives the extra information that there exists an increasing subsequence of length 3 (the maximal length) with all letters smaller or equal to 1.

This theorem already allows a better understanding of the LIS, but Schensted took it one step further: instead of just discarding the bumped letters, these are used for the next line, where we apply the same rules of patience sorting. The result of such a procedure, called the Schensted insertion, is not merely a list but an arrangement of n natural integers in boxes. See Figure 13 for an example of the Schensted insertion of $w = (1, 3, 2, 1, 1)$, the bold letters being the newly inserted ones.

1

1	**3**

1	**2**
3	

(3 was bumped)

1	**1**
2	
3	

(2 was bumped to the second line, bumping 3 to the third line)

1	1	**1**
2		
3		

Figure 13: The Schensted insertion procedure applied to the word $w = (1, 3, 2, 1, 1)$

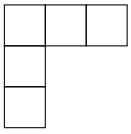


Figure 14: The Young diagram of the last tableau of Figure 13

Note that the lengths of the lines of the last tableau are $3, 1, 1$ and therefore are weakly decreasing. More generally, the output P of the Schensted insertion procedure always has weakly decreasing row lengths, so it is a Young tableau (an arrangement of natural integers with weakly decreasing row lengths - for a standard introduction on Young tableaux, see [21]). Furthermore, P also has each row weakly increasing, and each column strictly increasing, such tableaux are called Semistandard Young Tableaux (SSYT). One can prove by induction on n that any output of the Schensted insertion procedure is a SSYT. A Standard Young Tableau (SYT) is a Young tableau with strictly increasing rows and strictly increasing columns and containing exactly the numbers 1 through n (the total number of boxes). So when the input w is a permutation, P is also a SYT.

A partition λ of n is a weakly decreasing list of non-negative integers $(x_1, \ldots, x_\ell)$ such that $x_1 + \cdots + x_\ell = n$, and this is written $\lambda \vdash n$. We also denote by $\ell(\lambda)$ the length of λ, the number of nonzero elements in the list. For any partition λ, we may complete it with an arbitrary number of zeros, so that λ_k is well defined for all $k \geq 1$ (and it is zero when $k > \ell(\lambda)$). The shape of a Young tableau of size n (meaning the total number of boxes is n) is defined as the partition of n composed of the lengths of its rows.

A Young diagram is an arrangement of boxes, with weakly decreasing row lengths (in other words, a Young tableau with empty boxes). Figure 14 is the Young diagram of the last tableau of Figure 13.

Clearly, there is a canonical one-to-one correspondence between the Young diagrams of size n and the partitions of n. For any Young diagram T, if $c_1, \ldots, c_k$ are the lengths of its columns, the Young diagram T' with rows of length $c_1, \ldots, c_k$ is called the conjugate of T. One can also define the conjugate diagram as its reflexion about the line $y = -x$.

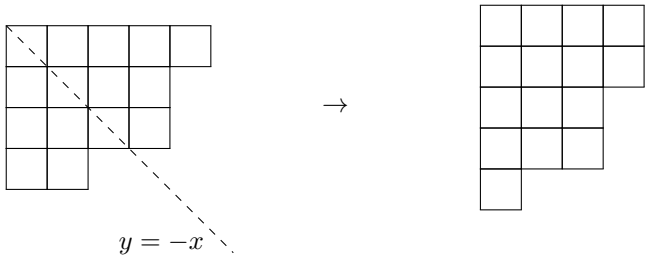


Figure 15: The correspondence between a Young diagram and its dual

Using the correspondence between partitions and Young diagrams, we may also consider for $\lambda \vdash n$ the conjugate partition λ'. In the Figure 15 example, the first partition is $(5, 4, 4, 2)$ and the conjugate on the right is $(4, 4, 3, 3, 1)$.

We can now state the following result, generalizing Schensted's theorem:

Theorem 0.3.1 (Greene, 1974 [25]). *Let $w \in \{1, \dots, m\}^n$, let $\lambda \vdash n$ be the shape of the output of the Schensted insertion procedure of w, and λ' its conjugate. Then for any $k \leq \ell(\lambda)$, $\lambda_1 + \cdots + \lambda_k$ is equal to the length of the longest k disjoint increasing subsequences of w, and for any $\ell \leq \ell(\lambda')$, $\lambda'_1 + \cdots + \lambda'_\ell$ is equal to the length of the longest ℓ disjoint strictly decreasing subsequences of w.*

As an example, this theorem states that the word $(1, 3, 2, 1, 1)$ has two disjoint increasing subsequences of total length $3+1 = 4$ (see Figure 13), and indeed one may consider the subsequences $(1, 1, 1)$ and (2). It also implies that the maximal length of a strictly decreasing subsequence is 3 (and indeed there is the subsequence $(3, 2, 1)$).

Note that given a SSYT, there may be different words giving this same output. For example, $(3, 2, 1, 1, 1)$ gives the same output as above. In order to make the procedure a bijection between the words and the outputs, the RSK procedure keeps track of the insertion order in another tableau, the recording tableau. The output is a couple (P, Q) of P the SSYT output from the Schensted insertion, and Q a tableau with the same shape (meaning the i-th line of P has same length as the i-th line of Q) recording the order of insertion. To construct Q, when the i-th letter is inserted, a cell i where a new cell appears in P. Still with our previous example, we get the pair (P, Q) as detailed in Figure 16.

The recording tableau Q is always a SYT. Here is the key fact establishing the bijection mentioned above:

Theorem 0.3.2. *The RSK procedure is a bijection between:*

- *The set of words $\{1, \dots, m\}^n$ and the pairs of Young tableaux (P, Q) such that P, Q have same shape, same size n, P is a* SSYT *with alphabet $\{1, \dots, m\}$, Q is a* SYT.
- *The symmetric group $\mathfrak{S}_n$ and the pairs of Young tableaux (P, Q) such that P, Q have same shape, same size n, P and Q are* SYT.

We now provide an informal proof of this theorem. The key is to proceed in reverse: given (P, Q), one knows the position of the letters added last (that is where the entry n is in Q), and from there, read in P the chain of insertions and bumps that lead to this insertion, so read the last letter of the input. For example, consider the following output (P, Q):

We read in the recording tableau on the right that the last letter inserted was at the bottom box. From the tableau on the left, we see it is a 4, and to get there, it must have been bumped from the line above, and necessarily by a 2, which in turn must have been bumped by the insertion

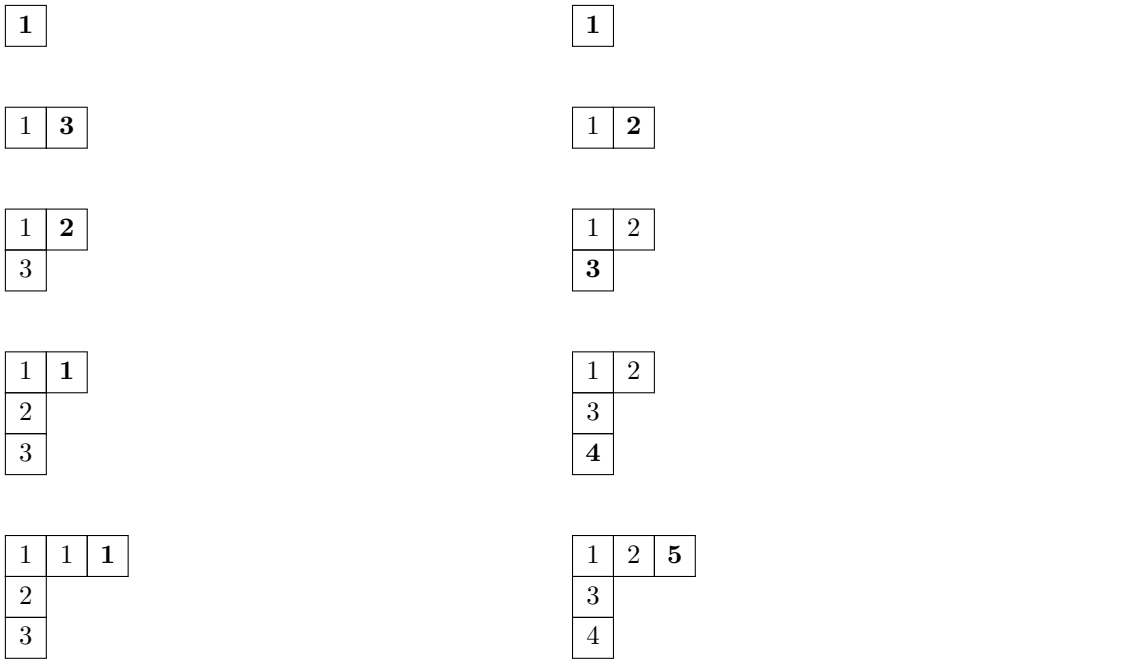


Figure 16: The RSK procedure applied to the word $w = (1, 3, 2, 1, 1)$

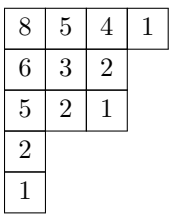
of a 1 in the line above. So the last letter of the input was a 1. We also know what were the tableaux P, Q before the last letter. So we can go on, and find the whole input, $(2, 4, 3, 6, 5, 8, 7, 1)$.

In the sequel, for any word w, we denote by RSKshape(w) the shape of the tableaux P, Q given by the RSK algorithm with input w.

The first corollary of the RSK correspondence is an explicit formula for the distribution of I_n, and more generally, the distribution of RSKshape(π) for π a random permutation of $\{1, \dots, n\}$ (recall that by Greene's Theorem, $I_n = \text{RSKshape}(\pi)_1$). Let $\lambda \vdash n$, by the RSK correspondence, the number of permutations σ such that RSKshape$(\sigma) = \lambda$ is equal to the number of pairs of SYTs of same shape λ. We denote by $\text{SYT}(\lambda)$ the set of SYTs of shape λ, and define $f^\lambda := |\,\text{SYT}(\lambda)|$ (this is also denoted in the literature by $\dim \lambda$). We then have:

$$\mathbb{P}\left(\text{RSKshape}(\pi) = \lambda\right) = \frac{(f^\lambda)^2}{n!}.$$

To make this formula more explicit, we now remind the Hook Length Formula for f^λ. For any box $\square$ in λ (we identify any partition with its Young diagram), we define $h(\square)$ to be the number of boxes either on the same column, but down $\square$, or on the same line, but to the right of $\square$. For example, here is a Young diagram with, in each box, its hook length:



More explicitly, let i, j be the coordinates of $\square$ (i is the number of the line, j is the number of the row, we start the numbering at 1), and let λ' be the conjugate of λ, we have

$$h(\square) = \lambda_i + \lambda'_j - i - j + 1.$$

The Hook Length Formula states that:

$$f^\lambda = \frac{n!}{\prod_{\square\in\lambda} h(\square)}.$$

One may also rewrites the term $\prod_{\square\in\lambda} h(\square)$ a bit differently. Let

$$h_i = \lambda_i + \ell(\lambda) - i,$$

this is the hook length of the first square of the i-th row. One shows

$$h_i! = \prod_{j>i}(h_i - h_j) \prod_{j\leq\lambda_i} h(\square_{i,j}),$$

where $\square_{i,j}$ is the box with coordinates i, j. Hence

$$\prod_{\square\in\lambda} h(\square) = \frac{\prod_{i=1}^{\ell(\lambda)} h_i!}{\Delta(h_1, \dots, h_{\ell(\lambda)})},$$

where $\Delta(h_1, \dots, h_{\ell(\lambda)})$ is the Vandermonde determinant $\prod_{1\leq i<j\leq\ell(\lambda)}(h_i - h_j)$, and finally we get an alternative Hook Length Formula

$$f^\lambda = \frac{n!\Delta(h_1, \dots, h_{\ell(\lambda)})}{\prod_{i=1}^{\ell(\lambda)} h_i!}.$$

Let us now turn to the distribution of RSKshape$(X_1, \dots, X_n)$ where $X_1, \dots, X_n$ are i.i.d. random variables taking values in $\{1, \dots, m\}$ with distribution $p_1, \dots, p_m$. For $\lambda \vdash n$, let $\mathrm{SSYT}(\lambda)$ be the set of SSYTs of shape λ and let $\mathrm{SSYT}_m(\lambda)$ be the set of SSYTs of shape λ with entries in $\{1, \dots, m\}$. The Schur polynomials are defined as

$$s_\lambda(x_1, x_2, \dots) = \sum_{T\in\mathrm{SSYT}(\lambda)} \prod_{i=1}^{\infty} x_i^{\text{number of entries } i \text{ in } T}.$$

Let, for any integer $k \geq 1$ (and with a slight abuse of notation):

$$s_\lambda(x_1, \dots, x_k) = s_\lambda(x_1, \dots, x_k, 0, 0, \dots).$$

We have, in particular:

$$s_\lambda(x_1, \dots, x_m) = \sum_{T\in\mathrm{SSYT}_m(\lambda)} \prod_{i=1}^{m} x_i^{\text{number of entries } i \text{ in } T}.$$

Note that if λ has more than m lines, $SSYT_m(\lambda) = \emptyset$ and $s_\lambda(x_1, \dots, x_m) = 0$. We now compute $\mathbb{P}\left(\mathrm{RSKshape}(X_1, \dots, X_m) = \lambda\right)$. We have

$$\begin{aligned}\mathbb{P}\left(\mathrm{RSKshape}(X_1 \dots, X_m) = \lambda\right) &= \sum_{P\in\mathrm{SSYT}_m(\lambda), Q\in\mathrm{SYT}(\lambda)} \mathbb{P}\left(\mathrm{RSK}(X_1, \dots, X_m) = (P, Q)\right) \\ &= \sum_{P\in\mathrm{SSYT}_m(\lambda), Q\in\mathrm{SYT}(\lambda)} \mathbb{P}\left(X_1, \dots, X_m = \mathrm{RSK}^{-1}(P, Q)\right),\end{aligned}$$

where $RSK^{-1}(P,Q)$ is the unique word giving the output (P,Q). Note that the number of letters i in this word is the number of entries i in P, so

$$\mathbb{P}\left(X_1,\dots,X_m = \mathrm{RSK}^{-1}(P,Q)\right) = \prod_{i=1}^{m} p_i^{\text{number of entries } i \text{ in } P},$$

and

$$\begin{aligned}\mathbb{P}\left(\mathrm{RSKshape}(X_1\dots,X_m)=\lambda\right) &= \sum_{P\in \mathrm{SSYT}_m(\lambda), Q\in \mathrm{SYT}(\lambda)} \prod_{i=1}^{m} p_i^{\text{number of entries } i \text{ in } P} \\ &= f^\lambda s_\lambda(p_1,\dots,p_m). \end{aligned} \tag{0.3.1}$$

In particular, this provides an explicit formula for the distribution of LI_n.

We will also use Cauchy's bialternant formula: for any $\lambda \vdash n$, for any $m \geq \ell(\lambda)$,

$$s_\lambda(x_1,\dots,x_m) = \frac{\det\left(x_i^{\lambda_j+m-j}\right)_{1\leq i,j\leq m}}{\Delta(x_1,\dots,x_m)}.$$

Note that this polynomial is well defined: the determinant is alternating, therefore it is divisible by $\Delta(x_1,\dots,x_m)$. Also, recall that when $m<\ell(\lambda)$, $s_\lambda(x_1,\dots,x_m)=0$. From this formula, it is clear that the Schur polynomials are symmetric. Using (0.3.1), this implies:

Proposition 0.3.3. *Let $\sigma \in \mathfrak{S}_m$, let $X_1,\dots,X_n$ be i.i.d. random variables with p.m.f. $p_1,\dots,p_m$, and let $Y_1,\dots,Y_n$ be i.i.d. random variables with p.m.f. $p_{\sigma(1)},\dots,p_{\sigma(m)}$. Then,* $\mathrm{RSKshape}(X_1,\dots,X_n)$ *and* $\mathrm{RSKshape}(Y_1,\dots,Y_n)$ *have the same law.*

The distribution of $\mathrm{RSKshape}(X_1,\dots,X_n)$ is called the Schur-Weyl distribution with parameters p,n, and it is denoted by $\mathrm{SW}^n(p)$. We also write SW^n_m in the special case of the uniform distribution $p=(1/m,\dots,1/m)$.

In the sequel, and especially in Chapter 3 of our work, to investigate the distribution of $\mathrm{RSKshape}(X_1,\dots,X_n)$ we will therefore assume, without loss of generality, that $p_1\geq p_2\geq\cdots\geq p_m$.

We now review limiting theorems, firstly for the random word model, secondly, the random permutation model.

0.3.2 The RSK shape for random words

We now review some known results on the asymptotic distribution of $LI_n = LIS(X_1,\dots,X_n)$, and more generally, $\mathrm{RSKshape}(X_1,\dots,X_n)$, for $X_1,\dots,X_m$ i.i.d. random variables taking values in $\{1,\dots,m\}$ with distribution $p_1,\dots,p_m$.

The asymptotic distribution of LI_n, and more generally of the RSKshape, turns out to be closely connected to the eigenvalues of certain random matrices: the matrices with the Gaussian unitary ensemble (GUE) distribution. We remind that the Gaussian unitary ensemble of size m, denoted by GUE_m, is the probability distribution of the $m\times m$ Hermitian matrices H defined as follows:

- For any $i\in\{1,\dots,m\}, H_{i,i}\sim\mathcal{N}(0,1)$;
- For any $i,j\in\{1,\dots,m\}$ such that $i<j$, $H_{i,j}\sim\mathcal{N}_{\mathbb{C}}(0,1)$ (the complex standard Gaussian, equal to $\mathcal{N}(0,1/2)+i\mathcal{N}(0,1/2)$);

- These entries are drawn independently.

One may compute the density of probability of the distribution GUE_m: for some normalizing constant $C > 0$, the density is $\mathrm{e}^{-\mathrm{Tr}(H^2)/2}/C$ (this is direct consequence of the definition, and an alternative equivalent definition of GUE_m).

Using the notations of [41], we also define the traceless GUE of size m, GUE_m^0, as the distribution of $H - (\mathrm{Tr}(H)/m)I_m$ where $H \sim \mathrm{GUE}_m$. This is the conditional distribution of GUE_m given that the trace is zero. The first asymptotic distribution result was obtained by Tracy and Widom, in the uniform case.

Theorem 0.3.4 (Tracy and Widom, 2001 [70]). *Let $H \sim \mathrm{GUE}_m^0$ and let $\mu_1(H)$ denote its largest eigenvalue. In the uniform case with m letters, we have the convergence in distribution:*

$$\frac{LI_n - n/m}{\sqrt{n/m}} \Longrightarrow_{n\to\infty} \mu_1(H).$$

It was conjectured in [70] that the convergence holds for the whole Young diagram, and later proved by Johansson:

Theorem 0.3.5 (Johansson, 2001 [44]). *Let $H \sim \mathrm{GUE}_m^0$ and $\mu_1(H) \geq \cdots \geq \mu_m(H)$ denote its eigenvalues. If $\lambda \sim \mathrm{SW}_m^n$, we have the convergence in distribution:*

$$\left(\frac{\lambda_1 - n/m}{\sqrt{n/m}}, \ldots, \frac{\lambda_m - n/m}{\sqrt{n/m}}\right) \Longrightarrow_{n\to\infty} (\mu_1(H), \ldots, \mu_m(H)).$$

This last result was proved earlier (1994) by Kerov [45, Chap. 3, Sec. 3.4, Theorem 2]. This result was then generalized to a non-uniform distribution. We follow [41] for the following notations. For any sorted p.m.f. p on $\{1, \ldots, m\}$, define the generalized traceless GUE distribution $\mathrm{GUE}^0(p)$ as the distribution of H, where H is defined as follows. Let $d_1, \ldots, d_k$ be the multiplicities of p, which means $p_1 = \cdots = p_{d_1} > p_{d_1+1} = \cdots = p_{d_1+d_2} > \ldots$.

- Let $H_1 \sim \mathrm{GUE}_{d_1}^0, \ldots, H_k \sim \mathrm{GUE}_{d_k}^0$ be independent random matrices;
- Let B be the $m \times m$ matrix defined by blocks with $H_1, \ldots, H_k$ on its diagonal;
- Let $T = \sum_{i=1}^m \sqrt{p_i} B_{i,i}$;
- Finally, for $i, j \in \{1, \ldots, m\}$, let $H_{i,j} = B_{i,j}$ if $i \neq j$ and $H_{i,i} = B_{i,i} - \sqrt{p_i}T$.

As shown in [41], $\mathrm{GUE}^0(p)$ is the the distribution of the direct sum of mutually independent $d_i \times d_i$ Gaussian unitary ensembles conditional on the eigenvalues $\mu_1, \ldots \mu_d$ satisfying $\sqrt{p_1}\mu_1 + \cdots + \sqrt{p_m}\mu_m = 0$. We may now state the generalization of the previous theorem:

Theorem 0.3.6 (Its, Tracy, Widom, 2001 [42]). *Let p be a sorted distribution on $\{1, \ldots, m\}$, let $H \sim \mathrm{GUE}^0(p)$ and let $\mu_1(H) \geq \cdots \geq \mu_m(H)$ denote its eigenvalues. If $\lambda \sim \mathrm{SW}^n(p)$, we have the convergence in distribution:*

$$\left(\frac{\lambda_1 - p_1 n}{\sqrt{p_1 n}}, \ldots, \frac{\lambda_m - p_m n}{\sqrt{p_m n}}\right) \Longrightarrow_{n\to\infty} (\mu_1(H), \ldots, \mu_m(H)).$$

Remarks.

(i) In particular, if $p_1 > \cdots > p_m$, which means $d_1 = \cdots = d_m = 1$, the limiting distribution is multivariate Gaussian.

(ii) Recalling Proposition 0.3.3, there is no loss in generality with the assumption that p is sorted.

The distribution $\mathrm{GUE}^0(p)$ may not be very intuitive, but the next theorem allows a more natural interpretation of the limit. We denote by $p^{(i)}$, for $1 \leq i \leq k$, the probability with multiplicity d_i (in other words, $(p_1, \ldots, p_m) = (p^{(1)}, \ldots, p^{(1)}, p^{(2)}, \ldots, p^{(2)}, \ldots, p^{(k)}, \ldots, p^{(k)})$).

Theorem 0.3.7 (Méliot 2012 [55], as stated by Wright [76]). *Let $H \in \mathrm{GUE}^0(p)$, let $g_1, \ldots, g_k$ be centered Gaussian random variables with covariance $\left(\mathbb{1}_{i=j} d_i - d_i d_j \sqrt{p^{(i)}}\sqrt{p^{(j)}}\right)_{1 \leq i,j \leq k}$, and let for each $i \in \{1, \ldots, k\}$, $H_i \sim \mathrm{GUE}^0_{d_i}$ (the vector g and the H_i's being independent). Then, we have the following equality in distribution:*

$$(\mu_1(H), \ldots, \mu_m(H)) \overset{d}{=} \left(\frac{g_1}{d_1} + \mu_1(H_1), \ldots, \frac{g_1}{d_1} + \mu_{d_1}(H_1), \frac{g_2}{d_2} + \mu_1(H_2), \ldots, \frac{g_k}{d_k} + \mu_{d_k}(H_k) \right).$$

The limiting distribution may also be written as a Brownian functional: it was done firstly in [32] for the LIS, then in [33] for the LIS in a Markovian framework (the letters are a Markov chain, generalizing the i.i.d. framework), and lastly in full generality in [34] for the whole Young diagram still in a Markovian framework (see also [41]). Broadly, the main idea to get such limits is to revise Greene's Theorem but with disjoint subsequences. Then, the Brownian motions appear as renormalized random walks counting the number of occurrences of each letter. The main advantage of this approach is not only the ability to generalize to a Markovian framework, but also to give non-asymptotic rates of convergence, as seen in Chapter 3. We recall a simple case: the limit of LI_n (the length of the first line of the Young diagram), when the distribution of the letters is p. Denoting by d_1 the multiplicity of k, and letting $B = (B_k(t))_{1 \leq k \leq d_1, t \in [0,1]}$ be a standard d_1-dimensional Brownian motion, the following convergence in distribution result holds true [32, Corollary 3.3]:

$$\frac{LI_n - p_1 n}{\sqrt{p_1 n}} \underset{n \to \infty}{\Longrightarrow} \frac{\sqrt{1 - d_1 p_1} - 1}{d_1} \sum_{j=1}^{d_1} B_j(1) + \max_{0 = t_0 \leq \cdots \leq t_{d_1} = 1} \left(B_j(t_j) - B_j(t_{j-1}) \right). \tag{0.3.2}$$

Since the limiting distribution of the Young diagram is already known (Theorem 0.3.6 and Theorem 0.3.7 above), the Brownian functional must have same distribution. This is not surprising given the connections between some Brownian functionals and the eigenvalues of the GUE ([5], [24]). More precisely, the following theorem from [6], which is a generalization of [5], makes the connection complete. Following [6], we first introduce some notations. Let $B = (B_k(t))_{1 \leq k \leq M, t \in [0,1]}$ be a standard M-dimensional Brownian motion. Let $\mathcal{P}$ be the set of càdlàg, non-decreasing functions from $[0,1]$ to $\{1, \ldots, M\}$. For $\pi \in \mathcal{P}$, π might be written as $\sum_{j=1}^{M-1} j \mathbb{1}_{[t_{j-1}, t_j)} + M \mathbb{1}_{[t_{M-1}, t_M)}$, and let

$$\begin{aligned} \Delta_\pi B &= \int_0^1 \mathrm{d}B_{\pi(t)}(t) \\ &= \sum_{j=1}^{m} \left(B_j(t_j) - B_j(t_{j-1}) \right). \end{aligned}$$

Let $H \sim \mathrm{GUE}_m$, and for $1 \leq k \leq M$, let $\mu_1^k \geq \cdots \geq \mu_k^k$ be the eigenvalues of the principal $k \times k$ minor of H.

Theorem 0.3.8 (Benaych-Georges and Houdré, 2013 [6]). *We have the following equality in distribution:*

$$\left(\mu_i^k\right)_{1 \leq \ell \leq k \leq M} \overset{d}{=} \left(\sup \left\{ \sum_{i=1}^{\ell} \Delta_{\pi_i} B; \pi_1, \ldots \pi_\ell \in \mathcal{P}, \pi_1 < \cdots < \pi_\ell \leq k \right\} \right)_{1 \leq \ell \leq k \leq M}.$$

Now, we may prove directly that the right-hand term in (0.3.2) has the same distribution as the limit given by Theorem 0.3.7, namely, $\mu_1(H_1) + (g_1/d_1)$. Firstly, [6, Corollary 2] asserts the following equality in distribution holds true:

$$\frac{\sqrt{1-d_1p_1}-1}{d_1}\sum_{j=1}^{d_1} B_j(1) + \max_{0=t_0\leq\cdots\leq t_{d_1}=1}\left(B_j(t_j) - B_j(t_{j-1})\right) \stackrel{d}{=} \frac{\sqrt{1-d_1p_1}-1}{d_1}\operatorname{Tr}(H) + \mu_1(H),$$

where $H \sim \mathrm{GUE}_{d_1}$. Secondly, using the independence of the projections of a Gaussian random variables on orthogonal spaces, $H - (\operatorname{Tr}(H)/d_1)I_{d_1}$ and $\operatorname{Tr}(H)$ are independent, and

$$\begin{aligned}\frac{\sqrt{1-d_1p_1}-1}{d_1}\operatorname{Tr}(H) + \mu_1(H) &= \mu_1\left(H - \frac{\operatorname{Tr}(H)}{d_1}I_{d_1}\right) + \frac{\sqrt{1-d_1p_1}}{d_1}\operatorname{Tr}(H)\\ &= \mu_1(H^0) + \frac{\sqrt{d_1-d_1^2p_1}}{d_1}Z\end{aligned}$$

where $H' = H - (\operatorname{Tr}(H)/d_1)I_{d_1} \sim \mathrm{GUE}^0_{d_1}$ and Z is an independent standard Gaussian, which is exactly the limit in Theorem 0.3.7.

Note that the limit of LCI_n, given in Chapter 1, is also a Brownian functional but to this date, the connection with GUE eigenvalues remains unknown.

0.3.3 The RSK shape for random permutations

We now review some known results on the asymptotic distribution of $I_n := LIS(\pi(1),\ldots,\pi(n))$, and more generally, $\lambda^{(n)} := \mathrm{RSKshape}(\pi(1),\ldots,\pi(n))$, for π a random permutation of $\{1,\ldots,m\}$ (following the uniform distribution). This problem, as mentioned before, is older than the random word model, and there would be much more to say, but we focus on the connection with the previous models.

We denote by TW the Tracy-Widom distribution, whose density is defined as the solution of a Painlevé equation, see e.g. [71]. Its mean is approximately -1.771 and its variance is approximately 0.813. Tracy and Widom proved the following convergence result:

Theorem 0.3.9 (Tracy and Widom, 1994 [69]). *Let $H_m \sim \mathrm{GUE}_m$, and $\mu_1(H_m)$ denotes its largest eigenvalue. We have the convergence in distribution:*

$$\sqrt{2}m^{1/6}\left(\mu_1(H_m) - 2\sqrt{m}\right) \Longrightarrow_{n\to\infty} TW.$$

It turned out to be also the limiting distribution of I_n (once rescaled):

Theorem 0.3.10 (Baik, Deift and Johansson, 1999 [4]). *We have:*

$$\frac{I_n - 2\sqrt{n}}{n^{1/6}} \Longrightarrow_{n\to\infty} TW.$$

Note that for any n, when m goes to infinity, LI_n tends to I_n in distribution. Therefore, the following theorem is a generalization of the former one.

Theorem 0.3.11 (Johansson, 2001 [44]). *Let $m \in \mathbb{N}^{\mathbb{N}}$ be such that $(\log n)^{1/6}/m_n \xrightarrow[n\to\infty]{} 0$, then in the uniform case, we have:*

$$\frac{LI_{m_n} - n/m_n - 2\sqrt{n}}{n^{1/6}\left(1+\sqrt{n}/m_n\right)^{2/3}} \Longrightarrow_{n\to\infty} TW.$$

Let us also note that similarly to the random word case, there are also theorems on the limiting shape of the Young diagrams. This time, the number of rows goes to infinity, so a renormalization is in order: let, for all $x \geq 0$,

$$\overline{\lambda^{(n)}}(x) = \frac{\lambda^{(n)}_{\lfloor x\sqrt{n}\rfloor}}{\sqrt{n}},$$

let $G^{(n)}$ be the set under this function, that is, $G^{(n)} = \left\{(x,y) : y < \overline{\lambda^{(n)}}(x), x \geq 0\right\}$, and finally let

$$C = \left\{(x,y) : x \leq \left(\frac{2\theta}{\pi}+1\right)\sin(\theta) + \frac{2}{\pi}\cos(\theta), y \leq \left(\frac{2\theta}{\pi}-1\right)\sin(\theta) + \frac{2}{\pi}\cos(\theta), \theta \in \left[-\frac{\pi}{2}, \frac{\pi}{2}\right]\right\}.$$

Then we have the following result:

Theorem 0.3.12 (Logan, Shep, Vershnik and Kerov, 1977, as stated in [59]). *For any $\varepsilon > 0$, we have*

$$\mathbb{P}\left((1-\varepsilon)C \subset G^{(n)} \subset (1+\varepsilon)C\right) \xrightarrow[n\to\infty]{} 0.$$

Remark. *This is a slight modification (with the usual Young diagram, instead of a tilted "Russian" diagram) of a theorem originally proved independently in 1977 in [73] and [51].*

There is, once again, a connection with the shape of random words: if we consider, instead of $\lambda^{(n)}$, the RSK shape of a word of length n with an alphabet $\{1,\dots,m(n)\}$ (with the uniform distribution), and $m(n)/\sqrt{n} \xrightarrow[n\to\infty]{} 0$, then [7] proved that Theorem 0.3.12 continues to hold. However, when $m(n)/\sqrt{n}$ tends to any other limit than zero, the limiting curve is different, and is given in [7]: this time, the fact that many letters are repeated makes it very different from the permutation case.

To conclude, let us mention two connections between this model and the longest common subsequences problem.

Firstly, $I_n = LCS(\pi; I_n)$ and if π' is another independent random permutation, I_n has the same distribution as $LCS(\pi;\pi')$, because $LCS(\pi;\pi') = LCS(\pi^{-1}\pi'; I_n) = LIS(\pi^{-1}\pi')$.

Secondly, the study of I_n has consequences on the study of γ_k. Although we do not know the exact value of any γ_k, we have the following theorem:

Theorem 0.3.13 (Kiwi, Loebl, Matoušek, 2005 [46]). $\lim_{k\to\infty} \gamma_k\sqrt{k} = 2$.

The idea is to study the case $k >> n$, where one can use the results on I_n, then to come back to this case by partitioning into blocks. It is based in particular on a very strong concentration inequality from [4].

0.4 Quantum statistics

We now give a brief introduction to quantum measurements, and explore connections with RSK. This is the object of Chapter 3 of this work.

A d-dimensional quantum system is a system that is in a mixed state of d quantum states, which means it has probability p_1 to be in the state $u_1 \in \mathbb{C}^d$, p_2 to be in the state $u_2 \in \mathbb{C}^d, \dots, p_d$ to be in the state $u_d \in \mathbb{C}^d$. The matrix $\rho := p_1 u_1 u_1^* + \cdots + p_d u_d u_d^*$ is positive semi-definite (p.s.d.) Hermitian with trace one, it is called the density matrix of the system. Reciprocally, for any p.s.d. Hermitian matrix ρ with trace one, there is (at least) one system with density matrix ρ: if $(p_1,\dots,p_d)$ are the eigenvalues of ρ and $u_1,\dots,u_d$ the unit eigenvectors, one may indeed consider the mixed state: u_1 with probability p_1, u_2 with probability p_2, ... with density matrix

$p_1 u_1 u_1^* + \cdots + p_d u_d u_d^* = \rho$. In the following, we call density matrix any p.s.d. Hermitian matrix with trace one.

Physically, two systems with the same density matrix are indistinguishable. So we may assume that the eigenvectors are the different states and the eigenvalues are the different probabilities of these states. Of course, it is impossible to measure directly ρ, but one may get a measurement of ρ, which is a random variable as defined next.

The simplest kind of measurement is the basis measurement: one provides an orthonormal basis $v_1, \ldots, v_d$ of $\mathbb{C}^d$, and the outcome of the measurement is a random variable N taking values in $\{1, \ldots, d\}$ with p.m.f.:

$$\begin{aligned}\mathbb{P}(N = i) &= \sum_{j=1}^{d} p_j \langle u_i, v_j \rangle^2 \\ &= \mathrm{Tr}(\rho E_i^*) \text{ where } E_i := v_i v_i^* \\ &= \langle \rho, E_i \rangle \text{ with the usual dot product on matrices.}\end{aligned}$$

More generally, to make a projective measurement, one provides $E_1, \ldots, E_d$ some self-adjoint projections (i.e. orthogonal projections) such that $E_1 + \cdots + E_d = I_d$, and the outcome of the measurement is a random variable N taking values in $\{1, \ldots, d\}$ with p.m.f.:

$$\mathbb{P}(N = i) = \langle \rho, E_i \rangle.$$

Note that if one knows an orthonormal basis of eigenvectors (i.e. the quantum states), then a basis measurement with this basis has p.m.f. $p_1, \ldots, p_d$, therefore estimating the p_i's this way amounts to classical statistics.

In quantum statistics, after each measurement the state collapses, so one needs n independent copies of the quantum system to complete n independent measurements. But instead of making n measurements one after the other, it is actually best to see the n copies as one single quantum system with density matrix $\rho^{\otimes n}$, and then make one single measurement, called entangled measurement, to estimate ρ. Out of all the measurements that one may perform on $\rho^{\otimes n}$, it turns out that one of them is optimal: a projective measurement called weak Schur sampling. By "optimal", we mean optimal to compute any property of the spectrum $p_1, \ldots, p_d$, in the sense of [76, Theorem 2.6.3]: If we have an algorithm for computing the property which has failure rate β on any density matrix, then there is a similar algorithm for doing so using only weak Schur sampling followed by classical post-processing. Therefore, in this work, we will focus on this measurement of $\rho^{\otimes n}$. The family of self-adjoint projectors of this projective measurement is given by the Schur-Weyl duality Theorem, and is indexed by the partitions of n. Therefore, the outcome of this measurement, the weak Schur sampling, is a random partition $\lambda \vdash n$, and it is well known (see e.g. [76]) that its law is $\mathrm{SW}^n(p)$. In other words, one may see the measurement λ as the shape of the RSK algorithm applied to a random word of length n and letters drawn with p.m.f. p. The problem is to find a good estimator of the p.m.f. p, given λ.

As previously seen, the limiting distributions of $\lambda \sim \mathrm{SW}^n(p)$ are well known, but in the first part of Chapter 3 we will revisit the rates of convergence, because of the need for non-asymptotic estimation results. Then, we will improve two results connected with the estimation of p, the first one is a bound on the sum of the variances of λ, the second one is a bound on the "excess" of λ.

Chapter 1

The Limiting Law of the Length of the Longest Common and Increasing Subsequences in Random Words

This chapter is taken from our publication [18]. We sincerely thank an Associate Editor and a referee for their detailed readings and numerous comments which greatly helped to improve this manuscript.

Let $(X_k)_{k\geq 1}$ and $(Y_k)_{k\geq 1}$ be two independent sequences of i.i.d. random variables, with values in a finite and totally ordered alphabet $\mathcal{A}_m := \{1, \ldots, m\}$, $m \geq 2$, having respective probability mass function $p_1^X, \ldots, p_m^X$ and $p_1^Y, \ldots, p_m^Y$. Let LCI_n be the length of the longest common and weakly increasing subsequences in $X_1, ..., X_n$ and $Y_1, ..., Y_n$. Once properly centered and normalized, LCI_n is shown to have a limiting distribution which is expressed as a functional of two independent multidimensional Brownian motions.

1.1 Introduction and preliminary results

1.1.1 Introduction

We analyze the asymptotic behavior of LCI_n, the length of the longest common subsequences in random words with an additional weakly increasing requirement. Throughout, $(X_k)_{k\geq 1}$ and $(Y_k)_{k\geq 1}$ are two independent sequences of i.i.d. random variables with values in the finite totally ordered alphabet $\mathcal{A}_m := \{1, \ldots, m\}$, $m \geq 2$, and respective p.m.f. $p_1^X, \ldots, p_m^X$, $p_i^X > 0$, $i = 1, \ldots, m$ and $p_1^Y, \ldots, p_m^Y$, $p_i^Y > 0$, $i = 1, \ldots, m$. Next, LCI_n, the length of the longest common and weakly increasing subsequences of the two random words $X_1 \cdots X_n$ and $Y_1 \cdots Y_n$, is the largest integer $r \in \{1, \ldots, n\}$ such that there exist $1 \leq i_1 < \cdots < i_r \leq n$ and $1 \leq j_1 < \cdots < j_r \leq n$ such that

- $\forall s \in \{1, \ldots, r\}$, $X_{i_s} = Y_{j_s}$,
- $X_{i_1} \leq X_{i_2} \leq \cdots \leq X_{i_r}$ and $Y_{j_1} \leq Y_{j_2} \leq \cdots \leq Y_{j_r}$,

and if no integer satisfies these two conditions, we set $LCI_n = 0$.

A thorough discussion of the study of LCI_n, with potential applications, and a more complete bibliography, is present in [12], where the following is further proved (below, as usual, $\wedge$ is short

for minimum):

Theorem 1.1.1. *Let X_k and Y_k ($k = 1, 2, \dots$) be uniformly distributed over $\{1, \dots, m\}$. Then,*

$$\frac{LCI_n - n/m}{\sqrt{n/m}} \overset{\Longrightarrow}{n\to\infty} \max_{0=t_0\le t_1\le\cdots\le t_m=1} \left[\left(-\frac{1}{m}\sum_{i=1}^{m} B_i^X(1) + \sum_{i=1}^{m} \left(B_i^X(t_i) - B_i^X(t_{i-1})\right) \right) \wedge \left(-\frac{1}{m}\sum_{i=1}^{m} B_i^Y(1) + \sum_{i=1}^{m} \left(B_i^Y(t_i) - B_i^Y(t_{i-1})\right) \right) \right],$$

where B^X and B^Y are two independent m-dimensional standard Brownian motions on $[0, 1]$.

The results of [12] extended (and corrected) the proof of the case $m = 2$ analyzed in [31] and also conjectured the following generalization:

Theorem 1.1.2. *Let X_k and Y_k ($k = 1, 2, \dots$) have the same distribution, let $p_{\max} = \max_{i\in\{1,\dots,m\}} p_i^X$ and let k^* be its multiplicity. Then*

$$\frac{LCI_n - np_{\max}}{\sqrt{np_{\max}}} \overset{\Longrightarrow}{n\to\infty} \max_{0=t_0\le t_1\le\cdots\le t_{k^*}=1} \left[\left(\frac{\sqrt{1-k^*p_{\max}}-1}{k^*}\sum_{i=1}^{k^*} B_i^X(1) + \sum_{i=1}^{k^*} \left(B_i^X(t_i) - B_i^X(t_{i-1})\right) \right) \wedge \left(\frac{\sqrt{1-k^*p_{\max}}-1}{k^*}\sum_{i=1}^{k^*} B_i^Y(1) + \sum_{i=1}^{k^*} \left(B_i^Y(t_i) - B_i^Y(t_{i-1})\right) \right) \right],$$

where B^X and B^Y are two independent k^*-dimensional standard Brownian motions on $[0, 1]$.

Clearly, in case $k^* = m$, the two limiting distributions in (1.1.1) and (1.1.2) are the same but they differ otherwise. Indeed, (1.1.1) involves two independent m-dimensional Brownian motions while (1.1.2) involves k^*-dimensional ones. So, in particular, if $k^* = 1$, then the right-hand side of (1.2) is just the minimum of two independent centered normal random variables. In view of the results obtained in the one-sequence case, e.g., see [32], [6], and the many references therein, it is tantalizing to conjecture that both the right-hand side of (1.1.1) and of (1.1.2) can be realized as maximal eigenvalues of some Gaussian random matrix models.

Below, we aim to obtain the limiting distribution of LCI_n, without assuming that the X_k and Y_k ($k = 1, 2, \dots$) have the same distribution; providing also an alternative proof of Theorem 1.1.1 as well as a proof of the conjectured (1.1.2). A brief description of the content of our notes is as follows: the rest of the current section is devoted to studying the asymptotic mean of LCI_n. This asymptotic mean result is already not so predictable and allows for the proper centering in the limiting theorem whose proof is provided in the next section. The third and final section is mainly devoted to studying extensions and complements, such as results for sequences with blocks and infinite countable alphabets.

1.1.2 Probability

For $i \in \{1, \dots, m\}$ and $j \in \{1, \dots, n\}$, let $\ell \in \mathbb{N} = \{0, 1, 2, \dots\}$ be such that $j + \ell \le n + 1$, and let

$$N_{j,\ell}^{X,i} = \sum_{k=0}^{\ell-1} \mathbb{1}_{X_{j+k}=i} \qquad \left(\text{resp. } N_{j,\ell}^{Y,i} = \sum_{k=0}^{\ell-1} \mathbb{1}_{Y_{j+k}=i}\right),$$

be simply the number of letters i between, and including, j and $j + \ell - 1$ in $X_1, ..., X_n$ (resp. $Y_1, ..., Y_n$), with the convention that the sum is zero in case $\ell = 0$. From the very definition of LCI_n, it is clear that

$$LCI_n = \max_{\substack{\ell^X,\ell^Y\in\mathbb{N}^m \\ \ell_1^X+\cdots+\ell_m^X=n \\ \ell_1^Y+\cdots+\ell_m^Y=n}} \left(N_{1,\ell_1^X}^{X,1} \wedge N_{1,\ell_1^Y}^{Y,1} + N_{\ell_1^X,\ell_2^X}^{X,2} \wedge N_{\ell_1^Y,\ell_2^Y}^{Y,2} + \cdots + N_{\ell_1^X+\cdots+\ell_{m-1}^X,\ell_m^X}^{X,m} \wedge N_{\ell_1^Y+\cdots+\ell_{m-1}^Y,\ell_m^Y}^{Y,m} \right).$$

Next, let $\Lambda = \{\lambda \in (\mathbb{R}_+)^m = [0,+\infty)^m \,:\, \lambda_1 + \cdots + \lambda_m = 1\}$. For $\lambda \in \Lambda$, let

$$\ell^n(\lambda)_i = \lfloor (\lambda_1 + \cdots + \lambda_i)n \rfloor - \lfloor (\lambda_1 + \cdots + \lambda_{i-1})n \rfloor,$$

where $\lfloor . \rfloor$ is the usual integer part, aka the floor, function. When λ runs through Λ, $\ell^n(\lambda) = (\ell^n(\lambda)_1, \ldots, \ell^n(\lambda)_m)$ runs exactly through $\{\ell \in \mathbb{N}^m \,:\, \ell_1 + \cdots + \ell_m = n\}$, so

$$LCI_n = \max_{\lambda^X, \lambda^Y \in \Lambda} \left(N^{X,1}_{1,\ell^n(\lambda^X)_1} \wedge N^{Y,1}_{1,\ell^n(\lambda^Y)_1} + N^{X,2}_{\ell^n(\lambda^X)_1,\ell^n(\lambda^X)_2} \wedge N^{Y,2}_{\ell^n(\lambda^Y)_1,\ell^n(\lambda^Y)_2} + \cdots \right.$$
$$\left. + N^{X,m}_{\ell^n(\lambda^X)_1+\cdots+\ell^n(\lambda^X)_{m-1},\ell^n(\lambda^X)_m} \wedge N^{Y,m}_{\ell^n(\lambda^Y)_1+\cdots+\ell^n(\lambda^Y)_{m-1},\ell^n(\lambda^Y)_m} \right).$$

For ease of notations, throughout the paper, for all $x \in (\mathbb{R}^m)^2$, we write $x = (x^X, x^Y)$ so, for example, above, $\lambda^X, \lambda^Y \in \Lambda$ becomes $\lambda \in \Lambda^2$.

For $i \in \{1, \ldots, m\}$ and $t \in [0,1]$, let now

$$\widetilde{B}^{n,X}_i(t) = \frac{N^{X,i}_{1,\lfloor tn \rfloor} - p^X_i tn}{\sqrt{p^X_i(1-p^X_i)n}}, \qquad \left(\text{resp. } \widetilde{B}^{n,Y}_i(t) = \frac{N^{Y,i}_{1,\lfloor tn \rfloor} - p^Y_i tn}{\sqrt{p^Y_i(1-p^Y_i)n}} \right),$$

and for $\lambda \in \Lambda^2$, let

$$\widetilde{V}^{n,X}_i(\lambda^X) = \sqrt{p^X_i(1-p^X_i)} \left(\widetilde{B}^{n,X}_i(\lambda^X_1 + \cdots + \lambda^X_i) - \widetilde{B}^{n,X}_i(\lambda^X_1 + \cdots + \lambda^X_{i-1}) \right),$$
$$\widetilde{V}^{n,Y}_i(\lambda^Y) = \sqrt{p^Y_i(1-p^Y_i)} \left(\widetilde{B}^{n,Y}_i(\lambda^Y_1 + \cdots + \lambda^Y_i) - \widetilde{B}^{n,Y}_i(\lambda^Y_1 + \cdots + \lambda^Y_{i-1}) \right),$$

so that (1.1.2) becomes

$$LCI_n = \max_{\lambda \in \Lambda^2} \sum_{i=1}^m \left[\left(np^X_i \lambda^X_i + \sqrt{n} \widetilde{V}^{n,X}_i(\lambda^X) \right) \wedge \left(np^Y_i \lambda^Y_i + \sqrt{n} \widetilde{V}^{n,Y}_i(\lambda^Y) \right) \right].$$

The above identity provides a representation of LCI_n as a maximum over the locations, $\lambda \in \Lambda^2$, where to pick in each word $X_1, \ldots, X_n$ and $Y_1, \ldots, Y_n$, the letters $1, 2, \ldots, m$ in order to form a common sub-word. This is different from the approach in [12], where the maximum is over the numbers of letters $1, 2, \ldots, m$ in a common sub-word. Of course the two representations are equivalent. However, the advantage of our approach is that λ takes its values in a deterministic set, as opposed to a random set.

In order to keep dealing with maxima it will be convenient to replace $\widetilde{B}^n_i$ in (1.1.2) by its continuous alternative: for $i \in \{1, \ldots, m\}$ and $t \in [0,1]$, let

$$B^{n,X}_i(t) = \frac{N^{X,i}_{1,\lfloor tn \rfloor} + (tn - \lfloor tn \rfloor) \mathbb{1}_{X_{\lfloor tn \rfloor + 1} = i} - p^X_i tn}{\sqrt{p^X_i(1-p^X_i)n}}$$

and

$$B^{n,Y}_i(t) = \frac{N^{Y,i}_{1,\lfloor tn \rfloor} + (tn - \lfloor tn \rfloor) \mathbb{1}_{Y_{\lfloor tn \rfloor + 1} = i} - p^Y_i tn}{\sqrt{p^Y_i(1-p^Y_i)n}}.$$

Next define $V^{n,X}$, $V^{n,Y}$ just as in (1.1.2) and (1.1.2), replacing $\widetilde{B}$ by B, and let

$$LCI^c_n = \max_{\lambda \in \Lambda^2} \sum_{i=1}^m \left[\left(np^X_i \lambda^X_i + \sqrt{n} V^{n,X}_i(\lambda) \right) \wedge \left(np^Y_i \lambda^Y + \sqrt{n} V^{n,Y}_i(\lambda) \right) \right].$$

Our analysis rests upon estimating the variations of $B_i^{n,X}$ and of $B_i^{n,Y}$. To do so, let $\eta \in (0, 1/6)$ and let A_n^η be the event:

$$\forall i \in \{1,\dots,m\}, \forall j \in \{1,\dots,n\}, \forall \ell \in \{0,\dots,n+1-j\}, \left|\frac{N_{j,\ell}^{X,i} - p_i^X \ell}{\sqrt{n}}\right| \leq \frac{n^\eta}{2}\sqrt{\frac{\ell}{n}},$$
$$\text{and} \quad \left|\frac{N_{j,\ell}^{Y,i} - p_i^Y \ell}{\sqrt{n}}\right| \leq \frac{n^\eta}{2}\sqrt{\frac{\ell}{n}}.$$

By Hoeffding's inequality,

$$1 - \mathbb{P}\left(A_n^\eta\right) \leq 2n(n+1)m \exp\left(-\frac{n^{2\eta}}{2}\right), \tag{1.1.1}$$

and so if A_n^η occurs, then for all x, y in $[0,1]$ and $i \in \{1,\dots,m\}$,

$$\left|\sqrt{p_i^X(1-p_i^X)}\left(B_i^{n,X}(y) - B_i^{n,X}(x)\right)\right| \leq \frac{n^\eta}{2}\sqrt{|y-x| + \frac{1}{n}}.$$

and in particular,

$$\left|\sqrt{p_i^X(1-p_i^X)}\left(B_i^{n,X}(y) - B_i^{n,X}(x)\right)\right| \leq \frac{n^\eta}{2}\sqrt{|y-x|} + \frac{n^{\eta-1/2}}{2} \leq n^\eta,$$

and the same applies to Y instead of X.

1.1.3 Asymptotic mean: distinct cases

Let us investigate the limiting behavior of LCI_n/n. From (1.1.2),

$$\frac{LCI_n}{n} = \max_{\lambda \in \Lambda^2} \sum_{i=1}^m \left[\left(p_i^X \lambda_i^X + \frac{\widetilde{V}_i^{n,X}(\lambda^X)}{\sqrt{n}}\right) \wedge \left(p_i^Y \lambda_i^Y + \frac{\widetilde{V}_i^{n,Y}(\lambda^Y)}{\sqrt{n}}\right)\right].$$

Note that $|\widetilde{V}_i^{n,X}(\lambda^X) - V_i^{n,X}(\lambda^X)| \leq 1/\sqrt{n}$ (and similarly for Y). Thus, using (throughout the paper) the following elementary inequality, valid for any $a, b, c, d \in \mathbb{R}$,

$$|a \wedge b - (a+c) \wedge (b+d)| \leq \max(|c|, |d|), \tag{1.1.2}$$

we get

$$\left|\frac{LCI_n}{n} - \frac{LCI_n^c}{n}\right| \leq \frac{m}{n}.$$

Moreover, if A_n^η occurs, then for all $\lambda \in \Lambda^2$,

$$\left|\sum_{i=1}^m \left[\left(p_i^X \lambda_i^X + \frac{V_i^{n,X}(\lambda^X)}{\sqrt{n}}\right) \wedge \left(p_i^Y \lambda_i^Y + \frac{V_i^{n,Y}(\lambda^Y)}{\sqrt{n}}\right)\right] - \sum_{i=1}^m \left[\left(p_i^X \lambda_i^X\right) \wedge \left(p_i^Y \lambda_i^Y\right)\right]\right| \leq \frac{m}{n^{1/2-\eta}},$$

so, letting $f : \left(\mathbb{R}^m\right)^2 \to \mathbb{R}$ be given via

$$f : (y^X, y^Y) \mapsto \sum_{i=1}^m \left[\left(p_i^X y_i^X\right) \wedge \left(p_i^Y y_i^Y\right)\right], \tag{1.1.3}$$

we have:

$$\left|\frac{LCI_n}{n} - \max_{\lambda \in \Lambda^2} f(\lambda)\right| \leq m n^{\eta - 1/2}.$$

By the Borel-Cantelli lemma (recalling (1.1.1)), almost surely, eventually A_n^η occurs so LCI_n^c/n and LCI_n/n both converge almost surely to

$$e_{\max} := \max_{\lambda \in \Lambda^2} f(\lambda). \tag{1.1.4}$$

From

$$\frac{LCI_n}{n} \xrightarrow[n\to\infty]{} e_{max}, \text{a.s.},$$

we also get by dominated convergence

$$\frac{\mathbb{E}LCI_n}{n} \xrightarrow[n\to\infty]{} e_{max}.$$

One can think of $e_{\max}$ as the length ratio of the longest common and increasing subsequences in a continuous, non-probabilistic setup: the letters have density masses $p_1^X, p_2^X, \ldots, p_m^X$ and $p_1^Y, p_2^Y, \ldots, p_m^Y$.

Now, let

$$U = \left\{ u \in (\mathbb{R}_+)^m \, : \, \frac{u_1}{p_1^X} + \cdots + \frac{u_m}{p_m^X} \le 1, \frac{u_1}{p_1^Y} + \cdots + \frac{u_m}{p_m^Y} \le 1 \right\},$$

and let $\phi : \mathbb{R}^m \to \mathbb{R}$ be given by $\phi : u \mapsto u_1 + \cdots + u_m$.

On U, there is a correspondence between f in (1.1.3), and the above ϕ. Indeed, for $\lambda \in \Lambda^2$, defining u by $u_i = \left(p_i^X \lambda_i^X\right) \wedge \left(p_i^Y \lambda_i^Y\right)$, $f(\lambda) = \phi(u)$, and for $u \in U$, there exists $\lambda \in \Lambda^2$, such that $\lambda_i^X \ge u_i/p_i^X$ and $\lambda_i^Y \ge u_i/p_i^Y$ so that $f(\lambda) \ge \phi(u)$. Therefore, $e_{\max} = \max_{u\in U} \phi(u)$. Also, let

$$K_{\Lambda^2} = f^{-1}\left(\{e_{\max}\}\right) \cap \Lambda^2, \text{ and } L_U = \phi^{-1}\left(\{e_{\max}\}\right) \cap U.$$

The above correspondence provides for each element of K_{Λ^2} an element of L_U, and for each element of L_U at least one element of K_{Λ^2} (if one of the two inequalities defining U is strict, then there is more than one way to define the corresponding λ). Next, let I be the set of integers $i \in \{1, \ldots, m\}$ such that there exists $u^i \in L_U$ with $u_i^i > 0$. One can think of I as the letters that can be used to maximize ϕ, or, equivalently, to maximize f. Let

$$u^I = \frac{1}{|I|} \sum_{i \in I} u^i, \tag{1.1.5}$$

so $u^I \in L_U$ and for all $i \in I$, $u_i^I > 0$. Thanks to the above correspondence, we define (and will use throughout the paper) $a \in \Lambda^2$ such that $a_i^X = a_i^Y = 0$ for all $i \notin I$ and $a_i^X \ge u_i^I/p_i^X$, $a_i^Y \ge u_i^I/p_i^Y$, for all $i \in I$ (a is a correspondent of u_I). Since $f(a) \ge \phi(u^I) = e_{\max}$, $a \in K_{\Lambda^2}$. We shall see, and use, that when restricting the alphabet to I, asymptotically (when properly centered and normalized) the distribution of LCI_n remains unchanged.

Two distinct cases need to be analyzed in order to study the limiting distribution of LCI_n.

Case a) There exists $u \in L_U$ such that $\frac{u_1}{p_1^X} + \cdots + \frac{u_m}{p_m^X} = 1$ and $\frac{u_1}{p_1^Y} + \cdots + \frac{u_m}{p_m^Y} < 1$.

For example, when $p^X = (3/8, 3/8, 1/4)$ and $p^Y = (1/2, 3/8, 1/8)$. Here the maximum is $3/8$, and $I = \{1, 2\}$.

Heuristically, this case indicates that the length of the common words is limited by the word $X_1 \cdots X_n$ and not by $Y_1 \cdots Y_n$. Using the correspondence between L_U and K_{Λ^2}, this case is equivalent to the following statement: there exists $\lambda \in K_{\Lambda^2}$ such that for all $i \in \{1, \ldots, m\}, p_i^X \lambda_i^X \le p_i^Y \lambda_i^Y$ with at least one strict inequality. In this case, one has:

Lemma 1.1.3. *Let $p^X_{\max} = \max_{i\in\{1,\dots,m\}} p^X_i$. Then $I = \{i \in \{1,\dots,m\} \,:\, p^X_i = p^X_{\max}\}$ and $e_{\max} = p^X_{\max}$. Moreover there exists $i_1 \in I$ such that $p^Y_{i_1} > p^X_{\max}$.*

Proof. Let $i, j \in \{1,\dots,m\}$ be such that $p^X_i < p^X_j$, and assume, by contradiction, that $i \in I$. Let $u \in L_U$ satisfying $\frac{u_1}{p^X_1} + \cdots + \frac{u_m}{p^X_m} = 1$ and $\frac{u_1}{p^Y_1} + \cdots + \frac{u_m}{p^Y_m} < 1$, and let $v = (u^i + u)/2$, so that $v \in U$, $v_i > 0$, $\frac{v_1}{p^X_1} + \cdots + \frac{v_m}{p^X_m} \le 1$ and $\frac{v_1}{p^Y_1} + \cdots + \frac{v_m}{p^Y_m} < 1$. Let, for $\varepsilon > 0$, $v(\varepsilon)$ be the vector v except at the coordinates i and j where $v(\varepsilon)_i := v_i - \varepsilon p^X_i$ and $v(\varepsilon)_j := v_j + \varepsilon p^X_j$. It is clear that, when ε is small enough, $v(\varepsilon) \in U$ and $\phi(v(\varepsilon)) = e_{\max} + \varepsilon(p^X_j - p^X_i) > e_{\max}$, leading to a contradiction. Hence $I \subset \{i \in \{1,\dots,m\} \,:\, p^X_i = p^X_{\max}\}$. Reciprocally, let $i \in \{1,\dots,m\}$ be such that $p^X_i = p^X_{\max}$ and let $j \in I$. If $i = j$ we are done. Otherwise, one can slightly change u by adding ε to the ith coordinate and subtracting ε to the jth coordinate so that $\phi(u)$ remains unchanged, and u is still in U (for ε small enough), so $I = \{i \in \{1,\dots,m\} \,:\, p^X_i = p^X_{\max}\}$.

Since $\frac{u_1}{p^X_1} + \cdots + \frac{u_m}{p^X_m} = \sum_{i\in I} \frac{u_i}{p^X_{\max}} > \sum_{i\in I} \frac{u_i}{p^Y_i}$, there exists $i_1 \in I$ such that $p^Y_{i_1} > p^X_{\max}$. It is finally clear that $e_{\max} = p^X_{\max}$, completing the proof. □

As a consequence of the above lemma, we prove next that

$$J := \left\{ \lambda^X \in \Lambda \,:\, \forall i \notin I,\ \lambda^X_i = 0,\ \sum_{i\in I} \frac{\lambda^X_i}{p^Y_i} \le \frac{1}{p^X_{\max}} \right\} = \left\{ \lambda^X \,:\, \lambda \in K_{\Lambda^2} \right\}, \tag{1.1.6}$$

(in particular, this set is non-empty which is all that is really needed in the rest of the proof). To show this equality, first note that $\left\{ \lambda^X \,:\, \lambda \in K_{\Lambda^2} \right\} \subset J$ since, indeed, when $\lambda \in K_{\Lambda^2}$, for every $i \in I$, $p^X_{\max}\lambda^X_i \le p^Y_i \lambda^Y_i$ and then take the sum. Conversely, if $\lambda^X \in J$, $\sum_{i\in I} p^X_{\max}\lambda^X_i/p^Y_i \le 1$, so let λ^Y be such that for every $i \in I$, $\lambda^Y_i \ge p^X_{\max}\lambda^X_i/p^Y_i$ and $\sum_{i\in I} \lambda^Y_i = 1$, while for $i \in I^c$, let $\lambda^Y_i = 0$. Clearly, $\lambda \in K_{\Lambda^2}$.

Case b) For all $u \in L_U$, $\frac{u_1}{p^X_1} + \cdots + \frac{u_m}{p^X_m} = \frac{u_1}{p^Y_1} + \cdots + \frac{u_m}{p^Y_m} = 1$.

Heuristically, this second case indicates that in order to form the longest common words, it is necessary to make full use of both words. Using the correspondence between L_U and K_{Λ^2}, this case is equivalent to the following: for all $\lambda \in K_{\Lambda^2}$, for all $i \in \{1,\dots,m\}, p^X_i\lambda^X_i = p^Y_i\lambda^Y_i$. We can further distinguish two subcases, namely, we are in Case b1) if each coordinate of $P^X := \left(1/p^X_i\right)_{i\in I} \in \mathbb{R}^I$ is equal to each coordinate of $P^Y = \left(1/p^Y_i\right)_{i\in I} \in \mathbb{R}^I$, and in Case b2) otherwise.

For example, if $p^X = (1/3, 1/3, 2/9, 1/9)$ and $p^Y = (1/3, 1/3, 1/9, 2/9)$, we are in Case b1) and $e_{\max} = 1/3$. If $p^X = (2/3, 1/6, 1/6)$ and $p^Y = (1/6, 2/3, 1/6)$, we are in Case b2) and $e_{\max} = 4/15$. In both of these examples, $I = \{1, 2\}$.

Below $\mathrm{Span}(P^X)$ (resp. $\mathrm{Span}(P^Y)$) is the linear span of P^X (resp. P^Y).

Lemma 1.1.4. *In Case b2), there exists a unique pair of reals s, t such that $sP^X + tP^Y = (1)_{i\in I}$*

Proof. The only alternatives to Case b1) are: P^X and P^Y are linearly independent, or P^X and P^Y are linearly dependent and $P^X \neq P^Y$. If the latter, given that P^X and P^Y have positive coordinates, $P^X < P^Y$ (coordinate by coordinate) or $P^Y < P^X$. But $P^X < P^Y$ clearly implies that Case a) occurs, and not Case b) leading to a contradiction (and similarly $P^Y < P^X$). Therefore, the only alternative to Case b1) is for P^X and P^Y to be linearly independent. We now prove that $H := (1)_{i\in I} \in \mathrm{Span}(P^X, P^Y)$. To do so, we use an elementary duality result: if E is a finite-dimensional space with dual E^*, and if $l_1, l_2, l_3 \in E^*$, then $\mathrm{Ker}(l_1) \cap \mathrm{Ker}(l_2) \subset \mathrm{Ker}(l_3)$ if and only if $l_3 \in \mathrm{Span}(l_1, l_2)$. Indeed, considering the restrictions $l_{2|\mathrm{Ker}(l_1)}$ and $l_{3|\mathrm{Ker}(l_1)}$ of l_2 and l_3 to the subspace $\mathrm{Ker}(l_1)$, we have $\mathrm{Ker}(l_{2|\mathrm{Ker}(l_1)}) \subset \mathrm{Ker}(l_{3|\mathrm{Ker}(l_1)})$. Therefore, $l_{3|\mathrm{Ker}(l_1)} = \lambda l_{2|\mathrm{Ker}(l_1)}$ for some $\lambda \in \mathbb{R}$, and if $u \notin \mathrm{Ker}(l_1)$, then $l_3 = \lambda l_2 + \frac{l_3(u) - \lambda l_2(u)}{l_1(u)} l_1$ (because this is true on $\mathrm{Ker}(l_1)$ and

on u). So, returning to our problem, $H \in \mathrm{Span}(P^X, P^Y)$ is equivalent to: $\mathrm{Ker}(P^{X^*}) \cap \mathrm{Ker}(P^{Y^*}) \subset \mathrm{Ker}(H^*)$, where for any $L \in \mathbb{R}^I$, L^* denotes the linear form defined by $L^*(y) = L \cdot y$. Let $x \in \mathrm{Ker}((P^X)^*) \cap \mathrm{Ker}((P^Y)^*)$. Clearly, there exists $\varepsilon > 0$ such that $u^I + \varepsilon x$ and $u^I - \varepsilon x$ have non-negative coordinates, and so they are in L_U, and $H^*(u^I + \varepsilon x) = H^*(u^I - \varepsilon x) = e_{\max}$ otherwise one of them would be greater than $e_{\max}$, hence $x \in \mathrm{Ker}(H^*)$. □

For instance, taking again $p^X = (2/3, 1/6, 1/6)$ and $p^Y = (1/6, 2/3, 1/6)$, we get $P^X = (3/2, 6), P^Y = (6, 3/2)$ and $s = t = 2/15$.

Without loss of generality (switching the roles of X and Y), one can thus assume that either Case a) or Case b) occurs.

In Case b), the following technical lemma, whose proof (given in the Appendix) is not crucial to understand the rest of this manuscript, is needed to state our main theorem. Let us define first, in Case b1),

$$s_X := \begin{cases} \max_{i \in I^c : p_i^X \geq e_{\max}} \frac{p_i^Y (p_i^X - e_{\max})}{e_{\max}(p_i^X - p_i^Y)} & \text{if } \{i \in I^c, p_i^X \geq e_{\max}\} \neq \emptyset, \\ 0, & \text{if } \{i \in I^c, p_i^X \geq e_{\max}\} = \emptyset, \end{cases} \quad t_X := 1 - s_X,$$

and, similarly,

$$s_Y := \begin{cases} \max_{i \in I^c, p_i^Y \geq e_{\max}} \frac{p_i^X (p_i^Y - e_{\max})}{e_{\max}(p_i^Y - p_i^X)}, & \text{if } \{i \in I^c : p_i^Y \geq e_{\max}\} \neq \emptyset, \\ 0, & \text{if } \{i \in I^c, p_i^Y \geq e_{\max}\} = \emptyset, \end{cases} \quad t_Y := 1 - s_Y.$$

It is clear, from the definition of I, that if $i \in I$ is such that $p_i^X \geq e_{\max}$, then $p_i^Y < e_{\max}$, therefore s_X and s_Y are well defined and one can check that $s_X, t_X, s_Y, t_Y \in [0, 1]$.

In order to state our next lemma, below let $E = \{x \in \mathbb{R}^m : x_1 + \cdots + x_m = 0\}$ and let $E' = \{x \in E : \forall i \in I^c, x_i \geq 0\}$.

Lemma 1.1.5. *Let $\nu \in (\mathbb{R}^m)^2$ be such that for all $i \in I^c, \nu_i^X = \nu_i^Y = 0$, then the following maximum is well defined:*

$$\mathfrak{m}(\nu) := \max_{x \in E'^2} \sum_{i=1}^m \left[\left(p_i^X x_i^X + \nu_i^X\right) \wedge \left(p_i^Y x_i^Y + \nu_i^Y\right) \right], \tag{1.1.7}$$

and

$$\mathfrak{m}(\nu) = \max_{\substack{x \in E'^2 \\ \|x\|_\infty \leq 2Cm\|\nu\|_\infty}} \sum_{i=1}^m \left[\left(p_i^X x_i^X + \nu_i^X\right) \wedge \left(p_i^Y x_i^Y + \nu_i^Y\right) \right], \tag{1.1.8}$$

for some constant $C > 0$, depending only on p^X and p^Y, as given in Lemma 1.2.3. In Case b1), writing $S^\bullet := \sum_{i \in I} \nu_i^\bullet$, then

$$\mathfrak{m}(\nu) = \begin{cases} s_X S^Y + t_X S^X, & if S^X \leq S^Y, \\ s_Y S^X + t_Y S^Y, & if S^X \geq S^Y. \end{cases}.$$

In Case b2), and recalling the notations of Lemma 1.1.4, then

$$\mathfrak{m}(\nu) = \sum_{i \in I} \left(\frac{s}{p_i^X} \nu_i^X + \frac{t}{p_i^Y} \nu_i^Y \right).$$

1.1.4 Representation of $e_{\max}$

We now aim to give a more explicit expression for e_{max} defined by (1.1.4). To do so, let us start with the following lemma which asserts that, in the non-probabilistic setup, "two letters are enough to reach the maximum".

Lemma 1.1.6. *There exist $i, j \in \{1, \dots, m\}$ and $\lambda \in K_{\Lambda^2}$ such that for all $k \notin \{i, j\}, \lambda_k^X = \lambda_k^Y = 0$.*

Proof. Let $u \in L_U$ having (at least) three non-zero coordinates. Then, recalling the correspondence between L_U and K_{Λ^2}, in order to prove the result it is enough to show that there exists a $v \in L_U$ having one less null coordinate. Without loss of generality, let $u_1, u_2, u_3 > 0$, and let

$$V = \left\{ x \in \mathbb{R}^m \, : \, \sum_{i=1}^m \frac{x_i}{p_i^X} = \sum_{i=1}^m \frac{x_i}{p_i^Y} = 0, x_4 = \cdots = x_n = 0 \right\}.$$

Since the dimension of V is at least one, let $x \in V \setminus \{0\}$. Then clearly, there exists $t \in \mathbb{R}$ such that $v := u + tx$ has non-negative coordinates and one more null coordinate than u. Moreover, $v \in L_U$, which completes the proof. □

If there exists $u \in L_U$ such all its coordinates except one, call it i, are zeros, then $e_{\max} = p_i^X \wedge p_i^Y$. Otherwise, let i, j be defined as in the statement of the lemma. At first, assume that $p_i^X = p_j^X$ and that $p_i^Y \leq p_j^Y$, then $e_{\max} \leq (\lambda_i^X p_i^X \wedge \lambda_i^Y p_j^Y) + (\lambda_j^X p_i^X \wedge \lambda_j^Y p_j^Y) \leq (\lambda_i^X p_i^X + \lambda_j^X p_i^X) \wedge (\lambda_i^Y p_j^Y + \lambda_j^Y p_j^Y) = p_i^X \wedge p_i^Y$, so $e_{\max} = p_i^X \wedge p_i^Y$ and we are actually in the first case, giving a contradiction. Similarly, if $p_i^X \leq p_j^X$ and $p_i^Y \leq p_j^Y$, using $\lambda_i^X p_i^X \wedge \lambda_i^Y p_i^Y \leq \lambda_i^X p_j^X \wedge \lambda_i^Y p_j^Y$ we get a contradiction as well. Therefore, in the second case, necessarily, possibly permuting i and j, $p_i^X < p_j^X$ and $p_i^Y > p_j^Y$. Additionally, it is necessary to have that $p_i^X < p_i^Y$, otherwise $e_{\max} = p_i^Y$ and we are in the first case. Similarly, $p_j^Y < p_j^X$. Then, in this case, the maximum is when the quantities in each minima are equal, and so one shows that

$$e_{\max} = e(i,j) := \frac{p_i^X p_i^Y (p_j^X - p_j^Y) + p_j^X p_j^Y (p_i^Y - p_i^X)}{p_i^Y p_j^X - p_i^X p_j^Y}.$$

Therefore,

$$e_{\max} = \max\Big(\max_{1 \leq i \leq m} \left(p_i^X \wedge p_i^Y\right), \max_{\substack{i,j \, : \, p_i^X < p_j^X \\ \wedge \quad \vee \\ p_i^Y > p_j^Y}} e(i,j) \Big). \tag{1.1.9}$$

Note that

$$\max_{1 \leq i \leq m} \left(p_i^X \wedge p_i^Y\right) \leq e_{\max} \leq \left(\max_{1 \leq i \leq m} p_i^X \right) \wedge \left(\max_{1 \leq i \leq m} p_i^Y \right),$$

where the left inequality is clear, while the right one is easily seen from the expression of f. Note also that above, e_{max} is equal to the lower bound when the second max in (1.1.9) is over the empty set, and is equal to the upper bound when there exists i such that $p_{\max}^X = p_i^X \leq p_i^Y$ or $p_{\max}^Y = p_i^Y \leq p_i^X$.

When $p^X = p^Y$ (same distribution for each word), we see that $e_{\max} = \max_{i \in \{1,\dots,m\}} p_i^X$ is minimal when p^X is uniform (for a given alphabet). This is to be contrasted with the case of the length of the longest common subsequences, LC_n (defined just as LCI_n, but without the increasing condition). Indeed, little is known about $\gamma^* := \lim_{n \to +\infty} \mathbb{E}LC_n/n$, for instance whether or not it is minimal (for a given alphabet) for the uniform distribution. Since LC_n is defined with one less constraint than LCI_n, clearly $e_{\max} \leq \gamma^*$ which is of potential interest since the exact value of γ^* is unknown, even in the uniform binary case. (This last inequality provides a lower bound on γ^*, no matter the distributions on the letters. For uniform letters, $e_{\max} = 1/m$, although it is known that, then, asymptotically, $\gamma^* \sim 2/\sqrt{m}$, see [46].)

1.1.5 A criterion to distinguish the three cases

For a given distribution, it is not completely apparent which situation is in play as far as the respective cases a), b1) and b2) are concerned. Our next result makes this more transparent. First, set

$$e_1 = \max_{1\le i\le m} \left(p_i^X \wedge p_i^Y\right), \qquad e_2 = \max_{\substack{i,j\,:\,p_i^X < p_j^X \\ \wedge \quad \vee \\ p_i^Y > p_j^Y}} e(i,j),$$

so that, by (1.1.9), $e_{\max} = \max(e_1, e_2)$.

Theorem 1.1.7. *Let $e_1 < e_2$, then Case b2) holds true. Let $e_1 \ge e_2$, then:*

(i) If for some $i \in \{1,\dots,m\}$ such that $p_i^X \wedge p_i^Y = e_1$, one has $p_i^X \neq p_i^Y$, then Case a) holds true or so does its symmetric version: there exists $u \in L_U$ such that $\frac{u_1}{p_1^Y} + \cdots + \frac{u_m}{p_m^Y} = 1$ and $\frac{u_1}{p_1^X} + \cdots + \frac{u_m}{p_m^X} < 1$.

(ii) Otherwise, i.e., if for all $i \in \{1,\dots,m\}$ such that $p_i^X \wedge p_i^Y = e_1$, one has $p_i^X = p_i^Y$, then if $e_1 > e_2$ Case b1) holds true, while if $e_1 = e_2$, then so does Case b2).

Proof. First, for any $0 < \delta < 1$, let $e_{\max,\delta}$, $e_{1,\delta}$, $e_{2,\delta}$ and $e_\delta(i,j)$ be defined just as $e_{\max}, e_1, e_2$ and $e(i,j)$ but replacing p_i^Y with δp_i^Y, for all $i \in \{1,\dots,m\}$. Next, from the very definition of Case a): There exists $u \in L_U$ such that $\frac{u_1}{p_1^X} + \cdots + \frac{u_m}{p_m^X} = 1$ and $\frac{u_1}{p_1^Y} + \cdots + \frac{u_m}{p_m^Y} < 1$. Letting $\delta_0 := \frac{u_1}{p_1^Y} + \cdots + \frac{u_m}{p_m^Y}$, we have $\frac{u_1}{\delta_0 p_1^Y} + \cdots + \frac{u_m}{\delta_0 p_m^Y} = 1$ so $e_{\max,\delta_0} \ge e_{\max}$ and therefore (clearly, $e_{\max,\delta}$ is non-decreasing in δ) $e_{\max,\delta_0} = e_{\max}$. So when Case a) occurs there exists $0 < \delta_0 < 1$, such that for all $\delta \in (\delta_0, 1], e_{\max,\delta} = e_{\max}$, and one can easily check the converse. A similar result continues to hold for the symmetric version of Case a).

We can now prove the statement of the theorem by distinguishing the following four occurrences.

(1) Let $e_1 < e_2$. Let $0 < \delta_0 < 1$ be close enough to 1 such that for any $\delta \in (\delta_0, 1]$, the set of pairs $i,j \in \{1,\dots,m\}$ such that $\begin{smallmatrix} p_i^X < p_j^X \\ \wedge \quad \vee \\ p_i^Y > p_j^Y \end{smallmatrix}$ is equal to the set of $i,j \in \{1,\dots,m\}$ such that $\begin{smallmatrix} p_i^X < p_j^X \\ \wedge \quad \vee \\ \delta p_i^Y > \delta p_j^Y \end{smallmatrix}$. Since for every i,j in this set, it is immediate to check that $e(i,j) > e_\delta(i,j)$, the maximums satisfy $e_2 > e_{\delta,2}$. Since $e_1 < e_2$, by continuity, for δ close enough to 1, $\max(e_{\delta,1}, e_{\delta,2}) = e_{\delta,2}$ so $e_{\delta,\max} < e_{max}$, hence we are in Case b). There are $i,j \in \{1,\dots,m\}$ such that $e_{\max} = e_2 = e(i,j)$, so i,j are in I, but $p_i^X < p_j^X$ so we are in Case b2).

(2) Let $e_1 \ge e_2$, and let there exists $i \in \{1,\dots,m\}$ such that $p_i^X \wedge p_i^Y = e_1$ and $p_i^X \neq p_i^Y$, say, $p_i^X < p_i^Y$. Then, the very definition of Case a) is verified with the vector $u \in \mathbb{R}^m$ having coordinates equal to zero except for $u_i = p_i^X$. If instead, $p_i^X > p_i^Y$ then the symmetric case holds true.

(3) Let $e_1 > e_2$ and let for all $i \in \{1,\dots,m\}$ such that $p_i^X \wedge p_i^Y = e_1$, $p_i^X = p_i^Y$. By continuity, for δ close enough to 1, $\max(e_{\delta,1}, e_{\delta,2}) = e_{\delta,1} = \delta e_{\max}$ so we are in Case b). Additionally, one verifies that under our assumptions I is restricted to the set of $i \in \{1,\dots,m\}$ such that $p_i^X = p_i^Y = e_{\max}$. Therefore, we are, in fact, in Case b1).

(4) Let $e_1 = e_2$ and let for all $i \in \{1,\dots,m\}$ such that $p_i^X \wedge p_i^Y = e_1$, $p_i^X = p_i^Y$. From what is done above, we see that for δ close enough to 1, $e_{\delta,\max} < e_{\max}$ hence we are in Case b). Once again, since there are $i,j \in \{1,\dots,m\}$ such that $e_{\max} = e_2 = e(i,j)$, we are in Case b2). □

To present another explicit example, let us fully corner the case $m = 2$, with p_1^X, p_2^X, p_1^Y, and p_2^Y. The following completely describes the various cases:

- If $p_1^X = p_1^Y$, then (since, necessarily, $p_2^X = p_2^Y$) $e_{max} = \max(p_1^X, p_2^X) = \max(p_1^X, 1 - p_1^X)$ and we are in Case b1).
- If $p_1^X \neq p_1^Y$ and $1/2 \in (\min(p_1^X, p_1^Y), \max(p_1^X, p_1^Y))$, then

$$e_{max} = \max(\min(p_1^X, p_1^Y), \min(p_2^X, p_2^Y)) = \max(\min(p_1^X, p_1^Y), \min(1 - p_1^X, 1 - p_1^Y)),$$

  and we are in Case a) or its symmetric.
- If $p_1^X \neq p_1^Y$ and $1/2 \notin (\min(p_1^X, p_1^Y), \max(p_1^X, p_1^Y))$, then

$$e_{max} = \frac{p_1^X p_1^Y (p_2^X - p_2^Y) + p_2^X p_2^Y (p_1^Y - p_1^X)}{p_1^Y p_2^X - p_1^X p_2^Y} = p_1^X p_1^Y + p_2^X p_2^Y = p_1^X p_1^Y + (1 - p_1^X)(1 - p_1^Y),$$

  and we are in Case b2).

1.2 The limiting law

It is clear, from the previous section, that the proper way to center (and normalize) LCI_n is via

$$\begin{aligned} Z_n &= \frac{LCI_n - ne_{\max}}{\sqrt{n}} \\ &= \max_{\lambda \in \Lambda^2} \sum_{i=1}^{m} \left[\left(\sqrt{n} p_i^X \lambda_i^X + \widetilde{V}_i^{n,X}(\lambda^X) \right) \wedge \left(\sqrt{n} p_i^Y \lambda_i^Y + \widetilde{V}_i^{n,Y}(\lambda^Y) \right) \right] - \sqrt{n} e_{\max}. \end{aligned}$$

Let also

$$\begin{aligned} Z_n^c &= \frac{LCI_n^c - ne_{\max}}{\sqrt{n}} \\ &= \max_{\lambda \in \Lambda^2} \sum_{i=1}^{m} \left[\left(\sqrt{n} p_i^X \lambda_i^X + V_i^{n,X}(\lambda^X) \right) \wedge \left(\sqrt{n} p_i^Y \lambda_i^Y + V_i^{n,Y}(\lambda^Y) \right) \right] - \sqrt{n} e_{\max}, \end{aligned}$$

from (1.1.3) we have

$$|Z_n - Z_n^c| \leq \frac{m}{\sqrt{n}},$$

and therefore the convergence in distribution of Z_n^c will imply the convergence, in distribution, of Z_n towards the same limit.

1.2.1 Statement of the theorem

Below is the main result of the paper. In this statement, the covariance matrices of the Brownian motions stem from the covariance matrix of the rescaled variables $(\mathbb{1}_{X_k=i})_{i \in I}$ (resp. $\mathbb{1}_{Y_k=i}, i \in I$) used to construct the polygonal approximations $B_i^{n,\bullet}$ (here, and throughout, $\bullet$ is short for either X or Y). Indeed, note that $\mathbb{E}\left(\frac{(\mathbb{1}_{X_k=i} - p_i^X)(\mathbb{1}_{X_k=j} - p_j^X)}{\sqrt{p_i^X(1-p_i^X)}\sqrt{p_j^X(1-p_j^X)}} \right) = -\sqrt{\frac{p_i^X p_j^X}{(1-p_i^X)(1-p_j^X)}}$ (with a similar result for Y).

Theorem 1.2.1. *Let B^X and B^Y be two independent $|I|$-dimensional Brownian motions defined on $[0,1]$ with respective covariance matrix C^X defined by $C_{i,i}^X = 1$ and $C_{i,j}^X = -\sqrt{\frac{p_i^X p_j^X}{(1-p_i^X)(1-p_j^X)}}$,*

for $i \neq j$ in I, and C^Y defined in a similar fashion, replacing p_i^X by p_i^Y and p_j^X by p_j^Y. For all $\lambda \in K_{\Lambda^2}$ and $i \in I$, set

$$V_i^X(\lambda^X) = \sqrt{p_i^X(1-p_i^X)}\left(B_i^X\left(\sum_{j=1}^{i}\lambda_j^X\right) - B_i^X\left(\sum_{j=1}^{i-1}\lambda_j^X\right)\right),$$

$$V_i^Y(\lambda^Y) = \sqrt{p_i^Y(1-p_i^Y)}\left(B_i^Y\left(\sum_{j=1}^{i}\lambda_j^Y\right) - B_i^Y\left(\sum_{j=1}^{i-1}\lambda_j^Y\right)\right).$$

If there exists $u \in L_U$ such that $\frac{u_1}{p_1^X} + \cdots + \frac{u_m}{p_m^X} = 1$ and $\frac{u_1}{p_1^Y} + \cdots + \frac{u_m}{p_m^Y} < 1$ (Case a)), then

$$\frac{LCI_n - ne_{\max}}{\sqrt{n}} \underset{n\to\infty}{\Longrightarrow} Z^a := \max_{\lambda^X \in J}\sum_{i\in I} V_i^X(\lambda^X),$$

where J is given by (1.1.6).

If for all $u \in L_U$, $\frac{u_1}{p_1^X} + \cdots + \frac{u_m}{p_m^X} = \frac{u_1}{p_1^Y} + \cdots + \frac{u_m}{p_m^Y} = 1$ (Case b)), then

$$\frac{LCI_n - ne_{\max}}{\sqrt{n}} \underset{n\to\infty}{\Longrightarrow} Z^b := \max_{\lambda\in K_{\Lambda^2}} \mathfrak{m}\left(V^X(\lambda^X), V^Y(\lambda^Y)\right),$$

where $\mathfrak{m}$ is given by (1.1.7).

At this point, one can remark that $e_{\max}$ is invariant with respect to the order in which the letters are chosen, and that both in Case a) and Case b1), the above limiting laws are invariant as well (to see this fact in Case a), recall Lemma 1.1.3). Therefore, in Case a) and Case b1), no matter the prescribed order (increasing, decreasing, etc..) the asymptotic behavior of the length of the corresponding optimal alignments is the same. We refer the reader to Section 1.3.2 for more general results of this flavor.

In Case b2) it is less clear that the limiting distribution is permutation-invariant as it might not just boil down to $\mathfrak{m}(\nu)$. Indeed, in Case b2) the limiting law can be written as the law of

$$Z = \max_{\lambda\in K_{\Lambda^2}} \sum_{\substack{i\in\{1,\ldots,m\}\\ \bullet\in\{X,Y\}}} V(\lambda)_i^\bullet,$$

where $V(\lambda)$ is in $(\mathbb{R}^m)^2$, and defined via

$$V^\bullet(\lambda)_i = B_i^\bullet\left(\sum_{j=1}^{i}\lambda_j^\bullet\right) - B_i^\bullet\left(\sum_{j=1}^{i-1}\lambda_j^\bullet\right),$$

where the $B_i^\bullet$ are Brownian motions which are, up to a multiplicative factor, as in our main theorem. Further introducing, for any permutation σ of $\{1,\ldots,n\}$, $V_\sigma(\lambda)$ defined via

$$V_\sigma^\bullet(\lambda)_i = B_i^\bullet\left(\sum_{j=1}^{\sigma^{-1}(i)}\lambda_{\sigma(j)}^\bullet\right) - B_i^\bullet\left(\sum_{j=1}^{\sigma^{-1}(i)-1}\lambda_{\sigma(j)}^\bullet\right),$$

we have $V(\lambda) = V_{\mathrm{Id}}(\lambda)$, where Id is the identity permutation. When the letters are not required to be non-decreasing, but instead follow an order given by σ, the limiting law is simply the law of $Z_\sigma := \max_{\lambda\in K_{\Lambda^2}} \sum_{\substack{i\in\{1,\ldots,m\}\\ \bullet\in\{X,Y\}}} V_\sigma(\lambda)_i^\bullet$. It is still not that clear whether or not this last quantity depends on σ. For example, if $m = 3$ and $K_{\Lambda^2} = \Lambda^2$ and B_1^X is a standard Brownian motion, while all others are null, define σ by $\sigma(1) = 2, \sigma(2) = 1, \sigma(3) = 3$, then with probability one $Z_\sigma > Z_{\mathrm{Id}}$. However, in Case b2) it is actually not possible to have $K_{\Lambda^2} = \Lambda^2$ (and also to have only one non null Brownian motion) but this shows that a general argument for the validity of the permutation-invariance is not that transparent.

1.2.2 Proof of Theorem 1.2.1

The proof of this theorem is based on a non-probabilistic lemma. First, let E_n^η be the set of all continuous functions b from $[0,1]$ into $\mathbb{R}$ such that: for all x, y in $[0,1]$, $|b(y) - b(x)| \leq \left(n^\eta \sqrt{|y-x|} + n^{\eta-1/2}\right)/2$. Then, for all $b \in (E_n^\eta)^m$, $i \in \{1, \dots, m\}$ and $\lambda \in \Lambda$, set $v_i^b(\lambda) = b_i(\lambda_1 + \cdots + \lambda_i) - b_i(\lambda_1 + \cdots + \lambda_{i-1})$, and for all $b^X, b^Y \in (E_n^\eta)^m$ and $\lambda \in \Lambda^2$ let

$$z_n(\lambda) = \sum_{i=1}^m \left[\left(\sqrt{n} p_i^X \lambda_i^X + v_i^{b^X}(\lambda^X) \right) \wedge \left(\sqrt{n} p_i^Y \lambda^Y + v_i^{b^Y}(\lambda^Y) \right) \right] - \sqrt{n} e_{\max}.$$

One can think of b_i^X (resp. b_i^Y) as $\sqrt{p_i^X(1-p_i^X)} B_i^{n,X}(\omega)$ (resp. $\sqrt{p_i^Y(1-p_i^Y)} B^{n,Y}(\omega)$) for a fixed $\omega \in A_n^\eta$, where the symbol b^X (resp. b^Y) is used for ease of notation and in order to emphasize the non-probabilistic nature of the proof. For further ease of notation, we omit the dependency in b^X and b^Y in the notation z_n. This omission is also present in v and v^X is just short for v^{b^X} (similarly with Y), and further write $v(\lambda) := \left(v^X(\lambda^X), v^Y(\lambda^Y)\right)$.

In Case a), for all $\lambda^X \in \Lambda$, let

$$z^a(\lambda^X) := \sum_{i \in I} v_i^X(\lambda^X).$$

In Case b), for all $\lambda \in \Lambda^2$, let

$$z^b(\lambda) = \mathfrak{m}\left(v^X(\lambda^X), v^Y(\lambda^Y)\right).$$

Next, let us finally present two simple inequalities stemming from the very definition of E_n^η, often used in the sequel, which are valid for all $b \in E_n^\eta$, $\lambda, \lambda' \in \Lambda$, $i \in \{1, \dots, m\}$, $\bullet \in \{X, Y\}$, namely,

$$|v_i^\bullet(\lambda^\bullet)| \leq \frac{n^\eta \sqrt{\lambda_i^\bullet} + n^{\eta-1/2}}{2} \quad \text{and in particular} \quad |v_i^\bullet(\lambda^\bullet)| \leq n^\eta, \tag{1.2.1}$$

$$\begin{aligned} |v_i^\bullet(\lambda^\bullet) - v_i^\bullet(\lambda'^\bullet)| &\leq n^\eta \sqrt{\max_{i \in \{1,\dots,m\}} |\lambda_1 + \cdots + \lambda_i - \lambda_1' - \cdots - \lambda_i'|} + n^{\eta-1/2} \\ &\leq n^\eta \sqrt{m \|\lambda - \lambda'\|_\infty} + n^{\eta-1/2}. \end{aligned} \tag{1.2.2}$$

Lemma 1.2.2. *There exists a sequence $(\varepsilon_n)_{n \geq 1}$ of positive reals converging to zero and such that for all $n \geq 1$ and $b^X, b^Y \in (E_n^\eta)^m$, either $|\max_{\lambda \in \Lambda^2} z_n(\lambda) - \max_{\lambda \in \Lambda^2} z^a(\lambda)| \leq \varepsilon_n$, or $|\max_{\lambda \in \Lambda^2} z_n(\lambda) - \max_{\lambda \in K_{\Lambda^2}} z^b(\lambda)| \leq \varepsilon_n$, in Case a) or b), respectively.*

The proof of this crucial lemma is delayed to the next subsections, and instead we turn our attention to the proof of the main theorem.

Proof of Theorem 1.2.1. Let us assume that Case b) is occurring. Let

$$Z_n^b = \max_{\lambda \in K_{\Lambda^2}} \mathfrak{m}\left(V^{n,X}(\lambda^X), V^{n,Y}(\lambda^Y)\right).$$

For all $\omega \in A_n^\eta$, $B^{n,X}(\omega)$ and $B^{n,Y}(\omega)$ are in E_n^η so by Lemma 1.2.2, $|Z_n^c(\omega) - Z_n^b(\omega)| \leq \varepsilon_n$. So $\left|Z_n^c - Z_n^b\right| \mathbb{1}_{A_n^\eta} \leq \varepsilon_n$, but $Z_n^c - Z_n^b = \left(Z_n^c - Z_n^b\right) \mathbb{1}_{A_n^\eta} + \left(Z_n^c - Z_n^b\right) \mathbb{1}_{(A_n^\eta)^c}$, where this second term tends to zero in probability, therefore so does $Z_n^c - Z_n^b$. Next, by Donsker's theorem and the continuity of $\mathfrak{m}$ (recalling Lemma 1.1.5), Z_n^b tends to Z^b in distribution, so does Z_n^c and finally so is the case for Z_n, recalling (1.2). The proof in the Case a) is analogous and therefore omitted. □

Let us now turn to the proof of Lemma 1.2.2. The method of proof goes as follows: Maximizing $z_n(\lambda)$ is equivalent to maximizing

$$z_n(\lambda)/\sqrt{n} = \sum_{i=1}^{m} \left[\left(p_i^X \lambda_i^X + v_i^{b^X}(\lambda^X)/\sqrt{n} \right) \wedge \left(p_i^Y \lambda^Y + v_i^{b^Y}(\lambda^Y)/\sqrt{n} \right) \right] - e_{\max},$$

which converges, as n goes to infinity, to $f(\lambda) - e_{\max}$. So one can expect that λ must "almost" be maximizing f, i.e., be in or "close to" the set K_{Λ^2}. In Case a), we bound the maximum by taking the maximum over two sets which are closer and closer to the set J. In Case b), first write $\lambda = \lambda^{K_{\Lambda^2}} + \lambda^r$ (actually dealing with a $\lambda - a$ in order to have a vector space, but the idea is the same), then ignore the small perturbation term λ^r in v, and the idea is (roughly) to fix $\lambda^{K_{\Lambda^2}}$ and to find the maximum over λ^r. In both cases, the end of the proof consists in showing how the maximum of the relevant function (z^a or z^b) over a set of parameters that "tends to" a limiting set goes to the maximum over this limiting set.

1.2.3 Proof of Lemma 1.2.2, Case a)

Restriction to I

First, fix $b = (b^X, b^Y) \in ((E_n^\eta)^m)^2$. Next, for ease of notation, omit in the sub-index b in z and v. Roughly speaking, we begin by proving that any λ maximizing z_n must have "small" coordinates outside of I, and therefore we can "replace" the variations v_i, for $i \notin I$, by zero.

Let

$$p_{\text{sec}}^X = \begin{cases} \max_{i \notin I} p_i^X & I \neq \{1, \dots, m\}, \\ 0 & I = \{1, \dots, m\} \end{cases}.$$

Let us assume first that $I \neq \{1, \dots, m\}$. Then by Lemma 1.1.3, $p_{\text{sec}}^X < p_{\max}^X$. Our first observation is that if λ maximizes z_n, i.e., if $z_n(\lambda) = \max_{\lambda \in \Lambda^2} z_n(\lambda)$, then

$$s := \sum_{i \notin I} \lambda_i^X \leq \frac{2mn^{\eta - 1/2}}{p_{\max}^X - p_{\text{sec}}^X}. \tag{1.2.3}$$

In words, the above indicates that the contribution of the letters not in I is, as expected, very limited. To prove this inequality, note that on the one hand (recalling Lemma 1.1.3 and (1.2.1)),

$$z_n(\lambda) \leq \sum_{i=1}^{m} \left(\sqrt{n} p_i^X \lambda_i^X + v_i^b(\lambda) \right) - \sqrt{n} p_{\max}^X \leq \sqrt{n} \left(p_{\max}^X (1 - s) + p_{\text{sec}}^X s \right) + mn^\eta - \sqrt{n} p_{\max}^X,$$

while on the other hand, for $\tilde{\lambda} \in K_{\Lambda^2}$, using (1.2.1) and the elementary inequality (1.1.2),

$$z_n(\lambda) \geq z_n(\tilde{\lambda}) \geq \sqrt{n} f(\tilde{\lambda}) - mn^\eta - \sqrt{n} p_{\max}^X = -mn^\eta. \tag{1.2.4}$$

The inequality (1.2.3) follows, and it therefore allows, for $i \notin I$, to replace the terms $v_i^X(\lambda^X)$ by zero. More precisely, let for all $\lambda \in \Lambda^2$,

$$z_n^I(\lambda) = \sum_{i \in I} \left[\left(\sqrt{n} p_i^X \lambda_i^X + v_i^X(\lambda^X) \right) \wedge \left(\sqrt{n} p_i^Y \lambda_i^Y + v_i^Y(\lambda^Y) \right) \right] + \sum_{i \notin I} \left[\left(\sqrt{n} p_i^X \lambda_i^X \right) \wedge \left(\sqrt{n} p_i^Y \lambda_i^Y + v_i^Y(\lambda^Y) \right) \right] - \sqrt{n} e_{\max},$$

then as shown next,

$$\left| \max_{\lambda \in \Lambda^2} z_n(\lambda) - \max_{\lambda \in \Lambda^2} z_n^I(\lambda) \right| \leq \frac{|I^c|}{2} \left(n^\eta \sqrt{\frac{2mn^{\eta - 1/2}}{p_{\max}^X - p_{\text{sec}}^X}} + n^{\eta - 1/2} \right),$$

and this inequality remains true when $I = \{1, \dots, m\}$ (since then $\max_{\lambda\in\Lambda^2} z_n(\lambda) = \max_{\lambda\in\Lambda^2} z_n^I(\lambda)$ and $|I^c| = 0$).

Indeed, let $\lambda \in \Lambda^2$ be such that $z_n(\lambda) = \max_{\lambda\in\Lambda^2} z_n(\lambda)$. Using (1.1.2) along with (1.2.1) ($\lambda_i^X \leq 2mn^{\eta-1/2}/(p^X_{\max} - p^X_{\sec})$, for all $i \notin I$), it follows that

$$\max_{\lambda\in\Lambda^2} z_n^I(\lambda) \geq z_n^I(\lambda) \geq \max_{\lambda\in\Lambda^2} z_n(\lambda) - \frac{|I^c|}{2}\left(n^\eta\sqrt{\frac{2mn^{\eta-1/2}}{p^X_{\max} - p^X_{\sec}}} + n^{\eta-1/2}\right).$$

Moreover, let $\tilde{\lambda} \in \Lambda^2$ be such that $\max_{\lambda\in\Lambda^2} z_n^I(\lambda) = z_n^I(\tilde{\lambda})$. Then, just as in proving (1.2.3), it follows that $\sum_{i\notin I} \tilde{\lambda}_i^X \leq 2|I|n^{\eta-1/2}/(p^X_{\max} - p^X_{\sec})$. Hence

$$\max_{\lambda\in\Lambda^2} z_n(\lambda) \geq z_n(\tilde{\lambda}) \geq \max_{\lambda\in\Lambda^2} z_n^I(\lambda) - \frac{|I^c|}{2}\left(n^\eta\sqrt{\frac{2mn^{\eta-1/2}}{p^X_{\max} - p^X_{\sec}}} + n^{\eta-1/2}\right),$$

which completes the proof.

Bounds on the maximum with different sets of constraints

Let us next define two sets "close" to J. To do so, let $S_n = 2|I|^2 n^{\eta-1/2}$, let $C_I = \sum_{i\in I} \frac{1}{p_i^Y}$, let $T_n = C_I 2n^{\eta-1/2}$, and finally let

$$J_n^+ = \left\{\lambda^X \in \Lambda \,:\, \sum_{i\in I} \frac{\lambda_i^X}{p_i^Y} \leq \frac{1+S_n}{p^X_{\max}}\right\},$$

and

$$J_n^- = \left\{\lambda^X \in \Lambda \,:\, \sum_{i\in I} \frac{\lambda_i^X}{p_i^Y} \leq \frac{1-T_n}{p^X_{\max}}\right\}.$$

Note that by Lemma 1.1.3, setting $\delta_{i_1} = (\mathbb{1}_{i=i_1})_{i\in\{1,\dots,m\}}$, $\delta_{i_1} \in J_n^-$ eventually. We show, in this part of the proof, that

$$\max_{\lambda\in J_n^-} z^a(\lambda) \leq \max_{\lambda\in\Lambda^2} z_n^I(\lambda) \leq \max_{\lambda\in J_n^+} z^a(\lambda). \tag{1.2.5}$$

Let us prove the upper bound first. Let $\lambda \in \Lambda^2$ be such that $z_n^I(\lambda) = \max_{\lambda\in\Lambda^2} z_n^I(\lambda)$, and let S be the unique real such that

$$\sum_{i\in I} \frac{\lambda_i^X}{p_i^Y} = \frac{1+S}{p^X_{\max}}.$$

Then, there exists $i_0 \in I$ such that,

$$\lambda_{i_0}^Y p_{i_0}^Y \leq \lambda_{i_0}^X p^X_{\max} - \frac{S}{|I|},$$

since otherwise, $\sum_{i\in I} \lambda_i^Y > 1$, which is a contradiction. Then, using the following inequalities,

$$\begin{aligned}
\forall i \in I\setminus\{i_0\} \quad \left(\sqrt{n}p_i^X\lambda_i^X + v_i^X(\lambda^X)\right) \wedge \left(\sqrt{n}p_i^Y\lambda_i^Y + v_i^Y(\lambda^Y)\right) &\leq \left(\sqrt{n}p_i^X\lambda_i^X + v_i^X(\lambda^X)\right),\\
\left(\sqrt{n}p_{i_0}^X\lambda_{i_0}^X + v_{i_0}^X(\lambda^X)\right) \wedge \left(\sqrt{n}p_{i_0}^Y\lambda_{i_0}^Y + v_{i_0}^Y(\lambda^Y)\right) &\leq \left(\sqrt{n}\left(\lambda_{i_0}^X p^X_{\max} - \frac{S}{|I|}\right) + v_{i_0}^Y(\lambda^Y)\right),\\
\forall i \notin I \quad \left(\sqrt{n}p_i^X\lambda_i^X\right) \wedge \left(\sqrt{n}p_i^Y\lambda_i^Y + v_i^Y(\lambda^Y)\right) &\leq \sqrt{n}p_i^X\lambda_i^X,
\end{aligned}$$

leads to

$$\begin{aligned} z_n^I(\lambda) &\le \sqrt{n}\sum_{i=1}^m p_i^X\lambda_i^X + \sum_{i\in I\setminus\{i_0\}}\left(v_i^X(\lambda^X)+v_{i_0}^Y(\lambda^Y)\right) - \sqrt{n}\frac{S}{|I|} - \sqrt{n}e_{\max} \\ &\le \sum_{i\in I\setminus\{i_0\}}\left(v_i^X(\lambda^X)+v_{i_0}^Y(\lambda^Y)\right) - \sqrt{n}\frac{S}{|I|} \\ &\le |I|n^\eta - \sqrt{n}\frac{S}{|I|}. \end{aligned}$$

Just as in obtaining the inequality (1.2.4), we have $-|I|n^\eta \le z_n^I(\lambda)$, hence $S \le 2|I|^2 n^{\eta-1/2}$, i.e., $\lambda^X \in J_n^+$, leading to conclude with the upper estimate:

$$\max_{\lambda\in\Lambda^2} z_n^I(\lambda) = z_n^I(\lambda) \le \sqrt{n}f(\lambda^X) + z^a(\lambda^X) - \sqrt{n}e_{\max} \le z^a(\lambda^X) \le \max_{\lambda\in J_n^+} z^a(\lambda).$$

Let us now turn our attention to the lower bound. Let $\lambda^X \in J_n^-$ be such that $z^a(\lambda^X) = \max_{\lambda\in J_n^-} z^a(\lambda)$. Since

$$\sum_{i\in I}\left(p_{\max}^X\lambda_i^X + 2n^{\eta-1/2}\right)/p_i^Y \le 1,$$

there exists $\lambda^Y \in \Lambda$ such that for $i \in I$, $\lambda_i^Y \ge \left(p_{\max}^X\lambda_i^X + 2n^{\eta-1/2}\right)/p_i^Y$ and for $i \notin I$, $\lambda_i^Y = 0$. For all $i \in I$,

$$\begin{aligned} \sqrt{n}p_i^Y\lambda_i^Y + v_i^Y(\lambda^Y) &\ge \sqrt{n}p_{\max}^X\lambda_i^X + 2n^\eta + v_i^Y(\lambda^Y) \\ &\ge \sqrt{n}p_{\max}^X\lambda_i^X + v_i^X(\lambda^X) = \sqrt{n}p_i^X\lambda_i^X + v_i^X(\lambda^X). \end{aligned}$$

Therefore,

$$\begin{aligned} z_n^I(\lambda) &= \sum_{i\in I}\left(\sqrt{n}p_i^X\lambda_i^X + v_i^X(\lambda^X)\right) + \sum_{i\notin I}\left[(\sqrt{n}p_i^X\lambda_i^X)\wedge 0\right] - \sqrt{n}p_{\max}^X \\ &= \sum_{i\in I} v_i^X(\lambda^X) = z^a(\lambda^X) = \max_{\lambda\in J_n^-} z^a(\lambda), \end{aligned}$$

and $\max_{\lambda\in J_n^-} z^a(\lambda) \le \max_{\lambda\in\Lambda^2} z_n^I(\lambda)$.

End of the proof

Both quantities $|\max_{\lambda\in J_n^-} z^a(\lambda) - \max_{\lambda\in J} z^a(\lambda)|$ and $|\max_{\lambda\in J_n^+} z^a(\lambda) - \max_{\lambda\in J} z^a(\lambda)|$ still need to be investigated. Let $C_1 = \left(1 - \frac{p_{\max}^X}{p_{i_1}^Y}\right) > 0$. For $\lambda^X \in \Lambda$ and $t \in (0,1)$, let $\lambda^{X,t} = t\delta_{i_1} + (1-t)\lambda^X$. It is straightforward to prove that for all n greater than some constant, depending only on η, p^X and p^Y, and for all $\lambda^X \in J$, $\lambda^{X,\frac{T_n}{C_1}}$ is well defined, and is in J_n^-, while for all $\lambda^X \in J_n^+$, $\lambda^{X,\frac{2S_n}{C_1}} \in J$.

This is useful since for all $i \in \{1,\dots,m\}$,

$$|\lambda_1^X + \cdots + \lambda_i^X - \lambda_1^{X,t} - \cdots - \lambda_i^{X,t}| \le 2t,$$

and therefore, using (1.1.2) along with (1.2.2),

$$\begin{aligned} \max_{\lambda\in J} z^a(\lambda) - \max_{\lambda\in J_n^-} z^a(\lambda) &\le |I|\left(n^\eta\sqrt{\frac{2T_n}{C_1}} + n^{\eta-1/2}\right), \\ \max_{\lambda\in J_n^+} z^a(\lambda) - \max_{\lambda\in J} z^a(\lambda) &\le |I|\left(n^\eta\sqrt{\frac{4S_n}{C_1}} + n^{\eta-1/2}\right). \end{aligned}$$

Putting these two inequalities, together with (1.2.5), leads to

$$\left|\max_{\lambda\in\Lambda^2} z_n^I(\lambda) - \max_{\lambda\in J} z^a(\lambda)\right| \leq C_2 n^{\frac{6\eta-1}{4}} + |I| n^{\eta-1/2},$$

for some constant C_2 depending only on the p's but need not be made explicit. The lemma is thus proved in this case.

1.2.4 Proof of Lemma 1.2.2, Case b)

Preliminaries

Fix $b = (b^X, b^Y) \in ((E_n^\eta)^m)^2$. Just as in Case a), we omit in the notation the sub-index b. Let $E = \{x \in \mathbb{R}^m : x_1 + \cdots + x_m = 0\}$, let K be the subspace of E^2 defined by

$$K = \left\{x \in E^2 : \forall i \in I, p_i^X x_i^X = p_i^Y x_i^Y, \forall i \notin I, x_i^X = y_i^Y = 0\right\},$$

and let P (recalling the definition of a following (1.1.5): $a \in K_{\Lambda^2}$, for all $i \in I, p_i^X a_i^X = p_i^Y a_i^Y > 0$, for $i \notin I, a_i^\bullet = 0$, and $f(a) = e_{\max}$) be given by:

$$P = \left\{x \in E^2 : \forall i \in \{1,\dots,m\}, x_i^X \geq -a_i^X, x_i^Y \geq -a_i^Y\right\}. \tag{1.2.6}$$

Note that $\Lambda^2 = a + P$. By definition of the case b), for all $\lambda \in K_{\Lambda^2}$, for all $i \in I$ $\lambda_i^X p_i^X = \lambda_i^Y p_i^Y$, while for all $i \notin I$, $\lambda_i^X = \lambda_i^Y = 0$. Reciprocally, let $\lambda \in \Lambda^2$ such that for all $i \in I$ $\lambda_i^X p_i^X = \lambda_i^Y p_i^Y$ and for all $i \notin I$, $\lambda_i^X = \lambda_i^Y = 0$, we show that $\lambda \in K_{\Lambda^2}$. Let $u \in \mathbb{R}^I$ be defined by $u_i = p_i^X \lambda_i^X - p_i^X a_i^X$ for all $i \in I$. We have that $u \cdot P^X = u \cdot P^Y = 1 - 1 = 0$ so by Lemma 1.1.4, $u \cdot (1)_{i\in I} = 0$, hence the result. This characterization of K_{Λ^2}, combined with $\Lambda^2 = a + P$, gives us

$$K_{\Lambda^2} = a + K \cap P. \tag{1.2.7}$$

Since $p_i^X a_i^X = p_i^Y a_i^Y$, for all $i \in \{1,\dots,m\}$,

$$z_n(a+x) = \sum_{i=1}^m \left[\left(\sqrt{n} p_i^X x_i^X + v_i^X (a^X + x^X)\right) \wedge \left(\sqrt{n} p_i^Y x_i^Y + v_i^Y (a^Y + x^Y)\right)\right].$$

Clearly,

$$\max_{\lambda\in\Lambda^2} z_n(\lambda) = \max_{x\in P} z_n(a+x).$$

Note also that for all $x \in (\mathbb{R}^m)^2$, $f(a+x) = f(a) + f(x)$ so by (1.2.7)

$$\forall x \in P,\ f(x) \leq 0 \quad \text{and} \quad (f(x) = 0) \iff (x \in K \cap P). \tag{1.2.8}$$

Our next result is an elementary projection result.

Lemma 1.2.3. *There exists $C > 0$ depending only on p^X and p^Y such that for all $x \in P$, there exist $x^{K\cap P} \in K \cap P$ and $x^r \in E^2$ such that $x = x^{K\cap P} + x^r$ and $\|x^r\|_\infty \leq -Cf(x)$.*

Proof. Let $K^\perp$ be the orthogonal complement of K in E^2 (for the usual Euclidean inner product defined on E^2 by, for $x, y \in E^2$, $x \cdot y := x_1^X y_1^X + \cdots + x_m^X y_m^X + x_1^Y y_1^Y + \cdots + x_m^Y y_m^Y$). Let $x \in P$ (so $x \in E^2$) and let $(x^K, x^{K^\perp})$ be its orthogonal decomposition, i.e., $x^K \in K$, $x^{K^\perp} \in K^\perp$ and $x = x^K + x^{K^\perp}$. Without loss of generality, assume $x^{K^\perp} \neq 0$. For ease of notation, set $g = -f$. Let

$$a_{\min} = \min_{i\in I} a_i.$$

In order to bound the image of $x^{K^\perp}$, we first rescale it to make it an element of P: it is easy to check that $y := \left(\frac{a_{\min}}{\|x^{K^\perp}\|_\infty}\right) x^{K^\perp} \in P$. Now, consider the sphere,

$$S_{a_{\min}} := \left\{ z \in K^\perp \, : \, \|z\|_\infty = a_{\min} \right\}.$$

Then, $S_{a_{\min}} \cap P$ is a non-empty compact set, so let

$$M = \min_{z \in S_{a_{\min}} \cap P} g(z).$$

Recalling (1.2.8), $M > 0$. Since $y \in S_{a_{\min}} \cap P$, $M \leq g(y)$ so that, using $g\left(x^{K^\perp}\right) = g(x)$,

$$\|x^{K^\perp}\|_\infty \leq \frac{a_{\min}}{M} g(x).$$

This is almost the desired result, except that x^K might not be in P. Let us assume, firstly, that $g(x) \leq M$ (and therefore that $\|x^{K^\perp}\|_\infty \leq a_{\min}$). Let $x^{K\cap P} = \left(1 - \frac{\|x^{K^\perp}\|_\infty}{a_{\min}}\right) x^K$ and let $x^r = \frac{\|x^{K^\perp}\|_\infty}{a_{\min}} x^K + x^{K^\perp}$. We next prove that $x^{K\cap P} \in K \cap P$. Since $x \in P$, for $i \in I$,

$$\left(1 - \frac{\|x^{K^\perp}\|_\infty}{a_{\min}}\right) x_i^K + \left(1 - \frac{\|x^{K^\perp}\|_\infty}{a_{\min}}\right) x_i^{K^\perp} \geq - \left(1 - \frac{\|x^{K^\perp}\|_\infty}{a_{\min}}\right) a_i$$

$$\begin{aligned} x_i^{K\cap P} &\geq -a_i + \frac{\|x^{K^\perp}\|_\infty}{a_{\min}} a_i - \left(1 - \frac{\|x^{K^\perp}\|_\infty}{a_{\min}}\right) x_i^{K^\perp} \\ &\geq -a_i + \|x^{K^\perp}\|_\infty - \left(1 - \frac{\|x^{K^\perp}\|_\infty}{a_{\min}}\right) \|x^{K^\perp}\|_\infty \\ &\geq -a_i, \end{aligned}$$

and for $i \notin I$, $x_i^{K\cap P} = 0$, since $x^{K\cap P} \in K$. So $x^{K\cap P} \in K \cap P$.

Let us turn to x^r. Since $a + x \in \Lambda^2$, $\|x\|_\infty \leq 1$. Moreover, x^K is the orthogonal projection of x so $\|x^K\|_\infty \leq \sqrt{2m}\|x\|_\infty \leq \sqrt{2m}$ and

$$\begin{aligned} \|x^r\|_\infty &\leq \left(\frac{\sqrt{2m}}{a_{\min}} + 1\right) \|x^{K^\perp}\|_\infty \\ &\leq \left(\frac{\sqrt{2m}}{a_{\min}} + 1\right) \frac{a_{\min}}{M} g(x). \end{aligned}$$

Setting $C := \left(\sqrt{2m} + a_{\min}\right)/M$, we have just proved that if $g(x) \leq M$, then there exist suitable $x^{K\cap P}$ and x^r satisfying the lemma. Finally, if $g(x) > M$, we let $x^{K\cap P} = 0$ and $x^r = x$, so that $\|x^r\|_\infty \leq 1 < g(x)/M < Cg(x)$ which completes the proof. □

Separation of the parameters

To begin with, we prove that $\max_{x\in P} z_n(a + x)$ can be written as a maximum over two kind of parameters, one belonging to K in the variations v_i, the other one being a small remaining term.

Let $x \in P$ be such that $z_n(a + x) = \max_{\lambda\in\Lambda^2} z_n(\lambda)$. Then,

$$-mn^\eta \leq z_n(a) \leq z_n(a + x) \leq \sqrt{n} f(x) + mn^\eta,$$

and so

$$-f(x) \leq 2mn^{\eta-1/2}. \tag{1.2.9}$$

Now, let

$$D = \left\{(x^{K\cap P}, x^r) \in (K\cap P)\times E^2 \,:\, x^{K\cap P} + x^r \in P\right\},$$

and, recalling the constant C from Lemma 1.2.3, let

$$D_n = \left\{(x^{K\cap P}, x^r) \in (K\cap P)\times E^2 \,:\, \|x^r\|_\infty \leq 2Cmn^{\eta-1/2}, x^{K\cap P} + x^r \in P\right\}.$$

Then, for all $(x^{K\cap P}, x^r) \in D$, set

$$\overline{z}_n(x^{K\cap P}, x^r) = z_n(a + x^{K\cap P} + x^r) =$$
$$\sum_{i=1}^{m}\left[\left(\sqrt{n}p_i^X x_i^{r,X} + v_i^X(a^X + x^{K\cap P,X} + x^{r,X})\right) \wedge \left(\sqrt{n}p_i^Y x_i^{r,Y} + v_i^Y(a^Y + x^{K\cap P,Y} + x^{r,Y})\right)\right].$$

Applying Lemma 1.2.3 to (1.2.9) gives $\max_{x\in D_n}\overline{z}_n(x) = \max_{x\in P} z_n(a+x)$.

Let us next define a slight modification of $\overline{z}_n$ by letting, for all $(x^{K\cap P}, x^r) \in D_n$,

$$\overline{z}'_n(x^{K\cap P}, x^r) = \sum_{i=1}^{m}\left[\left(\sqrt{n}p_i^X x_i^{r,X} + v_i^X(a^X + x^{K\cap P,X})\right) \wedge \left(\sqrt{n}p_i^Y x_i^{r,Y} + v_i^Y(a^Y + x^{K\cap P,Y})\right)\right].$$

The parameters are now "separated". For all $(x^{K\cap P}, x^r) \in D_n$, by (1.2.2),

$$\left|\overline{z}'_n(x^{K\cap P}, x^r) - \overline{z}_n(x^{K\cap P}, x^r)\right| \leq m\left(n^\eta\sqrt{2Cm^2n^{\eta-1/2}} + n^{\eta-1/2}\right),$$

so that

$$\left|\max_{x\in P} z_n(a+x) - \max_{x\in D_n}\overline{z}'_n(x)\right| = \left|\max_{x\in D_n}\overline{z}_n(x) - \max_{x\in D_n}\overline{z}'_n(x)\right|$$
$$\leq m\left(n^\eta\sqrt{2Cm^2n^{\eta-1/2}} + n^{\eta-1/2}\right). \tag{1.2.10}$$

Independence of the parameters

A major issue with D_n is the condition $x^{K\cap P} + x^r \in P$. We would rather have a set of possible values for x^r independent of the value of $x^{K\cap P}$. To try to achieve that goal, let

$$P_n \;=\; \left\{x \in E^2 \colon \forall i \in I, \forall \bullet \in \{X,Y\}, x_i^\bullet \geq -a_i^\bullet + 2Cmn^{\eta-1/2}, \forall i \notin I, x_i^X \geq 0, x_i^Y \geq 0\right\} \;\subset\; P,$$

and let $D'_n \subset D_n$ be given by

$$D'_n = \left\{(x^{K\cap P_n}, x^r) \in (K\cap P_n)\times E^2 \,:\, \|x^r\|_\infty \leq 2Cmn^{\eta-1/2}, x^{K\cap P_n} + x^r \in P\right\}.$$

Now, recalling the definition $E' = \{x \in E : \forall i \in I^c, x_i \geq 0\} \subset E$, we have that

$$D'_n = \left\{(x^{K\cap P_n}, x^r) \in (K\cap P_n)\times E'^2 \,:\, \|x^r\|_\infty \leq 2Cmn^{\eta-1/2}\right\}.$$

For $(x^{K\cap P}, x^r) \in D_n$, and for n large enough so that $\frac{2Cmn^{\eta-1/2}}{a_{\min}} \leq 1$, it follows that, letting $x'^{K\cap P} := \left(1 - \frac{2Cmn^{\eta-1/2}}{a_{\min}}\right)x^{K\cap P}$, $(x'^{K\cap P}, x^r) \in D'_n$, so by (1.2.2)

$$\left|\max_{x\in D'_n}\overline{z}'_n(x) - \max_{x\in D_n}\overline{z}'_n(x)\right| \leq |I|\left(n^\eta\sqrt{\frac{2Cm^2n^{\eta-1/2}}{a_{\min}}} + n^{\eta-1/2}\right). \tag{1.2.11}$$

Connections with the functions of Lemma 1.2.2

Let us now prove that for n large enough,

$$\max_{x\in D'_n} \overline{z}^{\eta}_n(x) = \max_{\lambda\in a+K\cap P_n} \mathfrak{m}\left(v^X(\lambda^X), v^Y(\lambda^Y)\right).$$

Fix $x^{K\cap P_n} \in K\cap P_n$. Applying the previous lemma to $\nu := v(a+x^{K\cap P_n})$, since $\|\nu\|_\infty \le n^\eta$, by Lemma 1.1.5

$$\begin{aligned}\max_{\substack{x^r\in E'^2\\ \|x^r\|_\infty\le 2Cmn^\eta}} \sum_{i=1}^m \left[\left(p_i^X x_i^{r,X}+\nu_i^X\right)\wedge\left(p_i^Y x_i^{r,Y}+\nu_i^Y\right)\right] &= \max_{x^r\in E'^2}\sum_{i=1}^m\left[\left(p_i^X x_i^{r,X}+\nu_i^X\right)\right.\\ &\qquad\left.\wedge\left(p_i^Y x_i^{r,Y}+\nu_i^Y\right)\right]\\ &= \mathfrak{m}(\nu),\end{aligned}$$

and so

$$\begin{aligned}\max_{\substack{x^r\in E'^2\\ \|x^r\|_\infty\le 2Cmn^{\eta-1/2}}} \overline{z}'_n(x^{K\cap P_n},x^r) &= \max_{\substack{x^r\in E'^2\\ \|x^r\|_\infty\le 2Cmn^{\eta-1/2}}}\sum_{i=1}^m\left[\left(\sqrt{n}p_i^X x_i^{r,X}+\nu_i^X\right)\wedge\left(\sqrt{n}p_i^Y x_i^{r,Y}+\nu_i^Y\right)\right]\\ &= \max_{\substack{x^r\in E'^2\\ \|x^r\|_\infty\le 2Cmn^{\eta}}}\sum_{i=1}^m\left[\left(p_i^X x_i^{r,X}+\nu_i^X\right)\wedge\left(p_i^Y x_i^{r,Y}+\nu_i^Y\right)\right]\\ &= \mathfrak{m}(\nu).\end{aligned}$$

Finally,

$$\max_{x\in D'_n}\overline{z}'_n(x) = \max_{x\in K\cap P_n}\max_{\substack{x\in E'^2\\ \|x\|_\infty\le 2Cmn^{\eta-1/2}}}\overline{z}'_n(x^{K\cap P_n},x^r) = \max_{\lambda\in a+K\cap P_n}\mathfrak{m}\left(v^X(\lambda^X),v^Y(\lambda^Y)\right). \quad (1.2.12)$$

End of the proof

Just as done with (1.2.11),

$$\left|\max_{\lambda\in a+K\cap P}\mathfrak{m}(v(\lambda)) - \max_{\lambda\in a+K\cap P_n}\mathfrak{m}(v(\lambda))\right| \le |I|\left(n^\eta\sqrt{\frac{2Cm^2n^{\eta-1/2}}{a_{\min}}}+n^{\eta-1/2}\right),$$

and so, using (1.2.10), (1.2.11) and (1.2.12) (recall that $a+K\cap P = K_{\Lambda^2}$),

$$\left|\max_{x\in P} z_n(a+x) - \max_{\lambda\in K_{\Lambda^2}}\mathfrak{m}(v(\lambda))\right| \le \left(\frac{2|I|}{\sqrt{a_{\min}}}+m\right)\sqrt{2Cm^2}n^{\frac{6\eta-1}{4}} + (2|I|+m)n^{\eta-1/2}.$$

1.3 Consistency with previous results and generalizations

1.3.1 Two words with identical distributions

As stated in the introductory section, Theorem 1.1.1 and the conjectured Theorem 1.1.2 are consequences of our main theorem. Indeed, let X_k and Y_k $(k=1,2,\dots)$ have the same distribution, then note that

$$I = \left\{i\in\{1,\dots,m\} \,:\, p_i^X = p_{\max}\right\},$$

and so the multiplicity k^* of $p_{\max}$ is equal to $|I|$ and we are in Case b1). It is also clear that

$$K_{\Lambda^2} = \left\{\lambda \in \Lambda^2 \,:\, \forall i \notin I, \lambda_i^X = \lambda_i^Y = 0\right\}^2.$$

In this case, Lemma 1.1.5 simplifies and gives $\mathfrak{m}(\nu) = S^X \wedge S^Y$, so our theorem states that the limiting distribution of $Z_n/\sqrt{p_{\max}(1-p_{\max})}$ is

$$\max_{\lambda \in K_{\Lambda^2}} \left[\left(\sum_{i\in I} B_i^X \left(\sum_{j=1}^{i} \lambda_j^X \right) - B_i^X \left(\sum_{j=1}^{i-1} \lambda_j^X \right) \right) \wedge \left(\sum_{i\in I} B_i^Y \left(\sum_{j=1}^{i} \lambda_j^Y \right) - B_i^Y \left(\sum_{j=1}^{i-1} \lambda_j^Y \right) \right) \right]$$
$$= \max_{0=t_0\le t_1 \le \cdots \le t_{k^*}=1} \left[\left(\sum_{i=1}^{k^*} \left(B_i^X(t_i) - B_i^X(t_{i-1}) \right) \right) \wedge \left(\sum_{i=1}^{k^*} \left(B_i^Y(t_i) - B_i^Y(t_{i-1}) \right) \right) \right],$$

where B^X and B^Y are two independent k^*-dimensional Brownian motions on $[0,1]$ with respective covariance matrix defined in Theorem 1.2.1. The proof of Corollary 3.3 in [32] shows that, by writing B^X and B^Y as linear combinations of independent standard Brownian motions, Z_n is identical in law to

$$\max_{0=t_0\le t_1 \le \cdots \le t_{k^*}=1} \sqrt{p_{\max}} \left[\left(\frac{\sqrt{1-k^*p_{\max}}-1}{k^*} \sum_{i=1}^{k^*} \overline{B}_i^X(1) + \sum_{i=1}^{k^*} \left(\overline{B}_i^X(t_i) - \overline{B}_i^X(t_{i-1}) \right) \right) \right.$$
$$\left. \wedge \left(\frac{\sqrt{1-k^*p_{\max}}-1}{k^*} \sum_{i=1}^{k^*} \overline{B}_i^Y(1) + \sum_{i=1}^{k^*} \left(\overline{B}_i^Y(t_i) - \overline{B}_i^Y(t_{i-1}) \right) \right) \right],$$

where now $\overline{B}^X$ and $\overline{B}^Y$ are two independent k^*-dimensional standard Brownian motions on $[0,1]$. Dividing both sides by $\sqrt{p_{\max}}$, one obtains the conjectured Theorem 1.1.2 which reduces to Theorem 1.1.1 when $k^* = m$.

1.3.2 Generalization to any fixed sequence of blocks

As pointed out by an Associate Editor, and also developed, for binary alphabets, in [77], a longest common increasing subsequence can be viewed as a longest common subsequence where letters are aligned in blocks. (For LCI_n, a non-void block only aligns a single type of letter and the first block consists of the letter $\alpha(1) := 1$, then the second one consists of $\alpha(2) := 2$ and so on, up to the last block eventually consisting of the letter $\alpha(m) := m$.) So, more generally, one could investigate the longest common subsequences where letters are aligned in blocks of letters $\alpha(1), \ldots, \alpha(l)$, for any $l \ge m$, and where $\alpha : \{1, \ldots, l\} \to \mathcal{A}_m$ is onto. For any fixed α, the length of the longest common subsequences where letters are aligned with blocks α is at most equal to LC_n, the length of the longest common subsequences, and moreover, LC_n is the maximum of these lengths over all the possible block-orders α (l is not fixed). To pass from the block version to LC_n, there is, however, a major issue of iterated limits. In what follows, at first, we merely give for any fixed α, the limiting law of the length of the (rescaled) longest common subsequences where letters are aligned in blocks $\alpha(1), \ldots, \alpha(l)$, and then the corresponding limiting laws, when allowing for a fixed numbers of such blocks.

Firstly, defining for any $k \in \mathbb{N}$, $k \ge 2$, $\Lambda_k := \{\lambda \in (\mathbb{R}_+)^k =: \lambda_1 + \cdots + \lambda_k = 1\}$, we claim that:

$$\max_{\lambda \in \Lambda_l^2} \sum_{i=1}^{l} \left[\left(p_{\alpha(i)}^X \lambda_{\alpha(i)}^X \right) \wedge \left(p_{\alpha(i)}^Y \lambda_{\alpha(i)}^Y \right) \right] = \max_{\lambda \in \Lambda_m^2} \sum_{i=1}^{m} \left[\left(p_i^X \lambda_i^X \right) \wedge \left(p_i^Y \lambda_i^Y \right) \right]. \tag{1.3.1}$$

Indeed to see the validity of this equality, note that above the left-hand side is greater or equal than the right-hand side since α is onto, while it is also less or equal since we can partition $\{1, \ldots, l\}$ via $\alpha^{-1}(\{1\}), \alpha^{-1}(\{2\}), \ldots, \alpha^{-1}(\{m\})$ and use the basic inequality $(a \wedge b) + (c \wedge d) \le (a+c) \wedge (b+d)$.

Next, to adapt the proof of our main theorem, we need to define the set U^α, as well as all other quantities which depended on m or p, with l instead of m and $p^\bullet_{\alpha(1)}, \dots, p^\bullet_{\alpha(l)}$ instead of $p^\bullet_1, \dots, p^\bullet_m$. Note also that, when $l > m$, the quantities $p^\bullet_{\alpha(1)}, \dots, p^\bullet_{\alpha(l)}$ do not form a probability mass function (their sum is not equal to one), but all their elements are positive which is enough to have everything well defined.

Formally, for example,

$$U^\alpha := \left\{ u \in \mathbb{R}^l_+ \, : \, \frac{u_1}{p^X_{\alpha(1)}} + \cdots + \frac{u_l}{p^X_{\alpha(l)}} \leq 1, \frac{u_1}{p^Y_{\alpha(1)}} + \cdots + \frac{u_l}{p^Y_{\alpha(l)}} \leq 1 \right\},$$

$\phi^\alpha : \mathbb{R}^l \to \mathbb{R}$ is given by

$$\phi^\alpha : u \mapsto u_1 + \cdots + u_l,$$

and I^α is now defined to be the set of integers $i \in \{1, \dots, l\}$ such that there exists $u^i \in L_{U^\alpha}$ with $u^i > 0$. Using almost the same proof as the one showing the equality of the two maxima in (1.3.1), we get $\alpha^{-1}(I) = I^\alpha$, where I is defined as before. There is no need to redefine the various cases a), b1), b2) here since they coincide with those previously defined when taking $p^\bullet_{\alpha(1)}, \dots, p^\bullet_{\alpha(l)}$ instead of $p^\bullet_1, \dots, p^\bullet_m$. For example, "there exists $u \in U^\alpha$ maximizing ϕ^α over U^α such that $\frac{u_1}{p^X_{\alpha(1)}} + \cdots + \frac{u_l}{p^X_{\alpha(l)}} = 1$ and $\frac{u_1}{p^Y_{\alpha(1)}} + \cdots + \frac{u_l}{p^Y_{\alpha(l)}} < 1$" is equivalent to Case a) defined in Section 1.1.3. Finally, the function $\mathfrak{m}$ defined in Lemma 1.1.5 can be extended naturally to $\left(\mathbb{R}^l\right)^2$.

Within this generalized setting, the proof of Lemma 1.2.2 carries over, giving us the following theorem for, LC^α_n, the length of the longest common subsequences with blocks $\alpha(1), \dots, \alpha(l)$.

Theorem 1.3.1. *Let B^X and B^Y be two independent $|I|$-dimensional Brownian motions defined on $[0,1]$ with respective covariance matrix C^X defined by $C^X_{i,i} = 1$ and $C^X_{i,j} = -\sqrt{\frac{p^X_{\alpha(i)} p^X_{\alpha(j)}}{(1-p^X_{\alpha(i)})(1-p^X_{\alpha(j)})}}$, for $i \neq j$ in I, and C^Y defined in a similar fashion. For all $\lambda \in K^\alpha_{\Lambda^2}$ and $i \in I^\alpha$, set*

$$V^{\alpha,X}_i(\lambda^X) = \sqrt{p^X_{\alpha(i)}(1-p^X_{\alpha(i)})} \left(B^X_{\alpha(i)} \left(\sum_{j=1}^{i} \lambda^X_j \right) - B^X_{\alpha(i)} \left(\sum_{j=1}^{i-1} \lambda^X_j \right) \right),$$

$$V^{\alpha,Y}_i(\lambda^Y) = \sqrt{p^Y_{\alpha(i)}(1-p^Y_{\alpha(i)})} \left(B^Y_{\alpha(i)} \left(\sum_{j=1}^{i} \lambda^Y_j \right) - B^Y_{\alpha(i)} \left(\sum_{j=1}^{i-1} \lambda^Y_j \right) \right).$$

If there exists $u \in L_{U^\alpha}$ such that $\frac{u_1}{p^X_{\alpha(1)}} + \cdots + \frac{u_l}{p^X_{\alpha(l)}} = 1$ and $\frac{u_1}{p^Y_{\alpha(1)}} + \cdots + \frac{u_l}{p^Y_{\alpha(l)}} < 1$, or equivalently if there exists $u \in L_U$ such that $\frac{u_1}{p^X_1} + \cdots + \frac{u_m}{p^X_1} = 1$ and $\frac{u_1}{p^Y_1} + \cdots + \frac{u_m}{p^Y_m} < 1$ (Case a)), then

$$\frac{LC^\alpha_n - n e_{\max}}{\sqrt{n}} \underset{n\to\infty}{\Longrightarrow} Z^a := \max_{\lambda^X \in J^\alpha} \sum_{i \in I^\alpha} V^{\alpha,X}_i(\lambda^X).$$

If for all $u \in L_{U^\alpha}$, $\frac{u_1}{p^X_{\alpha(1)}} + \cdots + \frac{u_l}{p^X_{\alpha(l)}} = 1$ and $\frac{u_1}{p^Y_{\alpha(1)}} + \cdots + \frac{u_l}{p^Y_{\alpha(l)}} = 1$, or equivalently if for all $u \in L_U$, $\frac{u_1}{p^X_1} + \cdots + \frac{u_m}{p^X_1} = 1$ and $\frac{u_1}{p^Y_1} + \cdots + \frac{u_m}{p^Y_m} = 1$ (Case b)), then

$$\frac{LC^\alpha_n - n e_{\max}}{\sqrt{n}} \underset{n\to\infty}{\Longrightarrow} Z^b := \max_{\lambda \in K^\alpha_{\Lambda^2}} \mathfrak{m}\left(V^{\alpha,X}(\lambda^X), V^{\alpha,Y}(\lambda^Y) \right),$$

where, again, now $\mathfrak{m}$ is defined on $\left(\mathbb{R}^l\right)^2$.

For instance, for $m = 2$ and in the uniform case, the order $\alpha(1) = 2, \alpha(2) = 1, \alpha(3) = 2$ gives the limiting distribution:

$$\frac{LC^\alpha_n - n e_{\max}}{\sqrt{n}} \underset{n\to\infty}{\Longrightarrow} Z^b := \max_{\substack{\lambda^X_1 + \lambda^X_2 + \lambda^X_3 = 1 \\ \lambda^Y_1 + \lambda^Y_2 + \lambda^Y_3 = 1}} \mathfrak{m}\left(V^{\alpha,X}(\lambda^X), V^{\alpha,Y}(\lambda^Y) \right),$$

i.e.,

$$\frac{LC_n^\alpha - ne_{\max}}{\sqrt{n}} \underset{n\to\infty}{\Longrightarrow} Z^b := \frac{1}{2} \max_{\substack{\lambda_1^X+\lambda_2^X+\lambda_3^X=1 \\ \lambda_1^Y+\lambda_2^Y+\lambda_3^Y=1}} \min_{\bullet\in\{X,Y\}} (B_2^\bullet(\lambda_1^\bullet) + B_1^\bullet(\lambda_1^\bullet+\lambda_2^\bullet) - B_1^\bullet(\lambda_1^\bullet) + B_2^\bullet(1) - B_2^\bullet(\lambda_1^\bullet+\lambda_2^\bullet)).$$

Also note that, sometimes, the limit in the above theorem is simply a normal random variable. Indeed, take $p_1^X = 1/3, p_2^X = 2/3, p_1^Y = 1/4, p_2^Y = 3/4$, and $\alpha(1) = 1, \alpha(2) = 2$, then we are in Case a), $I = \{2\}$ and:

$$\frac{LC_n^\alpha - ne_{\max}}{\sqrt{n}} \underset{n\to\infty}{\Longrightarrow} Z^a := \frac{\sqrt{2}}{3} B_2^X(1).$$

This is also, as one would expect, the limiting distribution of the number of 2's in the first word (which is almost equal to LC_n^α). However, if we take $\alpha(1) = 2, \alpha(2) = 1, \alpha(3) = 2$, the limit is more involved.

For $b \in \mathbb{N}$ such that $b \geq m$, let now F_m^b denote the set of all surjections from $\{1,\dots,b\}$ to $\{1,\dots,m\}$, and let $LC_n^{(b)}$ be the length of the longest common subsequences with $b \geq m$ blocks, with for each letter at least one block of this letter, and still allowing the blocks to have size zero. This is nothing but the maximum, over all the possible $\alpha \in F_m^b$, of LC_n^α, so, recalling the discussion preceding the statement of Theorem 1.3.1, we have:

Theorem 1.3.2. *In Case a),*

$$\frac{LC_n^{(b)} - ne_{\max}}{\sqrt{n}} \underset{n\to\infty}{\Longrightarrow} Z^a := \max_{\substack{\lambda^X\in J^\alpha \\ \alpha\in F_m^b}} \sum_{i\in I^\alpha} V_i^{\alpha,X}(\lambda^X).$$

In Case b),

$$\frac{LC_n^{(b)} - ne_{\max}}{\sqrt{n}} \underset{n\to\infty}{\Longrightarrow} Z^b := \max_{\substack{\lambda\in K_{\Lambda^2}^\alpha \\ \alpha\in F_m^b}} \mathfrak{m}\left(V^{\alpha,X}(\lambda^X), V^{\alpha,Y}(\lambda^Y)\right).$$

Proof. The proof of this theorem follows lines of the proof of our previous main result, considering $p_{\alpha(i)}^\bullet$ instead of $p_i^\bullet$. □

Note that LC_n, the length of the longest common subsequences without any conditions on blocks, corresponds to $LC_n^{(n+m)}$ (or to be more precise, $LC_n^{(b)}$ for any $b \geq m+n-2$: this is because when, say, there are only two kind of letters involved in the longest common word, we have to take $m-2$ additional empty blocks to make α onto). Although the above theorem requires a fixed number of blocks, say, b, it is nevertheless noteworthy that no matter this fixed number,

$$\lim_{n\to+\infty} \frac{\mathbb{E}LC_n^{(b)}}{n} = e_{\max}.$$

1.3.3 Countably infinite alphabet

To continue, let us consider, as in [32, Section 4], the generalization to countably infinite alphabets. Let the alphabet be $\mathbb{N}^* = \{1,2,\dots\}$, let $(p_i'^X)_{i\geq1}$ and $(p_i'^Y)_{i\geq1}$ be two probability mass functions

on this alphabet, we are now interested in LCI_n^∞, the length of the longest common and increasing subsequences over this countably infinite alphabet. Let

$$\Lambda^\infty = \left\{ \lambda \in (\mathbb{R}_+)^{\mathbb{N}^*} = [0,+\infty)^{\mathbb{N}^*} \ : \ \sum_{i=1}^{+\infty} \lambda_i = 1 \right\},$$

and let

$$e_{\max}^\infty = \sup_{\lambda \in (\Lambda^\infty)^2} \sum_{i=1}^{+\infty} \left[\left(p_i'^X \lambda_i^X \right) \wedge \left(p_i'^Y \lambda_i^Y \right) \right].$$

Let $m \in \mathbb{N}, m \geq 2$ be such that $\sum_{i=m}^{+\infty} p_i'^X < e_{\max}^\infty$ and $\sum_{i=m}^{+\infty} p_i'^Y < e_{\max}^\infty$. Let us consider the distributions over $\{1,\ldots,m\}$ obtained by replacing all the letters greater or equal to m by m, namely, let $p_i^X = p_i'^X$ for $i < m$ and $p_m^X := \sum_{i=m}^{+\infty} p_i'^X$, and let p_i^Y, $1 \leq i \leq m$, be defined in a similar fashion. Let now LCI_n be the length of the longest increasing subsequences formed by replacing all the letters greater or equal to m by m, i.e., the longest common and increasing subsequences on $\{1,\ldots,m\}$ associated with the probability mass functions p'^X and p'^Y. Next we argue, via a sandwiching argument, that when properly centered and scaled (note that $e_{\max}^\infty = e_{\max}$), LCI_n^∞ and LCI_n tend to the same limit. Indeed, let LCI_n^* be the length of the longest common and increasing subsequences not using the letter m, i.e., the length of the longest common and increasing subsequences on $\{1,\ldots,m-1\}$ associated with the probability mass functions p'^X and p'^Y or, equivalently, p^X and p^Y. Since $m \notin I$ (where I is defined with the distribution $(p_i^X)_{1\leq i\leq m}$ and $(p_i^Y)_{1\leq i\leq m}$), $(LCI_n^* - ne_{\max})/\sqrt{n}$ and $(LCI_n - ne_{\max})/\sqrt{n}$ converge to the same limiting distribution. But,

$$\frac{LCI_n^* - ne_{\max}}{\sqrt{n}} \leq \frac{LCI_n^\infty - ne_{\max}}{\sqrt{n}} \leq \frac{LCI_n - ne_{\max}}{\sqrt{n}},$$

completing the proof.

From the proofs presented above, the passage from two to three or more sequences is clear: the minimum over two Brownian functionals becomes a minimum over three or more Brownian functionals, and such a passage applies to the cases touched upon above and below.

Throughout the text, the two sequences $(X_k)_{k\geq 1}$ and $(Y_k)_{k\geq 1}$ are assumed to be independent with respective i.i.d. components. In view of [34] or [30], one expects that the i.i.d. assumption could be replaced by a Markovian one or even a hidden Markovian one. Moreover, one further expects that the independence of the two sequences is unnecessary and that a potential dependence structure between the two sequences would carry over to corresponding $2m$-dimensional Brownian functionals, another case at hand could be the hidden Markov framework. Finally, it should also be of interest (as already done in [12] for uniform letters) to study the ramifications/connections of our results with last passage percolation.

Appendix: proof of Lemma 1.1.5

Proof. Define $f_\nu : E'^2 \to \mathbb{R}$ by $f_\nu : x \mapsto \sum_{i=1}^m \left[\left(p_i^X x_i^X + \nu_i^X \right) \wedge \left(p_i^Y x_i^Y + \nu_i^Y \right) \right]$. In order to prove that $\mathfrak{m}(\nu)$ is well defined and (1.1.8), it is enough to prove that for all $x \in E'^2$, there exists $x' \in E'^2$ such that $\|x'\|_\infty \leq 2Cm\|\nu\|_\infty$ and $f_\nu(x') \geq f_\nu(x)$. Let $x \in E'^2$. Firstly, assume that $x \in P$ (recalling (1.2.6)). If $f_\nu(x) < f_\nu(0)$, taking $x' = 0$ works, so assume $f_\nu(x) \geq f_\nu(0)$. By (1.1.2) (applied twice),

$$-m\|\nu\|_\infty \leq f_\nu(0) \leq f_\nu(x) \leq m\|\nu\|_\infty + f(x)$$

hence $-f(x) \leq 2m\|\nu\|_\infty$ and, by Lemma 1.2.3, there exists $x^{K\cap P} \in K \cap P$ and $x^r \in E^2$ such that $x = x^{K\cap P} + x^r$ and $\|x^r\|_\infty \leq -Cf(x) \leq 2Cm\|\nu\|_\infty$. But from the definition of K, $f_\nu(x^{K\cap P} + x^r) = f(x^{K\cap P}) + f_\nu(x^r)$, and by (1.2.8), $f(x^{K\cap P}) = 0$ so $f_\nu(x) = f_\nu(x^r)$. Moreover, since $x \in P$ and $x_i^{K\cap P,\bullet} = 0$ for all $i \in I^c$, $x^r \in E'^2$.

Now, if we do not assume $x \in P$ anymore, observe that for $\varepsilon > 0$ small enough, $\varepsilon x \in P$, so $f_{\varepsilon\nu}(x') \geq f_{\varepsilon\nu}(\varepsilon x)$ for some $x' \in E'^2$ such that $\|x'\|_\infty \leq 2Cm\|\varepsilon\nu\|_\infty$. Finally, dividing by ε, $f_\nu((1/\varepsilon)x') \geq f_\nu(x)$ where $\|(1/\varepsilon)x'\|_\infty \leq 2Cm\|\nu\|_\infty$.

In Case b1), let us begin with the subcase $I = \{1\}$. In this instance, $p_1^X = p_1^Y = e_{\max}$, while for all $1 < i \leq m$, $p_i^X < e_{\max}$ or $p_i^Y < e_{\max}$ (otherwise i would be in I). We now show that "the maximum of f_ν is realized with the first letter plus one other letter", more precisely, there exists $x \in E'^2$ such that $f_\nu(x) = \mathfrak{m}(\nu)$ and $|\{i \in \{2,\dots,m\} : x_i^X \neq 0 \text{ or } x_i^Y \neq 0\}| \leq 1$. Indeed, using the same method than in the proof of Lemma 1.1.6, keeping in mind $\nu_2^\bullet = \cdots = \nu_m^\bullet = 0$, one can see that there exists some x maximizing f_ν such that $\{i \in \{1,\dots,m\} : x_i^X \neq 0 \text{ or } x_i^Y \neq 0\}$ has at most two elements, and they can't both belong to $\{2,\dots,m\}$ otherwise they would be null (by the definition of E').

Returning to the proof of the lemma, we have shown that

$$\max_{x\in E'^2} f_\nu(x) = \max_{i_0\in\{2,\dots,m\}} \sup_{\substack{x\in E'^2\\ \forall i\in\{2,\dots,m\}\setminus\{i_0\}, x_i^\bullet=0}} f_\nu(x).$$

Fixing $i_0 \in \{2,\dots,m\}$, we have

$$\sup_{\substack{x\in E'^2\\ \forall i\in\{2,\dots,m\}\setminus\{i_0\}, x_i^\bullet=0}} f_\nu(x) = \sup_{t^X,t^Y>0} \left[(\nu_1^X - e_{\max}t^X) \wedge (\nu_1^Y - e_{\max}t^Y) + (p_{i_0}^X t^X) \wedge (p_{i_0}^Y t^Y)\right].$$

It is then easily seen that this last supremum does not change with the additional condition $p_{i_0}^X t^X = p_{i_0}^Y t^Y$. (Indeed, if, for example, $p_{i_0}^X t^X > p_{i_0}^Y t^Y$, reducing t^X to transform this strict inequality into equality will only increase the sum of the two minima in the definition of f_ν.) Hence,

$$\begin{aligned}\sup_{\substack{x\in E'^2\\ \forall i\in\{2,\dots,m\}\setminus\{i_0\}, x_i^\bullet=0}} f_\nu(x) &= \sup_{t^X>0} \left[\left(\nu_1^X - e_{\max}t^X + p_{i_0}^X t^X\right) \wedge \left(\nu_1^Y - e_{\max}\frac{p_{i_0}^X}{p_{i_0}^Y}t^X + p_{i_0}^X t^X\right)\right] \\ &= \sup_{t^X>0} \left[\left(\nu_1^X + (p_{i_0}^X - e_{\max})t^X\right) \wedge \left(\nu_1^Y + \frac{p_{i_0}^X}{p_{i_0}^Y}(p_{i_0}^Y - e_{\max})t^X\right)\right].\end{aligned}$$

Since $i_0 \notin I$, it is impossible for both $p_{i_0}^X - e_{\max}$ and $p_{i_0}^Y - e_{\max}$ to be positive, so this last supremum is attained at $t^X = 0$ (and is equal to $\nu_1^X \wedge \nu_1^Y$) unless $\nu_1^X < \nu_1^Y$ and $p_{i_0}^X - e_{\max} > 0$, or $\nu_1^X > \nu_1^Y$ and $p_{i_0}^Y - e_{\max} > 0$, in which case the supremum is attained at $t^X = \frac{p_{i_0}^Y}{e_{\max}} \frac{\nu_1^Y - \nu_1^X}{p_{i_0}^X - p_{i_0}^Y}$, a value at which the two sides in the above minimum are equal to each other. So if $\nu_1^X < \nu_1^Y$ and $p_{i_0}^X - e_{\max} > 0$, or $\nu_1^X > \nu_1^Y$ and $p_{i_0}^Y - e_{\max} > 0$, then

$$\sup_{\substack{x\in E'^2\\ \forall i\in\{2,\dots,m\}\setminus\{i_0\}, x_i^\bullet=0}} f_\nu(x) = \frac{p_{i_0}^X(e_{\max} - p_{i_0}^Y)}{e_{\max}(p_{i_0}^X - p_{i_0}^Y)}\nu_1^X + \frac{p_{i_0}^Y(p_{i_0}^X - e_{\max})}{e_{\max}(p_{i_0}^X - p_{i_0}^Y)}\nu_1^Y.$$

Assuming that $\nu_1^X < \nu_1^Y$, we see that in this case $\mathfrak{m}(\nu^X, \nu^Y) = s_X S^Y + t_X S^X$. This remains true if $S^X = S^Y$ (in this case, $\mathfrak{m}(\nu^X, \nu^Y) = S^X = S^Y$), and, similarly, when $S^Y \leq S^X$. The proof of Case b1) is therefore done when $I = \{1\}$.

Still in Case b1), but without the assumption that $I = \{1\}$, assume, without loss of generality, that $I = \{1,\dots,k\}$, $k \geq 2$. Define $\tilde{\nu}$ by $\tilde{\nu}_1^\bullet = S^\bullet$ and $\tilde{\nu}_i^\bullet = 0$, for all $i \geq 2$. Let $x^0 \in E'^2$ be defined by $x^{0,Y} = 0$, $x_1^{0,X} = (S^X - S^Y + \nu_1^Y - \nu_1^X)/e_{\max}$, $x_i^{0,X} = (\nu_i^Y - \nu_i^X)/e_{\max}$, for all $i \in \{2,\dots,k\}$, and $x_i^{0,\bullet} = 0$ for all $i \in \{k+1,\dots,m\}$. Note that for all $x \in E'^2$, $f_\nu(x + x^0) = f_{\tilde{\nu}}(x)$, so $\mathfrak{m}(\nu) = \mathfrak{m}(\tilde{\nu})$. Moreover, defining x' via $x_1'^\bullet = x_1^\bullet + \cdots + x_k^\bullet$, $x_i'^\bullet = 0$, for $i \in \{2,\dots,k\}$, and $x_i'^\bullet = x_i^\bullet$ everywhere

else, we have $x' \in E'^2$, and

$$\big(e_{\max}(x_1^X + \cdots + x_k^X) + \tilde{\nu}_1^X\big) \wedge \big(e_{\max}(x_1^Y + \cdots + x_k^Y) + \tilde{\nu}_1^Y\big) \geq (e_{\max} x_1^X + \tilde{\nu}_1^X) \wedge (e_{\max} x_1^Y + \tilde{\nu}_1^Y) + e_{\max}(x_2^X + \cdots + x_k^X) \wedge (x_2^Y + \cdots + x_k^Y),$$

$$\big(e_{\max}(x_1^X + \cdots + x_k^X) + \tilde{\nu}_1^X\big) \wedge \big(e_{\max}(x_1^Y + \cdots + x_k^Y) + \tilde{\nu}_1^Y\big) \geq (e_{\max} x_1^X + \tilde{\nu}_1^X) \wedge (e_{\max} x_1^Y + \tilde{\nu}_1^Y) + e_{\max}(x_2^X \wedge x_2^Y) + \cdots + (x_k^X \wedge x_k^Y).$$

Hence, $f_{\tilde{\nu}}(x') \geq f_{\tilde{\nu}}(x)$, and therefore

$$\mathfrak{m}(\tilde{\nu}) = \max_{\substack{x \in E'^2 \\ \forall i \in \{2,\ldots,k\}, x_i^\bullet = 0}} f_{\tilde{\nu}}(x).$$

Now applying the subcase $I = \{1\}$ concludes the proof of Case b1).

In Case b2), again assume without loss of generality that $I = \{1, \ldots, k\}$, $k \geq 2$. Let $L_1 = (1, 0, \ldots, 0, -1, 0, \ldots, 0) \in \mathbb{R}^{2k}$, having $k-1$ zeros between the two non-zero coordinates, let $L_2 = (0, 1, 0, \ldots, 0, -1, 0, \ldots, 0)$ (still with $k-1$ zeros between the two non-zero coordinates), and iterate this process up to L_k. Let also $\widetilde{P^X}$ be the concatenation of $P^X \in \mathbb{R}^k$ with $0 \in \mathbb{R}^k$, and let $\widetilde{P^Y}$ be the concatenation of $0 \in \mathbb{R}^k$ with $P^Y \in \mathbb{R}^k$. The vectors $L_1, \ldots, L_k, \widetilde{P^X}, \widetilde{P^Y}$ are linearly independent since, as already seen in Lemma 1.1.4, P^X and P^Y are linearly independent. Now, let Q be a $2k \times 2k$ invertible matrix with first rows $L_1, \ldots, L_k, \widetilde{P^X}, \widetilde{P^Y}$ (for example, to form such a matrix Q, one could complete the first columns with vectors from the canonical basis), let $\Delta \in \mathbb{R}^{2k}$ be defined by

$$\Delta_i := \begin{cases} \nu_i^Y - \nu_i^X & \text{if } i \in \{1, \ldots, k\} \\ 0, & \text{if } i \in \{k+1, \ldots, 2k\}, \end{cases}$$

and let $u \in \mathbb{R}^{2k}$ be defined by

$$u_i := \begin{cases} (Q^{-1}\Delta)_i & \text{if } i \in \{1, \ldots, k\} \\ 0, & \text{if } i \in \{k+1, \ldots, 2k\}. \end{cases}$$

We have $u_i^X - u_i^Y = \nu_i^Y - \nu_i^X$ (where u^X is the vector of the first k coordinates of u and u^Y the vector of the last k coordinates of u) for all $i \in \{1, \ldots, k\}$: these conditions stem from the rows $L_1, \ldots, L_k$. Moreover, $u_1^X/p_1^X + \cdots + u_m^X/p_m^X = u_1^Y/p_1^Y + \cdots + u_m^Y/p_m^Y = 0$ (conditions stemming from the rows $\widetilde{P^X}, \widetilde{P^Y}$). Then, expand u^X and u^Y to $\mathbb{R}^m$ by filling with zeros, so that $u := (u^X, u^Y)$ is now in $(\mathbb{R}^m)^2$. Setting, for all $i \in \{1, \ldots, m\}$, $y_i^X := u_i^X/p_i^X, y_i^Y := u_i^Y/p_i^Y$, lead to $y \in (\mathbb{R}^m)^2$, more precisely $y \in E'^2$ such that for all $i \in \{1, \ldots, m\}, p_i^X y_i^X + \nu_i^X = p_i^Y y_i^Y + \nu_i^Y$, with moreover

$$\sum_{i=1}^m \big[\big(p_i^X y_i^X + \nu_i^X\big) \wedge \big(p_i^Y y_i^Y + \nu_i^Y\big)\big] = \sum_{i \in I} \left(\frac{p_i^X y_i^X + \nu_i^X + p_i^Y y_i^Y + \nu_i^Y}{2}\right).$$

Setting $U^X := (u_i^X)_{i \in I} \in \mathbb{R}^k$, $U^Y := (u_i^Y)_{i \in I}$, $R^X := (\nu_i^X)_{i \in I}$ and $R^Y := (\nu_i^Y)_{i \in I}$, the above expression becomes

$$\sum_{i=1}^m \big[\big(p_i^X y_i^X + \nu_i^X\big) \wedge \big(p_i^Y y_i^Y + \nu_i^Y\big)\big] = \frac{1}{2}(U^X + R^X + U^Y + R^Y) \cdot (1)_{i \in I}.$$

With the notations of Lemma 1.1.4,

$$\begin{aligned} U^X \cdot (1)_{i \in I} &= U^X \cdot (sP^X + tP^Y) \\ &= U^X \cdot tP^Y \\ &= (U^X - U^Y) \cdot tP^Y \\ &= (R^Y - R^X) \cdot tP^Y. \end{aligned}$$

Similarly, $U^Y \cdot (1)_{i\in I} = (R^X - R^Y) \cdot sP^X$. So,

$$\begin{aligned}\sum_{i=1}^{m} \left[\left(p_i^X y_i^X + \nu_i^X\right) \wedge \left(p_i^Y y_i^Y + \nu_i^Y\right)\right] &= \frac{1}{2}(R^X - R^Y) \cdot (sP^X - tP^Y) \\ &\quad + \frac{1}{2}(R^X + R^Y) \cdot (sP^X + tP^Y) \\ &= R^X \cdot sP^X + R^Y \cdot tP^Y \\ &= \sum_{i\in I} \left(\frac{s}{p_i^X}\nu_i^X + \frac{t}{p_i^Y}\nu_i^Y\right).\end{aligned}$$

This shows that

$$\max_{x\in E'^2} \sum_{i=1}^{m} \left[\left(p_i^X x_i^X + \nu_i^X\right) \wedge \left(p_i^Y x_i^Y + \nu_i^Y\right)\right] \geq \sum_{i\in I} \left(s\nu_i^X/p_i^X + t\nu_i^Y/p_i^Y\right).$$

Now let $x \in E'^2$,

$$\begin{aligned}&\sum_{i=1}^{m}\left[\left(p_i^X x_i^X + \nu_i^X\right) \wedge \left(p_i^Y x_i^Y + \nu_i^Y\right)\right] - \sum_{i\in I}\left(\frac{s}{p_i^X}\nu_i^X + \frac{t}{p_i^Y}\nu_i^Y\right) \\ &\qquad = \sum_{i=1}^{m}\left[\left(p_i^X x_i^X + \nu_i^X\right) \wedge \left(p_i^Y x_i^Y + \nu_i^Y\right)\right] - \sum_{i=1}^{m}\left[\left(p_i^X y_i^X + \nu_i^X\right) \wedge \left(p_i^Y y_i^Y + \nu_i^Y\right)\right] \\ &\qquad = \sum_{i=1}^{m}\left[\left(p_i^X (x-y)_i^X\right) \wedge \left(p_i^Y (x-y)_i^Y\right)\right] \\ &\qquad = f(x-y).\end{aligned}$$

We have $x - y \in E'^2$ (recall, also, that $y_i = 0$ for all $i \in I^c$), so for some $c > 0$, $(x-y)/c \in P$, and then $f((x-y)/c) \leq 0$, so $f(x-y) \leq 0$. Hence $\sum_{i=1}^{m}\left[\left(p_i^X x_i^X + \nu_i^X\right) \wedge \left(p_i^Y x_i^Y + \nu_i^Y\right)\right] - \sum_{i\in I}\left(s\nu_i^X/p_i^X + t\nu_i^Y/p_i^Y\right) \leq 0$ and, finally,

$$\max_{x\in E'^2} \sum_{i=1}^{m} \left[\left(p_i^X x_i^X + \nu_i^X\right) \wedge \left(p_i^Y x_i^Y + \nu_i^Y\right)\right] = \sum_{i\in I} \left(s\nu_i^X/p_i^X + t\nu_i^Y/p_i^Y\right).$$

□

Chapter 2

Variance Bounds: Some Old and Some New

As mentioned in the introduction of this work, the problem of finding the asymptotic behavior of the variance of the length of the longest common subsequences of random word was the motivation to introduce some tools to study the variance, that we now present in the most general framework. For functions of independent random variables, various upper and lower variance bounds are revisited in diverse settings. These are then specialized to the Bernoulli, Gaussian, infinitely divisible cases and to Banach space valued random variables. Frameworks and techniques vary from jackknives through semigroups and beyond. Some new applications are presented, recovering and improving, in particular, all the known estimates on the variance of the length of the longest common subsequences of two random words.

2.1 Introduction and preliminary results

We revisit below various lower and upper bounds on the variance of functions of independent random variables. Throughout and unless otherwise noted, $X_1, \ldots, X_n, X'_1, \ldots, X'_n$ are independent random variables such that for all $k \in \{1, \ldots, n\}$, X_k and X'_k are identically distributed, while $S : \mathbb{R}^n \to \mathbb{R}$ is a Borel function such that $\mathbb{E}S(X_1, \ldots, X_n)^2 < +\infty$. Next, and if S is short for $S(X_1, \ldots, X_n)$, for any $k \in \{1, \ldots, n\}$, let $S^k := S(X_1, \ldots, X_{k-1}, X'_k, X_{k+1}, \ldots, X_n)$ and more generally if $\alpha \subset \{1, \ldots, n\}$, let S^α be defined as $S(X_1, \ldots, X_n)$ but with X_k replaced by X'_k for all $k \in \alpha$. With these preliminary notations, we next recall the definitions of various quantities which will play an important role in the sequel.

Following [8], for $k \in \{1, \ldots, n\}$, let

$$B_k := \mathbb{E}\frac{1}{n!}\sum_{i \in \mathfrak{S}_n} S(S^{i_1,\ldots,i_{k-1}} - S^{i_1,\ldots,i_k}), \tag{2.1.1}$$

where $\mathfrak{S}_n$ is the symmetric group of degree n and where for $k = 1, S^{i_1,\ldots,i_{k-1}} = S$. As the following sum is telescopic:

$$\sum_{k=1}^{n} B_k = \mathbb{E}\frac{1}{n!}\sum_{i \in \mathfrak{S}_n} S(S - S^{i_1,\ldots,i_n}) = \operatorname{Var} S.$$

One key fact motivating the definition of the B_k's is that they can be rewritten as:

$$B_k = \mathbb{E}\frac{1}{2n!}\sum_{i \in \mathfrak{S}_n} (S - S^{i_k})(S^{i_1,\ldots,i_{k-1}} - S^{i_1,\ldots,i_k}).$$

Indeed, if $\alpha, \beta \subset \{1, \ldots, n\}$,

$$\mathbb{E}\left(S^{\alpha} S^{\beta}\right) = \mathbb{E}\left(S S^{\alpha \Delta \beta}\right), \tag{2.1.2}$$

where Δ denotes the symmetric difference operator, so

$$\mathbb{E}(S - S^{i_k})(S^{i_1,\ldots,i_{k-1}} - S^{i_1,\ldots,i_k}) = 2\mathbb{E}\left(S S^{i_1,\ldots,i_{k-1}} - S S^{i_1,\ldots,i_k}\right).$$

Next, for all $k \in \{1, \ldots, n\}$, let $\Delta_k S := S - S^k$, and iterating this operator: for $k \neq \ell$, let $\Delta_{k,\ell} S := \Delta_k(\Delta_\ell S) = S - S^k - S^\ell + S^{k,\ell}$ (note the commutativity property: $\Delta_k(\Delta_\ell S) = \Delta_\ell(\Delta_k S)$). Iterating this process, let $\Delta_{i_1,\ldots,i_k} S := \Delta_k(\Delta_{i_1,\ldots,i_{k-1}} S)$. Using this notation, we have

$$B_k = \mathbb{E}\frac{1}{2n!} \sum_{i \in \mathfrak{S}_n} (\Delta_{i_k} S)(\Delta_{i_k} S)^{i_1,\ldots,i_{k-1}}, \tag{2.1.3}$$

and so $B_k \geq 0$ since if U, U' and V are independent with U and U' identically distributed, then for any function F such that $F(U, V)$ is integrable, $\mathbb{E}\left(F(U,V)F(U',V)\right) = \mathbb{E}\left(\mathbb{E}\left(F(U,V)|V\right)^2\right) \geq 0$. We are now ready to generalize the approach used to go from (2.1.1) to (2.1.3), leading to novel properties of the $B_k's$.

Lemma 2.1.1. *Let α, β be two disjoint subsets of $\{1, \ldots, n\}$. Then*

$$\mathbb{E}\left(S(\Delta_\alpha S)^\beta\right) = \frac{1}{2^{|\alpha|}} \mathbb{E}\left(\Delta_\alpha S (\Delta_\alpha S)^\beta\right). \tag{2.1.4}$$

Proof. Firstly, by a straightforward induction on $k := |\alpha|$, note that $\Delta_\alpha S = \sum_{\alpha' \subset \alpha} (-1)^{|\alpha'|} S^{\alpha'}$. Then, for any $\alpha' \subset \alpha$,

$$(-1)^{|\alpha'|} S^{\alpha'} (\Delta_\alpha S)^\beta = \sum_{\alpha'' \subset \alpha} (-1)^{|\alpha'|+|\alpha''|} S^{\alpha'} S^{\alpha'' \cup \beta},$$

and so using (2.1.2) (α and β are disjoint and $\alpha' \subset \alpha$ so $\alpha' \Delta(\alpha \cup \beta) = (\alpha' \Delta \alpha) \cup \beta$),

$$\mathbb{E}\left((-1)^{|\alpha'|} S^{\alpha'} (\Delta_\alpha S)^\beta\right) = \mathbb{E}\left(\sum_{\alpha'' \subset \alpha} (-1)^{|\alpha'|+|\alpha''|} S S^{(\alpha' \Delta \alpha'') \cup \beta}\right).$$

Since $\alpha'' \mapsto \alpha' \Delta \alpha''$ is just a permutation of the subsets of α and $(-1)^{\alpha' \Delta \alpha''} = (-1)^{|\alpha'|+|\alpha''|}$,

$$\mathbb{E}\left((-1)^{|\alpha'|} S^{\alpha'} (\Delta_\alpha S)^\beta\right) = \mathbb{E}\left(\sum_{\alpha'' \subset \alpha} (-1)^{|\alpha''|} S S^{\alpha'' \cup \beta}\right) = \mathbb{E}\left(S(\Delta_\alpha S)^\beta\right),$$

and so

$$\frac{1}{2^{|\alpha|}} \mathbb{E}\left(\Delta_\alpha S (\Delta_\alpha S)^\beta\right) = \frac{1}{2^{|\alpha|}} \sum_{\alpha' \subset \alpha} \mathbb{E}\left((-1)^{|\alpha'|} S^{\alpha'} (\Delta_\alpha S)^\beta\right) = \mathbb{E}\left(S(\Delta_\alpha S)^\beta\right).$$

□

Let T be the forward shift operator, i.e., for $k \in \{1, \ldots, n-1\}$, let $TB_k := B_{k+1}$ and let D be the backward discrete derivative: $D := Id - T$ (so for $k \in \{1, \ldots, n-1\}$, $DB_k = B_k - B_{k+1}$), and denote by D^ℓ ($\ell \geq 0$) its ℓ-th iteration. It is known (see [8]) that the finite sequence $(B_k)_{1 \leq k \leq n}$ is non-increasing. More can be said.

Theorem 2.1.2. *For all $\ell \geq 0$ and $k \in \{1, \ldots, n - \ell\}$,*

$$D^\ell B_k = \mathbb{E}\frac{1}{2^{\ell+1} n!} \sum_{i \in \mathfrak{S}_n} (\Delta_{i_1,\ldots,i_{\ell+1}} S)(\Delta_{i_1,\ldots,i_{\ell+1}} S)^{i_{\ell+2},\ldots,i_{k+\ell}}. \tag{2.1.5}$$

In particular, $D^\ell B_k \geq 0$, i.e., $(B_k)_{1 \leq k \leq n}$ is completely monotone (recall that $D = Id - T$).

Proof. With the previous lemma, it is enough to prove that for all $\ell \in \{0, \ldots, n-1\}$ and $k \in \{1, \ldots, n-\ell\}$,

$$D^\ell B_k = \mathbb{E}\frac{1}{n!}\sum_{i\in\mathfrak{S}_n} S(\Delta_{i_1,\ldots,i_{\ell+1}}S)^{i_{\ell+2},\ldots,i_{k+\ell}}. \tag{2.1.6}$$

This is done by induction on ℓ. When $\ell = 0$, (2.1.6) is just the very definition of B_k. Assume next that (2.1.6) holds for $\ell \in \{0, \ldots, n-2\}$. Let $k \in \{1, \ldots, n-(\ell+1)\}$. Then,

$$\begin{aligned} D^{\ell+1}B_k &= \mathbb{E}\frac{1}{n!}\sum_{i\in\mathfrak{S}_n} S(\Delta_{i_1,\ldots,i_{\ell+1}}S)^{i_{\ell+2},\ldots,i_{k+\ell}} - \mathbb{E}\frac{1}{n!}\sum_{i\in\mathfrak{S}_n} S(\Delta_{i_1,\ldots,i_{\ell+1}}S)^{i_{\ell+2},\ldots,i_{k+1+\ell}} \\ &= \mathbb{E}\frac{1}{n!}\sum_{i\in\mathfrak{S}_n} S(\Delta_{i_1,\ldots,i_{\ell+1}}S)^{i_{\ell+3},\ldots,i_{k+1+\ell}} - \mathbb{E}\frac{1}{n!}\sum_{i\in\mathfrak{S}_n} S(\Delta_{i_1,\ldots,i_{\ell+1}}S)^{i_{\ell+2},\ldots,i_{k+1+\ell}} \\ &= \mathbb{E}\frac{1}{n!}\sum_{i\in\mathfrak{S}_n} S(\Delta_{i_1,\ldots,i_{\ell+2}}S)^{i_{\ell+3},\ldots,i_{k+1+\ell}}, \end{aligned}$$

where in getting the second equality, the terms are reindexed. □

We wish now to study potential connections between the B_k's and jackknives operators J_k and K_k previously studied in [10]. For $Y \in \sigma(X_1, \ldots, X_n)$, i.e., Y measurable with respect to the σ-field generated by $X_1, \ldots, X_n$ and $i \in \{1, \ldots, n\}$, let $\mathbb{E}^{(i)}Y := \mathbb{E}(Y|X_1, \ldots, X_{i-1}, X_{i+1}, \ldots, X_n)$ and more generally for a subset α of $\{1, \ldots, n\}$, let

$$\mathbb{E}^\alpha Y := \mathbb{E}(Y|(X_i)_{i\notin\alpha}).$$

For $i \in \{1, \ldots, n\}$, let

$$\mathrm{Var}^{(i)}\, Y := \mathbb{E}^{(i)}Y^2 - (\mathbb{E}^{(i)}Y)^2$$

and iterating, for $i \in \mathfrak{S}_n$, let

$$\mathrm{Var}^{(i_1,\ldots,i_k)}\, Y := \mathbb{E}^{(i_1)}(\mathrm{Var}^{(i_2,\ldots,i_k)}\, Y) - \mathrm{Var}^{(i_2,\ldots,i_k)}(\mathbb{E}^{(i_1)}Y).$$

For $k \in \{1, \ldots, n\}$, let

$$J_k := \sum_{i_1\neq i_2\ldots\neq i_k} \mathrm{Var}^{(i_1,\ldots,i_k)}\, S,$$

and

$$K_k := \sum_{i_1\neq i_2\ldots\neq i_k} \mathrm{Var}^{i_1,\ldots,i_k}\, \mathbb{E}^{\overline{(i_1,\ldots,i_k)}}S,$$

where $\overline{(i_1, \ldots, i_k)} = (i_{k+1}, \ldots, i_n)$. For ease of notation, set also $J'_k := J_k/k!$ and $K'_k := K_k/k!$. The next lemma provides relationships between these quantities and the B_k's, it will allow us to get easily, and in a unified fashion, many of the known expressions involving the variance, along with some new ones.

Lemma 2.1.3. *Let α, β be two disjoint subsets of $\{1, \ldots, n\}$. Then*

$$\mathbb{E}\left(\mathrm{Var}^\alpha\, \mathbb{E}^\beta S\right) = \mathbb{E}\left(S(\Delta_\alpha S)^\beta\right). \tag{2.1.7}$$

Proof. This is straightforward by induction on the cardinality of α. □

Recalling (2.1.6), we get from (2.1.7) that for all $k \in \{1, \ldots, n\}$,

$$J'_k = \binom{n}{k} D^{k-1}B_1 \quad \text{and} \quad K'_k = \binom{n}{k} D^{k-1}B_{n-k+1}. \tag{2.1.8}$$

It is easy to check that for any finite sequence $(a_k)_{1\le k\le n}$ and any positive integers $k \in \{1,\dots,n\}$,

$$a_k = \sum_{j=0}^{k-1}(-1)^j\binom{k-1}{j}D^j a_1 = \sum_{j=0}^{n-k}\binom{n-k}{j}D^j a_{n-j}. \tag{2.1.9}$$

In particular, for all $k \in \{1,\dots,n\}$,

$$B_k = \sum_{j=0}^{k-1}(-1)^j\frac{\binom{k-1}{j}}{\binom{n}{j+1}}J'_{j+1}, \tag{2.1.10}$$

$$B_k = \sum_{j=0}^{n-k}\frac{\binom{n-k}{j}}{\binom{n}{j+1}}K'_{j+1}. \tag{2.1.11}$$

We can now connect the J'_k's and K'_k's to the variance.

Lemma 2.1.4. *For all $k \in \{1,\dots,n\}$,*

$$\operatorname{Var} S - J'_1 + J'_2 - \cdots + (-1)^k J'_k = (-1)^k \sum_{1\le i_1<\cdots<i_{k+1}\le n} D^k B_{i_1}, \tag{2.1.12}$$

$$\operatorname{Var} S - K'_1 - K'_2 - \cdots - K'_k = \sum_{1\le i_1<\cdots<i_{k+1}\le n} D^k B_{i_{k+1}}. \tag{2.1.13}$$

Proof. Let us prove (2.1.12) by induction on $k \in \{1,\dots,n\}$. For the base case:

$$\operatorname{Var} S - J'_1 = B_1 + \cdots + B_n - nB_1 = \sum_{j=2}^{n}(B_j - B_1) = -\sum_{j=2}^{n}\sum_{i=1}^{j-1} DB_i.$$

For the inductive step: assume it is true for $k \in \{1,\dots,n\}$. Then,

$$\begin{aligned}
\operatorname{Var} S - J'_1 + J'_2 - \cdots + (-1)^k J'_k + (-1)^{k+1} J'_{k+1} &= (-1)^k \left(\left(\sum_{1\le i_1<\cdots<i_{k+1}\le n} D^k B_{i_1}\right) - J'_{k+1}\right) \\
&= (-1)^k \sum_{1\le i_1<\cdots<i_{k+1}\le n} \left(D^k B_{i_1} - D^k B_1\right) \\
&= (-1)^k \sum_{1\le i_1<\cdots<i_{k+1}\le n}\ \sum_{1\le i_0<i_1} -D^{k+1} B_{i_0} \\
&= (-1)^{k+1} \sum_{1\le i_0<i_1<\cdots<i_{k+1}\le n} D^{k+1} B_{i_0}.
\end{aligned}$$

The proof of (2.1.13) is very similar and so it is omitted. The following proposition recovers and extends some of the results obtained in [10]. □

Proposition 2.1.5.

$$\operatorname{Var} S = J'_1 - J'_2 + \cdots + (-1)^{n-1} J'_n, \tag{2.1.14}$$

$$\operatorname{Var} S = K'_1 + K'_2 + \cdots + K'_n,$$

and for all $k \in \{1,\dots,n\}$,

$$K'_{k+1} \le (-1)^k \left(\operatorname{Var} S - J'_1 + J'_2 - \cdots + (-1)^k J'_k\right) \le J'_{k+1},$$

$$K'_{k+1} \le \operatorname{Var} S - K'_1 - K'_2 - \cdots - K'_k \le J'_{k+1}.$$

$$\operatorname{Var} S = J_1' - J_2' + \cdots + (-1)^{k-1} J_k' + (-1)^k \sum_{j=k+1}^{n} \binom{j-1}{k} K_j'. \tag{2.1.15}$$

$$\operatorname{Var} S = \frac{\binom{k}{1}}{\binom{n}{1}} J_1' - \frac{\binom{k}{2}}{\binom{n}{2}} J_2' + \cdots + (-1)^{k-1} \frac{\binom{k}{k}}{\binom{n}{k}} J_k' + \frac{\binom{n-k}{1}}{\binom{n}{1}} K_1' + \frac{\binom{n-k}{2}}{\binom{n}{2}} K_2' + \cdots + \frac{\binom{n-k}{n-k}}{\binom{n}{n-k}} K_{n-k}'. \tag{2.1.16}$$

Proof. Above, the first two equalities simply follow from the fact that the right-hand terms in Lemma 2.1.4 are zero when $k = n$. Then, the first two inequalities follow from Lemma 2.1.4 and the complete monotonicity of the B_k's: for $1 \leq i_1 \leq n-k$, $D^k B_{n-k} \leq D^k B_{i_1} \leq D^k B_1$. Let us turn to the identity (2.1.15).

Applying the inversion formula (2.1.9) to $(D^k B_i)_{1 \leq i \leq n-k}$, with $i \leq n-k$, we get

$$\begin{aligned}
\sum_{1 \leq i_1 < \cdots < i_{k+1} \leq n} D^k B_{i_1} &= \sum_{i=1}^{n-k} \binom{n-i}{k} D^k B_i \\
&= \sum_{i=1}^{n-k} \sum_{j=0}^{n-k-i} \binom{n-i}{k} \binom{n-k-i}{j} D^{k+j} B_{n-k-j} \\
&= \sum_{i=1}^{n-k} \sum_{j=0}^{n-k-i} \binom{n-i}{k} \binom{n-k-i}{j} \frac{K_{k+j+1}'}{\binom{n}{k+j+1}} \\
&= \sum_{j=0}^{n-1} \sum_{i=1}^{n-k-j} \binom{n-i}{k+j} \binom{k+j}{k} \frac{K_{k+j+1}'}{\binom{n}{k+j+1}} \\
&= \sum_{j=0}^{n-1} \binom{k+j}{k} K_{k+j+1}',
\end{aligned}$$

where the last equality stems from the hockey-stick formula and reindexing.

To finish, let us prove (2.1.16) which will follow from $\operatorname{Var} S = B_1 + \cdots + B_k + B_{k+1} + \cdots + B_n$. Indeed, the equality (2.1.14) remains valid for any sequence $(a_n)_{n \geq 1}$, namely, the same proof shows that

$$a_1 + \cdots + a_n = \binom{n}{1} D^0 a_1 - \binom{n}{2} D^1 a_1 + \cdots + (-1)^{n-1} \binom{n}{n} D^{n-1} a_1.$$

In particular,

$$\begin{aligned}
B_1 + \cdots + B_k &= \binom{k}{1} D^0 B_1 - \binom{k}{2} D^1 B_1 + \cdots + (-1)^{k-1} \binom{k}{k} D^{k-1} B_1 \\
&= \frac{\binom{k}{1}}{\binom{n}{1}} J_1' - \frac{\binom{k}{2}}{\binom{n}{2}} J_2' + \cdots + (-1)^{k-1} \frac{\binom{k}{k}}{\binom{n}{k}} J_k'.
\end{aligned}$$

The second part, $B_{k+1} + \cdots + B_n$, is treated similarly. □

The equality (2.1.16) could be of use to find the order of $\operatorname{Var} S$ as n tends to infinity. For example, if there is a constant $C > 1$ (independent of n) such that $J_2'(n) \leq C J_1'(n)$, then, taking $k = \lfloor \frac{n}{2C} \rfloor$ will lead to

$$\liminf_{n \to \infty} \frac{\operatorname{Var} S(n)}{J_1'(n)} \geq \frac{1}{4C}.$$

We have proved that the finite sequence $(B_k)_{1 \leq k \leq n}$ is completely monotone and we already knew from [8] that it is non-increasing, so it is natural to wonder if one could find further properties of the B_k's. On the other hand, one may also wonder whether or not $(K_k)_{1 \leq k \leq n}$ does satisfy any further property except, of course, from being non-negative. Both answers appear to be negative:

Proposition 2.1.6. *For any $a_1, \dots, a_n \geq 0$, there exists $S : \mathbb{R}^n \to \mathbb{R}$ a Borel function such that for all $k \in \{1, \dots, n\}$, $K_k = a_k$.*

Corollary 2.1.7. *If $(b_k)_{1 \leq k \leq n}$ is completely monotone, then there exists $S : \mathbb{R}^n \to \mathbb{R}$ a Borel function such that for all $k \in \{1, \dots, n\}$, $B_k = b_k$.*

Proof of the Corollary. It is easy to see that $(b_k)_{1 \leq k \leq n}$ is completely monotone if and only if for all $k \in \{1, \dots, n\}$, $D^{k-1} b_{n-k+1} \geq 0$. From the statement of the proposition, there exists $S : \mathbb{R}^n \to \mathbb{R}$ a Borel function such that for all $k \in \{1, \dots, n\}$, $K_k = \frac{n!}{(n-k)!} D^{k-1} b_{n-k+1}$, and, recalling (2.1.8), since there is no choice for the B_k's knowing the K_k's, $B_k = b_k$. □

Proof of the proposition. This follows from using the link with the Hoeffding decomposition observed in [10]. Consider for example $A_1, \dots, A_n \geq 0$ and $S(X_1, \dots, X_n) := A_1 \sum_{1 \leq i_1 \leq n} (X_{i_1} - \mathbb{E}X_{i_1}) + A_2 \sum_{1 \leq i_1 < i_2 \leq n} (X_{i_1} - \mathbb{E}X_{i_1})(X_{i_2} - \mathbb{E}X_{i_2}) + \dots + A_n \sum_{1 \leq i_1 < \dots < i_n \leq n} (X_{i_1} - \mathbb{E}X_{i_1}) \dots (X_{i_n} - \mathbb{E}X_{i_n})$. Then, from [10],

$$\begin{aligned} K_k &= A_k^2 \, k! \sum_{1 \leq i_1 < \dots < i_k \leq n} \operatorname{Var}(X_{i_1} - \mathbb{E}X_{i_1}) \dots (X_{i_n} - \mathbb{E}X_{i_k}) \\ &= A_k^2 \, k! \sum_{1 \leq i_1 < \dots < i_k \leq n} \operatorname{Var}(X_{i_1}) \dots \operatorname{Var}(X_{i_k}), \end{aligned}$$

so it is possible to adjust the A_k's to have the K_k's as wanted. □

One could expect the J_k's to behave like the K_k's and to also be able to take any values, but this is unfortunately not the case, for example $2J_2/n = (n-1)(B_1 - B_2) \leq nB_1 = J_1$.

To conclude this section, we connect the B_k's and the quantities T_A introduced in [14]. For any subset A of $\{1, \dots, n\}$, T_A is defined as

$$T_A = \sum_{j \notin A} \Delta_j S (\Delta_j S)^A,$$

and then T is defined as

$$T = \sum_{A \subsetneq \{1, \dots, n\}} \frac{T_A}{2(n - |A|)\binom{n}{|A|}}.$$

It is easy to check that for all $k \in \{1, \dots, n\}$,

$$B_k = \sum_{A : |A| = k-1} \frac{T_A}{2(n - |A|)\binom{n}{|A|}},$$

hence $\mathbb{E}T = \sum_{k=1}^n B_k = \operatorname{Var} S$ (as expected).

Remark. *(i) One might wonder if the above variance results can be transferred to the Φ-entropy. Let Φ be a convex function of the real variable such that $\mathbb{E}|\Phi(S)| < +\infty$, and let the Φ-entropy H_Φ of S (e.g., see [9]) be defined as:*

$$H_\Phi(S) = \mathbb{E}\Phi(S) - \Phi(\mathbb{E}S).$$

Following [10], for $i \in \{1, \dots, n\}$, let

$$H_\Phi^{(i)}(S) = \mathbb{E}^{(i)}\Phi(S) - \Phi(\mathbb{E}^{(i)}(S)),$$

while for $i \neq j \in \{1, \dots, n\}$,

$$H_\Phi^{(j,i)}(S) := \mathbb{E}^{(j)} H_\Phi^{(i)}(S) - H_\Phi^{(i)}(\mathbb{E}^{(j)}S) = H_\Phi^{(i,j)}(S).$$

Still iterating, for $i_1 \neq \ldots \neq i_k \in \{1, \ldots, n\}$,

$$H_\Phi^{(i_1,\ldots,i_k)}(S) := \mathbb{E}^{(i_1)} H_\Phi^{(i_2,\ldots,i_k)}(S) - H_\Phi^{(i_2,\ldots,i_k)}(\mathbb{E}^{(i_1)} S).$$

Define the corresponding B_k*'s as,*

$$B_k := \mathbb{E}\frac{1}{n!}\sum_{i\in\mathfrak{S}_n} H_\Phi^{(i_k)}(\mathbb{E}^{(i_1,\ldots,i_{k-1})} S),$$

for all $k \in \{1, \ldots, n\}$*. Once again the sum is telescopic:*

$$\sum_{k=1}^n B_k = \mathbb{E}\frac{1}{n!}\sum_{i\in\mathfrak{S}_n} \mathbb{E}^{(i_k)}\Phi(\mathbb{E}^{(i_1,\ldots,i_{k-1})} S) - \Phi(\mathbb{E}^{(i_1,\ldots,i_k)} S) = H_\Phi S.$$

By the conditional Jensen inequality, the B_k*'s are non-negative. Just like in the variance case, it is clear by induction that for all* $\ell \in \{0, \ldots, n-1\}$*,*

$$D^\ell B_k = \mathbb{E}\frac{1}{n!}\sum_{i\in\mathfrak{S}_n} H_\Phi^{(i_1,\ldots,i_{\ell+1})}(\mathbb{E}^{(i_{\ell+2},\ldots,i_{k+\ell})} S).$$

Let us now look for the class of convex functions Φ *such that for any* S *and* $X_1, \ldots, X_n$ *satisfying the basic independence and integrability assumptions,* $(B_k)_{1\leq k\leq n}$ *is non-increasing. In particular, for any random variable* Z *defined on a product space* $\Omega_1 \times \Omega_2$ *satisfying the integrability conditions, choosing* S *and* $X_1, \ldots, X_n$ *such that* $S = Z$ *(*$S = f(X_1, X_2)$ *for some function* f*), we have that*

$$\begin{aligned} D^1 B_k &= \frac{1}{n!}\mathbb{E}\sum_{i\in\mathfrak{S}_n} H_\Phi^{(i_1,i_2)}(\mathbb{E}^{(i_3,\ldots,i_{k+1})} S) \\ &= \frac{2}{n!}\mathbb{E} H_\Phi^{(1,2)}(S) \\ &= \frac{2}{n!}\mathbb{E}\left(\Phi(Z) - \Phi(\mathbb{E}^{(1)} Z) - \Phi(\mathbb{E}^{(2)} Z) + \Phi(\mathbb{E}^{(1,2)} Z)\right), \end{aligned}$$

so $\mathbb{E}\left(\Phi(Z) - \Phi(\mathbb{E}^{(1)} Z) - \Phi(\mathbb{E}^{(2)} Z) + \Phi(\mathbb{E}^{(1,2)} Z)\right) \geq 0$*. Reciprocally, if for any random variable* Y *defined on a product space* $\Omega_1 \times \Omega_2$ *satisfying the integrability conditions,*

$$\mathbb{E}\left(\Phi(Y) - \Phi(\mathbb{E}^{(1)} Y) - \Phi(\mathbb{E}^{(2)} Y) + \Phi(\mathbb{E}^{(1,2)} Y)\right) \geq 0,$$

then clearly $D^1 B_k \geq 0$ *for all* $k \in \{1, \ldots, n-1\}$*. Theorem 1 in [75] tells us that this happens if and only if* Φ *is affine or is twice differentiable with* $\Phi'' > 0$ *and* $1/\Phi''$ *concave.*

(ii) One may further wonder what conditions on Φ *would guarantee* $(B_k)_{1\leq k\leq n}$ *to be completely monotone, or, at least, to have* $D^2 B_k \geq 0$ *for all* $k \in \{1, \ldots, n-2\}$*. Unfortunately, the variance is basically the only case for which this holds true. Indeed, if the condition* $D^2 B_k \geq 0$ *is satisfied for all* S*, then, as before, choosing* $S = f(X_1, X_2, X_3)$*, we get*

$$\begin{aligned} D^1 B_k &= \frac{1}{n!}\mathbb{E}\sum_{i\in\mathfrak{S}_n} H_\Phi^{(i_1,i_2,i_3)}(\mathbb{E}^{(i_3,\ldots,i_{k+2})} S) \\ &= \frac{6}{n!}\mathbb{E} H_\Phi^{(1,2,3)}(S) \\ &= \frac{6}{n!}\mathbb{E}\sum_{\alpha\subset\{1,2,3\}} (-1)^{|\alpha|}\Phi(\mathbb{E}^\alpha S). \end{aligned}$$

Therefore, for any random variable Y *defined on a product space* $\Omega_1 \times \Omega_2 \times \Omega_3$ *satisfying the integrability conditions,* $\sum_{\alpha\subset\{1,2,3\}}(-1)^{|\alpha|}\Phi(\mathbb{E}^\alpha Y) \geq 0$*. Reciprocally, this guarantees the non-negativity of* $D^2 B_k$*, for any* $k \in \{1, \ldots, n-2\}$ *and any* S*. According to [75, Theorem 2], this happens if and only if there exist* $a, b, c \in \mathbb{R}$ *with* $a \geq 0$ *and* $\Phi : x \mapsto ax^2 + bx + c$*. So for any function* Φ *that is not of this form, the* K_k*'s and the* J_k*'s (defined as the variations of* B_k*'s) are not always non-negative: for some functions* S *they are negative.*

(iii) It is tempting to use the representation of completely monotone functions for the B_k's. Unfortunately, a completely monotone finite sequence may not be the restriction of a completely monotone function.

2.2 Connection with a more general decomposition of the variance

Let $U_1, \dots, U_n$ be random variables taking values in $(0,1)$ and independent of $X_1, \dots, X_n, X_1', \dots, X_n'$. For any $\alpha \in [0,1]$, let $X^{(\alpha)}$ be the vector with coordinates, $X_i^{(\alpha)} := \mathbb{1}_{\alpha \le U_i} X_i + \mathbb{1}_{\alpha > U_i} X_i'$, $1 \le i \le n$. Then,

$$\operatorname{Var} S = \mathbb{E}\left(S(X^{(0)})\left(S(X^{(0)}) - S(X^{(1)})\right)\right),$$

and it is tempting to rewrite this last term as an integral. Let us assume that each U_i has a density ν_i. For any $0 \le \alpha < \alpha' \le 1$, denote by $A_{\alpha,\alpha'}$ the random set of indices $i \in \{1, \dots, n\}$ such that $\alpha \le U_i < \alpha'$. By the Cauchy-Schwarz inequality,

$$\begin{aligned}\left|\mathbb{E}\left(S(X^{(0)})S(X^{(\alpha')})\right) - \mathbb{E}\left(S(X^{(0)})S(X^{(\alpha)})\right)\right| &\le 2\mathbb{E}(S^2)\mathbb{P}\left(|A_{\alpha,\alpha'}| > 0\right)\\ &\le 2\mathbb{E}(S^2)\mathbb{E}|A_{\alpha,\alpha'}|\\ &\le 2\mathbb{E}(S^2)\sum_{i=1}^{n}\int_{\alpha}^{\alpha'} \mathrm{d}\nu_i,\end{aligned}$$

hence $\alpha \mapsto \mathbb{E}\left(S(X^{(0)})S(X^{(\alpha)})\right)$ is absolutely continuous, its derivative is well defined almost everywhere, integrable, and

$$\operatorname{Var} S = \mathbb{E}\left(S(X^{(0)})S(X^{(0)})\right) - \mathbb{E}\left(S(X^{(0)})S(X^{(1)})\right) = -\int_0^1 \frac{\mathrm{d}}{\mathrm{d}\alpha}\mathbb{E}\left(S(X^{(0)})S(X^{(\alpha)})\right)\mathrm{d}\alpha. \tag{2.2.1}$$

In order to compute the derivative in (2.2.1), fix $\alpha \in (0,1)$ and $\varepsilon \in (0, 1-\alpha)$. Conditioning on $A_{\alpha,\alpha+\varepsilon}$ and letting

$$\Delta_{\alpha,\varepsilon} := \frac{\mathbb{E}\left(S(X^{(0)})S(X^{(\alpha+\varepsilon)})\right) - \mathbb{E}\left(S(X^{(0)})S(X^{(\alpha)})\right)}{\varepsilon},$$

we get

$$\Delta_{\alpha,\varepsilon} = \sum_{1 \le i_1 < \dots < i_k \le n, k \le n} \frac{\mathbb{E}\left(S(X^{(0)})\left(S(X^{(\alpha+\varepsilon)}) - S(X^{(\alpha)})\right)|A_{\alpha,\alpha+\varepsilon} = \{i_1, \dots, i_k\}\right)}{\varepsilon}\mathbb{P}\left(A_{\alpha,\alpha+\varepsilon} = \{i_1, \dots, i_k\}\right),$$

so for almost every α,

$$\Delta_{\alpha,\varepsilon} \xrightarrow[\varepsilon \to 0]{} \sum_{i=1}^{n} \mathbb{E}\left(S(X^{(0)})(S(X^{(\alpha),\hat{i}}) - S(X^{(\alpha),i}))\right)\nu_i(\alpha),$$

where $X^{(\alpha),i}$ is defined like $X^{(\alpha)}$ but with X_i for its i-th coordinate, and $X^{(\alpha),\hat{i}}$ is defined like $X^{(\alpha)}$ but with X_i' for its i-th coordinate. So we get finally:

$$\operatorname{Var} S = \sum_{i=1}^{n}\int_0^1 \mathbb{E}\left(S(X^{(0)})(S(X^{(\alpha),i}) - S(X^{(\alpha),\hat{i}}))\right)\mathrm{d}\nu_i(\alpha). \tag{2.2.2}$$

Let us further define, for $i \in \{1, \dots, n\}$ and any $x_1, \dots, x_n \in \mathbb{R}^n$, $d_i S$ via,

$$d_i S(x_1, \dots, x_n) := S(x_1, \dots, x_n) - \mathbb{E}S(x_1, \dots, x_{i-1}, X_i, x_{i+1}, \dots, x_n).$$

Note that if Z_i is independent of all the other random variables and has same distribution as X_i, we have

$$d_i S(X) = \mathbb{E}_{Z_i}\left(S(X) - S(X_1, \dots, X_{i-1}, Z_i, X_{i+1}, \dots, X_n)\right).$$

Therefore we notice, conditioning on U_i, that

$$\mathbb{E}\left(d_i S(X^{(0)}) d_i S(X^{(\alpha)})\right) = \mathbb{P}(\alpha \leq U_i)\mathbb{E}\left(S(X^{(0)})(S(X^{(\alpha),i}) - S(X^{(\alpha),\hat{i}}))\right).$$

We can rewrite the variance as

$$\operatorname{Var} S = \sum_{i=1}^{n} \int_0^1 \mathbb{E}\left(d_i S(X^{(0)}) d_i S(X^{(\alpha)})\right) \frac{1}{\int_\alpha^1 \mathrm{d}\nu_i(\alpha)} \mathrm{d}\nu_i(\alpha). \tag{2.2.3}$$

Note that in the special case where U_i are uniformly distributed on $[0,1]$,

$$\operatorname{Var} S = \sum_{i=1}^{n} \int_0^1 \mathbb{E}\left(d_i S(X^{(0)}) d_i S(X^{(\alpha)})\right) \frac{1}{1-\alpha} \mathrm{d}\alpha,$$

and a simple change of variables allows us to recover again (2.2.3). Therefore, we will focus on the uniformly distributed case.

2.2.1 Connection with the B_k's

From (2.2.2),

$$\begin{aligned}
\operatorname{Var} S &= \sum_{i=1}^{n} \int_0^1 \mathbb{E}\left(S(X^{(0)})(S(X^{(\alpha),i}) - S(X^{(\alpha),\hat{i}}))\right) \mathrm{d}\alpha \\
&= \sum_{i=1}^{n} \int_0^1 \sum_{k=0}^{n-1} \mathbb{E}\left(S(X^{(0)})(S(X^{(\alpha),i}) - S(X^{(\alpha),\hat{i}}))\right) \mathbb{1}_{|A_{0,\alpha}\setminus\{i\}|=k} \mathrm{d}\alpha \\
&= \int_0^1 n \sum_{k=0}^{n-1} \frac{1}{n} \sum_{i=1}^{n} \mathbb{E}\left(S\left(\Delta_i S\right)^{\beta_{k,i}}\right) \mathbb{P}\left(|A_{0,\alpha} \setminus \{i\}| = k\right) \mathrm{d}\alpha,
\end{aligned}$$

where $\beta_{k,i}$ is a random set of k elements chosen in $\{1, \dots, n\} \setminus \{i\}$. Clearly $\mathbb{P}\left(|A_{0,\alpha} \setminus \{i\}| = k\right) = \binom{n-1}{k}\alpha^k(1-\alpha)^{n-1-k}$, and from the representations (2.1.3) and (2.1.4), we get

$$\sum_{k=0}^{n-1} \frac{1}{n} \sum_{i=1}^{n} \mathbb{E}\left(S\left(\Delta_i S\right)^{\beta_{k,i}}\right) = B_{k+1},$$

hence

$$\operatorname{Var} S = \sum_{k=0}^{n-1} \int_0^1 n \binom{n-1}{k} \alpha^k (1-\alpha)^{n-1-k} B_{k+1} \mathrm{d}\alpha = \sum_{k=0}^{n-1} B_{k+1}.$$

2.2.2 Connection with a semigroup approach

The semigroup approach, as developed in [43] for the hypercube, boils down to the same integration trick along α. We need first to rewrite our results in a more general setup: we assume the X_i's are i.i.d. discrete variables, taking a finite number of values and this time, S takes values in a Banach space $(E, \|\cdot\|_E)$. We also consider a continuous convex function $\Phi : E \to \mathbb{R}^+$, so instead of considering $\operatorname{Var} S = \mathbb{E}\|S - \mathbb{E}S\|_E^2 = \|S - \mathbb{E}S\|_{E,2}^2$, we consider $\mathbb{E}\left(\Phi(S - \mathbb{E}S)\right)$. The price to pay is a suboptimal constant, as seen next, and the lack of connection with the B_k's, which do not seem

to have any equivalent in this setup. We hope that making this connection casts a new light on the breakthrough [43], but also gives prospects to generalize it: indeed, while it is not clear what would be the adequate semigroup when the X_i's are not binary variables, our theorem works for all discrete distributions with finite support (and it is straightforward to generalize to all discrete distributions or even bounded continuous distributions).

Theorem 2.2.1. *For any $\alpha \in (0,1)$, let $\varepsilon_1(\alpha), \dots, \varepsilon_n(\alpha)$ be i.i.d. random variable such that $\mathbb{P}(\xi_i(\alpha) = 1) = 1-\alpha, \mathbb{P}(\xi_i(\alpha) = -1) = \alpha$, and let $\delta_i(\alpha) = \frac{\xi_i(\alpha) - \mathbb{E}\xi_i(\alpha)}{\sqrt{\operatorname{Var}\xi_i(\alpha)}}$. Then, with the notations above,*

$$\mathbb{E}(\Phi(S - \mathbb{E}S)) \leq \int_0^1 \mathbb{E}\Phi\left(\pi \sum_{i=1}^n \delta_i(\alpha) d_i S(X)\right) \frac{\mathrm{d}\alpha}{\pi\sqrt{\alpha(1-\alpha)}}.$$

Proof. Firstly, without loss of generality, we may assume $\mathbb{E}S = 0$ (one may check all the following results are true when one adds a constant to S). Following [43], denoting by Φ^* the convex conjugate of Φ, we note that for any $x \in E$,

$$\Phi(x) = \sup_{y \in E^*} \left(\langle y, x\rangle - \Phi^*(y)\right),$$

and therefore, since the X_i's only take a finite number of values,

$$\mathbb{E}(\Phi(S - \mathbb{E}S)) = \sup_{T \text{ is } \sigma(X_1,\dots,X_n)-\text{measureable, taking values in } \in E^*} \mathbb{E}(\langle T, S\rangle - \Phi^*(T)). \tag{2.2.4}$$

Now we bound the term $\mathbb{E}(\langle T, S\rangle - \Phi^*(T))$. We write, as in (2.2.1),

$$\mathbb{E}(\langle T, S\rangle - \Phi^*(T)) = -\int_0^1 \frac{\mathrm{d}}{\mathrm{d}\alpha}\left(\mathbb{E}(\langle T, S(X^{(\alpha)})\rangle\right)\mathrm{d}\alpha - \mathbb{E}\Phi^*(T)),$$

and just like we obtained (2.2.2), we get

$$\mathbb{E}(\langle T, S\rangle - \Phi^*(T)) = \int_0^1 \mathbb{E}(\langle T, \sum_{i=1}^n S(X^{(\alpha),i}) - S(X^{(\alpha),\hat{i}}))\rangle \mathrm{d}\alpha - \mathbb{E}\Phi^*(T).$$

Note that

$$S(X^{(\alpha),i}) - S(X^{(\alpha),\hat{i}}) = d_i S(X^{(\alpha),i}) - d_i S(X^{(\alpha),\hat{i}}),$$

and by independence,

$$\mathbb{E}(\langle T, d_i S(X^{(\alpha),\hat{i}})\rangle) = 0,$$

so

$$\mathbb{E}(\langle T, S\rangle - \Phi^*(T)) = \sum_{i=1}^n \int_0^1 \mathbb{E}(\langle T, d_i S(X^{(\alpha),i})\rangle \mathrm{d}\alpha - \mathbb{E}\Phi^*(T). \tag{2.2.5}$$

Now, let

$$\delta_i(\alpha) := \frac{\mathbb{1}_{U_i \geq \alpha} - (1-\alpha)}{\sqrt{\alpha(1-\alpha)}} = \frac{2(\mathbb{1}_{U_i \geq \alpha} - 1/2) - (1-2\alpha)}{2\sqrt{\alpha(1-\alpha)}},$$

where the last equality is here to show that this is just a renormalized random variable taking values in $\{-1, 1\}$, much like the $\xi_i(t)$'s, random variables with $\mathbb{P}(\xi_i(t) = 1) = (1 + \mathrm{e}^{-t})/2$ and $\mathbb{P}(\xi_i(t) = -1) = (1 - \mathrm{e}^{-t})/2$ introduced in [43]. We have:

$$\mathbb{E}(\langle T, \delta_i(\alpha) d_i S(X^{(\alpha)})\rangle) = \mathbb{E}\left(\langle T, \frac{\mathbb{1}_{U_i \geq \alpha} d_i S(X^{(\alpha),i}) - (1-\alpha)(\mathbb{1}_{U_i \geq \alpha} d_i S(X^{(\alpha),i}) + \mathbb{1}_{U_i < \alpha} d_i S(X^{(\alpha),\hat{i}}))}{\sqrt{\alpha(1-\alpha)}}\rangle\right)$$

$$= \mathbb{E}\left(\langle T, \frac{\alpha \mathbb{1}_{U_i \geq \alpha} d_i S(X^{(\alpha),i})}{\sqrt{\alpha(1-\alpha)}}\rangle\right)$$
$$= \mathbb{E}\left(\langle T, \sqrt{\alpha(1-\alpha)} d_i S(X^{(\alpha),i})\rangle\right),$$

hence, with (2.2.5) we get:

$$\begin{aligned}\mathbb{E}(\langle T,S\rangle-\Phi^*(T)) &= \sum_{i=1}^n \int_0^1 \frac{\mathbb{E}(\langle T,\delta_i(\alpha)d_iS(X^{(\alpha)})\rangle)}{\sqrt{\alpha(1-\alpha)}}\mathrm{d}\alpha - \mathbb{E}\Phi^*(T)\\ &= \int_0^1 \mathbb{E}(\langle T,\pi\sum_{i=1}^n \delta_i(\alpha)d_iS(X^{(\alpha)})\rangle) - \mathbb{E}\Phi^*(T)\frac{\mathrm{d}\alpha}{\pi\sqrt{\alpha(1-\alpha)}}\\ &\leq \int_0^1 \mathbb{E}\Phi\left(\pi\sum_{i=1}^n \delta_i(\alpha)d_iS(X^{(\alpha)})\right)\frac{\mathrm{d}\alpha}{\pi\sqrt{\alpha(1-\alpha)}}.\end{aligned}$$

Note that $(U_1,\dots,U_n,X_1^{(\alpha)},\dots,X_n^{(\alpha)})$ has the same distribution as $(U_1,\dots,U_n,X_1,\dots,X_n)$, so

$$\mathbb{E}(\langle T,S\rangle-\Phi^*(T)) \leq \int_0^1 \mathbb{E}\Phi\left(\pi\sum_{i=1}^n \delta_i(\alpha)d_iS(X)\right)\frac{\mathrm{d}\alpha}{\pi\sqrt{\alpha(1-\alpha)}}.$$

Recalling (2.2.4), the result follows. □

Let us see how the above allows us to recover the main results of [43], for Rademacher random variables. We recall the notation in use in [43]: for $x\in\{-1,1\}^n$, let

$$D_iS(x) := \frac{S(x_1,\dots,x_i,\dots,x_n)-S(x_1,\dots,-x_i,\dots,x_n)}{2}.$$

We may now state the corollary, in the Rademacher case:

Corollary 2.2.2. *In particular, if X_i's follow a Rademacher distribution,*

$$\mathbb{E}(\Phi(S-\mathbb{E}S)) \leq \int_0^1 \mathbb{E}\Phi\left(\pi\sum_{i=1}^n \delta_i(\alpha)D_iS(X)\right)\frac{\mathrm{d}\alpha}{\pi\sqrt{\alpha(1-\alpha)}},$$

so with a change of variable we get [43, Theorem 1.2], with a different ξ and the constant π instead of $\pi/2$ in Φ:

$$\mathbb{E}(\Phi(S-\mathbb{E}S)) \leq \int_0^{+\infty} \mathbb{E}\Phi\left(\pi\sum_{i=1}^n \delta_i(\mathrm{e}^{-2t})D_iS(X)\right)\frac{2\mathrm{d}t}{\pi\sqrt{\mathrm{e}^{2t}-1}}.$$

Proof. Since $\mathbb{E}^{(i)}S$ does not depend on x_i,

$$\begin{aligned}D_iS(x) &= \frac{(S-\mathbb{E}^{(i)}S)(x_1,\dots,x_i,\dots,x_n)-(S-\mathbb{E}^{(i)}S)(x_1,\dots,-x_i,\dots,x_n)}{2}\\ &= \frac{d_iS(x_1,\dots,x_i,\dots,x_n)-d_iS(x_1,\dots,-x_i,\dots,x_n)}{2}\\ &= d_iS(x),\end{aligned}$$

the last equality coming from the fact that $d_iS(x_1,\dots,1,\dots,x_n) = -d_iS(x_1,\dots,-1,\dots,x_n)$ since $\mathbb{E}^{(i)}(d_iS(X)) = 0$.

□

The above implies a slightly weaker [43, Theorem 1.2], i.e., with a different absolute constant, but the fact that Enflo type and Rademacher type coincide still follows from Theorem 2.2.1 just as it follows from [43, Theorem 1.4] with, as indicated there, a routine symmetrization argument.

To make the connection complete, recall the additional notations in [43]: the operator Δ is defined by

$$\Delta := \sum_{i=1}^{n} D^i,$$

and the semigroup P_t is defined as

$$P_t := \mathrm{e}^{-t\Delta}.$$

In the case where the X_i's are Rademacher random variables, the crucial observation in [43] is that (we denote by ξ', δ' the variables ξ, δ introduced there, to avoid any confusion with δ previously defined):

$$-\frac{\mathrm{d}P_t S}{\mathrm{d}t} = \frac{1}{\sqrt{\mathrm{e}^{2t}-1}}\mathbb{E}_{\xi'(t)}\left(\sum_{i=1}^{n}\delta_i'(t)D_i S(\xi'(t)X)\right), \tag{2.2.6}$$

where $\xi'(t)X$ is defined as $(\xi_1'(t)X_1, \ldots, \xi_n'(t)X_n)$.

Something similar holds in a more general framework (when the $X_i's$ are random variables taking a finite number of values):

Theorem 2.2.3. *With the same assumptions as in Theorem 2.2.1,*

$$-\frac{\mathrm{d}\mathbb{E}_{X',U}S(X^{(\alpha)})}{\mathrm{d}\alpha} = \frac{1}{\sqrt{\alpha(1-\alpha)}}\mathbb{E}_{X',U}\left(\sum_{i=1}^{n}\delta_i(\alpha)d_i S(X^{(\alpha)})\right).$$

Proof. This is essentially the same proof as Theorem 2.2.1. □

We conclude this section with a remark on the Talagrand $L_1 - L_2$ inequality in Banach spaces of Rademacher type 2.

As noted in [16], it is natural, to try and understand for which Banach spaces $(E, \|\cdot\|_E)$ there exists $C = C(E) > 0$ such that for any function function S of Rademacher random variables $X_1, \ldots, X_n$ taking values in E,

$$\|S - \mathbb{E}S\|_{E,2}^2 \leq C\sigma(S)\sum_{i=1}^{n}\frac{\|D_i S\|_{E,2}^2}{1+\log\left(\frac{\|D_i S\|_{E,2}}{\|D_i S\|_{E,1}}\right)}, \tag{2.2.7}$$

where $\|\cdot\|_{E,k} = \left(\mathbb{E}\|\cdot\|_E^k\right)^{1/k}$, which is a generalization of Talagrand's $L_1 - L_2$ inequality (see Theorem 2.3.6) to Banach spaces.

Clearly, if a Banach space satisfies (2.2.7), it must be Rademacher type 2. It is still unknown whether or not the converse is true. The best result, to date, is:

Theorem 2.2.4 ([16, Theorem 1])**.** *Let $(E, \|\cdot\|_E)$ be a Banach space with Rademacher type 2. Then there there exists $C = C(E) > 0$ such that for any function function S of Rademacher random variables $X_1, \ldots, X_n$ taking values in E,*

$$\|S - \mathbb{E}S\|_{E,2}^2 \leq C\sigma(S)\sum_{i=1}^{n}\frac{\|D_i S\|_{E,2}^2}{1+\log\left(\frac{\|D_i S\|_{E,2}}{\|D_i S\|_{E,1}}\right)},$$

where $\sigma(S) = \max_{i\in\{1,\ldots,n\}}\log\left(1+\log\left(\frac{\|D_i S\|_{E,2}}{\|D_i S\|_{E,1}}\right)\right)$.

It is still unclear whether or not the logarithmic term $\sigma(S)$ is needed or not, but we now show how hypercontractivity comes short to removing it.

As noted in [16], one may apply (2.2.6) to P_tS instead of S (for a fixed t), while the chain rule and semigroup property give:

$$-\frac{\mathrm{d}P_{2t}S}{\mathrm{d}t} = \frac{2}{\sqrt{\mathrm{e}^{2t}-1}}\mathbb{E}_{\xi'(t)}\left(\sum_{i=1}^n \delta_i'(t) D_i P_t S(\xi'(t)X)\right).$$

Hence, using the fact that E is Rademacher type 2, if we denote by K its constant, we get (see e.g. [16, (57)]):

$$\|S-\mathbb{E}S\|_{E,2} \leq 4K\int_0^{+\infty}\left(\sum_{i=1}^n \|D_iP_tS\|_{E,2}^2\right)^{1/2}\frac{\mathrm{d}t}{\sqrt{\mathrm{e}^{2t}-1}}. \tag{2.2.8}$$

We now show that in some cases hypercontractivity may not be enough to get rid of the factor $\sigma(S)$. More precisely, let

$$I = \int_0^{+\infty}\left(\sum_{i=1}^n \|D_iS\|_{E,2}^2\left(\frac{\|D_iS\|_{E,1}}{\|D_iS\|_{E,2}}\right)^{2\frac{1-\mathrm{e}^{-2t}}{1+\mathrm{e}^{-2t}}}\right)^{1/2}\frac{\mathrm{d}t}{\sqrt{\mathrm{e}^{2t}-1}},$$

which is the upper bound on the right term of (2.2.8) one gets using hypercontractivity.

We let $L_i = \log\left(\mathrm{e}\frac{\|D_iS\|_{E,2}}{\|D_iS\|_{E,1}}\right)$, $d_i = \|D_iS\|_{E,2}$ and $\theta(t) = \frac{1-\mathrm{e}^{-2t}}{1+\mathrm{e}^{-2t}}$, so

$$I \sim \int_0^{+\infty}\left(\sum_{i=1}^n d_i^2\mathrm{e}^{-2L_i\theta(t)}\right)^{1/2}\frac{\mathrm{d}t}{\sqrt{\mathrm{e}^{2t}-1}}.$$

With a change of variable,

$$\begin{aligned} I &\sim \int_0^1\left(\sum_{i=1}^n d_i^2\mathrm{e}^{-2L_i\theta}\right)^{1/2}\frac{\mathrm{d}\theta}{\sqrt{\theta(1-\theta)}} \\ &\sim \int_0^{1/2}\left(\sum_{i=1}^n d_i^2\mathrm{e}^{-2L_i\theta}\right)^{1/2}\frac{\mathrm{d}\theta}{\sqrt{\theta}} + \frac{\mathrm{e}^{-1}}{\sqrt{2}}\int_{1/2}^1\frac{\mathrm{d}\theta}{\sqrt{1-\theta}}\sqrt{\sum_{i=1}^n\frac{d_i^2}{L_i}}, \end{aligned}$$

so bounding $\frac{I^2}{\sum_{i=1}^n d_i^2/L_i}$ (we already know it is bounded by $\sigma(S)$) is equivalent to bounding

$$R := \frac{\int_0^{1/2}\left(\sum_{i=1}^n d_i^2\mathrm{e}^{-2L_i\theta}\right)^{1/2}\frac{\mathrm{d}\theta}{\sqrt{\theta}}}{\sqrt{\sum_{i=1}^n d_i^2/L_i}}.$$

Letting $\lambda_i := \frac{d_i^2/L_i}{\sum_{i=1}^n d_i^2/L_i}$, we get

$$R = \sqrt{2}\int_0^{1/2}\left(\sum_{i=1}^n \lambda_i L_i\mathrm{e}^{-L_i\theta}\right)^{1/2}\frac{\mathrm{d}\theta}{\sqrt{\theta}}.$$

Assume $L_i = 2^{i-1}$, $\lambda_i = 1/n$. Then for any $\theta \in (1/2^n, 1)$, there exists $i_0 \in \{1,\dots,n\}$ such that $1/2^{i_0} \leq \theta \leq 1/2^{i_0-1}$, and

$$\left(\sum_{i=1}^n \lambda_i L_i\mathrm{e}^{-L_i\theta}\right)^{1/2} \geq \left(\lambda_{i_0}L_{i_0}\mathrm{e}^{-L_{i_0}\theta}\right)^{1/2} \geq \left(\frac{L_{i_0}\theta\mathrm{e}^{-L_{i_0}\theta}}{n\theta}\right)^{1/2} \geq \frac{\sqrt{2\mathrm{e}^{-2}}}{\sqrt{n\theta}},$$

so

$$R \geq \frac{2\sqrt{\mathrm{e}^{-2}}}{\sqrt{n}} \int_{1/2^n}^{1} \frac{\mathrm{d}\theta}{\theta} \geq 2\sqrt{\mathrm{e}^{-2}} \log(2)\sqrt{n} \geq \frac{2\sqrt{\mathrm{e}^{-2}}}{\log(2)} \sqrt{\max_{i\in\{1,\ldots,n\}} L_i}.$$

Thus in this case, $\frac{I^2}{\sum_{i=1}^n d_i^2/L_i}$ is lower bounded by $C\sigma(S)$, for some constant $C > 0$.

2.3 Some applications

To finish these notes, we present some applications of the above inequalities to various contexts, in particular to lower-bounding the variance of the length of the longest common subsequences between two random words. For $(x_1, \ldots, x_s)$, $(y_1, \ldots, y_t)$ two sequences taking values in a finite set $\mathcal{A}$, Recall that we denote by $LCS(x_1 \ldots x_s; y_1 \ldots y_t)$ the largest integer k such that there exists $1 \leq i_1 < \cdots < i_k \leq s$, $1 \leq j_1 < \cdots < j_k \leq t$ satisfying $a_{i_1} = b_{j_1}, \ldots, a_{i_k} = b_{j_k}$, or 0 if there is no such integer. In the sequel, we take $\mathcal{A} = \{1, \ldots, m\}$ (for some m we specify in each case), $X_1, \ldots, X_n, Y_1, \ldots, Y_n$ i.i.d. random variables taking values in $\mathcal{A}$ (according to a distribution we specify), and consider the length of the longest common subsequences of these two random words, $LCS(X_1 \ldots X_n; Y_1 \ldots Y_n)$, written simply LC_n.

2.3.1 Iterated gradients and Gaussian (in)equalities

It is well known that one can transfer the finite samples results of the previous section to functions of normal random variables, somehow reversing the analogies between iterated jackknives and iterated gradients first unveiled in [27]. This transfer is then followed by a study of the infinitely divisible framework and by the semigroup approach to these inequalities.

Let Z be a standard random variable and G be an absolutely continuous function. As well known, the Gaussian Poincaré inequality asserts that

$$\operatorname{Var} G(Z) \leq \mathbb{E}\left(G'(Z)^2\right),$$

while in [29], this inequality is generalized with higher order gradients.

Lemma 2.1.4 and Proposition 2.1.5 allows us to quickly recover Gaussian results. Indeed, e.g., see [9] in the case $k = 1$, one can infer from the discrete decomposition of the variance a decomposition for $\operatorname{Var} G(Z)$.

Lemma 2.3.1. *Let G be a real-valued m-times continuously differentiable function, such that $\mathbb{E}\left(G^{(k)}(Z)^2\right) < +\infty$, $k = 0, \ldots, m$. Let $X_1, \ldots, X_n, X_1', \ldots, X_n'$ be independent Rademacher random variables and let $S(X_1, \ldots, X_n) := G\left(\frac{X_1+\cdots+X_n}{\sqrt{n}}\right)$. Then for all $k \in \{1, \ldots, m\}$,*

$$J_k(n) \xrightarrow[n\to+\infty]{} \mathbb{E}\left(G^{(k)}(Z)^2\right) \quad \text{and} \quad K_k(n) \xrightarrow[n\to+\infty]{} \left(\mathbb{E}\left(G^{(k)}(Z)\right)\right)^2.$$

Proof. It is enough to prove the theorem for G $m+1$-times continuously differentiable with compact support. From (2.1.6), we have

$$J_k = k! \binom{n}{k} D^{k-1} B_1,$$

so (using (2.1.5)),

$$J_k = k! \binom{n}{k} \mathbb{E} \frac{1}{2^k n!} \sum_{i\in\mathfrak{S}_n} \left(\Delta_{i_1,\ldots,i_k} S\right)^2.$$

By symmetry of the function $S(X_1,\dots,X_n) = G\left(\frac{X_1+\dots+X_n}{\sqrt{n}}\right)$, this simplifies to

$$J_k = k!\binom{n}{k}\mathbb{E}\frac{1}{2^k}\left(\Delta_{1,\dots,k}S\right)^2. \tag{2.3.1}$$

For any $i \in \{1,\dots,k\}$,

$$\Delta_i S = (D_i S) 2\mathbb{1}_{X_i = X_i'}$$

with

$$D_i S(x) := \frac{S(x_1,\dots,x_i,\dots,x_n) - S(x_1,\dots,-x_i,\dots,x_n)}{2}.$$

Iterating,

$$\Delta_{1,\dots,k}S = (D_{1,\dots,k}S)2^k \mathbb{1}_{X_1=X_1',\dots,X_k=X_k'},$$

hence

$$\mathbb{E}\frac{1}{2^k}\left(\Delta_{1,\dots,k}S\right)^2 = \mathbb{E}\left(D_{1,\dots,k}S\right)^2. \tag{2.3.2}$$

We now expand, for any $x \in \{-1,1\}^n$, $D_{1,\dots,k}S(x)$. Let us denote for $A \subset \{1,\dots,k\}$,

$$x^A := (2\mathbb{1}_{1\in A} - 1,\dots,2\mathbb{1}_{k\in A} - 1, x_{k+1},\dots,x_n).$$

It is straightforward to prove by induction that

$$D_{1,\dots,k}S(x) = (-1)^{|i\in\{1,\dots,k\}:x_i=1|}\frac{1}{2^k}\sum_{A\subset\{1,\dots,k\}}(-1)^{|A|}S(x^A),$$

which simplifies to

$$D_{1,\dots,k}S(x) = (-1)^{|i\in\{1,\dots,k\}:x_i=1|}\frac{1}{2^k}\sum_{i=0}^{k}\binom{k}{i}(-1)^i G\left(\frac{2i-k+x_{k+1}+\dots+x_n}{\sqrt{n}}\right).$$

By Taylor's formula, and using the fact that $\sum_{i=0}^k \binom{k}{i}(-1)^i i^\ell/\ell! = (-1)^k\mathbb{1}_{\ell=k}$ for any $\ell \in \{0,\dots,k\}$, we get that

$$|D_{1,\dots,k}S(x)| = \left|\frac{1}{\sqrt{n}^k}G^{(k)}\left(\frac{-k+x_{k+1}+\dots+x_n}{\sqrt{n}}\right)\right| + \mathcal{O}\left(\frac{1}{\sqrt{n}^{k+1}}\right),$$

with $\mathcal{O}$ uniform in x (thanks to the compact support assumption). This leads to

$$\mathbb{E}n^k\left(D_{1,\dots,k}S\right)^2 \xrightarrow[n\to\infty]{} \mathbb{E}\left(G^{(k)}(Z)\right)^2,$$

and using (2.3.1) and (2.3.2), we get the desired result

$$\mathbb{E}J_k(n) \xrightarrow[n\to\infty]{} \mathbb{E}\left(G^{(k)}(Z)\right)^2.$$

The other limit in the theorem is obtained in a very similar fashion. □

We now see, using (2.1.10) and (2.1.11), that for any fixed $k \geq 1$,

$$B_k(n) \sim_{n\to+\infty} \frac{1}{n}\mathbb{E}\left(G'(Z)^2\right) \quad \text{and} \quad B_{n-k}(n) \sim_{n\to+\infty} \frac{1}{n}\left(\mathbb{E}\left(G'(Z)\right)\right)^2.$$

More generally, for any $a \in (0,1)$,

$$B_{\lfloor an\rfloor}(n) \sim_{n\to+\infty} \frac{1}{n}\sum_{i=0}^{\infty}\frac{a^j(-1)^j}{j!}\mathbb{E}\left(G^{(j+1)}(Z)^2\right).$$

Note that

$$\int_0^1 B_{\lfloor an \rfloor}(n)\mathrm{d}a \xrightarrow[n\to+\infty]{} \sum_{i=0}^{\infty} \frac{(-1)^j}{(j+1)!}\mathbb{E}\left(G^{(j+1)}(Z)^2\right) = \operatorname{Var} G(Z),$$

as one could expect.

Proposition 2.3.2. *Under the same assumptions on G, for all $k \in \{1, \ldots, m-1\}$,*

$$\frac{\left(\mathbb{E}\left(G^{(k+1)}(Z)\right)\right)^2}{(k+1)!} \leq (-1)^k \left(\operatorname{Var} G(Z) - \mathbb{E}\left(G'(Z)^2\right) + \cdots + (-1)^k \frac{\mathbb{E}\left(G^{(k)}(Z)^2\right)}{k!} \right) \leq \frac{\mathbb{E}\left(G^{(k+1)}(Z)^2\right)}{(k+1)!}.$$

$$\frac{\left(\mathbb{E}\left(G^{(k+1)}(Z)\right)\right)^2}{(k+1)!} \leq \operatorname{Var} G(Z) - \left(\mathbb{E}\left(G'(Z)\right)\right)^2 - \frac{\left(\mathbb{E}\left(G''(Z)\right)\right)^2}{2} - \cdots - \frac{\left(\mathbb{E}\left(G^{(k)}(Z)\right)\right)^2}{k!} \leq \frac{\mathbb{E}\left(G^{(k+1)}(Z)^2\right)}{(k+1)!}.$$

The above indicates that the difference between the variance and each partial sum is squeezed between the Cauchy-Schwarz inequality. We may also get equalities, when G is infinitely differentiable, with additional conditions. Indeed,

Corollary 2.3.3. *Let G be a real-valued infinitely-differentiable function, such that, for all $k \geq 0$, $\mathbb{E}(G^{(k)}(Z))^2 < +\infty$. Then,*

$$\operatorname{Var} G(Z) = \sum_{i=1}^{+\infty} (-1)^{i-1} \frac{\mathbb{E}\left(G^{(i)}(Z)^2\right)}{i!},$$

if and only if $\lim_{k\to\infty} \mathbb{E}(G^{(k)}(Z))^2/k! = 0$, and under such a condition,

$$\operatorname{Var} G(Z) = \sum_{i=1}^{+\infty} \frac{\left(\mathbb{E}\left(G^{(i)}(Z)\right)\right)^2}{i!}.$$

For any $k \geq 1$,

$$Var G(Z) = \mathbb{E}\left(G'(Z)^2\right) - \frac{\mathbb{E}\left(G''(Z)^2\right)}{2} + \cdots + (-1)^{k-1} \frac{\mathbb{E}\left(G^{(k)}(Z)^2\right)}{k!} + (-1)^k \sum_{j=k+1}^{\infty} \binom{j-1}{k} \frac{\left(\mathbb{E}\left(G^{(j)}(Z)\right)\right)^2}{j!}. \tag{2.3.3}$$

For any $a \in [0,1]$,

$$\operatorname{Var} G(Z) = \sum_{i=1}^{+\infty} \left((-1)^{i-1} a^i \frac{\mathbb{E}\left(G^{(i)}(Z)^2\right)}{i!} + (1-a)^i \frac{\left(\mathbb{E}\left(G^{(i)}(Z)\right)\right)^2}{i!} \right). \tag{2.3.4}$$

Proof. This is nothing but Lemma 2.3.1 together with Proposition 2.1.5. To get the last equality, apply (2.1.16) to $k = \lfloor an \rfloor$. □

The equality (2.3.3) is a generalization of the equality in [10], where $k = 1$. Note that (2.3.4) can be rewritten as

$$\operatorname{Var} G(Z) = \sum_{i=1}^{+\infty} \frac{\left(\mathbb{E}\left(G^{(i)}(Z)\right)\right)^2}{i!} + \sum_{i=1}^{+\infty} \left((-1)^{i-1} \frac{\mathbb{E}\left(G^{(i)}(Z)^2\right)}{i!} + \sum_{j\geq i} (-1)^i \binom{j}{i} \frac{\left(\mathbb{E}\left(G^{(j)}(Z)\right)\right)^2}{j!} \right) a^i,$$

which gives us the additional equality: for all $i \geq 1$,

$$\frac{\mathbb{E}\left(G^{(i)}(Z)^2\right)}{i!} = \sum_{j\geq i} \binom{j}{i} \frac{\left(\mathbb{E}\left(G^{(j)}(Z)\right)\right)^2}{j!}.$$

This gives an alternative way to find (2.3.1) again:

$$\begin{aligned}\sum_{i=k+1}^{+\infty}(-1)^{i-1}\frac{\mathbb{E}\left(G^{(i)}(Z)^2\right)}{i!} &= \sum_{j\geq i\geq k+1}(-1)^{i-1}\binom{j}{i}\frac{\left(\mathbb{E}\left(G^{(j)}(Z)\right)\right)^2}{j!}\\ &= \sum_{j=k+1}^{+\infty}\left(\sum_{i=k+1}^{j}(-1)^{i-1}\binom{j}{i}\right)\frac{\left(\mathbb{E}\left(G^{(j)}(Z)\right)\right)^2}{j!}\\ &= \sum_{j=k+1}^{+\infty}\binom{j-1}{k}\frac{\left(\mathbb{E}\left(G^{(j)}(Z)\right)\right)^2}{j!},\end{aligned}$$

where the last equality comes from a simple formula for the partial alternate sum of binomial coefficients.

Multivariable versions of the above results remain true, and in fact, so do infinite-dimensional ones on Wiener space or Poisson space or even Fock space. In each case, what is needed is a proper definition of the gradient, e.g., see [38] for some infinite dimensional setting (Wiener and Poisson spaces). In the multivariate setting here is a small sample of results which can be easily obtained via the techniques developed to this point: Let $m \geq 1$, let $G : \mathbb{R}^m \to \mathbb{R}$ be a smooth function (for the sake of simplicity, just assume differentiability up to the correct order, as above), and let $Z_1, \ldots, Z_m$ be i.i.d. standard normal random variables. Now, for $k \geq 1$, let

$$\theta_k = \sum_{1\leq i_1,\ldots,i_k\leq m}\left(\mathbb{E}\left(\frac{\partial^k G}{\partial x_{i_1}\ldots x_{i_k}}(Z_1,\ldots,Z_m)\right)\right)^2,$$

and let

$$\eta_k = \sum_{1\leq i_1,\ldots,i_k\leq m}\left(\mathbb{E}\left(\frac{\partial^k G}{\partial x_{i_1}\ldots x_{i_k}}(Z_1,\ldots,Z_m)\right)^2\right).$$

Let further $(X_{i,j})_{i\in\{1,\ldots,m\},j\in\{1,\ldots,n\}}$ be independent Rademacher random variables and let

$$S(X_{1,1},\ldots,X_{m,n}) := G\left(\frac{X_{1,1}+\cdots+X_{1,n}}{\sqrt{n}},\ldots,\frac{X_{m,1}+\cdots+X_{m,n}}{\sqrt{n}}\right).$$

Then for all $k \geq 1$,

$$J_k(n) \xrightarrow[n\to+\infty]{} \eta_k \quad \text{and} \quad K_k(n) \xrightarrow[n\to+\infty]{} \theta_k.$$

Moreover, for all $k \geq 1$,

$$\frac{\theta_{k+1}}{(k+1)!} \leq (-1)^k\left(\operatorname{Var} G(Z_1,\ldots,Z_m) - \eta_1 + \frac{\eta_2}{2} - \cdots + (-1)^k\frac{\eta_k}{k!}\right) \leq \frac{\eta_{k+1}}{(k+1)!}.$$

$$\frac{\theta_{k+1}}{(k+1)!} \leq \operatorname{Var} G(Z_1,\ldots,Z_m) - \theta_1 - \frac{\theta_2}{2} - \cdots - \frac{\theta_k}{k!} \leq \frac{\eta_{k+1}}{(k+1)!}.$$

Remark. *It is well known that if $Z_1, Z_2, \ldots, Z_m$ are iid standard normal random variables, and if $\|Z\|_2^2 := \sum_{k=1}^m Z_k^2$, then $(Z_1/\|Z\|_2, \ldots, Z_m/\|Z\|_2)$ is uniformly distributed on the $m-1$-dimensional unit sphere. Therefore, the above multivariate Gaussian case allows to recover and extend various variance bounds and covariance representations on the high-dimensional sphere.*

2.3.2 The Infinitely divisible case

Let Y be an infinitely divisible real-valued random variable, and $G : \mathbb{R} \to \mathbb{R}$ be a smooth function such that its derivatives of all order are well defined and $\mathbb{E}\left(G^{(k)}(Y)\right)^2 < +\infty$ for all $k \geq 0$. We are interested in the decomposition of the variance of $G(Y)$.

We let $(Y_t)_{t\geq 1}$ be the corresponding Lévy process (i.e. Y_1 has the same distribution as Y), we denote by (b,σ,ν) its generator (from the Lévy–Khintchine representation), and let $(Y'_t)_{t\geq 0}$,$(Y''_t)_{t\geq 0}$ be independent copies of $(Y_t)_{t\geq 0}$. For $1\leq \ell, m$, let

$$X_{\ell,m} = Y_{\ell/m} - Y_{(\ell-1)/m},$$
$$X'_{\ell,m} = Y'_{\ell/m} - Y'_{(\ell-1)/m},$$

and let

$$S_n = G(X_{1,n} + \cdots + X_{n,n}).$$

We now study, for any fixed $\alpha\in(0,1)$, the limit when m goes to infinity of $nB_{\lfloor \alpha n\rfloor}$ (the B_k's of S_n) where $n=2^m+1$, which allows us to recover in another way the representation of the variance from [39].

Theorem 2.3.4. *Let $\alpha\in(0,1)$. Then with the notations above,*

$$2^m B_{\lfloor \alpha(2^m)\rfloor+1} \xrightarrow[m\to\infty]{} \mathbb{E}\left(\sigma G'(Y_\alpha + Y''_{1-\alpha})G'(Y'_\alpha + Y''_{1-\alpha}) + \int_{\mathbb{R}} \Delta_u G(Y_\alpha + Y''_{1-\alpha})\Delta_u G(Y'_\alpha + Y''_{1-\alpha})\mathrm{d}\nu\right),$$

where $\Delta_u G(x) = G(x+u) - G(x)$.

Proof. We first prove this fact for α a dyadic rational number, $\alpha = a/2^b \in (0,1)$. Let $m\geq b$ and $n=2^m$, with $m\geq b$. The proof is more convenient to write for a slightly different function: instead of computing the B_k's of S_n ($n=2^m$), we compute the B_k's of

$$T_n := G(X_{1,n} + \cdots + X_{n,n} + X_{n+1,n}).$$

Since the difference between the former and latest has order $\mathcal{O}(1/n)$, this is enough to get the desired result. We have

$$\begin{aligned} B_{\lfloor \alpha n\rfloor+1} &= \mathbb{E}\Big(G\left(\textstyle\sum_{i=1}^{n+1} X_{i,n}\right)\left(G\left(X_1 + \sum_{i=2}^{2^{m-b}a+1} X'_{i,n} + \sum_{i=2^{m-b}a+2}^{n+1} X_{i,n}\right) - G\left(\sum_{i=1}^{2^{m-b}a+1} X'_{i,n} + \sum_{i=2^{m-b}a+2}^{n+1} X_{i,n}\right)\right)\Big) \\ &= \mathbb{E}\left(G(Z_{1/n} + Y'_\alpha + Y''_{1-\alpha})\left(G(Z_{1/n} + Y_\alpha + Y''_{1-\alpha}) - G(Z'_{1/n} + Y'_\alpha + Y''_{1-\alpha})\right)\right), \end{aligned}$$

where Z and Z' are two independent copies of Y, since $(Z_{1/n}, Z'_{1/n}, Y_\alpha, Y'_\alpha, Y''_{1-\alpha})$ has the same distribution as $(X_1, X'_1, \sum_{i=2}^{2^{m-b}a+1} X_{i,n}, \sum_{i=2}^{2^{m-b}a+1} X'_{i,n}, \sum_{i=2^{m-b}a+2}^{n+1} X_{i,n})$.

Let $G_{\alpha,1}(\cdot) = \mathbb{E}(G(\cdot + Y_\alpha + Y''_\alpha)|Y_\alpha, Y''_{1-\alpha})$ and $G_{\alpha,2}(\cdot) = \mathbb{E}(G(\cdot + Y_\alpha + Y''_\alpha)|Y'_\alpha, Y''_{1-\alpha})$, we then have

$$B_{\lfloor \alpha n\rfloor+1} = \mathbb{E}\left(G_{\alpha,1}(Z_{1/n})G_{\alpha,2}(Z_{1/n}) - G_{\alpha,1}(0)G_{\alpha,2}(0) - (G_{\alpha,1}(Z_{1/n})G_{\alpha,2}(Z'_{1/n}) - G_{\alpha,1}(0)G_{\alpha,2}(0))\right),$$

so if A_1 be the infinitesimal generator of $(Y_t, Y_t)_{t\geq 0}$ and A_0 is the infinitesimal generator of $(Y_t, Y'_t)_{t\geq 0}$, then

$$\begin{aligned} nB_{\lfloor \alpha n\rfloor+1} &\xrightarrow[n\to\infty]{} \mathbb{E}\left((A_1 - A_0)G_{\alpha,1}\otimes G_{\alpha,2}(0,0)\right) \\ &= \mathbb{E}\left(\sigma G'(Y_\alpha + Y''_{1-\alpha})G'(Y'_\alpha + Y''_{1-\alpha}) + \int_{\mathbb{R}} \Delta_u G(Y_\alpha + Y''_{1-\alpha})\Delta_u G(Y'_\alpha + Y''_{1-\alpha})\mathrm{d}\nu\right), \end{aligned} \tag{2.3.5}$$

since $(A_1 - A_0)(f\otimes g)(0,0) = \sigma f'(0)g'(0) + \int_{\mathbb{R}} \Delta_u f(0)\Delta_u g(0)\mathrm{d}\nu$. (This computation is in [39, Proposition 2].)

Since the finite sequence B_k is non-decreasing, a routine density argument shows that for any $\alpha\in(0,1)$, $nB_{\lfloor \alpha n\rfloor}$ has limit (2.3.5), which is the desired result. □

Corollary 2.3.5.

$$\operatorname{Var} G(Y) = \int_0^1 \mathbb{E}\left(\sigma G'(Y_\alpha + Y''_{1-\alpha})G'(Y'_\alpha + Y''_{1-\alpha}) + \int_{\mathbb{R}} \Delta_u G(Y_\alpha + Y''_{1-\alpha})\Delta_u G(Y'_\alpha + Y''_{1-\alpha})\mathrm{d}\nu\right)\mathrm{d}\alpha.$$

This above representation of the variance stems from the decomposition $\operatorname{Var} G(Y) = \sum_{k=1}^n B_k$ and it can also be found, with a different approach, in [39]. Although we have only been concerned with representations of the variance, similar representations continue to hold for covariances in the spirit of the work just cited.

For example, we note that in the Poisson case, the limit (2.3.5) is simply

$$\mathbb{E}\left(DG(Y_\alpha + Y''_{1-\alpha})DG(Y'_\alpha + Y''_{1-\alpha})\right),$$

where $DG(x) = G(x+1) - G(x)$, $Y_\alpha, Y'_\alpha, Y''_{1-\alpha}$ are Poisson distributed independent random variables (with respective parameter α, α, and $1-\alpha$). In the Gaussian case, it is

$$\mathbb{E}\left(G'(Z_{1,\alpha})G'(Z_{2,\alpha})\right),$$

where $Z_{1,\alpha}$, $Z_{2,\alpha}$ are Gaussian random variables centered with variance one and covariance $1-\alpha$.

2.3.3 A weaker Talagrand $L_1 - L_2$ inequality

Let us focus on the special case where the X_i's are Bernoulli with parameter $1/2$. For $i \in \{1, \dots, n\}$, let

$$\tau_i S(X_1, \dots, X_n) := S(X_1, \dots, X_{i-1}, 0, X_{i+1}, \dots, X_n) - S(X_1, \dots, X_{i-1}, 1, X_{i+1}, \dots, X_n)$$

(so this does not depend on X_i). Then, Talagrand's $L_1 - L_2$ inequality can be stated as follows:

Theorem 2.3.6 ([67, Theorem 1.5])**.** *There exists $C > 0$ such that for any function $f : \{0,1\}^n \to \mathbb{R}$, the following inequality holds*

$$\operatorname{Var} S \leq C \sum_{i=1}^n \frac{\|\tau_i S\|_2^2}{1 + \log\left(\frac{\|\tau_i S\|_2}{\|\tau_i S\|_1}\right)},$$

where $S = f(X_1, \dots, X_n)$.

We now prove a weaker form of this inequality using the $B'_k s$, in the special case where there exists $a > 0$ such that for all $i \in \{1, \dots, n\}$, $|\tau_i S| \in \{0, a\}$. We can further assume without loss of generality by rescaling that $a = 1$. Note that this particular case includes LC_n (changing a letter can only change LC_n by at most one).

Firstly, conditioning on whether $X_{i_k} = X'_{i_k}$ or $X_{i_k} \neq X'_{i_k}$, we can rewrite (2.1.3) as

$$B_k = \mathbb{E}\frac{1}{4n!}\sum_{i \in \mathfrak{S}_n} (\tau_{i_k} S)(\tau_{i_k} S)^{i_1, \dots, i_{k-1}}, \tag{2.3.6}$$

so

$$\begin{aligned}\operatorname{Var} S &= \sum_{k=1}^n \mathbb{E}\frac{1}{4n!}\sum_{i \in \mathfrak{S}_n} (\tau_{i_k} S)(\tau_{i_k} S)^{i_1, \dots, i_{k-1}} \\ &= \frac{1}{4n!}\sum_{i \in \mathfrak{S}_n}\sum_{k=1}^n \mathbb{E}(\tau_{i_1} S)(\tau_{i_1} S)^{i_2, \dots, i_k}.\end{aligned}$$

Let us fix $i \in \mathfrak{S}_n$ and bound $\sum_{k=1}^n \mathbb{E}(\tau_{i_1} S)(\tau_{i_1} S)^{i_2,\dots,i_k}$. For ease of notation, by reindexing the X_i's, we may assume $i = Id$, also, let us write $X := (X_2, \dots, X_n)$. Since, by assumption, $\tau_1 S$ is boolean, there exists $m \leq 2^{n-1}$ and $x^1, \dots, x^m \in \{0,1\}^{n-1}$ pairwise distinct such that $|\tau_1 S| = \sum_{i=1}^m \mathbb{1}_{X=x^i}$. Let, for $\alpha \subset \{2, \dots, n\}$, $N(\alpha) := |\{(i,j) \in \{1, \dots, m\}^2 : \forall k \in \alpha, x_k^i = x_k^j\}|$. We have

$$\mathbb{E}(\tau_1 S)(\tau_1 S)^{2,\dots,k} \leq \mathbb{E}|\tau_1 S||\tau_1 S|^{2,\dots,k} = \frac{N(\{k+1, \dots, n\})}{2^{n+k-2}}.$$

Let $\ell \in \{1, \dots, n-1\}$ be such that $2^{\ell-1} \leq m \leq 2^\ell$ (we may exclude the trivial case $m = 0$). Using that for any $k \in \{1, \dots, \ell\}$, $N(\{k+1, \dots, n\}) \leq m 2^{k-1}$, and the trivial bound $N(\{k+1, \dots, n\}) \leq m^2$ when $k > \ell$, we get

$$\begin{aligned}
\sum_{k=1}^n \mathbb{E}(\tau_1 S)(\tau_1 S)^{2,\dots,k} &\leq \sum_{k=1}^{\ell} \frac{m}{2^{n-1}} + \sum_{k=1+1}^{n} \frac{m^2}{2^{n+k-2}} \\
&\leq \ell \frac{m}{2^{n-1}} + 2\frac{m^2}{2^{n-1+\ell}} \\
&\leq (\ell+2)\frac{m}{2^{n-1}} = (\ell+2)\|\tau_1 S\|_2^2.
\end{aligned}$$

Note that $\log\left(\frac{\|\tau_1 S\|_2}{\|\tau_1 S\|_1}\right) = \log\left(\sqrt{\frac{2^{n-1}}{m}}\right) = \log(2)(n - 1 - \log_2(m))/2$ so $\ell + 2 \leq n + 2 - \frac{2}{\log(2)} \log\left(\frac{\|\tau_1 S\|_2}{\|\tau_1 S\|_1}\right)$ hence

$$\sum_{k=1}^n \mathbb{E}(\tau_1 S)(\tau_1 S)^{2,\dots,k} \leq \left(n + 2 - \frac{2}{\log(2)} \log\left(\frac{\|\tau_1 S\|_2}{\|\tau_1 S\|_1}\right)\right) \|\tau_1 S\|_2^2.$$

Finally,

$$\begin{aligned}
\operatorname{Var} S &= \frac{1}{4n!} \sum_{i \in \mathfrak{S}_n} \sum_{k=1}^n \mathbb{E}(\tau_{i_1} S)(\tau_{i_1} S)^{i_2,\dots,i_k} \\
&\leq \frac{1}{4n!} \sum_{i \in \mathfrak{S}_n} \left(n + 2 - \frac{2}{\log(2)} \log\left(\frac{\|\tau_{i_1} S\|_2}{\|\tau_{i_1} S\|_1}\right)\right) \|\tau_{i_1} S\|_2^2 \\
&\leq \sum_{j=1}^n \left(1 + \frac{2}{n} - \frac{2}{n \log(2)} \log\left(\frac{\|\tau_j S\|_2}{\|\tau_j S\|_1}\right)\right) \|\tau_j S\|_2^2.
\end{aligned}$$

To see that it is weaker than $L_1 - L_2$ Talagrand's inequality, consider for example $X_1, \dots, X_n$ independent Bernoulli variables of parameter $1/2$, and S defined on $\{0,1\}^n$ by $S(x_1, \dots, x_n) := x_1 \dots x_{n/2}$ (assuming n is even). Then, for any $j \in \{1, \dots, n/2\}$, $\|\tau_j S\|_1 = (1/2)^{\frac{n}{2}-1}$ and $\|\tau_j S\|_2 = \sqrt{\|\tau_j S\|_1}$. So on the one hand, Talagrand's inequality gives a bound of order $(1/2)^{\frac{n}{2}-1}$, which is optimal, while on the other hand, our weaker bound gives an upper bound of order $n(1/2)^{\frac{n}{2}-1}$.

2.3.4 An upper bound on the variance of the length of the longest common subsequences

We got the upper bound for the variance $\operatorname{Var} S \leq nB_1$, which was already known in [65]. Let us apply it to LC_n, and then improve it.

Let $Z_1, \dots, Z_{2n}$ be *i.i.d.* Bernoulli random variables of parameter $1/2$, and consider the B_k's of the function $S(Z_1, \dots, Z_{2n}) := LCS(Z) = LC_n$. We know that $\operatorname{Var} LC_n \leq 2nB_1(2n)$. Using

(2.3.6), $B_1(2n) \leq 1/4$ so $\operatorname{Var} LC_n \leq n/2$ (see also [36]). But this bound can be improved: note that by symmetry of the zeros and ones in LC_n (that is, if $\bar{Z}_i := 1 - Z_i, i \in \{1, \dots, 2n\}$, $S(Z) = S(\bar{Z})$), $\mathbb{E}\tau_i S{=}0$ so $B_{2n}(2n) = 0$. By convexity of B, $B_1(2n) + \cdots + B_{2n}(2n) \leq 2n\frac{B_1(2n)+B_{2n}(2n)}{2}$, so $\operatorname{Var} LC_n \leq n/4$.

More generally, in the case of an alphabet $\{1, \dots, m\}$, conditioning on $X_i \neq X_i'$ we get $B_1(2n) \leq \left(1 - \sum_{k=1}^m p_k^2\right)/2$, and when additionally $B_{2n}(2n) = 0$, then $\operatorname{Var} LC_n \leq \left(1 - \sum_{k=1}^m p_k^2\right) n/2$, which improves, by a factor of two, on Steele's bound [65], $\operatorname{Var} LC_n \leq \left(1 - \sum_{k=1}^m p_k^2\right) n$. The condition $B_{2n}(2n) = 0$ is realized when $p_1 = \cdots = p_m = 1/m$, for instance (by symmetry).

In the remaining part of this section, we focus on lower bounds for the variance of LC_n. By Theorem 2.1.2, $(B_k)_{1 \leq k \leq 2n}$ is, in particular, non-decreasing, so

$$\operatorname{Var} LC_n \geq 2nB_{2n}, \tag{2.3.7}$$

which we will use throughout this section to lower bound the variance. [48, Theorem 2.1] provides a lower bound on the variance of LC_n, proving that when p is smaller than some universal (but extremely small) constant, the variance has order n, see also [36] for more explicit bounds (we already know by Efron-Stein that the variance is less than n). To obtain this bound, the authors first show Theorem 2.2 there, and then prove that it implies that the variance has order n. The proof of this implication is long and we aim to show that the jackknives tools we developed greatly simplifies it. We also generalize the case where one letter is omitted, and then proceed to prove, in the binary case, another slightly weaker bound: for some $p_1 \in (0.096, 0.5)$ (so not as small as in [48] or [36]), the limit superior of the variance over n is not zero. Finally, we give further partial results on the order of the variance in the uniform case.

2.3.5 On the order of the variance under a hypothesis on a modification of LC_n

In this section we prove how Theorem 2.2 in [48] or Theorem 2.1 in [36] imply their main theorem, namely the linear order of the variance. This shows how the use of the B_k's greatly simplify some proofs, and it is of interest to infer, more generally, a lower bound on the variance from a random perturbation that has an effect on the expectation. More specifically, here, the random perturbation is to pick a random 1 from the letters (if there is at least one), and turn it into a 0. The original letters are denoted by $Z_1, \dots, Z_{2n}$, the new letters (with a 1 turned into a 0) by $\tilde{Z}_1, \dots, \tilde{Z_2}n$. We refer to [48] and [36] for a more formal definition of $\tilde{Z}$. Theorem 2.2/Theorem 2.1 there implies, in particular, that for any $\delta \in (0, \alpha_1 - \alpha_2)$, where α_1, α_2 are constants defined there such that $\alpha_1 > \alpha_2$, for n large enough,

$$\mathbb{E}\left(LCS(Z) - LCS(\tilde{Z})\right) \geq \delta.$$

From this, it is natural to try to prove that $B_{2n}(2n)$ is greater than some absolute constant, to infer that the variance has linear order. Let, for all $z \in \{0,1\}^{2n}$, $x \in \{0,1\}$ and $k \in \{1, \dots, 2n\}$, $z^{k,x} := (z_1, \dots, z_{k-1}, x, z_{k+1}, \dots, z_{2n})$. Consider the modifications of Z, $Z^{N,1}$ and $Z^{N,0}$, with N picked in $\{1, \dots, n\}$ uniformly. Intuitively, this is "almost" like the previous pair $(Z, \tilde{Z})$. But it is easier to write $B_{2n}(2n)$ in terms of $\mathbb{E}\left(LCS(Z^{N,1}) - LCS(Z^{N,0}\right)$. Indeed, we have

$$\begin{aligned} B_{2n}(2n) &= \mathbb{E}\frac{1}{2(2n)!}\sum_{i \in \mathfrak{S}_{2n}} (S - S^{i_{2n}})(S^{i_1,\dots,i_{2n-1}} - S^{i_1,\dots,i_{2n}}) \\ &= \mathbb{E}\frac{1}{2(2n)}\sum_{k=1}^{2n}(S - S^k)(S^{\{1,\dots,2n\}\setminus\{k\}} - S^{\{1,\dots,2n\}}), \end{aligned}$$

conditioning on $(Z_{i_{2n}}, Z'_{i_{2n}})$ (first term when its $(0,1)$, second term $(1,0)$, the other terms are null)

we get

$$B_{2n}(2n) = \mathbb{E}\frac{1}{2(2n)}\sum_{k=1}^{2n}\left(LCS(Z^{k,0}) - LCS(Z^{k,1})\right)\left(LCS(Z'^{k,0}) - LCS(Z'^{k,1})\right)p(1-p)+$$
$$\mathbb{E}\frac{1}{2(2n)}\sum_{k=1}^{2n}\left(LCS(Z^{k,1}) - LCS(Z^{k,0})\right)\left(LCS(Z'^{k,1}) - LCS(Z'^{k,0})\right)p(1-p),$$

and by independence,

$$B_{2n}(2n) = \frac{1}{2n}\sum_{k=1}^{2n}\left(\mathbb{E}\left(LCS(Z^{k,1}) - LCS(Z^{k,0})\right)\right)^2 p(1-p),$$

so by the Cauchy-Schwarz inequality,

$$B_{2n}(2n) \geq \left(\mathbb{E}\left(LCS(Z^{N,1}) - LCS(Z^{N,0})\right)\right)^2 p(1-p). \tag{2.3.8}$$

We now give a lower bound on $\mathbb{E}\left(LCS(Z^{N,0}) - LCS(Z^{N,1})\right)$. First note that if N_1 denotes the number of ones, for any $\ell \in \{1,\ldots,2n\}$, $(Z^{N,1}, Z^{N,0})$ conditionally on $N_1(Z^{N,1}) = \ell$ has the same distribution as $(Z, \tilde{Z})$ conditionally on $N_1(Z) = \ell$. Indeed, this is the uniform distribution on all the possible pairs of $2n$ bits, the first one having k ones and the second one being identical except exactly for a 1 turned into a 0. To simplify the notations, for $\ell \in \{0,\ldots,2n\}$, let

$$f(\ell) := \mathbb{E}\left(LCS(Z) - LCS(\tilde{Z})|N_1(Z)\right).$$

We have

$$\begin{aligned}\mathbb{E}\left(LCS(Z) - LCS(\tilde{Z})\right) &= \sum_{\ell=1}^{2n} f(\ell)\mathbb{P}(N_1(Z) = \ell)\\ &= \mathbb{E}\left(f(N_1(Z))\right),\end{aligned}$$

while, since $f(0) = 0$,

$$\begin{aligned}\mathbb{E}\left(LCS(Z^{N,1}) - LCS(Z^{N,0})\right) &= \sum_{\ell=1}^{2n}\mathbb{E}\left(LCS(Z^{N,1}) - LCS(Z^{N,0})|N_1(Z)\right)\mathbb{P}(N_1(Z^{N,1}) = \ell)\\ &= \sum_{\ell=1}^{2n} f(\ell)p^{\ell-1}(1-p)^{n-\ell}\binom{n-1}{\ell-1}\\ &= \sum_{\ell=1}^{2n} f(\ell)\frac{\ell}{pn}\mathbb{P}(N_1(Z) = \ell)\\ &= \mathbb{E}\left(f(N_1(Z))\frac{N_1(Z)}{pn}\right),\end{aligned}$$

so by dominated convergence,

$$\mathbb{E}\left(LCS(Z^{N,1}) - LCS(Z^{N,0})\right) \xrightarrow[n\to\infty]{} \mathbb{E}\left(LCS(Z) - LCS(\tilde{Z})\right)$$

and so for any $\delta \in (0, \alpha_1 - \alpha_2)$, for n large enough,

$$\mathbb{E}\left(LCS(Z^{N,1}) - LCS(Z^{N,0})\right) \geq \delta$$

so using (2.3.7) and (2.3.8),

$$\frac{\operatorname{Var} LC_n}{n} \geq 2p(1-p)\delta^2.$$

2.3.6 On the order of the variance when one letter is omitted

As in [35], we consider the letters $X_1, \dots, X_n$ drawn from an alphabet $\alpha_1, \dots, \alpha_{m+1}$ and the letters $Y_1, \dots, Y_n$ drawn from an alphabet $\alpha_1, \dots, \alpha_m$: so α_{m+1} is an omitted letter, not belonging to any longest common subsequence. We let $p = \mathbb{P}(X_i = \alpha_{m+1})$ and assume $p > 0$, but in contrast to [35], we only make minimal assumptions on $p_{X,1} := \mathbb{P}(X_i = \alpha_1), \dots, p_{X,m} := \mathbb{P}(X_i = \alpha_m), p_{Y,1} := \mathbb{P}(Y_i = \alpha_1), \dots, p_{Y,m} := \mathbb{P}(Y_i = \alpha_m)$: we assume that there are all strictly positive, but these letters are no longer equiprobable, and we assume $m > 1$ (the case $m = 1$ is trivial and may be dealt with separately). Using

$$\operatorname{Var} LC_n \geq 2nB_{2n}(2n), \tag{2.3.9}$$

we see that it is enough to find a constant lower bound on $B_{2n}(2n)$. Firstly, we write

$$\begin{aligned}
B_{2n}(2n) &= \frac{1}{4n} \sum_{j=1}^{2n} \mathbb{E}\left(\Delta_j LC_n (\Delta_j LC_n)^{1,\dots,j-1,j+1,\dots,n}\right) \\
&= \frac{1}{4n} \sum_{j=1}^{2n} \sum_{i,i'=1}^{m} \left(\mathbb{E}\Delta_j LC_n^{Z_j=\alpha_i, Z'_j=\alpha_{i'}}\right)^2 \mathbb{P}(Z_j = \alpha_i)\mathbb{P}(Z'_j = \alpha_{i'}) \quad \text{conditioning} \\
&\geq \frac{1}{4n} \sum_{j=1}^{n} \sum_{i=1}^{m} \left(\mathbb{E}\Delta_j LC_n^{X_j=\alpha_i, X'_j=\alpha_{m+1}}\right)^2 p_{X,i} p \\
&\geq \frac{1}{4n} \sum_{j=1}^{n} \left(\sum_{i=1}^{m} \mathbb{E}\Delta_j LC_n^{X_j=\alpha_i, X'_j=\alpha_{m+1}} p_{X,i}\right)^2 p.
\end{aligned}$$

Writing $LC_{n-1,n} := LCS(X_1 \dots X_{n-1}; Y_1 \dots Y_n)$, we have for any $j \in \{1, \dots, n\}$,

$$\sum_{i=1}^{m} \mathbb{E}\Delta_j LC_n^{X_j=\alpha_i, X'_j=\alpha_{m+1}} p_{X,i} = \mathbb{E}(LC_n) - \mathbb{E}(LC_{n-1,n}),$$

hence

$$B_{2n}(2n) \geq \frac{1}{4} \left(\mathbb{E}(LC_n) - \mathbb{E}(LC_{n-1,n})\right) p. \tag{2.3.10}$$

Let (π, η) be the alignment of $(X_1 \dots, X_{n-1}), (Y_1, \dots, Y_n)$ that is minimal for the lexicographic order, so (π, η) is well defined as a (measurable) function of $X_1 \dots, X_{n-1}, Y_1, \dots, Y_n$. Let F_n be the event "$\eta_{LC_n} < n$", in other words, Y_n does not contribute to the longest common subsequences, then $\sum_{i=1}^{n} \Delta_n LC_n^{X_n=\alpha_i, X'_n=\alpha_{m+1}} \geq \mathbb{1}_{F_n}$, hence

$$\mathbb{E}(LC_n) - \mathbb{E}(LC_{n-1,n}) \geq p_{X,\min} \mathbb{P}(F_n), \tag{2.3.11}$$

where $p_{X,\min} := \min_{1 \leq i \leq m} p_{X,i}$.

We are now going to combine this bound with another one with some elements already present in [35]. Let $V_1 = \pi_1 - 1, V_2 = \pi_2 - \pi_1 - 1, \dots, V_{LC_n} = \pi_{LC_n} - \pi_{LC_n - 1} - 1$, and let M be the number of indices i such that $V_i > 0$. In terms of [35], M is the number of nonempty matches (except that there is also the term V_1). We denote by $I_{i,j}$ the event: "inserting α_i at the j-th position in

$(X_1 \ldots, X_{n-1}), (Y_1, \ldots, Y_n)$ increases the longest common subsequence". Observe that

$$\begin{aligned}
\mathbb{E}(LC_n) - \mathbb{E}(LC_{n-1,n}) &= \mathbb{E}\left(LCS(X_1 \ldots X_{j-1} X_1' X_j \ldots X_{n-1}; Y_1 \ldots, Y_n) - \right. \\
&\qquad \left. LCS((X_1 \ldots X_{n-1}; Y_1 \ldots Y_n)\right) \\
&= \sum_{i=1}^{m} p_{X,i} \mathbb{P}(I_{i,j}) \\
&= \frac{1}{n} \sum_{j=1}^{n} \sum_{i=1}^{m} p_{X,i} \mathbb{P}(I_{i,j}) \\
&\geq \frac{p_{X,\min}}{n} \mathbb{E} \sum_{j=1}^{n} \sum_{i=1}^{m} I_{i,j} \\
&\geq p_{X,\min} \frac{\mathbb{E}M}{n}. \qquad (2.3.12)
\end{aligned}$$

From (2.3.11) and (2.3.12), we get

$$\mathbb{E}(LC_n) - \mathbb{E}(LC_{n-1,n}) \geq \frac{p_{X,\min}}{2} \left(\mathbb{P}(F_n) + \frac{\mathbb{E}M}{n} \right). \qquad (2.3.13)$$

Let γ^* be the limit of $\mathbb{E}(LC_n)/n$, we have $\gamma^* \leq 1 - p < 1$. Fix $k_0 > 0$ such that

$$\sum_{k > k_0} mk(1 - p_{Y,\min})^k \leq \frac{1 - \gamma^*}{2}.$$

When F_n does not hold, that is, $\pi_n = LC_n$, we have

$$\sum_{i=1}^{LC_n} V_i = n - LC_n,$$

so

$$\mathbb{E}\left(\sum_{i=1}^{LC_n} V_i \right) \geq \mathbb{E}\left((n - LC_n) \mathbb{1}_{F_n^c} \right) \geq \mathbb{E}\left(n - LC_n\right) - \mathbb{P}(F_n) n \geq (1 - \gamma^*) n - \mathbb{P}(F_n) n.$$

Furthermore,

$$k_0 \mathbb{E}M \geq \mathbb{E}\left(\sum_{i=1}^{LC_n} V_i \mathbb{1}_{V_i \leq k_0} \right).$$

On the other hand, (π, η) is minimal, so any unmatched gap has (at least) one letter of the alphabet not used, namely, the letter used in the next match. Therefore the average number of indices i such that $V_i = k$ is no more than $nm(1 - p_{Y,\min})^k$, and

$$\mathbb{E}\left(\sum_{i=1}^{LC_n} V_i \mathbb{1}_{V_i > k_0} \right) \leq n \sum_{k > k_0} mk(1 - p_{Y,\min})^k \leq \frac{1 - \gamma^*}{2} n.$$

Finally we get

$$k_0 \mathbb{E}M \geq \frac{1 - \gamma^*}{2} n - \mathbb{P}(F_n) n,$$

and

$$\mathbb{P}(F_n) + \frac{\mathbb{E}M}{n} \geq \frac{k_0 \mathbb{E}M + \mathbb{P}(F_n) n}{k_0 n} \geq \frac{1 - \gamma^*}{2k_0},$$

so putting it together with (2.3.9), (2.3.10) and (2.3.13), we get

$$\operatorname{Var} LC_n \geq \frac{p p_{X,\min} (1 - \gamma^*)}{8k_0} n.$$

2.3.7 A weaker kind of lower bound

Let us return to the Bernoulli framework with parameter $0 < p < 1$, and let $\gamma_n(p) = \mathbb{E}LC_n/n$ and $\gamma(p) = \lim_{n\to\infty} \gamma_n(p)$. It seems reasonable to expect that $\operatorname{Var} LC_n/n$ converges when n tends to infinity, but unfortunately a proof of this result has been elusive so far. Actually little is known on the variance: to the best of our knowledge, it is still an open problem to determine whether or not the variance tends to infinity in the uniform case. The function γ is clearly symmetric around $1/2$, and it is expected to be strictly convex with a minimum at $1/2$, but besides numerical simulations there is no proof of this fact yet. The goal of this section is to prove:

Theorem 2.3.7. *Let $p_0 \in (0, 1/2)$ be such that $\gamma(p_0) > \gamma(1/2)$. Then there exists $p_1 \in (p_0, 1/2)$ such that when $p = p_1$,*

$$\limsup_{n\to\infty} \frac{\operatorname{Var} LC_n}{n} \geq 2p_0(1-p_0)\left(\frac{\gamma(p_0)-\gamma(1/2)}{1/2-p_0}\right)^2.$$

Remark. *Using the bound $\gamma(1/2) < 0.8263$ from [52], and since $\gamma(p) \geq p^2 + (1-p)^2$, we can apply the above theorem with $p_0 = 0.096$, to get for some $p_1 \in (0.096, 0.5)$, $\limsup_{n\to\infty} \operatorname{Var} LC_n/n \geq 1.8/10^8$. Clearly, by symmetry, this limsup result is also valid for some $p_2 \in (0.5, 0.904)$.*

Proof. We have

$$\gamma_n(p_0) - \gamma_n(1/2) = -\int_{p_0}^{1/2} \frac{\mathrm{d}\gamma_n}{\mathrm{d}p}(p)\mathrm{d}p = \int_{p_0}^{1/2} \frac{1}{2n}\sum_{k=1}^{2n} \mathbb{E}_p\left(LC_n^{k,0} - LC_n^{k,1}\right)\mathrm{d}p,$$

where we used a Russo-Margulis kind of formula. This is not strictly the Russo-Margulis lemma since LC_n is not monotone, but the proof of this version is elementary: as in [22], we rewrite γ_n as a function of $2n$ parameters, the parameters of each letter (Bernoulli random variables):

$$\frac{\mathrm{d}\gamma_n}{\mathrm{d}p}(p) = \frac{\mathrm{d}\gamma_n}{\mathrm{d}p}(p, p, \ldots, p) = \sum_{k=1}^{2n} \frac{\mathrm{d}\gamma_n}{\mathrm{d}p_k}(p, p, \ldots, p),$$

which yields the result. Hence,

$$\begin{aligned}\gamma(p_0) - \gamma(1/2) &= \limsup_{n\to\infty} \gamma_n(p_0) - \gamma_n(1/2) \\ &\leq \int_{p_0}^{1/2} \limsup_{n\to\infty} \frac{1}{2n}\sum_{k=1}^{2n} \mathbb{E}_p\left(LC_n^{k,0} - LC_n^{k,1}\right)\mathrm{d}p,\end{aligned}$$

so there exists $p_1 \in (p_0, 1/2)$ such that

$$\limsup_{n\to\infty} \frac{1}{2n}\sum_{k=1}^{2n} \mathbb{E}_{p_1}\left(LC_n^{k,0} - LC_n^{k,1}\right) \geq \frac{\gamma(p_0)-\gamma(1/2)}{1/2-p_0}.$$

Let us fix $p = p_1$. As seen previously,

$$B_{2n}(2n) = \frac{1}{2n}\sum_{k=1}^{2n} \left(\mathbb{E}\left(LC_n(Z^{k,0}) - LC_n(Z^{k,1})\right)\right)^2 p_1(1-p_1),$$

so

$$\begin{aligned}\operatorname{Var} LC_n &\geq \sum_{k=1}^{2n} \left(\mathbb{E}\left(LC_n(Z^{k,0}) - LC_n(Z^{k,1})\right)\right)^2 p_0(1-p_0) \\ &\geq 2n\left(\frac{1}{2n}\sum_{k=1}^{2n} \mathbb{E}_{p_1}\left(LC_n^{k,0} - LC_n^{k,1}\right)\right)^2 p_0(1-p_0),\end{aligned}$$

and finally

$$\limsup_{n\to\infty} \frac{\operatorname{Var} LC_n}{n} \geq 2p_0(1-p_0)\left(\frac{\gamma(p_0)-\gamma(1/2)}{1/2-p_0}\right)^2.$$

□

Remark. *As already mentioned, it is expected that the function γ is strictly convex, but even proving that γ is non-increasing on $[0,1/2]$ and non-decreasing on $[1/2,1]$ seems to be lacking. It also seems reasonable that for a fixed alphabet, say binary uniform, the sequence $(\mathbb{E}LC_n/n)_{n\geq 1}$ is non-decreasing, but again a proof is lacking.*

2.3.8 On the order of the variance in the uniform case

A long-standing open problem is to find the order of the variance of LC_n when the distribution is uniform. In this section, we focus on the uniform binary case, so $\lim_{n\to\infty} \mathbb{E}LC_n/n = \gamma(1/2) := \gamma_2$. We recall, from [37], the definition of the function $\widetilde{\gamma}$: for any $p > 0$,

$$\widetilde{\gamma}(p) := \lim_{n\to\infty} \frac{\mathbb{E}\left(LCS(X_1\dots X_n; Y_1\dots Y_{\lfloor np\rfloor})\right)}{n(1+p)/2}.$$

By a superadditivity argument, this limit is well defined and $\widetilde{\gamma}$ is concave, non-decreasing on $[0,1]$ and non-increasing on $[1,+\infty)$ (for the details, and more properties, we refer to [37]).

By symmetry, in this case, $B_{2n}(2n) = 0$. However, letting $Z_1 = (X_1, Y_1), Z_2 = (X_2, Y_2), \dots, Z_n = (X_n, Y_n)$, then the last B_k is $B_n(n)$ which can also be written as

$$B_n(n) = \frac{1}{n}\sum_{k=1}^{n} \frac{1}{4}\left(\mathbb{E}\left(LCS(Z^{k,0,0}) - LCS(Z^{k,0,1})\right)\right)^2,$$

with transparent notations (the proof is similar). So it is enough to find a lower bound for this quantity, which is doable for the terms on the edge (1 or n) but seems tricky for the terms in the middle.

We may also fix $b \geq 2$ and let $Z_1 = X_1, \dots, X_b$, $Z_2 = X_{b+1}, \dots, X_{2b}$, ... In this case, one gets that lower bounding $B_n(n)$ amounts to finding $w_1, w_2 \in \{0,1\}^b$ and $\delta > 0$ such that for all $n \geq 1$,

$$\frac{1}{n}\sum_{k=1}^{n} \left(\mathbb{E}\left(LCS(Z^{k,w_1}) - LCS(Z^{k,w_2})\right)\right)^2 \geq \delta.$$

For example, intuitively, it is likely to get a larger LCS with $w_1 = (1,0)$ than with $w_2 = (1,1)$, and with $w_1 = (1,0,1,0,1)$ than with $w_2 = (1,1,1,1,1)$. Running simulations in Python, Figure 2.1 seems to indicate that $B_n(n)$ is lower bounded by a strictly positive constant (which would yield the linearity of the variance).

We now pick again $Z_1 = X_1, Z_2 = X_2, \dots, Z_{2n} = Y_n$, and study $B_1(2n)$. Note that if $B_1(2n)$ was converging to zero, this would rule out the possibility of a linear lower bound on the variance. In the following, we study $B_1(2n)$, and find that it is lower bounded by a constant.

Let $X_1, \dots, X_n, Y_1, \dots, Y_n$ be independent Bernoulli random variables with parameter $1/2$, and let $v \leq n$. We may assume, for ease of notations, that $n = vm$ is a multiple of v, but it is not hard to adapt all the following proofs to the general case. Let $\mathcal{R} := \{\overrightarrow{r} \in \mathbb{N}^m : 1 = r_0 \leq r_1 \leq \dots \leq r_m = n\}$, and, for any $\overrightarrow{r} \in \mathcal{R}$, let

$$LC_n(\overrightarrow{r}) = \sum_{i=0}^{m-1} LCS\left(P_i\right),$$

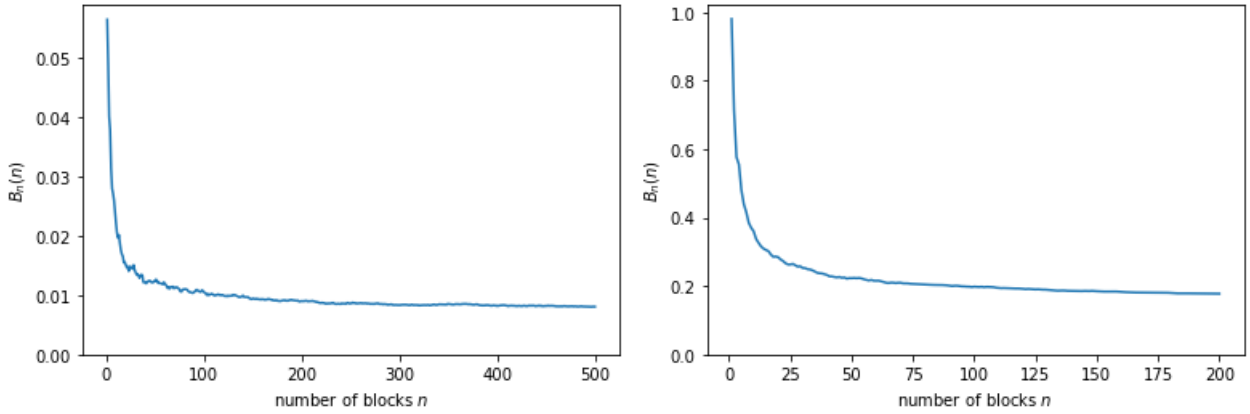


Figure 2.1: $B_n(n)$ for $w_1 = (1,0), w_2 = (1,1)$ (left), and $w_1 = (1,0,1,0,1)$, $w_2 = (1,1,1,1,1)$ (right), with the empirical measure over 1000 simulations.

where $P_i := \big((X_{\upsilon i+1}, \dots, X_{(\upsilon+1)i}), (Y_{r_i}, \dots, Y_{r_{i+1}-1})\big)$ (with the convention $(Y_{r_i}, \dots, Y_{r_{i+1}-1}) = ()$ if $r_i = r_{i+1}$). For any $\overrightarrow{r} \in \mathcal{R}$, call $\overrightarrow{r}$ an alignment if $LC_n = LC_n(\overrightarrow{r})$.

Denote by N_i the number of letters in the cell P_i, that is, $\upsilon + r_{i+1} - r_i$. For any $\overrightarrow{r} \in \mathcal{R}$, let $I_{p_1,p_2}(\overrightarrow{r}) = \{i \in \{0, \dots, m-1\}; r_{i+1} - r_i \in [\upsilon p_1, \upsilon p_2]\}$, and $\overline{I_{p_1,p_2}(\overrightarrow{r})}$ its complementary in $\{0, \dots, m-1\}$. Next, let $B^n_{\varepsilon,p_1,p_2}$ be the event that: for any alignment $\overrightarrow{r}$,

$$\sum_{i \in I_{p_1,p_2}(\overrightarrow{r})} N_i \geq \left(1 - \frac{\varepsilon}{2}\right) 2n.$$

Forgetting about the slight difference in notations for the alignments, as we define them following [28] with non-strict inequalities $r_0 \leq r_1 \leq \dots$ rather than the strict inequalities in [37], we have $B^n_{\varepsilon,p_1,p_2} \subset A^n_{\varepsilon,p_1,p_2}$, where this event is defined in [37]. Indeed, if $B^n_{\varepsilon,p_1,p_2}$ is not satisfied, there is an alignment $\overrightarrow{r}$ such that

$$\sum_{i \in \overline{I_{p_1,p_2}(\overrightarrow{r})}} N_i > \varepsilon n.$$

which implies $\mathrm{Card}\,(I_{p_1,p_2}(\overrightarrow{r})) > \varepsilon m$, which means $A^n_{\varepsilon,p_1,p_2}$ is not satisfied. Therefore, the following is a strengthening of [37, Theorem 2.2].

Lemma 2.3.8. *Let $\varepsilon > 0$. Let $0 < p_1 < 1 < p_2$ be such that $\widetilde{\gamma}(p_1) < \widetilde{\gamma}(1) = \gamma_2$ and $\widetilde{\gamma}(p_2) < \gamma_2$ and let $\delta \in (0, \min(\gamma_2 - \widetilde{\gamma}(p_1), \gamma_2 - \widetilde{\gamma}(p_2)))$.*

Fix the integer υ to be such that $(1 + \ln(1+\upsilon))/\upsilon \leq \delta^2\varepsilon^2/16$, then

$$\mathbb{P}\left(B^n_{\varepsilon,p_1,p_2}\right) \geq 1 - \exp\left(-n\left(\frac{\delta^2\varepsilon^2}{16} - \frac{1+\ln(1+\upsilon)}{\upsilon}\right)\right),$$

for all n large enough.

Proof. Let $\overrightarrow{r} \in \mathcal{R}$ be such that $\sum_{i \in \overline{I_{p_1,p_2}(\overrightarrow{r})}} N_i > \varepsilon n$. We first prove that

$$\mathbb{E}\left(LC_n(\overrightarrow{r}) - LC_n\right) \leq -\frac{\delta\varepsilon n}{2}$$

for all n large enough. We follow the proof of [37, Lemma 3.1]. Let $\delta^* = \min(\gamma_2 - \widetilde{\gamma}(p_1), \gamma_2 - \widetilde{\gamma}(p_2))$.

Using the superadditivity of $\widetilde{\gamma}$, we get

$$\begin{aligned}\mathbb{E}\left(LC_n(\overrightarrow{r})\right) &\leq \gamma_2 \left(\sum_{i\in I_{p_1,p_2}(\overrightarrow{r})} \frac{N_i}{2}\right) + (\gamma_2 - \delta^*)\left(\sum_{i\in\overline{I_{p_1,p_2}(\overrightarrow{r})}} \frac{N_i}{2}\right)\\ &\leq \left(\gamma_2 - \frac{\delta^*\varepsilon}{2}\right) n.\end{aligned}$$

Moreover, for n is large enough,

$$-\mathbb{E}\left(LC_n\right) \leq -\left(\gamma_2 - \frac{(\delta^* - \delta)\varepsilon}{2}\right) n,$$

so combining together these two inequalities, we get the desired result:

$$\mathbb{E}\left(LC_n(\overrightarrow{r}) - LC_n\right) \leq -\frac{\delta\varepsilon n}{2}.$$

The end of the proof is exactly like in [37], the only difference is, as pointed out in [28, Remark 2.2], that the cardinality of $\mathcal{R}$ is now $\binom{n+\upsilon}{\upsilon}$ instead of $\binom{n}{\upsilon}$ so $\ln\upsilon$ becomes $\ln(1+\upsilon)$.

□

Theorem 2.3.9. *There exists $C > 0$ such that for all n large enough, $B_1(2n) \geq C$.*

Proof. For any $\overrightarrow{r} \in \mathcal{R}$, let $S(\overrightarrow{r}) = \{i \in \{0,\dots,m-1\}; LCS(P_i) = \min(\upsilon, r_{i+1} - r_i)\}$ the set of the indices of "saturated" cells, meaning that $LCS(P_i)$ is maximal given the size of the cell. We first show that for some $\varepsilon > 0$, with high probability, for any alignment $\overrightarrow{r}$, $\mathrm{Card}(S(\overrightarrow{r})) \leq (1-\varepsilon)m$. The idea behind is that the εm non-saturated cells will guarantee the lower bound on $B_1(2n)$, as changing their coordinates might increase LC_n.

Let $x = 0.28, p_1 = 1 - x, p_2 = 1/p_1$, we know from [37] that $\widetilde{\gamma}(p_1) < \gamma_2$ and $\widetilde{\gamma}(p_2) < \gamma_2$. Let $\eta = \frac{2(1-x)}{2-x} - \gamma_2$, from the upper bound $\gamma_2 \leq 0.8263$, see [52], it that $\eta > 0$. Let $\varepsilon \in \left(0, \frac{\eta}{2(\gamma_2+\eta)}\right)$, and, lastly, let $\delta \in (0, \min(\gamma_2 - \widetilde{\gamma}(p_1), \gamma_2 - \widetilde{\gamma}(p_2)))$ and fix υ to be such that $(1+\ln(1+\upsilon))/\upsilon < \delta^2\varepsilon^2/16$.

Let C_ε^n be the event: for any alignment $\overrightarrow{r}$, $\mathrm{Card}(S(\overrightarrow{r})) \leq (1-\varepsilon)m$. If $(C_\varepsilon^n)^c \cap B_{\varepsilon,p_1,p_2}^n$ is realized, then there is some alignment $\overrightarrow{r}$ such that $\mathrm{Card}(S(\overrightarrow{r})) > (1-\varepsilon)m$, and

$$LC_n \geq \sum_{i\in S(\overrightarrow{r})\cap I_{p_1,p_2}(\overrightarrow{r})} \frac{N_i}{2} \frac{\min(\upsilon, r_{i+1} - r_i)}{\frac{N_i}{2}}.$$

For any $i \in I_{p_1,p_2}(\overrightarrow{r})$, $r_{i+1} - r_i \in [\upsilon p_1, \upsilon p_2]$ so

$$\frac{\min(\upsilon, r_{i+1} - r_i)}{\frac{N_i}{2}} \geq \frac{2}{1+p_2} = \frac{2p_1}{1+p_1} = \frac{2(1-x)}{2-x} = \gamma_2 + \eta$$

so

$$LC_n \geq \sum_{i\in S(\overrightarrow{r})\cap I_{p_1,p_2}(\overrightarrow{r})} \frac{N_i}{2}(\gamma_2 + \eta).$$

Furthermore,

$$\begin{aligned}\sum_{i\in\overline{S(\overrightarrow{r})\cap I_{p_1,p_2}(\overrightarrow{r})}} \frac{N_i}{2} &= \sum_{i\in\overline{S(\overrightarrow{r})}\cap I_{p_1,p_2}(\overrightarrow{r})} \frac{N_i}{2} + \sum_{i\in\overline{I_{p_1,p_2}(\overrightarrow{r})}} \frac{N_i}{2}\\ &\leq \upsilon\frac{1+p_2}{2}\varepsilon m + \frac{\varepsilon n}{2}\\ &\leq 2\varepsilon n,\end{aligned}$$

so we get

$$LC_n \geq (1-2\varepsilon)(\gamma_2+\eta)n,$$

but given the choice of ε, $(1-2\varepsilon)(\gamma_2+\eta) > \gamma_2$, so by concentration, this has probability exponentially small to happen. Therefore, $\mathbb{P}\left((C_\varepsilon^n)^c\right) \leq \mathbb{P}\left((C_\varepsilon^n)^c \cap B^n_{\varepsilon,p_1,p_2}\right) + \mathbb{P}\left((B^n_{\varepsilon,p_1,p_2})^c\right)$ goes to zero (exponentially fast) as n goes to infinity.

Now for $i \in \{1,\dots,m\}$, let

$$V_i = \max_{x\in\{0,1\}^\upsilon} \left| LCS(X_1 \dots X_{\upsilon(i-1)} x_1 \dots, x_\upsilon X_{\upsilon i+1} \dots X_n; Y_1 \dots Y_n) - LCS(X_1 \dots X_n; Y_1 \dots Y_n) \right|.$$

For any $i \in \overline{S(\overrightarrow{r})}$, $V_i \geq 1$, hence

$$\mathbb{E}\left(\frac{1}{m}\sum_{i=1}^m V_i^2\right) > \varepsilon \mathbb{P}(C_\varepsilon^n).$$

Now let for $x \in \{0,1\}^\upsilon$ and $j \in \{\upsilon(i-1)+1,\dots,\upsilon i\}$,

$$\begin{aligned}\delta_j(x) =& LCS(X_1 \dots X_{\upsilon(i-1)} x_1 \dots x_{j-\upsilon(i-1)} X_{j+1} \dots X_n; Y_1 \dots Y_n) \\ & - LCS(X_1 \dots X_{\upsilon(i-1)} x_1 \dots, x_{j-\upsilon(i-1)-1} X_j \dots X_n; Y_1 \dots Y_n),\end{aligned}$$

so that

$$\begin{aligned} V_i^2 &= \max_{x\in\{0,1\}^\upsilon} \left| \sum_{j=\upsilon(i-1)+1}^{\upsilon i} \delta_j(x) \right|^2 \\ &\leq \max_{x\in\{0,1\}^\upsilon} \upsilon \sum_{j=\upsilon(i-1)+1}^{\upsilon i} \delta_j(x)^2 \\ &\leq \upsilon \sum_{j=\upsilon(i-1)+1}^{\upsilon i} \max_{x\in\{0,1\}^\upsilon} \delta_j(x)^2. \end{aligned}$$

Note that $\mathbb{E}\Delta_j^2 = \mathbb{E}\delta_j(X_1',\dots,X_\upsilon')^2$ (see the next section to recall the definition of Δ_j), and

$$\mathbb{E}_{X_1',\dots,X_\upsilon'} \Delta_j^2 \geq \frac{1}{2^\upsilon} \max_{x\in\{0,1\}^\upsilon} \delta_j(x)^2.$$

Hence,

$$\mathbb{E}\Delta_j^2 \geq \frac{1}{2^\upsilon} \mathbb{E} \max_{x\in\{0,1\}^\upsilon} \delta_j(x)^2,$$

so

$$\begin{aligned} V_i^2 &\leq \upsilon 2^\upsilon \sum_{j=\upsilon(i-1)+1}^{\upsilon i} \mathbb{E}\Delta_j^2, \\ \mathbb{E}\left(\frac{1}{m}\sum_{i=1}^m V_i^2\right) &\leq \frac{\upsilon 2^\upsilon}{m} \sum_{j=1}^n \mathbb{E}\Delta_j^2 \\ \varepsilon \mathbb{P}(C_\varepsilon^n) &< \upsilon^2 2^\upsilon B_1(2n), \end{aligned}$$

and when n is large enough, $B_1(2n) > \varepsilon/2\upsilon^2 2^\upsilon$. □

Remark. *The above result is a necessary condition (certainly not sufficient, though) to have* $\operatorname{Var} LC_n$ *asymptotically linear. This implies that there exists* $C' > 0$*, such that for all* n*,* $B_1(2n) \geq C'$*, as for all* n*,* $B_1(2n) > 0$.

2.3.9 A note on a potential implication of [28]

In this section, $\alpha \in (0,1)$, $\upsilon = n^\alpha$, and $\overrightarrow{r}$ is a random alignment. Let $X'_1, \dots, X'_n, Y'_1, \dots, Y'_n$ be independent Bernoulli variables with parameter $1/2$, independent from all the previous variables. As previously, we write $Z = (Z_1, \dots, Z_{2n}) := (X_1, \dots, X_n, Y_1, \dots, Y_n)$, and as in [28], for $j \in \{1, \dots, 2n\}$, let

$$\Delta_j := LCS(Z) - LCS(Z_1 \dots Z'_j \dots Z_{2n})$$
$$\widetilde{\Delta_j} := LCS\left(P_i\right) - LCS\left(P'_i\right)$$

where P_i is the cell of length υ containing Z_j and P'_i is the same cell but with Z'_j instead of Z_j. We also write for $j, k \in \{1, \dots, m\}$:

$$LC_n^j := LCS(Z_1 \dots Z'_j \dots Z_{2n})$$
$$LC_n^{j,k} := LCS(Z_1 \dots Z'_j \dots Z'_k \dots Z_{2n})$$
$$\Delta_{j,k} := LC_n - LC_n^j - LC_n^j + LC_n^{j,k}.$$

It is claimed in [28] that $\mathbb{E}\left|\widetilde{\Delta_j} - \Delta_j\right| = \mathbb{E}\left(\widetilde{\Delta_j} - \Delta_j\right)$ is exponentially small in n. The equality comes from the fact that $\widetilde{\Delta_j} - \Delta_j \geq 0$ (as explained in [28]). Furthermore, $\mathbb{E}\Delta_j = 0$, so the problem boils to controlling $\mathbb{E}\widetilde{\Delta_j}$. Let us assume, in this section, that $\mathbb{E}\widetilde{\Delta_j} \leq \exp(-tn)$ for some $t > 0$ not depending on j, n, and let us denote by A_j the event $\widetilde{\Delta_j} - \Delta_j = 0$. Of course, $\mathbb{P}(A_j^c) \leq \exp(-tn)$. Finally, let $C_{j,k}$ be the event "Z_j and Z_k are not in the same cell". Let $j, k \in \{1, \dots, n\}$ and suppose A_j, A_k and $C_{j,k}$ are all realized, then when X_j is flipped to X'_j, the alignment $\overrightarrow{r'} = \overrightarrow{r}(Z)$ is still an alignment for $(Z_1, \dots, Z'_j, \dots, Z_{2n})$, so

$$LC_n^{j,k} - LC_n^j \geq -\widetilde{\Delta_k} = LC_n^k - LC_n$$

so, in other terms,

$$\Delta_{j,k} \mathbb{1}_{A_j} \mathbb{1}_{A_k} \mathbb{1}_{C_{j,k}} \geq 0. \tag{2.3.14}$$

Let us write $\Delta_{j,k} = \Delta_{j,k}^+ - \Delta_{j,k}^-$ (the positive and negative parts), using the bounds $|\Delta_{j,k}| \leq 2$ and (2.3.14) we get

$$\Delta_{j,k}^- \leq 2(1 - \mathbb{1}_{A_j} \mathbb{1}_{A_k} \mathbb{1}_{C_{j,k}})$$

so $(\Delta_{j,k}^-)^2 \leq 4(1 - \mathbb{1}_{A_j} \mathbb{1}_{A_k} \mathbb{1}_{C_{j,k}})$, and

$$\mathbb{E}(\Delta_{j,k}^-)^2 \leq 4\left(\mathbb{P}(A_j^c) + \mathbb{P}(A_k^c) + \mathbb{P}(C_{j,k}^c)\right),$$
$$\mathbb{E}(\Delta_{j,k}^+)^2 \leq 2\mathbb{E}\Delta_{j,k}^+ = 2\mathbb{E}\Delta_{j,k}^- \leq 4\left(\mathbb{P}(A_j^c) + \mathbb{P}(A_k^c) + \mathbb{P}(C_{j,k}^c)\right),$$

hence

$$\mathbb{E}(\Delta_{j,k})^2 \leq 8\left(\mathbb{P}(A_j^c) + \mathbb{P}(A_k^c) + \mathbb{P}(C_{j,k}^c)\right).$$

We may now give an upper bound on $B_1(2n) - B_2(2n)$:

$$\begin{aligned}
B_1(2n) - B_2(2n) &= \frac{1}{4(2n)(2n-1)} \sum_{\substack{j \neq k \\ j,k \in \{1,\dots,2n\}}} \mathbb{E}(\Delta_{j,k})^2 \\
&= \frac{2}{4(2n)(2n-1)} \sum_{\substack{j \neq k \\ j \in \{1,\dots,n\}, k \in \{1,\dots,2n\}}} \mathbb{E}(\Delta_{j,k})^2 \quad \text{(by symmetry)} \\
&\leq \frac{2}{n(2n-1)} \mathbb{E}\left(\sum_{\substack{j \neq k \\ j \in \{1,\dots,n\}}} \mathbb{1}_{C_{j,k}^c} \right) + \frac{2}{n(2n-1)} \sum_{\substack{j \neq k \\ j \in \{1,\dots,n\}}} \mathbb{P}(A_j^c) + \mathbb{P}(A_k^c) \\
&\leq \frac{2}{n(2n-1)}(2n\upsilon - n) + 2\exp(-tn)
\end{aligned}$$

So when n is large enough,

$$B_1(2n) - B_2(2n) \leq \frac{2\upsilon}{n}$$

and by convexity of B, and using the lower bound $0 < C \leq B_1(2n)$ (see Theorem 2.3.9),

$$\operatorname{Var} LC_n = B_1(2n) + \cdots + B_{2n}(2n) \geq \sum_{i=1}^{\frac{Cn}{2\upsilon}} C - \frac{2\upsilon(i-1)}{n}$$

which is equivalent to $C^2 n/(4\upsilon)$. So for some constant $C' > 0$,

$$\operatorname{Var} LC_n \geq C' n^{1-\alpha}.$$

Once again, this is under the assumption that $\mathbb{E}\widetilde{\Delta_j} \leq \exp(-tn)$. If, additionally, this assumption holds for some $\alpha < 1/10$, then by [28] there is convergence of the properly rescaled LC_n to a Gaussian.

There is also a somewhat weaker assumption that would guarantee the linearity of the variance. Recalling the percolation interpretation of the LCS as seen in Section 0.1, we denote by Geo the (random) set of geodesics, and for any $a, b \in \{1, \ldots, 2n\}$, Geo^a the set of geodesics when the Z_a is turned into Z'_a, and $\mathrm{Geo}^{a,b}$ the set of geodesics when Z_a is turned into Z'_a and Z_b is turned into Z'_b. For $j, k \in \{1, \ldots, m\}$, let $A_{j,k}$ be the event: there exists (p, q) such that $j < p < k$ and there exist $(g_1, g_2, g_3, g_4) \in \mathrm{Geo} \cap \mathrm{Geo}^j \cap \mathrm{Geo}^k \cap \mathrm{Geo}^{j,k}$ such that $(p, q) \in g_1 \cap g_2 \cap g_3 \cap g_4$. In words, this is the event that it is possible to find X_p aligned with Y_q no matter the values of X_j and X_k. Similarly, let $B_{j,k}$ be the event: there exists (p, q) such that $j < p$ and $k > q$ or $j > p$ and $k < q$ and there exist $(g_1, g_2, g_3, g_4) \in \mathrm{Geo} \cap \mathrm{Geo}^j \cap \mathrm{Geo}^{k+n} \cap \mathrm{Geo}^{j,k+n}$ such that $(p, q) \in g_1 \cap g_2 \cap g_3 \cap g_4$. In words, this is the event that it is possible to find X_p aligned with Y_q no matter the values of X_j and X_k, and such that X_j, Y_k are not both "on the same side". Now suppose that $\mathbb{P}(A^c_{j,k}), \mathbb{P}(B^c_{j,k}) \leq \exp(\alpha|k-j|)$ for some constant $\alpha > 0$. Then an adaptation of the proof above shows that the variance is lower bounded by $C'n$ for some constant $C' > 0$.

Chapter 3

Quantum Statistics

In this chapter, we are interested in quantum spectrum tomography, which consists in estimating the spectrum of the density matrix of a quantum system, given n independent copies of this system. As seen in Section 0.4, the Weak Schur sampling allows to reformulate this problem as follows: given $\lambda \sim \mathrm{SW}^n(p)$, how to estimate p? Note that from now on, we assume $p_1 \geq p_2 \geq \cdots \geq p_d$, as by Proposition 0.3.3, there is no loss in generality.

If one works with the L_2-loss function (the L_1-loss and others may also be considered), the goal is to find a function g_n such that $\mathbb{E}_{\lambda\sim\mathrm{SW}^n(p)}\|g_n(\lambda) - p\|^2$ is as small as possible . To date, the best estimator is simply the empirical Young diagram, originally introduced in [2]:

$$g_n(\lambda) = \underline{\lambda} := \left(\frac{\lambda_1}{n}, \ldots, \frac{\lambda_d}{n}\right). \tag{3.0.1}$$

This is the quantum equivalent of the empirical distribution, except that the expected L_2-loss (the risk) is not in $1/n$, but d/n (see, e.g., [76]):

$$\mathbb{E}_{\lambda\sim\mathrm{SW}^n(p)}\|g_n(\lambda) - p\|^2 = \mathcal{O}\left(\frac{d}{n}\right),$$

and one may check that in the uniform case, asymptotically, the risk has order d/n (using Theorem 0.3.5), therefore this upper bound is tight.

The question of the existence of an estimator with a risk of order $1/n$ (as in classical statistics), or at least a "better" estimator than the empirical Young diagram, is an important open question in quantum statistics.

This is made difficult by the fact that that none of the limiting theorems for the shapes of Young diagrams had explicit rates. Our first goal is to compute explicit rates of convergence of $\lambda \sim \mathrm{SW}^n(p)$. In this first part, we revisit, beyond the uniform case, some aspects of the convergence of the cumulative shape of the RSK Young diagrams associated with random words, obtaining rates of convergence in Kolmogorov's distance. Since the length of the top row of the diagrams is the length of the longest increasing subsequences of the word, a corresponding rate result follows.

It turns out that the rates of convergence are not sharp enough to infer an estimator better than the empirical Young diagram, so in a second part we introduce different estimators and the conditions that would make them better than the empirical Young diagram.

Finally, the difference between the shape of the RSK diagram and the "classical" histogram, which is called the excess, is investigated.

3.1 Rates of convergence of the shape of Young diagrams

Let $X_1, \ldots, X_n$ be i.i.d. random variables with values in a finite alphabet $\{1, \ldots, m\}$ and with probability mass function given by $p_1, \ldots, p_m$. For $i \in \{1, \ldots, m\}$ and $j \in \{0, \ldots, n\}$, and with elements of the notation in [32, 12, 18], let $N_j^i = \sum_{l=1}^{j} \mathbb{1}_{X_l=i}$ be the number of letters i within the first j letters, and for $t \in [0, 1]$, let

$$\widetilde{B_i^n}(t) = \frac{N_{\lfloor tn \rfloor}^i - p_i \lfloor tn \rfloor}{\sigma_i \sqrt{n}},$$

where $\sigma_i = \sqrt{p_i(1-p_i)}$. Let $\Lambda = \{\lambda \in [0,1]^m : \lambda_1 + \cdots + \lambda_m = 1\}$, and for $\lambda \in \Lambda$, let

$$\widetilde{V_i^n}(\lambda) = \sigma_i \left(\widetilde{B_i^n}(\lambda_1 + \cdots + \lambda_i) - \widetilde{B_i^n}(\lambda_1 + \cdots + \lambda_{i-1}) \right),$$

$i \in \{1, \ldots, m\}$ with for $i = 1$, the convention that $\widetilde{B_i^n}(\lambda_1 + \cdots + \lambda_{i-1}) = 0$.

Hence, LI_n, the length of the longest increasing subsequences of $X_1 \cdots X_n$ is given by:

$$LI_n = \max_{0=k_0 \le k_1 \le \cdots \le k_m = n} \sum_{i=1}^{m} \left(N_{k_i}^i - N_{k_{i-1}}^i \right) = \max_{\lambda \in \Lambda_d} \sum_{i=1}^{m} \left(np_i\lambda_i + \sqrt{n}\widetilde{V_i^n}(\lambda) \right),$$

where now $\Lambda_d := \{(j_1/n, \ldots, j_m/n) : j_1, \ldots, j_m \in \mathbb{N}, j_1 + \cdots + j_m = n\}$.

For $\lambda \in \Lambda$ let, finally,

$$Z_n(\lambda) := \sum_{i=1}^{m} \left(\sqrt{n}(p_i - p_{\max})\lambda_i + \widetilde{V_i^n}(\lambda) \right),$$

where $p_{\max} = \max_{i=1,\ldots,m} p_i$, so that

$$\frac{LI_n - np_{\max}}{\sqrt{n}} = \max_{\lambda \in \Lambda_d} Z_n(\lambda). \tag{3.1.1}$$

It is known (see [32, 12, 18] and the references therein) that the limiting distribution of (3.1.1) is the distribution of $\max_{\lambda \in \Lambda} Z'_n(\lambda)$ where Z'_n is defined as Z_n but with B_i^n, a Brownian motion, instead of $\widetilde{B_i^n}$ (as stated more precisely in the sequel). Note that Z'_n has the same distribution for all n, so that the limiting distribution above is well defined. Below, our main goal is to provide a rate of convergence for this result. To do so, the strategy is to use a KMT approximation to build a coupling: we will define on the same probability space $\widetilde{B}$ and B that are very close. Then, when the letters are uniformly distributed, $Z_n(\lambda)$ simplifies to $Z_n(\lambda) = \sum_{i=1}^{m} \widetilde{V_i^n}(\lambda)$, and therefore it is straightforward to infer from the coupling the rate of convergence bound. This can also be done for the other lines of the RSK Young diagrams associated with the word. When the distribution is not uniform, the strategy is to first approximate $\max_{\lambda \in \Lambda_d} Z_n(\lambda)$ by $\max_{\lambda \in \Lambda'_d} Z_n(\lambda)$, where $\Lambda'_d = \{\lambda \in \Lambda_d : p_i \neq p_{\max} \implies \lambda_i = 0\}$, because then for each $\lambda \in \Lambda'_d$, $Z_n(\lambda) = \sum_{i=1}^{m} \widetilde{V_i^n}(\lambda)$, as previously.

3.1.1 A coupling via KMT and rates of convergence

To start with, we prove the following key coupling lemma:

Lemma 3.1.1. *Let $\alpha \geq 1$. For every $m \geq 2$, every probability mass function $p_1, \ldots, p_m$, and every $n \geq 2$, there exists a probability space with $X_1, \ldots, X_n, \widetilde{B^n}$, as above, and B^n an m-dimensional Brownian motion with covariance matrix $\Sigma := \mathrm{Cov}\left((\mathbb{1}_{X_1=i}/\sigma_i)_{1\leq i\leq m}\right)$, defined on it, such that*

$$\mathbb{P}\left(\sup_{\substack{i\in\{1,\ldots,m\}\\0\leq t\leq 1}}\left|\sigma_i\widetilde{B_i^n}(t)-\sigma_i B_i^n(t)\right|\geq C\frac{\alpha(\log n)^2}{\sqrt{n}}\right)\leq\frac{2}{n^\alpha},\tag{3.1.2}$$

where C is a universal constant.

Proof. By the KMT approximation [47] (see also [64] for further details and extensive references, in particular on Kiefer processes), there exists a probability space with $(U_i)_{i\geq 1}$ i.i.d. uniform on $[0,1]$ random variables and a Kiefer process $(K(s,t))_{\substack{s\in[0,1]\\t\in[0,\infty)}}$ such that for all $x>0$,

$$\mathbb{P}\left(\sup_{\substack{0\leq s\leq 1\\l\in\{1,\ldots,n\}}}\left|\frac{1}{\sqrt{n}}\sum_{k=1}^{l}(\mathbb{1}_{U_k\leq s}-s)-\frac{K(s,l)}{\sqrt{n}}\right|\geq C_1\log n\frac{\log n+x}{\sqrt{n}}\right)\leq \mathrm{e}^{-x},$$

where C_1 is a universal constant (throughout, $C_2, C_3, \ldots$ are universal constants). In particular, for $x=\alpha\log n$ one gets

$$\mathbb{P}\left(\sup_{\substack{0\leq s\leq 1\\l\in\{1,\ldots,n\}}}\left|\frac{1}{\sqrt{n}}\sum_{k=1}^{l}(\mathbb{1}_{U_k\leq s}-s)-\frac{K(s,l)}{\sqrt{n}}\right|\geq 2C_1\alpha\frac{(\log n)^2}{\sqrt{n}}\right)\leq\frac{1}{n^\alpha}.\tag{3.1.3}$$

To replace the discrete parameter $l\in\{1,\ldots,n\}$ by a continuous one $l'\in[0,n]$, note that

$$\begin{aligned}\mathbb{P}\left(\sup_{\substack{0\leq s\leq 1\\l\in\{1,\ldots,n\}}}\sup_{l'\in[l-1,l]}|K(s,l)-K(s,l')|\geq 2C_1\alpha(\log n)^2\right)&\leq\sum_{l=1}^{n}\mathbb{P}\left(\sup_{\substack{0\leq s\leq 1\\l'\in[l-1,l]}}|K(s,l)-K(s,l')|\geq 2C_1\alpha(\log n)^2\right)\\&\leq n\mathbb{P}\left(\sup_{\substack{0\leq s\leq 1\\0\leq t\leq 1}}|K(s,t)|\geq 2C_1\alpha(\log n)^2\right)\\&\leq n\mathbb{P}\left(2\sup_{\substack{0\leq s\leq 1\\0\leq t\leq 1}}|W(s,t)|\geq 2C_1\alpha(\log n)^2\right),\end{aligned}$$

where W is a two-dimensional Brownian sheet (using the facts that $(s,t)\mapsto K(s,l)-K(s,l-t)$ and $(s,t)\mapsto W(s,t)-sW(1,t)$ are Kiefer processes on $[0,1]^2$). From [23, Theorem 3], and if Φ is the standard normal cumulative distribution function,

$$\mathbb{P}\left(\sup_{\substack{0\leq s\leq 1\\0\leq t\leq 1}}|W(s,t)|\geq C_1\alpha(\log n)^2\right)\leq 4\Phi(-C_1\alpha(\log n)^2).\tag{3.1.4}$$

If C_1 is large enough, which can be assumed without loss of generality, for all $\alpha\geq 1$ and all $n\geq 2$, $4\Phi(-C_1\alpha(\log n)^2)\leq 1/n^{1+\alpha}$. Therefore,

$$\mathbb{P}\left(\sup_{\substack{0\leq s\leq 1\\l\in\{1,\ldots,n\}}}\sup_{l'\in[l-1,l]}|K(s,l)-K(s,l')|\geq 2C_1\alpha(\log n)^2\right)\leq\frac{n}{n^{1+\alpha}}=\frac{1}{n^\alpha},$$

and using (3.1.3),

$$\mathbb{P}\left(\sup_{\substack{0\le s\le 1\\ 0\le t\le 1}}\left|\frac{1}{\sqrt{n}}\sum_{k=1}^{\lfloor tn\rfloor}(\mathbb{1}_{U_k\le s}-s)-\frac{K(s,tn)}{\sqrt{n}}\right|\ge 4C_1\alpha\frac{(\log n)^2}{\sqrt{n}}\right)\le\frac{2}{n^\alpha}. \tag{3.1.5}$$

For $i\in\{1,\dots,n\}$, let $X_i:=\min_{k\in\{1,\dots,m\}}\{k:U_i\le p_1+\cdots+p_k\}$. Clearly, the X_i are i.i.d random variables with values in $\{1,\dots,m\}$ and probability mass function $p_1,\dots,p_m$. So, with the notations above,

$$\widetilde{B_i^n}(t)=\frac{N^i_{1,\lfloor tn\rfloor}-p_i\lfloor tn\rfloor}{\sigma_i\sqrt{n}}=\frac{\sum_{k=1}^{\lfloor tn\rfloor}\left(\mathbb{1}_{U_k\le p_1+\cdots+p_i}-(p_1+\cdots+p_i)-\mathbb{1}_{U_k\le p_1+\cdots+p_{i-1}}+p_1+\cdots+p_{i-1}\right)}{\sigma_i\sqrt{n}}.$$

For $i\in\{1,\dots,m\}$ and $t\in[0,1]$, then

$$B_i^n(t):=\frac{K(p_1+\cdots+p_i,tn)-K(p_1+\cdots+p_{i-1},tn)}{\sigma_i\sqrt{n}},$$

are Brownian motions with covariance matrix $\Sigma:=\mathrm{Cov}\left((\mathbb{1}_{X_1=i}/\sigma_i)_{1\le i\le m}\right)$. Note that

$$\sup_{\substack{i\in\{1,\dots,m\}\\ 0\le t\le 1}}\left|\sigma_i\widetilde{B_i^n}(t)-\sigma_iB_i^n(t)\right|\le 2\sup_{\substack{0\le s\le 1\\ 0\le t\le 1}}\left|\frac{1}{\sqrt{n}}\sum_{k=1}^{\lfloor tn\rfloor}(\mathbb{1}_{U_k\le s}-s)-\frac{K(s,tn)}{\sqrt{n}}\right|,$$

and so from (3.1.5) the following coupling inequality:

$$\mathbb{P}\left(\sup_{\substack{i\in\{1,\dots,m\}\\ 0\le t\le 1}}\left|\sigma_i\widetilde{B_i^n}(t)-\sigma_iB_i^n(t)\right|\ge 8C_1\frac{\alpha(\log n)^2}{\sqrt{n}}\right)\le\frac{2}{n^\alpha},$$

is valid and this gives the desired result (letting $C=8C_1$).

□

From now on, our setting is the probability space introduced in Lemma 3.1.1 with its notation.

To start with, we address the case of uniformly distributed letters and study the rate of convergence, in Kolmogorov distance, for the cumulative shape of the RSK Young diagrams associated with the random word.

Let $(R_k(n,m))_{1\le k\le n}$ be the shape of the RSK Young diagram associated with the random words $X_1\cdots X_n$ with uniformly distributed letters over $\{1,\dots,m\}$, and define for $1\le k\le m$, $V_k(n,m)=\sum_{l=1}^k R_l(n,m)$ (so that, for example, $V_1(n,m)=R_1(n,m)=LI_n$). From [34, Corollary 3.1], for properly defined $I_{k,m}$ (keeping the notations of [34]), $(V_k(n,m))_{1\le k\le n}$ is such that:

$$\left(\frac{V_k(n,m)-kn/m}{\sqrt{n/m}}\right)_{1\le k\le m}\overset{\longrightarrow}{\scriptstyle n\to\infty}\left(\max_{\mathbf{t}\in I_{k,m}}\sum_{j=1}^{k}\sum_{l=j}^{m-k+j}\sqrt{(m-1)/m}\,(B_l(t_{j,l})-B_l(t_{j,l-1}))\right)_{1\le k\le m}, \tag{3.1.6}$$

where the convergence is in distribution, and B is a m-dimensional Brownian motion with covariance matrix having diagonal terms equal to 1 and off-diagonal terms equal to $-1/(m-1)$, i.e., the covariance matrix Σ of Lemma 3.1.1. Recall finally that the limiting law in (3.1.6) is the spectra of a traceless $m\times m$ GUE matrix.

To simplify notations, let

$$(J_{1,m},\dots,J_{m,m})=\left(\max_{\mathbf{t}\in I_{k,m}}\sum_{j=1}^{k}\sum_{l=j}^{m-k+j}\sqrt{(m-1)/m}\,(B_l^n(t_{j,l})-B_l^n(t_{j,l-1}))\right)_{1\le k\le m}$$

and

$$(T_{1,m},\dots,T_{m,m}) = \left((V_k(n,m) - kn/m)/\sqrt{n/m}\right)_{1\le k\le m}.$$

Theorem 3.1.2. *For every $n,m \ge 2$ and $1 \le k \le m$,*

$$\sup_{x\in\mathbb{R}} |\mathbb{P}(T_{k,m} \ge x) - \mathbb{P}(J_{k,m} \ge x)| \le C(m)\frac{(\log n)^2}{\sqrt{n}},$$

where $C(m)$ is a constant only depending on m.

Proof. As shown in [34, (3.7)],

$$T_{k,m} = \max_{\mathbf{t}\in I_{k,m}} \sum_{j=1}^{k}\sum_{l=j}^{m-k+j} \sqrt{(m-1)/m}\left(\widetilde{B_l^n}(t_{j,l}) - \widetilde{B_l^n}(t_{j,l-1})\right),$$

so applying (3.1.2), since $|T_{k,m} - J_{k,m}| \le 2k(m-k+1)\sqrt{m}\sup_{\substack{i\in\{1,\dots,m\}\\ 0\le t\le 1}} \left|\sigma_i\widetilde{B_i^n}(t) - \sigma_i B_i^n(t)\right|$,

$$\mathbb{P}\left(|T_{k,m} - J_{k,m}| \ge 2k(m-k+1)\sqrt{m}C\alpha\frac{(\log n)^2}{\sqrt{n}}\right) \le \frac{2}{n^\alpha}. \tag{3.1.7}$$

Recall next that with $\Theta_k : \mathbb{R}^k \to \mathbb{R}^k$ given by $(\Theta_k(x))_j = \sum_{i=1}^j x_i$, $1 \le j \le k$, it follows that $\Theta_m^{-1}(J_{1,m},\dots,J_{m,m})$ has the same distribution as the ordered spectrum of an $m\times m$ traceless GUE matrix. Moreover, by a bound on the joint density of the eigenvalues (see (3.5) in [40]), there exists $D(m) > 0$ bounding the supremum of the density of $J_{k,m}$ (for $1 \le k \le m$).

So, for any $x \in \mathbb{R}$,

$$\begin{aligned}|\mathbb{P}(T_{k,m} \ge x) - \mathbb{P}(J_{k,m} \ge x)| &\le \mathbb{P}\left(|T_{k,m} - J_{k,m}| \ge 2k(m-k+1)\sqrt{m}C\alpha\frac{(\log n)^2}{\sqrt{n}}\right)\\ &\quad + \mathbb{P}\left(|J_{k,m} - x| \le 2k(m-k+1)\sqrt{m}C\alpha\frac{(\log n)^2}{\sqrt{n}}\right)\\ &\le \frac{2}{n^\alpha} + D(m)2k(m-k+1)\sqrt{m}C\alpha\frac{(\log n)^2}{\sqrt{n}},\end{aligned}$$

so $\sup_{x\in\mathbb{R}} |\mathbb{P}(T_{k,m} \ge x) - \mathbb{P}(J_{k,m} \ge x)|$ is upper bounded as stated.

□

As a corollary, we can also study the speed of convergence of $T_{\cdot,m} = (T_{k,m})_{1\le k\le m}$ towards $J_{\cdot,m} = (J_{k,m})_{1\le k\le m}$, in the Kolmogorov distance, rather than coordinate by coordinate. Just as before, we have

$$\mathbb{P}\left(\|T_{\cdot,m} - J_{\cdot,m}\|_\infty := \max_{1\le k\le m}|T_{\cdot,m} - J_{\cdot,m}| \ge 2m^{5/2}C\alpha\frac{(\log n)^2}{\sqrt{n}}\right) \le \frac{2}{n^\alpha},$$

and

$$\mathbb{P}\left(\max_{1\le k\le m}|J_{k,m} - x| \le 2m^{5/2}C\alpha\frac{(\log n)^2}{\sqrt{n}}\right) \le mD(m)2m^{5/2}C\alpha\frac{(\log n)^2}{\sqrt{n}},$$

so for any $x_1,\dots,x_m \in \mathbb{R}$,

$$\left|\mathbb{P}\left(\max_{1\le k\le m} T_{k,m}\ge x_k\right)-\mathbb{P}\left(\max_{1\le k\le m} J_{k,m}\ge x_k\right)\right|\le \mathbb{P}\left(\|T_{.,m}-J_{.,m}\|_\infty\ge 2m^{5/2}C\alpha\tfrac{(\log n)^2}{\sqrt{n}}\right)$$
$$+\mathbb{P}\left(\max_{1\le k\le m}|J_{k,m}-x|\le 2m^{5/2}C\alpha\tfrac{(\log n)^2}{\sqrt{n}}\right)$$
$$\le \tfrac{2}{n^\alpha}+mD(m)2m^{5/2}C\alpha\tfrac{(\log n)^2}{\sqrt{n}}.$$

As in [11], for m no longer fixed, let us consider conditions on the sequences $(m(n))_{n\ge1}$ (writing just m for $m(n)$) under which for all $\varepsilon>0$,

$$\mathbb{P}\left(\left|\left(T_{1,m}-2\sqrt{m}\right)m^{1/6}-\left(J_{1,m}-2\sqrt{m}\right)m^{1/6}\right|\ge\varepsilon\right)\xrightarrow[n\to\infty]{}0. \tag{3.1.8}$$

Then, since $(J_{1,m}-2\sqrt{m})\,m^{1/6}$ converges in distribution to the Tracy–Widom distribution F_{TW}, this implies that $(T_{1,m}-2\sqrt{m})\,m^{1/6}$, that is, $\left(\left(LI_n-n/m\right)/\sqrt{n/m}-2\sqrt{m}\right)m^{1/6}$, converges to F_{TW} as well.

Applying (3.1.7) gives

$$\mathbb{P}\left(\left|\left(T_{1,m}-2\sqrt{m}\right)m^{1/6}-\left(J_{1,m}-2\sqrt{m}\right)m^{1/6}\right|\ge Cm^{1/6}2k(m-k+1)\sqrt{m}\frac{\alpha(\log n)^2}{\sqrt{n}}\right)\le\frac{2}{n^\alpha},$$

so that if $m^{1/6}2k(m-k+1)\sqrt{m}C\alpha(\log n)^2/\sqrt{n}$ converges to zero, that is, if $m=o((n/(\log n)^4)^{3/10})$, then (3.1.8) follows, and therefore the convergence in distribution of the properly centered and scaled LI_n to the Tracy-Widom distribution also follows.

Beyond the uniform case, in order to evaluate the rate of convergence to the limiting law for arbitrary distributions, we first need to control how close $\max_{\lambda\in\Lambda_d} Z_n(\lambda)$ is to $\max_{\lambda\in\Lambda'_d} Z_n(\lambda)$ where again,

$$\Lambda'_d:=\{(j_1/n,\dots,j_m/n): j_1+\cdots+j_m=n \text{ and } p_i\ne p_{\max}\implies j_i=0\}.$$

Lemma 3.1.3. *Let $n,m\ge2,\alpha\ge1,a=6+3\alpha$ and $\Delta=p_{\max}-p_{2nd}$, where p_{2nd} is the second highest of the p_i's. Let $a\log n\le2\sqrt{n}\Delta$. Then,*

$$\mathbb{P}\left(\left|\frac{\max_{\lambda\in\Lambda_d}Z_n(\lambda)}{\sqrt{p_{\max}}}-\frac{\max_{\lambda\in\Lambda'_d}Z_n(\lambda)}{\sqrt{p_{\max}}}\right|>\frac{(a\log n)^2}{4\Delta\sqrt{np_{\max}}}+\frac{am\log n}{\sqrt{np_{\max}}}\right)\le\frac{2m}{n^\alpha}.$$

Proof. Our analysis of this result rests upon estimating the variations of $\widetilde{B_i^n}$. To do so, let A_n be the event:

$$\forall i\in\{1,\dots,m\},\forall j\in\{0,\dots,n-1\},\forall\ell\in\{1,\dots,n-j\},\left|\frac{N^i_{j+\ell}-N^i_j-p_i^X\ell}{\sqrt{n}}\right|\le\frac{\sigma_i\sqrt{\ell}+1}{\sqrt{n}}a\log n.$$

By Bernstein's inequality:

$$1-\mathbb{P}(A_n)\le2m\frac{n(n+1)}{2}\exp\left(-\frac{\frac12(\sigma_i\sqrt{\ell}+1)^2(a\log n)^2}{\ell\sigma_i^2+\frac13(\sigma_i\sqrt{\ell}+1)a\log n}\right)$$

and since

$$\frac{\frac12(\sigma_i\sqrt{\ell}+1)^2(a\log n)^2}{\ell\sigma_i^2+\frac13(\sigma_i\sqrt{\ell}+1)a\log n}\ge\frac{\frac12(\sigma_i\sqrt{\ell}+1)^2(a\log n)^2}{\frac23(\sigma_i\sqrt{\ell}+1)a\log n+\frac13(\sigma_i\sqrt{\ell}+1)a\log n}\wedge\frac{\frac12(\sigma_i\sqrt{\ell}+1)^2(a\log n)^2}{\ell\sigma_i^2+\frac12\ell\sigma_i^2}\ge\frac{a}{3}\log n,$$

it follows that

$$1 - \mathbb{P}(A_n) \leq 2mn^{2-\frac{a}{3}} \leq \frac{2m}{n^\alpha}.$$

If A_n occurs, then for all $\lambda \in \Lambda_d$ and $i \in \{1, \dots, m\}$,

$$\widetilde{V_i^n}(\lambda) \leq a \log n \left(\sigma_i \sqrt{\lambda_i} + \frac{1}{\sqrt{n}} \right). \tag{3.1.9}$$

Let A_n occur. Let $\lambda \in \Lambda_d$. Let $s = \sum_{j:p_j \neq p_{\max}} \lambda_j$, let $i \in \{1, \dots, m\}$ be such that $p_i = p_{\max}$ and let λ' be defined as: $\lambda'_j = 0$ for all j such that $p_j \neq p_{\max}$, $\lambda'_i = \lambda_i + s$, and $\lambda'_j = \lambda_j$ elsewhere. So $\lambda' \in \Lambda'_d$ and from (3.1.9),

$$Z_n(\lambda') \geq Z_n(\lambda) + \sqrt{n} s \Delta - \left(\sigma_i \sqrt{s} + \frac{1}{\sqrt{n}} \right) a \log n - \sum_{j:p_j \neq p_{\max}} \left(\sigma_j \sqrt{\lambda_j} + \frac{1}{\sqrt{n}} \right) a \log n,$$

which leads by the Cauchy-Schwarz inequality to

$$Z_n(\lambda') \geq Z_n(\lambda) + \sqrt{n} s \Delta - \sqrt{s} a \log n - \frac{am \log n}{\sqrt{n}}.$$

Hence, since $a \log n \leq 2\sqrt{n}\Delta$,

$$Z_n(\lambda') \geq Z_n(\lambda) - \frac{(a \log n)^2}{4\Delta\sqrt{n}} - \frac{am \log n}{\sqrt{n}},$$

and finally

$$0 \leq \max_{\lambda \in \Lambda_d} Z_n(\lambda) - \max_{\lambda \in \Lambda'_d} Z_n(\lambda) \leq \frac{(a \log n)^2}{4\Delta\sqrt{n}} + \frac{am \log n}{\sqrt{n}}.$$

Therefore,

$$\mathbb{P}\left(\left| \frac{\max_{\lambda \in \Lambda_d} Z_n(\lambda)}{\sqrt{p_{\max}}} - \frac{\max_{\lambda \in \Lambda'_d} Z_n(\lambda)}{\sqrt{p_{\max}}} \right| > \frac{(a \log n)^2}{4\Delta\sqrt{np_{\max}}} + \frac{am \log n}{\sqrt{np_{\max}}} \right) \leq 1 - \mathbb{P}(A_n) \leq \frac{2m}{n^\alpha}.$$

□

We can now deduce our theorem, where below J_k is defined as in [40, Theorem 4.1] (J_k has same law than $J_{1,k} + \sqrt{(1 - kp_{\max})/k} Z$, where $Z \sim \mathcal{N}(0,1)$ is independent from the other variables)

Theorem 3.1.4. *Let $n, m \geq 2$ and let k be the multiplicity of $p_{\max}$, then,*

$$\left| \mathbb{P}\left(\frac{LI_n - np_{\max}}{\sqrt{np_{\max}}} \geq x \right) - \mathbb{P}(J_k \geq x) \right| \leq 2D(k, p_{\max}) \frac{(\log n)^2}{\sqrt{np_{\max}}} \left(\frac{21}{\Delta} + C_2 m \right), \tag{3.1.10}$$

where C_2 is a universal constant, where $D(1, p_{\max}) := 1/\sqrt{2\pi(1 - p_{\max})}$, and where for $k \geq 2$, $D(k, p_{\max}) := \min\left\{ \sqrt{k/2\pi(1 - kp_{\max})}, k^{3k}(2\pi e^2)^{k/2}\sqrt{e/\pi} \right\}$.

Proof. We apply Lemma 3.1.1, and for $\lambda \in \Lambda' := \{\lambda \in \Lambda : p_i \neq p_{\max} \implies \lambda_i = 0\}$, let $Z'_n(\lambda)$ be defined as $Z_n(\lambda)$ but with B_i^n instead of $\widetilde{B_i^n}$. For any $\lambda \in \Lambda'$, $|Z_n(\lambda) - Z'_n(\lambda)| \leq 2m \sup_{\substack{i \in \{1,\dots,m\} \\ 0 \leq t \leq 1}} \left| \sigma_i \widetilde{B_i^n}(t) - \sigma_i B_i^n(t) \right|$, hence

$$\mathbb{P}\left(\left| \sup_{\lambda \in \Lambda'} Z_n(\lambda) - \sup_{\lambda \in \Lambda'} Z'_n(\lambda) \right| \geq \frac{2C\alpha m (\log n)^2}{\sqrt{n}} \right) \leq \frac{2}{n^\alpha}. \tag{3.1.11}$$

For all $\lambda \in \Lambda'$, $Z_n(\lambda) = \sum_{i=1}^m \widetilde{V_i^n}(\lambda)$, so $\sup_{\lambda \in \Lambda'} Z_n(\lambda) = \max_{\lambda \in \Lambda'_d} Z_n(\lambda)$. So

$$\mathbb{P}\left(\left|\max_{\lambda\in\Lambda_d'} Z_n(\lambda) - \max_{\lambda\in\Lambda'} Z_n'(\lambda)\right| \geq \frac{2C\alpha m(\log n)^2}{\sqrt{n}}\right) \leq \frac{2}{n^\alpha}.$$

For any $x \in \mathbb{R}$, $\alpha \geq 1$, and $a = 6 + 3\alpha$,

$$\begin{aligned}\left|\mathbb{P}\left(\frac{\max_{\lambda\in\Lambda_d} Z_n(\lambda)}{\sqrt{p_{\max}}} \geq x\right) - \mathbb{P}\left(\frac{\max_{\lambda\in\Lambda'} Z_n'(\lambda)}{\sqrt{p_{\max}}} \geq x\right)\right| &\leq \mathbb{P}\left(\left|\frac{\max_{\lambda_d\in\Lambda} Z_n(\lambda)}{\sqrt{p_{\max}}} - \frac{\max_{\lambda\in\Lambda_d'} Z_n(\lambda)}{\sqrt{p_{\max}}}\right| > \frac{(a\log n)^2}{4\Delta\sqrt{np_{\max}}} + \frac{am\log n}{\sqrt{np_{\max}}}\right)\\ &+\mathbb{P}\left(\left|\frac{\max_{\lambda\in\Lambda_d'} Z_n(\lambda)}{\sqrt{p_{\max}}} - \frac{\max_{\lambda\in\Lambda'} Z_n'(\lambda)}{\sqrt{p_{\max}}}\right| > \frac{2C\alpha m(\log n)^2}{\sqrt{np_{\max}}}\right)\\ &+\mathbb{P}\left(\left|\frac{\max_{\lambda\in\Lambda'} Z_n'(\lambda)}{\sqrt{p_{\max}}} - x\right| \leq \frac{(a\log n)^2}{4\Delta\sqrt{np_{\max}}} + \frac{am\log n}{\sqrt{np_{\max}}} + \frac{2C\alpha m(\log n)^2}{\sqrt{np_{\max}}}\right).\end{aligned}$$

Now, with the notations of [40, Theorem 4.1], we see that $\max_{\lambda\in\Lambda'} Z_n'(\lambda)/\sqrt{p_{\max}}$ has the same law than J_k. Indeed, one can check that in [40], $\tilde{B}$ is an $(m-1)$-dimensional Brownian motion with the same covariance as $\left((\sigma_r B_r^n - \sigma_{r+1}B_{r+1}^n)/\sqrt{p_r + p_{r+1} - (p_r - p_{r+1})^2}\right)_{1\leq r\leq m}$, and rewriting $\tilde{B}$ that way gives exactly $\max_{\lambda\in\Lambda'} Z_n'(\lambda)/\sqrt{p_{\max}}$. Assuming $a\log n \leq 2\sqrt{n}\Delta$, from 3.1.3, [40, Proposition 3.1 (ii)] and the inequality $a = 6 + 3\alpha \leq 9\alpha$, one gets

$$\left|\mathbb{P}\left(\frac{LI_n - np_{\max}}{\sqrt{np_{\max}}} \geq x\right) - \mathbb{P}\left(J_k \geq x\right)\right| \leq \frac{4m}{n^\alpha} + 2D(k, p_{\max})\frac{(\log n)^2}{\sqrt{np_{\max}}}\left(\frac{21\alpha^2}{\Delta} + \left(2C + \frac{9}{\log 2}\right)\alpha m\right), \tag{3.1.12}$$

where we refer to [40, Proposition 3.1] for a proof that $D(k, p_{\max})$ is a bound on the infinity norm of the density of the limiting distribution $\max_{\lambda\in\Lambda'} Z_n'(\lambda)$. Taking $\alpha = 1$ and using $D(k, p_{\max}) \geq 1/\sqrt{2\pi}$, one gets, in particular, (3.1.10). Now if we no longer assume that $a\log n \leq 2\sqrt{n}\Delta$, the bound (3.1.12) (and therefore (3.1.10)) above is still valid because its right-hand side is then greater than 1. □

3.1.2 Some remarks

(i) Theorem 3.1.4 is slightly weaker than Theorem 4.1 in [40], because of the $(\log n)^2$ factor instead of $\log n$ and of the extra term $21/\Delta$.

The supplementary $\log n$ comes from the different version of the KMT Theorem (strong embedding) that is used compared to the one in [40] (weak embedding for each coordinate). Moreover, the proof that they can be built on a same probability space with the right covariance has been in question. More precisely, in [40], the authors use the KMT Theorem to get for each $i \in \{1, \ldots, m\}$ a probability space with a version of the random walk $\widetilde{B_i^n}$ and a Brownian motion approximating it, but it is not clear how this implies that on some probability space, there are versions of $\widetilde{B_1^n}, \ldots, \widetilde{B_m^n}$ satisfying the additional covariance requirement $\widetilde{B_1^n} + \cdots + \widetilde{B_m^n} = 0$. However, using Kiefer's version of the KMT approximation, we have a probability space with $(U_i)_{1\leq i\leq n}$ uniform on $[0, 1]$ and with $\widetilde{B_1^n}, \ldots, \widetilde{B_m^n}$ well defined on it (and satisfying the right covariance relation) as shown in the proof of Lemma 3.1.2.

Note that to the best of our knowledge, it is still an open problem to know whether or not the extra $\log n$ or $(\log n)^2$ factors in the strong embedding could be improved. Note, nevertheless, the rate $1/\sqrt{n}$ in the uniform binary case (see [40, Theorem 5.1]).

The extra term $21/\Delta$ is needed because if one picks $\Delta = p_{\max} - p_{2nd}$ very small, there is no hope, for any fixed m, to have a sequence $(A_n^m)_{n\geq 2}$ converging to zero such that for any distribution $p_1, \ldots, p_m$, any $n \geq 2$ and $x \in \mathbb{R}$,

$$\left|\mathbb{P}\left(\frac{LI_n - np_{\max}}{\sqrt{np_{\max}}} \geq x\right) - \mathbb{P}\left(J_k \geq x\right)\right| \leq A_n^m.$$

Indeed, assume it is the case for, say, $m = 2$. Let us apply this bound to

- Case a) $p_1 = p_2 = 1/2$, and denote $\left(LI_n - np_{\max}\right)/\sqrt{np_{\max}}$ by Z_n
- Case b) $p_1 = 1/2 + 1/2^n, p_2 = 1/2 - 1/2^n$, and denote $\left(LI_n - np_{\max}\right)/\sqrt{np_{\max}}$ by Z'_n.

We get

$$\mathbb{P}\left(Z_n \geq x\right) - \mathbb{P}\left(J_2 \geq x\right) - \mathbb{P}\left(Z'_n \geq x\right) + \mathbb{P}\left(J_1 \geq x\right) \xrightarrow[n\to\infty]{} 0.$$

There is a coupling such that with high probability, the letters $X_1, \ldots, Y_n$ in case a) are all equal to the letters in case b), hence

$$\mathbb{P}\left(Z_n \geq x\right) - \mathbb{P}\left(Z'_n \geq x\right) \xrightarrow[n\to\infty]{} 0.$$

Furthermore, J_1 has distribution $\mathcal{N}\left(0, \frac{1}{4} - \frac{1}{2^{n+1}}\right)$ so

$$\mathbb{P}\left(J_1 \geq x\right) \xrightarrow[n\to\infty]{} \mathbb{P}\left(Z_{1/4} \geq x\right),$$

where $Z_{1/4} \sim \mathcal{N}\left(0, 1/4\right)$. Putting together these three limits, we get

$$\mathbb{P}\left(J_2 \geq x\right) = \mathbb{P}\left(Z_{1/4} \geq x\right),$$

which means that the limiting law for the uniform binary case is normal. However, this is known to be false. So, $(A_n^m)_{n\geq 2}$ has to depend on the distribution. We can actually find a contradiction as soon as $\Delta = o\left(1/\sqrt{n}\right)$. Our bound (3.1.10), on the other hand, is not exposed to this kind of cases because if the right-hand side is less than 1 then $\Delta \geq \log n/\sqrt{n}$. So Theorem 4.1 in [40] only holds for n large enough, and not for all $n \geq 2$.

(ii) As before, let us no longer consider m, k and the distribution $p_1, \ldots, p_m$ to be fixed. We assume that both m and k converge to infinity with n, and aim to find for which sequences we have for all $\varepsilon > 0$,

$$\mathbb{P}\left(\left|\left(\frac{LI_n - np_{\max}}{\sqrt{np_{\max}}} - 2\sqrt{k}\right)k^{1/6} - \left(J_k - 2\sqrt{k}\right)k^{1/6}\right| \geq \varepsilon\right) \xrightarrow[n\to\infty]{} 0. \tag{3.1.13}$$

Then, since $\left(J_k - 2\sqrt{k}\right)k^{1/6}$ has the same distribution as $\left(J_{1,k} - 2\sqrt{k}\right)k^{1/6} + k^{1/6}\sqrt{\left(1 - kp_{\max}\right)/k}Z$, it will converge in distribution to F_{TW}, implying that $\left(\left(LI_n - np_{\max}\right)/\sqrt{np_{\max}} - 2\sqrt{k}\right)k^{1/6}$ converges to F_{TW} as well.

Applying the Lemma 3.1.3 and the bound (3.1.11) lead to

$$\mathbb{P}\left(\left|\left(\frac{LI_n - np_{\max}}{\sqrt{np_{\max}}} - 2\sqrt{k}\right)k^{1/6} - \left(J_k - 2\sqrt{k}\right)k^{1/6}\right| \geq \frac{(\log n)^2}{\sqrt{np_{\max}}}\left(\frac{21}{\Delta} + C_2 m\right)\right) \leq \frac{4m}{n^\alpha}.$$

Taking $m = o((n/(\log n)^4)^{3/10})$ (as previously done in the uniform case) and $m = o(\sqrt{np_{\max}}\Delta/(\log n)^2)$, then (3.1.13) follows.

Note that in particular, when the conditions of Theorem 6 in [11] are satisfied, the condition $m = o(\sqrt{np_{\max}}\Delta/\log n)$ follows. This is not enough to conclude, first, we need $(\log n)^2$ instead of $\log n$, and more importantly, the first condition $m = o((n/(\log n)^4)^{3/10})$ is missing. This is an omission in [11], because as it is, there is no condition on m and this leads to a counterexample. Indeed, let $k = n^{1/9}, p_{\max} = n^{-2/3}, m = 2^n + k, p_{2nd} = \left(1 - n^{5/9}\right)/2^n$, it is easy to check that the conditions of Theorem 6 there, hold true but its conclusion does not. The two conditions on m we give above do fix this issue.

(iii) Let us investigate the convergence of moments which, in particular, will provide a speed of convergence result in the distance W_p, $p \geq 1$, given by

$$W_p(X,Y) := \inf_{\substack{X' \text{ has same law than } X \\ Y' \text{ has same law than } Y}} \left(\mathbb{E}\,|X'-Y'|^p\right)^{1/p}.$$

Let us start with the uniform case. We have seen that

$$\mathbb{P}\left(|T_{k,m}-J_{k,m}| \geq 2k(m-k+1)\sqrt{m}C\frac{\alpha(\log n)^2}{\sqrt{n}}\right) \leq \frac{2}{n^\alpha}.$$

In particular, taking $\alpha = p$, and setting $\varepsilon_{n,m,k} := 2k(m-k+1)\sqrt{m}Cp(\log n)^2/\sqrt{n}$, we have

$$\begin{aligned}\mathbb{E}\,|T_{k,m}-J_{k,m}|^p &\leq \mathbb{E}\Big(|T_{k,m}-J_{k,m}|^p\,\mathbb{1}_{|T_{k,m}-J_{k,m}|\leq\varepsilon(n,m,k)} + |T_{k,m}-J_{k,m}|^p\,\mathbb{1}_{\varepsilon(n,m,k)<|T_{k,m}-J_{k,m}|<2m\sqrt{n}} \\ &\qquad + |T_{k,m}-J_{k,m}|^p\,\mathbb{1}_{2m\sqrt{n}\leq|T_{k,m}-J_{k,m}|}\Big) \\ &\leq \varepsilon(n,m,k)^p + (2m\sqrt{n})^p\frac{2}{n^p} + \mathbb{E}\,|2J_{k,m}|^p\,\mathbb{1}_{m\sqrt{n}\leq|J_{k,m}|},\end{aligned}$$

(since $|T_{k,m}| \leq \sqrt{n} \leq m\sqrt{n}$). Using, once more, (3.1.4) and an integration by parts, one gets for some constant $C(p)$

$$\mathbb{E}\,|2J_{k,m}|^p\,\mathbb{1}_{m\sqrt{n}\leq|J_{k,m}|} \leq C(p)\left(\frac{m}{\sqrt{n}}\right)^p,$$

and therefore, for some constant $C'(p)$,

$$W_p(T_{k,m},J_{k,m}) \leq \mathbb{E}\,|T_{k,m}-J_{k,m}|^p \leq C'(p)\varepsilon(m,n,k).$$

(iv) Theorem 3.1.4 can be generalized to longest common and increasing subsequences. We use the notations and framework in [18]. Additionally, let $Z_n = (LCI_n - ne_{\max})/\sqrt{n}$ and let Z be its limiting distribution, that is, either Z^a or Z^b depending on the distributions p^X, p^Y. The proof of Lemma 2.2 in [18], taking $\eta = 1/12$, holds for the sequence $(\varepsilon_n) = C_1(m,p^X,p^Y)n^{-1/8}$, where $C_1(m,p^X,p^Y)$ is a constant depending on m, p^X, p^Y. To study the cases a) and b) under the same umbrella, let us define the function $L : (\mathbb{R}^m)^2 \to \mathbb{R}$ by: for all $(v^X,v^Y) \in (\mathbb{R}^m)^2$, in case a), $L(v^X,v^Y) = \sum_{i=1}^m v^X$, and in case b), $L(v^X,v^Y) = \mathfrak{m}(v^X,v^Y)$. It is not hard to see, from the expression of $\mathfrak{m}$ in Lemma 1.5 in [18], that in each case, there is a constant $C_2(m,p^X,p^Y)$ such that L is $C_2(m,p^X,p^Y)$-Lipschitz for the ℓ^1-distance on $(\mathbb{R}^m)^2$. Now, as a consequence of Lemma 2.2 there, it follows that if $B^{n,X}, B^{n,Y}$ are in $E_n^{1/12}$, then

$$\left|Z_n - \max_{\lambda\in U} L(V^{n,X},V^{n,Y})\right| \leq \frac{C_1(m,p^X,p^Y)}{n^{1/8}},$$

where $U = J$ in case a) and $U = K_{\Lambda^2}$ in case b). Therefore,

$$\mathbb{P}\left(\left|Z_n - \max_{\lambda\in U} L(V^{n,X},V^{n,Y})\right| > \frac{C_1(m,p^X,p^Y)}{n^{1/8}}\right) \leq 1 - \mathbb{P}\left(A_n^{1/12}\right) \leq \frac{C(m)}{n}, \qquad (3.1.14)$$

where $C(m)$ is a constant (recalling (1.9) in [18]). Furthermore, from the expression of the limiting distribution in Theorem 2.1, we see that in any case a) or b), the limiting distribution may be written as $L(B^X,B^Y)$. We construct, with our Lemma 2.1, two Brownian motions $\widehat{B_n^X}, \widehat{B_n^Y}$ "close" to $B^{n,X}, B^{n,Y}$, on the same probability space (this is possible by applying the lemma twice, and then taking the product space). Let $\widehat{Z_n} = \max_{\lambda\in U} L(\widehat{V^{n,X}}, \widehat{V^{n,Y}})$, it

has the same distribution as the limiting distribution Z, and

$$\mathbb{P}\left(\left|\max_{\lambda\in U} L(V^{n,X},V^{n,Y}) - \widehat{Z_n}\right| > 2mC_2(m,p^X,p^Y)\varepsilon\right) \leq \mathbb{P}\left(\sup_{\substack{i\in\{1,\dots,m\}\\0\leq t\leq 1}}\left|\sigma_i\widehat{B_i^{n,X}}(t) - \sigma_i B_i^{n,X}(t)\right| \geq \varepsilon\right) + \mathbb{P}\left(\sup_{\substack{i\in\{1,\dots,m\}\\0\leq t\leq 1}}\left|\sigma_i\widehat{B_i^{n,Y}}(t) - \sigma_i B_i^{n,Y}(t)\right| \geq \varepsilon\right),$$

so in particular, applying Lemma 2.1,

$$\mathbb{P}\left(\left|\max_{\lambda\in U} L(V^{n,X},V^{n,Y}) - \widehat{Z_n}\right| > 2mC_2(m,p^X,p^Y)C\frac{\alpha(\log n)^2}{\sqrt{n}}\right) \leq \frac{4}{n^\alpha}. \tag{3.1.15}$$

Putting together (3.1.14) and (3.1.15), with $\alpha = 1$, gives

$$\mathbb{P}\left(\left|Z_n - \widehat{Z_n}\right| > \frac{C_1(m,p^X,p^Y)}{n^{1/8}} + 2mC_2(m,p^X,p^Y)C\frac{(\log n)^2}{\sqrt{n}}\right) \leq \frac{C(m)+4}{n}.$$

If, just as in the single subsequence case, a bound $D(m,p^X,p^Y)$ on the density of Z is possible, then we can conclude in the same way: For all $x\in\mathbb{R}$,

$$\begin{aligned}|\mathbb{P}(Z_n\geq x) - \mathbb{P}(Z\geq x)| &\leq \mathbb{P}\left(\left|Z_n - \widehat{Z_n}\right| > \frac{C_1(m,p^X,p^Y)}{n^{1/8}} + 2mC_2(m,p^X,p^Y)C\frac{(\log n)^2}{\sqrt{n}}\right)\\ &\quad + \mathbb{P}\left(\left|\widehat{Z_n} - x\right| \leq \frac{C_1(m,p^X,p^Y)}{n^{1/8}} + 2mC_2(m,p^X,p^Y)C\frac{(\log n)^2}{\sqrt{n}}\right)\\ &\leq \frac{C(m)+4}{n} + D(m,p^X,p^Y)\left(\frac{C_1(m,p^X,p^Y)}{n^{1/8}} + 2mC_2(m,p^X,p^Y)C\frac{(\log n)^2}{\sqrt{n}}\right)\\ &\leq \frac{C_3(m,p^X,p^Y)}{n^{1/8}}.\end{aligned}$$

Note that the exponent $-1/8$ can be improved taking a smaller η, but only up to $-1/4$. This is because in [18], the focus was to get convergence in distribution, rather than a tight bound. It might even be possible to get $(\log n)^2/\sqrt{n}$ instead.

3.2 Two other estimators

In this section, we go back to our main motivation of finding an estimator with lower risk than the empirical Young diagram (we refer to Definition (3.0.1)).

3.2.1 Bootstrapping

We introduce a bootstrap estimator and compute the empirical risk, which seems to be upper bounded by C/n in all the simulations we ran, and beats the classical empirical Young diagram estimator (although there is no formal proof at this point). The idea, given a Young diagram λ, is to take $\hat{p}_0 := \underline{\lambda}$ as the first estimate of p, and draw $\lambda_1 \sim \mathrm{SW}^n(\hat{p}_0)$. Then, we assume that $\underline{\lambda_1} - \hat{p}_0$ (which is related to the "excess", developed in Section 3.3) is "close" to $\underline{\lambda} - p$. Therefore, a "better" estimate for p is

$$\hat{p}_1 := \texttt{descending}(\underline{\lambda} - (\underline{\lambda_1} - \hat{p}_0)),$$

where `descending` sorts a list in descending order (in order to avoid absurd results), and we iterate this process: assume $\hat{p}_n$ is defined, draw $\lambda_{n+1}\sim \mathrm{SW}^n(\hat{p}_n)$ and let

$$\hat{p}_{n+1} := \texttt{descending}(\underline{\lambda} - (\underline{\lambda_{n+1}} - \hat{p}_n)).$$

The following plot gives the empirical mean (over 100 simulations) of the L_2-loss of $\hat{p}_0$ (the empirical young diagram estimator) and $\hat{p}_1, \hat{p}_1, \ldots, \hat{p}_{50}$ for $n = 100000$, $d = 100$, and various distributions:

- $(1/100, \ldots, 1/100)$ (uniform)
- $(3/200, \ldots, 3/200, 1/200, \ldots, 1/200)$ (a "fifty-fifty" distribution)
- $(C/1, C/2, \ldots, C/100)$ (Zipf distribution)

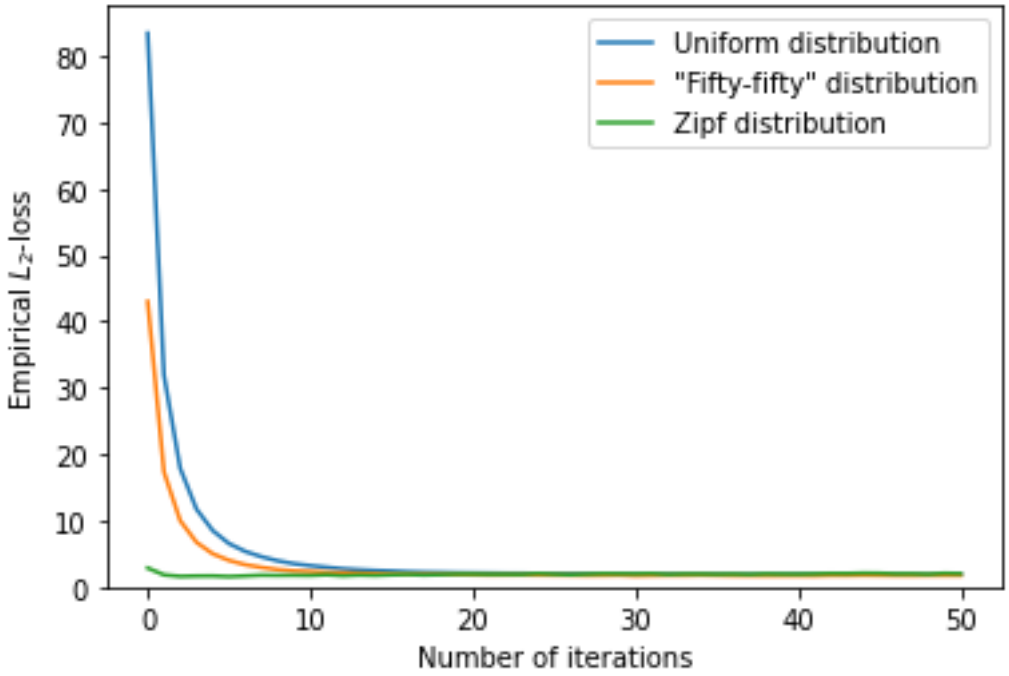


Figure 3.1: The risk of the bootstrap estimator as a function of the number of iterations

3.2.2 Minimum mean square error estimator; improvement on the sum of the variances

Let $\Delta_d = \{(p_1, \ldots, p_d) : p_1, \ldots, p_d \geq 0, p_1 + \cdots + p_d = 1\}$ and let g_n be defined, for any $\lambda^0 \vdash n$, by

$$g_n(\lambda^0) = \operatorname*{argmin}_{p \in \Delta_d} \mathbb{E}_{\lambda^1 \sim \mathrm{SW}^n(p)} \|\lambda^1 - \lambda^0\|^2. \tag{3.2.1}$$

In order to prove that this estimator has a risk of order $1/n$, we make the following conjecture:

Conjecture 3.2.1. *There exists a universal constant $C > 0$, for any $p_1, p_2 \in \Delta_d$,*

$$\|p_1 - p_2\|^2 \leq C \frac{\mathbb{E}_{(\lambda^1, \lambda^2) \sim \mathrm{SW}^n(p_1) \otimes \mathrm{SW}^n(p_2)} \|\lambda^1 - \lambda^2\|^2}{n \mathbb{E}_{\lambda^1 \sim \mathrm{SW}^n(p_1)} \left\|\lambda^1 - \mathbb{E}\lambda^1\right\|^2}$$

This conjecture is, once again, consistent with numerical simulations. Let $R_n(p_1, p_2)$ denote the ratio

$$R_n(p_1, p_2) = \frac{\|p_1 - p_2\|^2 n \mathbb{E}_{\lambda^1 \sim \mathrm{SW}^n(p_1)} \left\|\lambda^1 - \mathbb{E}\lambda^1\right\|^2}{\mathbb{E}_{(\lambda^1, \lambda^2) \sim \mathrm{SW}^n(p_1) \otimes \mathrm{SW}^n(p_2)} \|\lambda^1 - \lambda^2\|^2},$$

some numerical simulations tending to show that R_n is upper bounded: below, we fix $n = 50000$ and draw, for various d, two distributions p_1, p_2 by picking the probabilities randomly uniformly between 0 and 1, renormalizing to get a probability distribution and sorting in descending order.

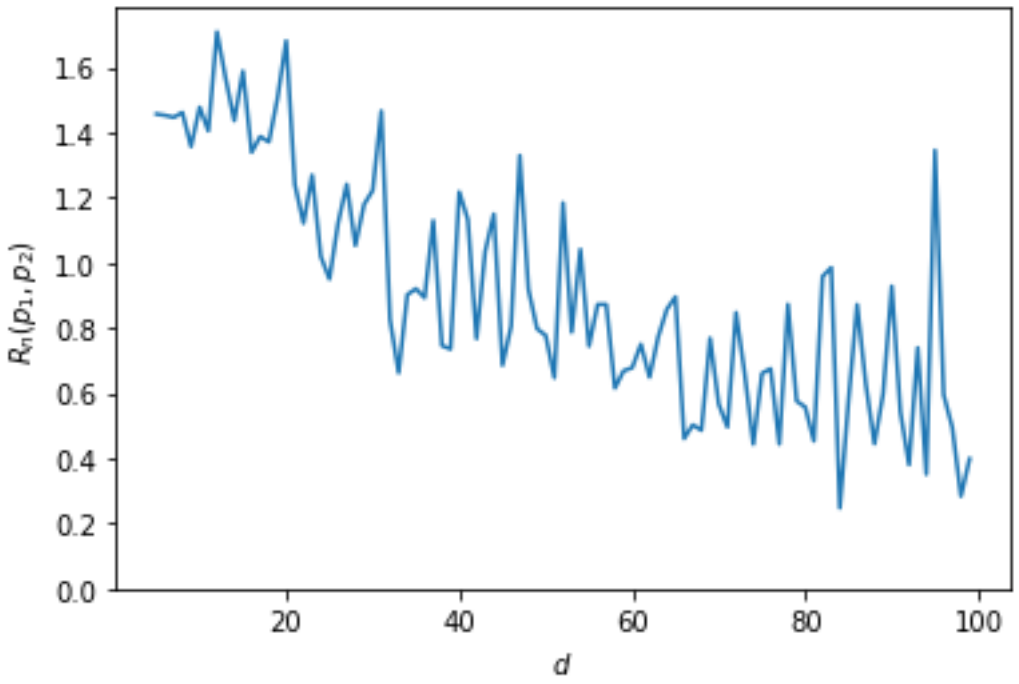


Figure 3.2: $R_n(p_1, p_2)$ for various distributions p_1, p_2

Proposition 3.2.2. *If Conjecture 3.2.1 holds, then the estimator g_n defined in* (3.2.1) *has a risk lower than $8C/n$.*

Proof. Assume Conjecture 3.2.1. Fix $\lambda \vdash n$. Then,

$$\begin{aligned}
\|p - g_n(\lambda)\|^2 &\leq C \frac{\mathbb{E}_{(\lambda^1,\lambda^2)\sim \mathrm{SW}^n(p)\otimes \mathrm{SW}^n(g_n(\lambda))} \|\lambda^1 - \lambda^2\|^2}{n \mathbb{E}_{\lambda^1 \sim \mathrm{SW}^n(p)} \left\|\lambda^1 - \mathbb{E}\lambda^1\right\|^2} \\
&\leq C \frac{\left(\sqrt{\mathbb{E}_{\lambda^1\sim \mathrm{SW}^n(p)} \|\lambda^1 - \lambda\|^2} + \sqrt{\mathbb{E}_{\lambda^2 \sim \mathrm{SW}^n(g_n(\lambda))} \|\lambda - \lambda^2\|^2}\right)^2}{n \mathbb{E}_{\lambda^1 \sim \mathrm{SW}^n(p)} \left\|\lambda^1 - \mathbb{E}\lambda^1\right\|^2} \\
&\leq C \frac{\left(2\sqrt{\mathbb{E}_{\lambda^1\sim \mathrm{SW}^n(p)} \|\lambda^1 - \lambda\|^2}\right)^2}{n \mathbb{E}_{\lambda^1 \sim \mathrm{SW}^n(p)} \left\|\lambda^1 - \mathbb{E}\lambda^1\right\|^2},
\end{aligned}$$

hence

$$\mathbb{E}_{\lambda \sim \mathrm{SW}^n(p)} \|p - g_n(\lambda)\|^2 \leq \frac{8C}{n}.$$

□

In an effort to prove the conjecture, a better understanding of the quantity $\mathbb{E}_{\lambda^1 \sim \mathrm{SW}^n(p)} \left\|\lambda^1 - \mathbb{E}\lambda^1\right\|^2$ is useful. In the following, $\lambda \sim \mathrm{SW}^n(p)$, so this quantity may be rewritten as $\sum_{k=1}^d \operatorname{Var} \lambda_k$. Numerical simulations tend to indicate that $\sum_{k=1}^d \operatorname{Var} \lambda_k \leq n$, which would also be consistent with the asymptotics of λ (Theorem 0.3.7). We prove below a slightly weaker result.

It is known [56, Proposition 4.8] that for any $k \in \{1, \ldots, d\}$,

$$\operatorname{Var} \lambda_k \leq 16n. \tag{3.2.2}$$

We also remind the useful result [56, Proposition 2.2], with the notation $\lambda_{\leq k} = \sum_{j=1}^{k} \lambda_j$:

Proposition 3.2.3. *Assume $w, w' \in \{1, \dots, d\}^n$ differ in exactly one coordinate and let $k \in \{1, \dots, d\}$, $\lambda = \mathrm{RSKshape}(w)$, $\lambda' = \mathrm{RSKshape}(w')$ Then,*

$$\left|\lambda_{\leq k} - \lambda'_{\leq k}\right| \leq 1,$$

and

$$|\lambda_k - \lambda'_k| \leq 2.$$

Note that Proposition 3.2.3 follows from Greene's Theorem 0.3.1, while the bound (3.2.2) follows from Proposition 3.2.3, using Hoeffding's inequality (Theorem 0.1.4).

The variance bound (3.2.2) provides a weak upper bound on $\sum_{k=1}^{d} \mathrm{Var}\, \lambda_k$:

$$\sum_{k=1}^{d} \mathrm{Var}\, \lambda_k \leq 16dn.$$

We leave open the conjecture that the sum of the variances is bounded by n, but we prove the following:

Theorem 3.2.4. *We have*

$$\sum_{k=1}^{d} \mathrm{Var}\, \lambda_k \leq 30 n^{5/4} \sqrt{\log n}.$$

Proof. For any $j \in \{1, \dots, d\}$, let $\lambda'_j = \lambda_j - \mathbb{E}\lambda_j$, and $\lambda'_{\leq j} = \sum_{i=1}^{j} \lambda'_i$. Using Proposition 3.2.3 and Hoeffding's inequality gives

$$\mathbb{P}\left(|\lambda'_{\leq j}| \geq t\right) \leq 2\mathrm{e}^{-\frac{t^2}{2n}}. \tag{3.2.3}$$

Now rewrite the sum of the variances, for any $k \in \{1, \dots, d \wedge n\}$ (to be chosen later):

$$\begin{aligned}
\sum_{j=1}^{d} \mathrm{Var}\, \lambda_j &= \mathbb{E}\left(\sum_{j=1}^{d\wedge n} \lambda_j \lambda'_j\right) \\
&= \mathbb{E}\left(\sum_{j=1}^{k} \lambda_j \lambda'_j + \sum_{j=k+1}^{d\wedge n} \lambda_j (\lambda'_{\leq j} - \lambda'_{\leq j-1})\right) \\
&= \mathbb{E}\left(\sum_{j=1}^{k} \lambda_j \lambda'_j + \sum_{j=k+1}^{d\wedge n} (\lambda_j - \lambda_{j+1}) \lambda'_{\leq j} - \lambda_{k+1} \lambda'_{\leq k}\right) \\
&\leq \sum_{j=1}^{k} \mathrm{Var}\, \lambda_j + \mathbb{E}\left(\sum_{j=k+1}^{d\wedge n} (\lambda_j - \lambda_{j+1}) \max_{k\leq j\leq d\wedge n} |\lambda'_{\leq j}| + \lambda_{k+1} \max_{k\leq j\leq d\wedge n} |\lambda'_{\leq j}|\right) \\
&\leq k(16n) + 2\mathbb{E}\left(\max_{k\leq j\leq d\wedge n} |\lambda'_{\leq j}| \lambda_{k+1}\right),
\end{aligned}$$

with the convention $\lambda_{d+1} = 0$, and where we use the facts that $\lambda_j - \lambda_{j+1} \geq 0$ for any $j \geq 1$ and $\lambda_j = 0$ for any $j > d \wedge n$. Since $(k+1)\lambda_{k+1} \leq n$ (the λ_j's are decreasing with sum n),

$$\mathbb{E}\left(\max_{k\leq j\leq d\wedge n} |\lambda'_{\leq j}| \lambda_{k+1}\right) \leq \frac{n}{k+1} \mathbb{E}\left(\max_{k\leq j\leq d\wedge n} |\lambda'_{\leq j}|\right),$$

and by the concentration inequality (3.2.3), we get for any $\alpha > 0$

$$\mathbb{P}\left(\max_{k\le j\le d\wedge n}\left|\lambda'_{\le j}\right| > \sqrt{2\alpha n\log n}\right) \le 2n\frac{1}{n^\alpha},$$

so

$$\mathbb{E}\left(\max_{k\le j\le d\wedge n}\left|\lambda'_{\le j}\right|\right) \le \sqrt{2\alpha n\log n} + \left(2n\frac{1}{n^\alpha}\right)n$$

and taking $\alpha = 3/2$,

$$\mathbb{E}\left(\max_{k\le j\le d\wedge n}\left|\lambda'_{\le j}\right|\right) \le 5\sqrt{n\log n}.$$

Now letting $k = \lfloor n^{1/4}\rfloor$ (we may assume without loss of generality $k \le d$, otherwise the upper bound $\sum_{j=1}^d \operatorname{Var}\lambda_j \le 16n^{5/4}$ is trivial), we get

$$\begin{aligned}\sum_{j=1}^d \operatorname{Var}\lambda_j &\le 16n^{5/4} + 10n^{3/4}\sqrt{n\log n}\\ &\le 30n^{5/4}\sqrt{\log n}.\end{aligned}$$

□

To conclude, note that if $p_1 = 1/2$, then [42] proved that $\operatorname{Var}\lambda_1 = p_1(1-p_1)n + \mathcal{O}(1) = n/4 + \mathcal{O}(n)$. So there is no universal upper bound on the variance (not depending on the distribution p) of order lower than n. In some particular cases, for example the uniform case with $n = m^2$, $(\sum_{j=1}^d \operatorname{Var}\lambda_j)/n$ may go to zero.

3.3 A generalization of a result on the excess of the RSKshape

When $p_1 = \cdots = p_d = 1/d$ (the uniform case), the asymptotic behavior of $\lambda_1(n), \ldots, \lambda_d(n)$ (where $\lambda(n) \sim \mathrm{SW}^n(p)$, or equivalently $\lambda(n)$ is the RSK shape of the word $X_1, \ldots, X_n$, with letters drawn according to p) is well known, it converges in distribution to the eigenvalues of the traceless GUE. We remind the well known result (see ,e.g., [54]) that these eigenvalues have a density in $\mathbb{R}^d$ proportional to

$$\mathrm{e}^{-\frac{1}{2}\sum_{i=1}^d x_i^2}\prod_{i<j}(x_i - x_j)^2,$$

the term $\mathrm{e}^{-\frac{1}{2}\sum_{i=1}^d x_i^2}$ being the density of independent Gaussian random variables, and the term $\prod_{i<j}(x_i - x_j)^2$ reveals the "eigenvalues repulsion". This last term explains the difference between the diagram $\lambda(n)$ and $N_1(n), \ldots, N_d(n)$ the numbers of occurrences of $1, 2, \ldots, d$ in the random word $X_1 \ldots X_n$, introduced in [56] and called the "excess": for $k_0 \in \{1, \ldots, d\}$, let

$$E_{k_0}^{(n)}(p) = \mathbb{E}\left(\sum_{k=1}^{k_0}\lambda_k(n) - \sum_{k=1}^{k_0} N_k(n)\right).$$

The goal of this section is to better understand the behavior of this "excess" in a non-uniform setting. In what follows, we assume $p_1 > \cdots > p_d$, and let $\lambda(n)$ be the RSK shape of the word $X_1, \ldots, X_n$, with letters drawn according to p (so $\lambda(n) \sim \mathrm{SW}^n(p)$). In [57], it is proven that $(\lambda(n))_{n\ge 1}$ is a Markov chain, but we are going to consider a slightly different one, that will have the advantages of being irreducible, aperiodic, positive recurrent. Let us define for $n \in \mathbb{N}$ and $1 \le k \le \ell \le d$, $m_{k,\ell}(n)$ the number of occurrences of the letter ℓ in the k-th row of $P(n)$. Clearly,

$$\sum_{1\le k\le \ell\le d} m_{k,\ell}(n) = n.$$

Moreover, for any $k_0 \in \{1, \dots, d\}$,

$$\sum_{k=1}^{k_0} \sum_{\ell=k}^{k_0} m_{k,\ell}(n) = \sum_{k=1}^{k_0} N_k(n)$$

where $N_k(n)$ is the number of occurrences of the letter k in $P(n)$ (or equivalently, in $X_1, \dots, X_n$), hence

$$\sum_{k=1}^{k_0} \lambda_k(n) = \sum_{k=1}^{k_0} N_k(n) + \sum_{1 \le k \le k_0 < \ell \le d} m_{k,\ell}(n),$$

hence

$$E_{k_0}^{(n)}(p) = \mathbb{E}\left(\sum_{1 \le k \le k_0 < \ell \le d} m_{k,\ell}(n) \right). \tag{3.3.1}$$

Theorem 1.13 in [56] states that

$$E_{k_0}^{(n)}(p) \xrightarrow[n\to\infty]{} \sum_{k \le k_0 < \ell \le d} \frac{p_\ell}{p_k - p_\ell}. \tag{3.3.2}$$

We will see that this theorem comes from a more general convergence result on the $m_{k,\ell}(n)$. From now on, we will simply write $m(n)$ for the vector $(m_{k,\ell}(n))_{1 \le k < \ell \le d}$. Note that we drop the diagonal terms $m_{k,k}$, as they can simply be written as $N_k - m_{1,k}(n) - \cdots - m_{k-1,k}(n)$, which also means there was no hope of convergence.

Theorem 3.3.1. *$(m(n))_{n \ge 0}$ is an irreducible, aperiodic, positive recurrent Markov chain with invariant measure*

$$\mu_p := \bigotimes_{1 \le k < \ell \le d} \textsc{Geo}\left(\frac{p_\ell}{p_k}\right).$$

Furthermore,

$$\mathbb{E}\left(m(n)\right) \xrightarrow[n\to\infty]{} \left(\frac{p_\ell}{p_k - p_\ell}\right)_{1 \le k < \ell \le d}$$

Corollary 3.3.2. *Using* (3.3.1)*, the previous result* (3.3.2) *follows immediately.*

Proof. Firstly, the fact that m is a Markov chain comes from the construction of the RSK algorithm: there is a deterministic function f_d such that $P(n+1) = f_d(P(n), X_{n+1})$, since X_{n+1} and $P(n)$ are independent (and the X_k's are i.i.d.), P is a Markov chain, and in particular, m is a Markov chain.

Let us check that it is irreducible. To do so, it suffices to show that there is a path leading from the state $0 = (0)_{1 \le k < \ell \le d}$ to any state $a \in \mathbb{N}^{\frac{n(n-1)}{2}}$, and reciprocally. Fix $a \in \mathbb{N}^{\frac{n(n-1)}{2}}$. We construct the letters $x_1, \dots, x_n$ that will produce the state a, which means more precisely that the resulting SSYD $P(n)$ from the RSK satisfies $(P(n)_{k,\ell})_{1 \le k < \ell \le d} = a$. Let us first see how to produce a state a with only one non-zero coordinate, say $a_{k,\ell} = 1$. We see that the sequence of letters $\ell, k-1, k-2, \dots, 1$ works. Now, we may apply the same treatment to the other letters, taking care of them in order: we start with $x_{d-1,d}$, and take the sequence $d, d-2, \dots, 1$ and repeat this block $x_{d-1,d}$ times. Then we proceed similarly for $x_{d-2,d-1}$, then $x_{d-2,d-1}$, and so on. We get a sequence of $n \le (d-1)\sum_{1 \le k < \ell \le d} a_{k,\ell}$ letters, which has the desired property. In order to go from the state a to the state 0, simply consider $M = \max_{1 \le k < \ell \le d} a_{k,\ell}$ and the sequence of letters of a block of M d's, followed by a block of $2M$ $d-1$'s, ... up to a block of dM 1's. It "clears out" the upper diagonal of P, in other terms, this brings the term 0. So the chain m is irreducible.

The aperiodicity comes from the fact that in the state 0, the letter 1 makes it stay in the state 0.

Let us now prove, by induction on $d \leq 2$, that μ_p is an invariant distribution (which implies the chain is positive recurrent).

Base case: When $d = 2$, m is one-dimensional, this is a well known birth-death Markov chain.

Inductive step: Assume the result holds at rank $d-1 \geq 2$, we prove it holds at rank d. Let $M \sim \mu_p$, we need to check that $f_d(M, X_1) \sim \mu_p$. Let $y \in \mathbb{N}^{\frac{n(n-1)}{2}}$, and let $k_1 < \cdots < k_c$ be the integers in $\{2, \ldots, d\}$ such that $y_{1,k_i} \neq 0$ (if there is any), let also $k_0 = 1, k_{c+1} = d+1$ and for $0 \leq i \leq c$, let $A_i = \{k_i + 1, \ldots, k_{i+1}\}$ (so $\cup_{i=0}^{c} A_i = \{2, \ldots, d\}$).

We have

$$\mathbb{P}\left(f_d(M, X_1) = y\right) = \sum_{k=1}^{d} \sum_{x : f_d(x,k) = y} \mu_p(x) p_k$$

Let $k \in \{2, \ldots, d\}$. If $k \notin \{k_0, \ldots, k_c\}$, there exists no x such that $f_d(x, k) = y$ (because then $y_{1,k}$ would be non zero, a contradiction with the definition of the k_i's). So let $i \in \{0, \ldots, c\}$, we see that x is such that $f_d(x, k_i) = y$ if and only if all one of the two conditions below is satisfied:

- (No letter is bumped to the second line) $i = c$, $x_{1,k_c} = y_{1,k_c} - 1$ and for any $\ell \neq k_c$, $x_{1,\ell} = y_{1,\ell}$.
- (One letter is bumped to the second line) If $i \neq 0$, then $x_{1,k_i} = y_{1,k_i} - 1$; there exists an integer $\ell \in A_i$, $x_{1,\ell} = y_{1,\ell} + 1$, and for any $\ell' \notin \{k_i, \ell\}$, $x_{1,\ell} = y_{1,\ell}$.

We will denote $(x_{i,j})_{2 \leq i < j \leq d}$ by x^{2+}. Let

$$p^{2+} = \left(\frac{p_2}{1-p_1}, \ldots, \frac{p_d}{1-p_1}\right).$$

We see that $m^{2+} \sim \mu_{p^{2+}}$, and therefore by the induction hypothesis,

$$\sum_{k=2}^{d} \sum_{x^{2+} : f_{d-1}(x^{2+}, k) = y^{2+}} \mu_{p^{2+}}(x^{2+}) p_k^{2+} = \mu_{p^{2+}}(y^{2+}),$$

and

$$\begin{aligned}
\mathbb{P}\left(f_d(M, X_1) = y\right) &= \sum_{k=1}^{d} \sum_{x : f_d(x,k) = y} \mu_p(x) p_k \\
&= \sum_{i=0}^{c} \sum_{x : f_d(x, k_i) = y} \mu_p(x) p_{k_i} \\
&= \sum_{i=0}^{c} \sum_{\ell \in A_i} \sum_{\substack{x : f_{d-1}(x^{2+}, \ell) = y \\ x_{1,\ell} = y_{1,\ell} + 1, i \neq 0 \Rightarrow x_{1,k_i} = y_{1,k_i} - 1 \\ \forall \ell' \notin \{k_i, \ell\}, x_{1,\ell} = y_{1,\ell}}} \mu_p(x) p_{k_i} + \mu_p(y - \delta_{(k,\ell) = (1,k_c)}) p_{k_c},
\end{aligned}$$

which we may rewrite, with the notation $\mu_p^1(x) := \prod_{\ell=2}^{d} \left(\frac{p_\ell}{p_1}\right)^{x_{1,\ell}} \left(1 - \frac{p_\ell}{p_1}\right)$ (the restriction of μ_p to the first line) as:

$$\begin{aligned}\mathbb{P}\left(f_d(M,X_1)=y\right) &= \sum_{i=0}^{c}\sum_{\ell\in A_i}\sum_{x^{2+}:f_{d-1}(x^{2+},\ell)=y}\mu_p^1(y)\frac{p_\ell}{p_1}\frac{p_1}{p_{k_i}}\mu_{p^{2+}}(x^{2+})p_{k_i}+\mu_p(y)\frac{p_1}{p_{k_c}}p_{k_c}\\ &= \mu_p^1(y)\sum_{\ell=2}^{d}\sum_{x^{2+}:f_{d-1}(x^{2+},\ell)=y}p_\ell\mu_{p^{2+}}(x^{2+})p_{k_i}+p_1\mu_p(y)\\ &= \mu_p^1(y)(1-p_1)\mu_{p^{2+}}(y^{2+})+p_1\mu_p(y)\\ &= \mu_p(y),\end{aligned}$$

which concludes the inductive step.

All there is left to prove is the convergence of the expectation, but this is a very general result: for any $K>0$,

$$\mathbb{E}\left(M_{k,\ell}\mathbb{1}_{M_{k,\ell}(n)>K}\right)=\mathbb{E}_{m(0)=M}\left(m_{k,\ell}\mathbb{1}_{m_{k,\ell}(n)>K}\right)=\sum_{x\in\mathbb{N}^{\frac{d(d-1)}{2}}}\mathbb{E}_{m(0)=x}\left(m_{k,\ell}(n)\mathbb{1}_{m_{k,\ell}(n)>K}\right)\mu_p(x)$$

so

$$\mathbb{E}_{m(0)=0}\left(m_{k,\ell}(n)\mathbb{1}_{m_{k,\ell}(n)>K}\right)\leq\frac{1}{\mu_p(0)}\mathbb{E}\left(M_{k,\ell}\mathbb{1}_{M_{k,\ell}(n)>K}\right),$$

which indicates that the Markov chain m starting at 0 is uniformly integrable. Since there is convergence in total variation, there is also convergence in L_1, and

$$\mathbb{E}\left(m_{k,\ell}(n)\right)\xrightarrow[n\to\infty]{}\mathbb{E}\left(M_{k,\ell}\right)=\frac{p_\ell}{p_k-p_\ell}.$$

□

Bibliography

[1] Kenneth S Alexander. The rate of convergence of the mean length of the longest common subsequence. *The Annals of Applied Probability*, pages 1074–1082, 1994.

[2] Robert Alicki, Sl/awomir Rudnicki, and Sl/awomir Sadowski. Symmetry properties of product states for the system of n n-level atoms. *Journal of mathematical physics*, 29(5):1158–1162, 1988.

[3] Hironobu Aoki, Ryuhei Uehara, and Koichi Yamazaki. Expected length of longest common subsequences of two biased random strings and its application. *RIMS Kôkyûroku*, 1185:1–10, 2001.

[4] Jinho Baik, Percy Deift, and Kurt Johansson. On the distribution of the length of the longest increasing subsequence of random permutations. *Journal of the American Mathematical Society*, 12(4):1119–1178, 1999.

[5] Yu Baryshnikov. Gues and queues. *Probability Theory and Related Fields*, 119:256–274, 2001.

[6] Florent Benaych-Georges and Christian Houdré. A note on GUE minors, maximal Brownian functionals and longest increasing subsequences. *Markov Processes and Related Fields*, 21, 12 2013.

[7] Philippe Biane. Approximate factorization and concentration for characters of symmetric groups. *International Mathematics Research Notices*, 2001(4):179–192, 2001.

[8] Charles Bordenave, Gábor Lugosi, and Nikita Zhivotovskiy. Noise sensitivity of the top eigenvector of a Wigner matrix. *Probability Theory and Related Fields*, pages 1–33, 2020.

[9] Stéphane Boucheron, Gábor Lugosi, and Pascal Massart. *Concentration inequalities: A nonasymptotic theory of independence*. Oxford university press, 2013.

[10] Olivier Bousquet and Christian Houdré. Iterated jackknives and two-sided variance inequalities. In *High Dimensional Probability VIII*, pages 33–40. Springer, 2019.

[11] Jean-Christophe Breton and Christian Houdré. Asymptotics for random Young diagrams when the word length and alphabet size simultaneously grow to infinity. *Bernoulli*, 16(2):471–492, 2010.

[12] Jean-Christophe Breton and Christian Houdré. On the limiting law of the length of the longest common and increasing subsequences in random words. *Stochastic Processes and their Applications*, 127(5):1676–1720, 2017.

[13] Wun-Tat Chan, Yong Zhang, Stanley PY Fung, Deshi Ye, and Hong Zhu. Efficient algorithms for finding a longest common increasing subsequence. In *Algorithms and Computation: 16th International Symposium, ISAAC 2005, Sanya, Hainan, China, December 19-21, 2005. Proceedings 16*, pages 665–674. Springer, 2005.

[14] Sourav Chatterjee. A new method of normal approximation. *The Annals of Probability*, 36(4):1584–1610, 2008.

[15] Václáv Chvátal and David Sankoff. Longest common subsequences of two random sequences. *Journal of Applied Probability*, pages 306–315, 1975.

[16] Dario Cordero-Erausquin and Alexandros Eskenazis. Talagrand's influence inequality revisited. *Analysis & PDE*, 16(2):571–612, 2023.

[17] Arthur L. Delcher, Simon Kasif, Robert D Fleischmann, Jeremy Peterson, Owen White, and Steven L. Salzberg. Alignment of whole genomes. *Nucleic acids research*, 27(11):2369–2376, 1999.

[18] Clément Deslandes and Christian Houdré. On the limiting law of the length of the longest common and increasing subsequences in random words with arbitrary distribution. *Electronic Journal of Probability*, 26:1–27, 2021.

[19] John D. Dixon. Longest common subsequences in binary sequences. *arXiv preprint arXiv:1307.2796*, 2013.

[20] Bradley Efron and Charles Stein. The jackknife estimate of variance. *The Annals of Statistics*, pages 586–596, 1981.

[21] William Fulton. *Young tableaux: with applications to representation theory and geometry.* Cambridge University Press, 1997.

[22] Christophe Garban and Jeffrey E Steif. *Noise sensitivity of Boolean functions and percolation*, volume 5. Cambridge University Press, 2014.

[23] Victor Goodman. Distribution estimates for functionals of the two-parameter Wiener process. *The Annals of Probability*, 4(6):977–982, 1976.

[24] Janko Gravner, Craig A Tracy, and Harold Widom. Limit theorems for height fluctuations in a class of discrete space and time growth models. *Journal of Statistical Physics*, 102:1085–1132, 2001.

[25] Curtis Greene. An extension of Schensted's theorem. *Advances in Mathematics*, 14(2):254–265, 1974.

[26] John M Hammersley. A few seedlings of research. In *Proc. Sixth Berkeley Symp. Math. Statist. and Probability*, volume 1, pages 345–394, 1972.

[27] Christian Houdré. The iterated jackknife estimate of variance. *Statistics & probability letters*, 35(2):197–201, 1997.

[28] Christian Houdré and Ümit Işlak. A central limit theorem for the length of the longest common subsequences in random words. *Electronic Journal of Probability*, 28:1–24, 2023.

[29] Christian Houdré and Abram Kagan. Variance inequalities for functions of gaussian variables. *Journal of Theoretical Probability*, 8(1):23–30, 1995.

[30] Christian Houdré and George Kerchev. On the rate of convergence for the length of the longest common subsequences in hidden Markov models. *Journal of Applied Probability*, 56(2):558–573, 2019.

[31] Christian Houdré, Jüri Lember, and Heinrich Matzinger. On the longest common increasing binary subsequence. *Comptes Rendus Mathematique*, 343(9):589–594, 2006.

[32] Christian Houdré and Trevis J. Litherland. On the longest increasing subsequence for finite and countable alphabets. In *High dimensional probability V: the Luminy volume*, volume 5, pages 185–213. Institute of Mathematical Statistics, 2009.

[33] Christian Houdré and Trevis J Litherland. Asymptotics for the length of the longest increasing subsequence of a binary markov random word. In *Malliavin Calculus and Stochastic Analysis: A Festschrift in Honor of David Nualart*, pages 511–524. Springer, 2013.

[34] Christian Houdré and Trevis J. Litherland. On the limiting shape of Young diagrams associated with Markov random words. *Markov Processes and Related Fields*, 26(5):779–838, 2020.

[35] Christian Houdré and Qingqing Liu. On the variance of the length of the longest common subsequences in random words with an omitted letter. *arXiv preprint arXiv:1812.09552*, 2018.

[36] Christian Houdré and Jinyong Ma. On the order of the central moments of the length of the longest common subsequences in random words. In *High dimensional probability VII*, pages 105–136. Springer, 2016.

[37] Christian Houdré and Heinrich Matzinger. Closeness to the diagonal for longest common subsequences in random words. *Electronic Communications in Probability*, 21, 2016.

[38] Christian Houdré and Victor Pérez-Abreu. Covariance identities and inequalities for functionals on Wiener and Poisson spaces. *The Annals of Probability*, pages 400–419, 1995.

[39] Christian Houdré, Victor Pérez-Abreu, and Donatas Surgailis. Interpolation, correlation identities, and inequalities for infinitely divisible variables. *Journal of Fourier Analysis and Applications*, 4(6):651–668, 1998.

[40] Christian Houdré and Zsolt Talata. On the rate of approximation in finite-alphabet longest increasing subsequence problems. *The Annals of Applied Probability*, 22(6):2539–2559, 2012.

[41] Christian Houdré and Hua Xu. On the limiting shape of Young diagrams associated with inhomogeneous random words. In *High Dimensional Probability VI: The Banff Volume*, pages 277–302. Springer, 2013.

[42] Alexander R Its, Craig A Tracy, and Harold Widom. Random words, Toeplitz determinants and integrable systems. I. In *Random matrix models and their applications*, pages 245–258. Cambridge University Press, 2001.

[43] Paata Ivanisvili, Ramon van Handel, and Alexander Volberg. Rademacher type and Enflo type coincide. *Annals of Mathematics*, 192(2):665–678, 2020.

[44] Kurt Johansson. Discrete orthogonal polynomial ensembles and the plancherel measure. *Annals of Mathematics*, pages 259–296, 2001.

[45] Sergei Vasilévich Kerov and NS Tsilevich. *Asymptotic representation theory of the symmetric group and its applications in analysis*, volume 219. American Mathematical Society Providence, RI, 2003.

[46] Marcos Kiwi, Martin Loebl, and Jiří Matoušek. Expected length of the longest common subsequence for large alphabets. *Advances in Mathematics*, 197(2):480–498, 2005.

[47] János Komlós, Péter Major, and Gábor Tusnády. An approximation of partial sums of independent rv's, and the sample df. I. *Zeitschrift für Wahrscheinlichkeitstheorie und verwandte Gebiete*, 32(1):111–131, 1975.

[48] Jüri Lember and Heinrich Matzinger. Standard deviation of the longest common subsequence. *The Annals of Probability*, 37(3):1192–1235, 2009.

[49] Jüri Lember, Heinrich Matzinger, and Felipe Torres. The rate of the convergence of the mean score in random sequence comparison. *The Annals of Applied Probability*, pages 1046–1058, 2012.

[50] Qingqing Liu and Christian Houdré. Simulations, computations, and statistics for longest common subsequences. *arXiv preprint arXiv:1705.06826*, 2017.

[51] Benjamin F Logan and Larry A Shepp. A variational problem for random young tableaux. *Advances in mathematics*, 26(2):206–222, 1977.

[52] George S. Lueker. Improved bounds on the average length of longest common subsequences. *Journal of the ACM (JACM)*, 56(3):1–38, 2009.

[53] Colin McDiarmid. On the method of bounded differences. *Surveys in combinatorics*, 141(1):148–188, 1989.

[54] Madan Lal Mehta. *Random matrices*. Academic Press, 1991.

[55] Pierre-Loïc Mèliot. Fluctuations of central measures on partitions. In *24th International Conference on Formal Power Series and Algebraic Combinatorics (FPSAC 2012)*, volume DMTCS Proceedings vol. AR, 24th International Conference on Formal Power Series and Algebraic Combinatorics (FPSAC 2012), pages 385–396. Discrete Mathematics and Theoretical Computer Science, 2012.

[56] Ryan O'Donnell and John Wright. Efficient quantum tomography ii. In *Proceedings of the 49th Annual ACM SIGACT Symposium on Theory of Computing*, pages 962–974, 2017.

[57] Neil O'Connell. A path-transformation for random walks and the robinson-schensted correspondence. *Transactions of the American Mathematical Society*, 355(9):3669–3697, 2003.

[58] WanSoo T Rhee. On rates of convergence for common subsequences and first passage time. *The Annals of Applied Probability*, pages 44–48, 1995.

[59] Dan Romik. *The surprising mathematics of longest increasing subsequences*. Number 4. Cambridge University Press, 2015.

[60] Yoshifumi Sakai. A linear space algorithm for computing a longest common increasing subsequence. *Information Processing Letters*, 99(5):203–207, 2006.

[61] David Sankoff and RJ Cedergren. A test for nucleotide sequence homology. *Journal of molecular biology*, 77(1):159–164, 1973.

[62] David Sankoff and Joseph B Kruskal. Time warps, string edits, and macromolecules: the theory and practice of sequence comparison. *Reading: Addison-Wesley Publication, 1983, edited by Sankoff, David; Kruskal, Joseph B.*, 1, 1983.

[63] Craige Schensted. Longest increasing and decreasing subsequences. *Canadian Journal of mathematics*, 13:179–191, 1961.

[64] Galen R Shorack and Jon A Wellner. *Empirical processes with applications to statistics*. SIAM, 2009.

[65] J Michael Steele. An Efron-Stein inequality for nonsymmetric statistics. *The Annals of Statistics*, 14(2):753–758, 1986.

[66] J Michael Steele. *Probability theory and combinatorial optimization*. SIAM, 1997.

[67] Michel Talagrand. On Russo's approximate zero-one law. *The Annals of Probability*, pages 1576–1587, 1994.

[68] Alexander Tiskin. The Chvátal-Sankoff problem: Understanding random string comparison through stochastic processes. *arXiv preprint arXiv:2212.01582*, 2022.

[69] Craig A Tracy and Harold Widom. Level spacing distributions and the Bessel kernel. *Communications in mathematical physics*, 161(2):289–309, 1994.

[70] Craig A Tracy and Harold Widom. On the distributions of the lengths of the longest monotone subsequences in random words. *Probability theory and related fields*, 119:350–380, 2001.

[71] Craig A Tracy and Harold Widom. The distributions of random matrix theory and their applications. In *New Trends in Mathematical Physics: Selected contributions of the XVth International Congress on Mathematical Physics*, pages 753–765. Springer, 2009.

[72] Stanislaw M Ulam. Monte Carlo calculations in problems of mathematical physics. *Modern Mathematics for the Engineers*, pages 261–281, 1961.

[73] Anatolii Moiseevich Veršik and Sergei V Kerov. Asymptotic behavior of the plancherel measure of the symmetric group and the limit form of young tableaux. In *Dokl. Akad. Nauk SSSR*, volume 233, pages 1024–1027, 1977.

[74] Michael S Waterman. *Introduction to computational biology: maps, sequences and genomes*. CRC Press, 1995.

[75] Paweł Wolff. Some remarks on functionals with the tensorization property. *Bulletin of the Polish Academy of Sciences. Mathematics*, 3(55):279–291, 2007.

[76] John Wright. *How to learn a quantum state*. PhD thesis, Carnegie Mellon University, 2016.

[77] Y. Zhang. *Topics on the length of the longest common subsequences with blocks in binary random words*. PhD thesis, Georgia Institute of Technology, 2019.

ECOLE
DOCTORALE
DE MATHEMATIQUES
HADAMARD

Titre : Plus longues sous-séquences de mots aléatoires: limites, variance, et statistiques quantiques.

Mots clés : Probabilités, statistiques quantiques, sous-séquences

Résumé : Nous considérons des problèmes de ”mots aléatoires”, et leurs applications. Le point de départ est le problème suivant : étant donné deux mots alétoires, ”combien ont-ils en commun” ? Bien qu'il soit fondamental en biologie, informatique, linguistique, il reste largement irrésolu. Nous commençons par l'étude des plus longs sous-mots communs croissants : cela signifie que l'on considère un alphabet ordonné, disons 1,...,m, et les sous-mots qui sont simplement faits d'un bloc de 1's, suivi d'un bloc de 2's, ... et ainsi de suite. Ensuite, nous nous intéressons au problème de la variance de la longueur des plus longs sous-mots communs. En introduisant des outils plus généraux, nous faisons des progrès sur la compréhension de cette variance, et nous revisitons des bornes supérieures et inférieures de la variance dans d'autres cadres. Enfin, nous considérons la longueur maximale d'un sous-mot croissant d'un seul mot aléatoire, et le lien étonnant avec les statistiques quantiques.

Title : Random subsequences problems: asymptotics, variance, and quantum statistics.

Keywords : Probability, quantum statistics, subsequences

Abstract : This work considers some random words combinatorial problems and their applications. The starting point of this endeavor is the following question : given two random words, ”how much do they have in common” ? Even if this question has emerged independently in various fields, including computer science, biology, linguistics, it remains mostly unsolved. Firstly, we study the asymptotic distribution of the length of the longest common and increasing subsequences. There we consider a totally ordered alphabet with an order, say 1,...,m, and the subsequences are simply made of a block of 1's, followed by a block of 2's, ... and so on (such a subsequence is increasing, but not strictly). Secondly, we deal with the problem of the variance of the LCS. By introducing a general framework going beyond this problem, partial results in this direction are presented, and various upper and lower variance bounds are revisited in diverse settings. Lastly, we consider the Longest Increasing Subsequences (LIS) of one random word, and the surprising connection with quantum statistics.

Institut Polytechnique de Paris
91120 Palaiseau, France