



HAL
open science

Graph inference by multiple testing with application to Neuroimaging

Marine Roux

► **To cite this version:**

Marine Roux. Graph inference by multiple testing with application to Neuroimaging. Signal and Image processing. Université Grenoble Alpes, 2018. English. NNT : 2018GREAT058 . tel-01971574

HAL Id: tel-01971574

<https://theses.hal.science/tel-01971574v1>

Submitted on 7 Jan 2019

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

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



THÈSE

pour obtenir le grade de

DOCTEUR DE LA COMMUNAUTÉ UNIVERSITÉ GRENOBLE ALPES

Spécialité : **Signal, Image, Parole, Télécoms (SIPT)**

Arrêté ministériel : 25 mai 2016

Présentée par

Marine ROUX

Thèse dirigée par **Sophie ACHARD**, CNRS

codirigée par **Pierre BORGNAT**, CNRS

préparée au sein du **Laboratoire Grenoble Images Parole Signal
Automatique**

dans l'**École Doctorale Electronique, Electrotechnique, Au-
tomatique, Traitement du Signal (EEATS)**

Inférence de graphes par une procé- dure de test multiple avec applica- tion en Neuroimagerie

Graph inference by multiple testing with ap- plication to Neuroimaging

Thèse soutenue publiquement le **24 septembre 2018**,
devant le jury composé de:

Madame SOPHIE ACHARD

DIRECTEUR DE RECHERCHE, CNRS, Directeur de thèse

Madame MAGALIE FROMONT

PROFESSEUR, UNIVERSITE RENNES 2, Rapporteur

Monsieur SAID MOUSSAOUI

PROFESSEUR, ECOLE CENTRALE NANTES, Rapporteur

Monsieur PIERRE BORGNAT

DIRECTEUR DE RECHERCHE, CNRS, Co-directeur de thèse

Madame IRENE GANNAZ

MAITRE DE CONFERENCES, INSA LYON, Examineur, Co-directrice
de thèse

Monsieur ETIENNE ROQUAIN

MAITRE DE CONFERENCES, UNIVERSITE PIERRE ET MARIE CURIE,
Co-directeur de thèse

Madame FLORENCE FORBES

DIRECTRICE DE RECHERCHE, INRIA CENTRE DE GRENOBLE
RHÔNE-ALPES, Présidente

Monsieur PIERRE NEUVIAL

CHARGE DE RECHERCHE, CNRS, Examineur

A toutes les petites filles du monde,
à celle que mon coeur aime déjà, Lena.

Résumé. Cette thèse est motivée par l'analyse des données issues de l'imagerie par résonance magnétique fonctionnelle (IRMf). La nécessité de développer des méthodes capables d'extraire la structure sous-jacente des données d'IRMf constitue un challenge mathématique attractif. A cet égard, nous modélisons les réseaux de connectivité cérébrale par un graphe et nous étudions des procédures permettant d'inférer ce graphe.

Plus précisément, nous nous intéressons à l'inférence de la structure d'un modèle graphique non orienté par une procédure de test multiple. Nous considérons deux types de structure, à savoir celle induite par la corrélation et celle induite par la corrélation partielle entre les variables aléatoires. Les statistiques de tests basées sur ces deux dernières mesures sont connues pour présenter une forte dépendance et nous les supposons être asymptotiquement gaussiennes. Dans ce contexte, nous analysons plusieurs procédures de test multiple permettant un contrôle des arêtes incluses à tort dans le graphe inféré.

Dans un premier temps, nous questionnons théoriquement le contrôle du *False Discovery Rate* (FDR) de la procédure de Benjamini et Hochberg dans un cadre gaussien pour des statistiques de test non nécessairement positivement dépendantes. Nous interrogeons par suite le contrôle du FDR et du *Family Wise Error Rate* (FWER) dans un cadre gaussien asymptotique. Nous présentons plusieurs procédures de test multiple, adaptées aux tests de corrélations (resp. corrélations partielles), qui contrôlent asymptotiquement le FWER. Nous proposons de plus quelques pistes théoriques relatives au contrôle asymptotique du FDR.

Dans un second temps, nous illustrons les propriétés des procédures contrôlant asymptotiquement le FWER à travers une étude sur simulation pour des tests basés sur la corrélation. Nous concluons finalement par l'extraction de réseaux de connectivité cérébrale sur données réelles.

Abstract. This thesis is motivated by the analysis of the functional magnetic resonance imaging (fMRI). The need for methods to build such structures from fMRI data gives rise to exciting new challenges for mathematics. In this regards, the brain connectivity networks are modeled by a graph and we study some procedures that allow us to infer this graph.

More precisely, we investigate the problem of the inference of the structure of an undirected graphical model by a multiple testing procedure. The structure induced by both the correlation and the partial correlation are considered. The statistical tests based on the latter are known to be highly dependent and we assume that they have an asymptotic Gaussian distribution. Within this framework, we study some multiple testing procedures that allow a control of false edges included in the inferred graph.

First, we theoretically examine the False Discovery Rate (FDR) control of Benjamini and Hochberg's procedure in Gaussian setting for non necessary positive dependent statistical tests. Then, we explore both the FDR and the Family Wise Error Rate (FWER) control

in asymptotic Gaussian setting. We present some multiple testing procedures, well-suited for correlation (resp. partial correlation) tests, which provide an asymptotic control of the FWER. Furthermore, some first theoretical results regarding asymptotic FDR control are established.

Second, the properties of the multiple testing procedures that asymptotically control the FWER are illustrated on a simulation study, for statistical tests based on correlation. We finally conclude with the extraction of cerebral connectivity networks on real data set.

Key words: multiple testing, FWER control, FDR control, Gaussian setting, Benjamini-Hochberg procedure, neuroscience, fMRI data analysis, brain connectivity networks, graph inference, correlation test.



Remerciements

Mes premiers remerciements s'adressent à toutes ces personnes, qui depuis si longtemps, construisent les mathématiques. Je reste naïvement émerveillée par l'incroyable efficacité de ce langage.

La solidarité et le partage sont incontestablement deux valeurs nécessaires à l'évolution des mathématiques. I'm very grateful to Yoav Benjamini for sharing the unpublished work of Cohen with us. Je souhaite également remercier Fanny Villers pour la transmission de son code informatique relatif à l'article [81].

Tout aussi précieux à toute activité de recherche, l'échange est. Je remercie Magalie Fromont et Saïd Moussaoui d'avoir accepté de rapporter mon travail. Je remercie Pierre Neuvial d'avoir répondu présent à mi-parcours et d'accepter de revenir pour la conclusion de cette thèse. Je remercie également Florence Forbes de se joindre à mon jury.

Je remercie Sophie, Pierre, Irène et Etienne pour leurs encadrements et leurs collaborations. Merci pour vos relectures respectives du manuscrit. Sophie, merci pour tes éclaircissements neuroscientifiques. Pierre, merci pour le regard un peu plus extérieur mais non moins pertinent sur la partie test multiple. Irène, merci pour ton enthousiasme mathématique, ta patience et ton optimisme à toute épreuve. Etienne, je ne pouvais pas espérer meilleur tuteur pour mettre les mains dans cette passionnante théorie des tests multiples, merci.

Ce manuscrit, c'est un sprint à la fin d'un marathon. Et il y a les cris des amis en bord de route pour vous faire avancer.

Patricia, tes qualités d'enseignante, qui n'ont d'égales que tes qualités humaines, sont à l'origine de mon attrait pour les statistiques. Merci de m'avoir initié aux tests multiples. Cette thèse n'existerait pas sans toi. Merci pour tout ce que tu as fait et continues de faire pour moi.

Moeanui, ce sentiment d'achèvement impose sa retrospective : ça fait 10 ans qu'on s'est croisées... Merci pour ton incroyable soutien à la hauteur de ton humanité rayonnante, venue de ton île lointaine. Paraita, Delphine te salue cordialement !

Nathalie, Christophe, de la théorie de Grothendieck au karaoké revisité, il n'y a qu'un pas que nous avons toujours franchi en riant. Merci pour votre légèreté. Merci de toujours croire en moi.

Anne-Laure, au-delà de toute statistique, tu m'as permis de croire en cette thèse. Merci pour tes précieux conseils, ton soutien et ces moments de poésie partagés.

Guillaume, de la Voie Romaine à bout de souffle à l'aisance respiratoire, il y a cet oxygène que tu m'apportes. Merci. Sans en oublier tes corrections d'anglais, merci pour ça aussi.

Adeline, ta générosité et ta profonde ouverture sur le monde me permettent de dépasser les limites du mien. Merci pour cette énergie. Merci pour tous ces défis, à la raison d'un samedi soir, relevés à mes côtés.

Et il y a ces personnes, qui d'une façon particulière, vous permettent de vous réaliser. Merci à A.Bescond, Buridane et I.Dufossez.

Et il y a ces personnes qui font notre identité. Gu, merci pour tout. Tout autant de pensées pour Danièle. Merci maman. Merci Hélène pour ton incroyable intelligence de vie.

Chères lectrices, chers lecteurs, bon voyage dans un monde où le hasard s'espère, se quantifie.

Marine.
1er Juillet 2018.

Contents

1	Introduction (français)	13
1.1	La problématique des tests multiples	14
1.2	Critères de décision	18
1.3	Quelques domaines d'applications des tests multiples	21
1.4	Reconstruction des réseaux de connectivité cérébrale	25
1	Introduction (english)	31
1.1	The multiple testing problem	31
1.2	Criteria and decisions	35
1.3	Multiple testing : some contemporary applications	38
1.4	Estimation of cerebral connectivity	42
2	Graphical models	47
2.1	Introduction	47
2.2	Well-posed problems	49
2.2.1	Preliminaries	49
2.2.2	Concentration graph	50
2.2.3	Covariance graph	53
2.3	How to infer the concentration or covariance graph?	54
2.3.1	A brief review	54
2.3.2	Multiple testing approaches	55
3	Multiple testing framework	63
3.1	Setting	64
3.1.1	Model	64
3.1.2	Criteria and decisions	65
3.2	Dependence assumptions	66
3.2.1	MTP_2 : multivariate total positivity of order 2	66
3.2.2	PRD: positive regression dependence	67
3.2.3	PA: positive association	68
3.2.4	PLOD: positive lower orthant dependence	69

CONTENTS

3.3	Classical FWER control	71
3.4	Classical FDR control	72
3.4.1	The BH procedure	73
3.4.2	Bounds of FDR of BH procedure: known results	73
3.5	A new result: relaxation of the PRDS condition	79
4	On BH procedure in the two-sided Gaussian setting	83
4.1	Statistical setting	83
4.2	Background	84
4.3	Known results and new formalized proofs	85
4.3.1	Result under the full null	85
4.3.2	Results for perfect correlation	86
4.4	New results	89
4.4.1	On sharpness of (CohenBound)	89
4.4.2	Extending scope of (ReinerBound)	91
4.4.3	Numerical study of (ReinerBound) with exact calculation	92
4.5	Summary and conjectures	95
5	Asymptotic results	97
5.1	Asymptotic Gaussian setting	97
5.2	Asymptotic control of the FWER	99
5.2.1	Subset pivotality fails	99
5.2.2	Single step methods	100
5.2.3	Step-down methods	106
5.2.4	Application to correlation tests	109
5.3	Towards an asymptotic control of the FDR	110
5.3.1	On BH procedure in the one-sided asymptotic Gaussian setting	110
5.3.2	Correlation tests	113
6	Application to neuroscientific framework	115
6.1	Simulation study	115
6.1.1	Statistical setting	115
6.1.2	FWER control	117
6.1.3	Power with fixed alternative	120
6.1.4	Power study for the chain graph	123
6.2	Application to estimation of cerebral connectivity	126
6.2.1	Description of the data and methodological context	126
6.2.2	Results	127
6.2.3	Discussion on power	130
	Conclusion and perspectives	133

Bibliography	135
A Supplementary results	145
A.1 Proofs of Chapter 3	145
A.1.1 Proof of Theorem 3.4	145
A.1.2 Proof of Theorem 3.3	148
A.1.3 Proof of Theorem 3.7	150

Notation

m	the number of null hypotheses to be tested.
n	the sample size.
X	the observed random variable.
$\mathbb{X}^{(n)}$	a n -sample of X .
$\mathbb{E}[Y], \text{Var}(Y)$	the expectation and the variance of a real random variable Y .
\mathcal{H}	the index set of pairwise correlation coefficients of X .
$H_{0,i}$ (resp. $H_{1,i}$), $1 \leq i \leq m$	the null (resp. alternative) hypothesis to be tested.
$\mathcal{H}_0(P)$ (resp. $\mathcal{H}_1(P)$)	the index set of the true (resp. false) null hypotheses for the distribution P .
$m_0(P)$ (resp. $m_1(P)$)	the number of true (resp. false) null hypotheses for the distribution P .
$(p_i(X))_{1 \leq i \leq m}$	a family of p -values based on the distribution of X .
$(p_{n,i}(\mathbb{X}^{(n)}))_{1 \leq i \leq m}$	a family of p -values based on the asymptotic distribution of $\mathbb{X}^{(n)}$.
$(\tau_k)_{1 \leq k \leq m}$	a sequence of critical values.
$R \subset \{1, \dots, m\}$	a multiple testing procedure.
μ (resp. Σ)	the mean (resp. covariance matrix) of the observed random variable X , when $X \in \mathbb{R}^m$ is m -multivariate Gaussian.
Γ	the covariance matrix of the observed random variable X , when $X \in \mathbb{R}^d$ is d -multivariate Gaussian.
I_m	the identity matrix of size m .
$\ \cdot\ _\infty$	the infinity norm defined for all $y \in \mathbb{R}^m$ by $\sup_{1 \leq i \leq m} y_i $.
$\mathcal{N}_m(\mu, \Sigma)$	the m -multivariate Gaussian distribution with mean μ and covariance matrix Σ . When $m = 1$, m is dropped in the notation.
$\Phi(\cdot)$	the cumulative distribution function of a standard Gaussian distribution, e.g., $\Phi(z) = \mathbb{P}(Z \leq z)$, $Z \sim \mathcal{N}(0, 1)$.
$\bar{\Phi}(\cdot)$	the upper-tail function of a standard Gaussian distribution, i.e., $\bar{\Phi}(\cdot) = 1 - \Phi(\cdot)$.
$\mathcal{U}([a, b])$	the uniform distribution on $[a, b]$.
$\xrightarrow{\mathbb{P}}, \xrightarrow{\mathcal{L}}$	the convergence in probability and in distribution.

Chapitre 1

Introduction (français)

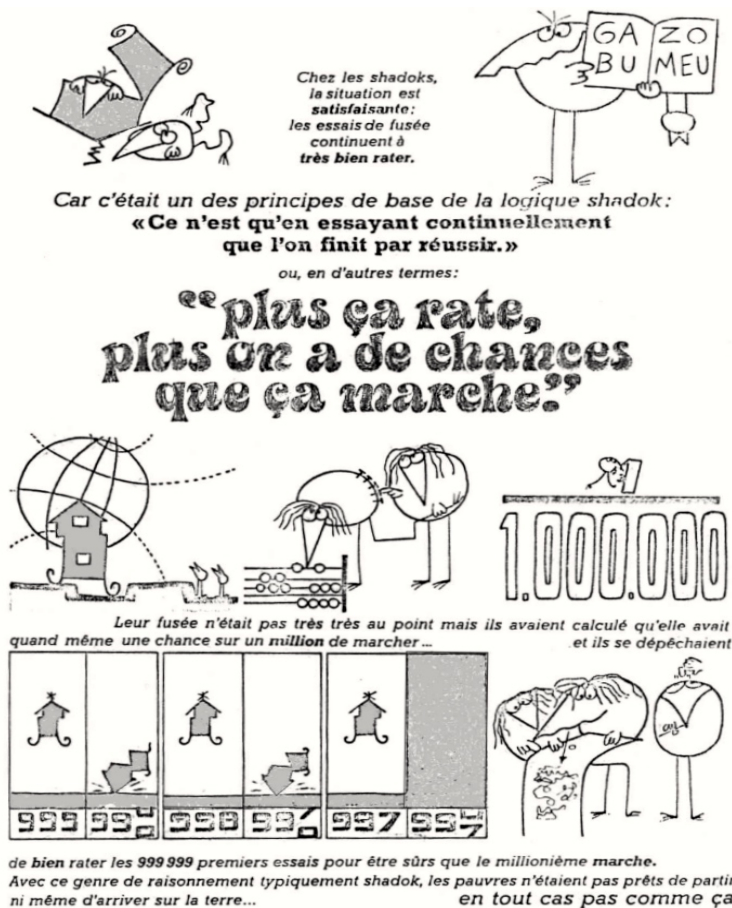


FIGURE 1.1 – Illustration artistique de la problématique des tests multiples par une devise Shadoks.

Cette introduction est volontairement informelle et propose un éclairage le plus intuitif possible sur la problématique des tests multiples. Nous situons également la place de cette problématique dans les enjeux contemporains. Nous détaillons plus particulièrement celui qui motive cette thèse, à savoir l'analyse des données issues de l'imagerie par résonance magnétique fonctionnelle (IRMf).

Sommaire

1.1 La problématique des tests multiples	14
1.2 Critères de décision	18
1.3 Quelques domaines d'applications des tests multiples	21
1.4 Reconstruction des réseaux de connectivité cérébrale	25

1.1 La problématique des tests multiples

Test simple

Informellement, nous appelons test statistique une démarche de la statistique inférentielle consistant à considérer comme vraie a priori une hypothèse dite "nulle", généralement notée H_0 et à confronter cette supposition à la réalité expérimentale (les observations dont nous disposons en pratique). En cas d'incompatibilité nous admettons une hypothèse différente dite "alternative", généralement notée H_1 . Deux types d'erreur peuvent survenir : rejeter H_0 alors qu'elle est vraie (erreur de type I), ou accepter H_0 alors qu'elle est fautive (erreur de type II). Traditionnellement, l'erreur de type I est considérée comme la plus problématique et un test statistique est construit de sorte que la probabilité de l'erreur de type I soit au plus $\alpha \in]0, 1[$, appelé le niveau du test. Autrement dit, lorsque les observations dont nous disposons ne sont pas suffisamment informatives, un test de niveau α acceptera H_0 avec une probabilité supérieure à $1 - \alpha$. En ce sens l'hypothèse H_0 est privilégiée et un test n'est informatif que lorsqu'il rejette.

Nous illustrons cette dissymétrie par l'exemple classique d'un n -échantillon de loi $\mathcal{N}(\mu, 1)$ où $\mu \in \mathbb{R}$ inconnu est le paramètre d'intérêt. Considérons un premier test, $H_0^1 : \mu < 0$ contre $H_1^1 : \mu \geq 0$. Alors le test qui rejette H_0^1 lorsque $n^{1/2}\bar{X}_n = n^{-1/2} \sum_{i=1}^n X_i$ est supérieur à $\bar{\Phi}^{-1}(\alpha)$ est un test de niveau α , où $\bar{\Phi}(\cdot) = 1 - \Phi(\cdot)$ et $\Phi(\cdot)$ est la fonction de répartition de la loi normale centrée réduite. Inversons l'hypothèse nulle et l'alternative *i.e.* considérons le second test, $H_0^2 : \mu \geq 0$ contre $H_1^2 : \mu < 0$. Alors le test qui rejette H_0^2 lorsque $n^{1/2}\bar{X}_n$ est inférieur à $-\bar{\Phi}^{-1}(\alpha)$ est un test de niveau α . La Figure 1.2 résume les conclusions des deux tests précédents. Dans la zone grisée, les deux hypothèses " $\mu \geq 0$ " et " $\mu < 0$ " peuvent être envisagées selon le choix de l'hypothèse nulle.

1.1. LA PROBLÉMATIQUE DES TESTS MULTIPLES

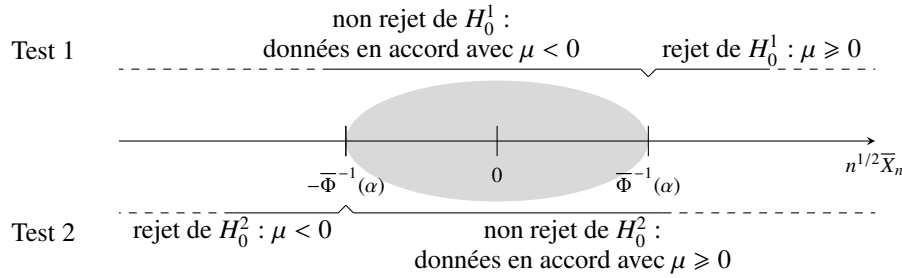


FIGURE 1.2 – Dissymétrie d’un test : cas d’un test de moyenne pour des variables gaussiennes *i.i.d.*

Un premier exemple de test multiple

Lorsque nous testons plusieurs hypothèses nulles simultanément, nous réalisons un test multiple. Une procédure de test multiple peut être vue comme un procédé permettant de prendre une décision pour chaque hypothèse nulle : la rejeter ou ne pas la rejeter. Bien que la démarche soit similaire pour chaque test individuellement, les spécificités dues à la pluralité des tests diffèrent du test simple. En particulier, lorsque les hypothèses nulles forment une partition de l’espace d’état du paramètre d’intérêt, nous ne pouvons influencer la conclusion du test. Autrement dit, le test multiple permet dans ce cas de ne pas privilégier les hypothèses nulles, la caractéristique dissymétrie d’un test d’hypothèses simple n’est pas retrouvée. En considérant un n -échantillon de loi $\mathcal{N}(\mu, 1)$ où $\mu \in \mathbb{R}$ inconnu est le paramètre d’intérêt, nous illustrons ce fait par le test multiple :

$$\left\{ \begin{array}{l} H_0 : “\mu = 0” \\ H_{0^+} : “\mu > 0” \\ H_{0^-} : “\mu < 0” \end{array} \right. . \quad (1.1)$$

Ce test est un cas particulier du tri-test réalisé dans [1] et peut être vu comme une “formulation test multiple” des tests de la section précédente. Le test qui rejette H_0 lorsque $n^{1/2}\bar{X}_n$ est supérieur à $\bar{\Phi}^{-1}(\alpha/2)$ ou inférieur à $-\bar{\Phi}^{-1}(\alpha/2)$ est de niveau α . Le test qui rejette H_{0^+} lorsque $n^{1/2}\bar{X}_n$ est inférieur à $-\bar{\Phi}^{-1}(\alpha)$ est de niveau α . De même le test qui rejette H_{0^-} lorsque $n^{1/2}\bar{X}_n$ est supérieur à $\bar{\Phi}^{-1}(\alpha)$ est de niveau α . Comme les trois hypothèses nulles forment une partition de \mathbb{R} (espace d’état de μ), seule une hypothèse nulle peut être rejetée à tort et dans ce cas, le “contrôle global” de l’erreur de type I est simplement assuré par le contrôle individuel de l’erreur de type I des trois tests. Autrement dit, réaliser ces trois tests simultanément au niveau α garantit un contrôle global de l’erreur de type I au niveau α . De la même façon que pour les tests simples, la Figure 1.3 schématise la conclusion du test (1.1).

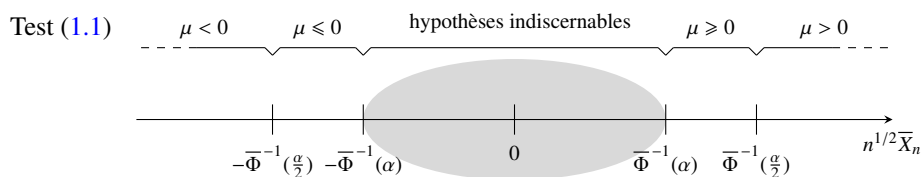


FIGURE 1.3 – Tri-test de moyenne pour des variables gaussiennes *i.i.d.*

Sur la Figure 1.3, la zone grisée représente une zone d’indécision, aucun rejet ne peut être fait. Les trois hypothèses nulles formant une partition de \mathbb{R} , la seule conclusion possible du test (dans la zone grisée) est que les données ne permettent pas de privilégier une hypothèse nulle par rapport à une autre.

Concrètement le test (1.1) répond typiquement à une problématique issue de l’industrie pharmaceutique. Dans [1] il s’agissait de comparer deux traitements pour le contrôle du diabète de type 2. Par exemple, en considérant que les données représentent la différence entre le taux de sucre dans le sang sous l’insuline B et le taux de sucre dans le sang sous l’insuline A, le tri-test se reformule comme :

$$\begin{cases} H_0 : \text{“Les deux insulines ont le même effet”} \\ H_{0^+} : \text{“L’insuline A est plus efficace que l’insuline B”} \\ H_{0^-} : \text{“L’insuline A est moins efficace que l’insuline B”} \end{cases} . \quad (1.2)$$

Le test multiple permet dans ce cas, d’éviter l’influence de l’industrie pharmaceutique sur les conclusions d’une étude.

Jusqu’alors nous avons interprété les tests statistiques en se basant sur une zone de rejet déterminée en fonction du niveau souhaité pour le test. Si nous changeons le niveau, il faut recalculer cette zone de rejet, peu commode en pratique. De plus, le choix du niveau du test est totalement arbitraire, alors pourquoi choisir $\alpha = 5\%$ et pas $\alpha = 4\%$ ou $\alpha = 6\%$? Un moyen alternatif possible pour interpréter un test est d’utiliser la notion de p -valeurs. Il s’agit du plus grand α pour lequel un test rejette l’hypothèse nulle au niveau α . Ainsi, une p -valeur donne par essence la conclusion d’un test à tous les niveaux possibles. Une p -valeur est une normalisation de la statistique de test, une variable aléatoire suivant une loi uniforme sous l’hypothèse nulle. Intuitivement, elle indique la “probabilité” d’observer nos réalisations si l’hypothèse nulle est vraie. Une “petite” p -valeur fournira la preuve que l’hypothèse nulle est fautive alors qu’une “grande” p -valeur indiquera qu’il n’y a pas de contradiction entre nos réalisations et l’hypothèse nulle.

Le problème de multiplicité

Lorsque nous testons plusieurs hypothèses nulles simultanément et que nous rejetons une hypothèse nulle alors qu'elle est vraie, nous réalisons un faux positif (ou encore appelé fausse découverte). Si les hypothèses nulles ne forment plus une partition de l'espace d'état du paramètre d'intérêt, le problème de multiplicité va se poser. Contrôler le niveau de chaque test individuellement ne va plus suffire à garantir un contrôle de l'erreur globale. En effet, si nous testons indépendamment toutes les hypothèses (potentiellement très nombreuses) au même niveau (et sans correction multiple), le cumul d'erreurs va entraîner une perte du contrôle global et la conclusion sera erronée. Par exemple, supposons tester (indépendamment) m vraies hypothèses nulles simultanément au niveau 5%. Quelle est la probabilité d'obtenir au moins un faux positif uniquement dû au hasard ? Nous avons,

$$\begin{aligned}\mathbb{P}(\text{“obtenir au moins un faux positif”}) &= 1 - \mathbb{P}(\text{“obtenir aucun faux positifs”}) \\ &= 1 - (1 - 0.05)^m,\end{aligned}$$

où la dernière égalité est obtenue par l'hypothèse d'indépendance.

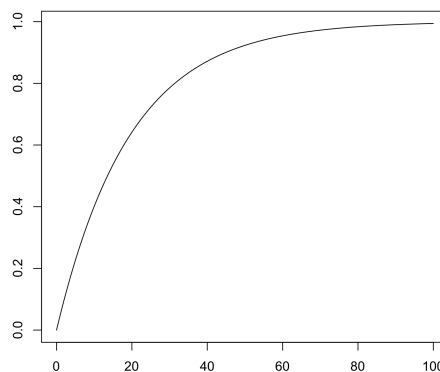


FIGURE 1.4 – Probabilité d'obtenir au moins un faux positif uniquement par hasard en fonction du nombre d'hypothèses nulles ($m \mapsto 1 - (1 - 0.05)^m$).

La figure 1.4 illustre un principe propre à la répétition des tests “plus nous cherchons, plus nous gagnons” (principe exploité par les Shadocks pour faire décoller une fusée en dépit d'en maîtriser sa construction, voir Figure 1.1). Même si toutes les hypothèses nulles testées étaient vraies, plus leur nombre augmente, plus nous avons de chances de faire au moins une (fausse) découverte. Cette problématique se retrouve en neuroscience. Plus concrètement, la Figure 1.5 illustre parfaitement le genre de conclusion absurde que nous pouvons obtenir en ignorant le problème de multiplicité dans ce domaine. Comme montré dans [2], en explorant suffisamment de zones du cerveau, il est possible de détecter une activité cérébrale chez un saumon mort.

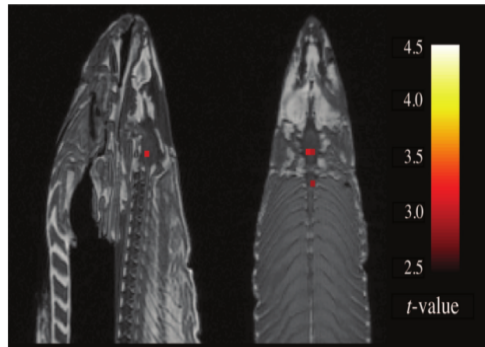


FIGURE 1.5 – Problème de multiplicité : détection d’activité cérébrale chez un saumon mort [2].

1.2 Critères de décision

Comme nous venons de le voir, le problème de multiplicité est l’enjeu majeur d’une procédure de test multiple. Comment s’assurer que les hypothèses nulles rejetées par une procédure de test multiple ne le soient pas uniquement par hasard ? Comment se mesure la qualité d’une procédure de test multiple ? En pratique, le mathématicien apporte une réponse probabiliste à ces problématiques via deux principaux (au sens rencontrés dans la littérature) critères : le *Family-Wise Error Rate* (FWER) et le *False Discovery Rate* (FDR). De la même façon qu’un test simple est construit de sorte que l’erreur de type I soit contrôlée à un niveau choisi, un test multiple sera généralement construit de sorte que l’un des deux critères (FWER ou FDR) soit contrôlé à un niveau choisi. Ainsi les rejets faits par une procédure de test multiple dépendent totalement du critère que l’on souhaite contrôler. Nous introduisons désormais ces deux critères plus précisément.

FWER

L’erreur de type I s’étend naturellement au cadre des tests multiples via le FWER. Pour une procédure de test multiple donnée, le FWER se définit comme la probabilité de rejeter au moins une hypothèse nulle à tort ou autrement dit d’obtenir au moins un faux positif à l’issue de la procédure. Lorsque nous testons plusieurs hypothèses nulles simultanément, nous avons

$$\text{FWER} = \mathbb{P}(\text{“il existe au moins une hypothèse nulle rejetée à tort”}).$$

Comme nous l’avons vu, si les hypothèses nulles ne forment pas une partition de l’espace d’état du paramètre d’intérêt, pour une procédure de test donnée, contrôler chaque test individuellement au niveau α ne sera pas suffisant pour assurer le contrôle du FWER au

même niveau α . Pour obtenir ce contrôle, il est alors nécessaire de modifier le niveau de chaque test. Cette opération se retrouve dans la littérature sous le terme “seuils critiques corrigés/ajustés” ou encore de façon équivalente “ p -valeurs corrigées/ajustées”. Historiquement, la correction de Bonferroni [3] est la première méthode proposant de corriger le niveau de chaque test afin de contrôler le FWER. Il s’agit d’ajuster le niveau α de chaque test, en fonction du nombre d’hypothèses testées par α/m . Certes très intuitive, cette méthode est assez conservative lorsque le nombre d’hypothèses fausses est grand devant le nombre d’hypothèses testées. En bref, c’est-à-dire que par construction, la méthode de Bonferroni ne permettra pas de rejeter “beaucoup” d’hypothèses nulles. Cette méthode devient par ailleurs très conservative en cas de dépendance entre les tests. Depuis, de nombreuses procédures moins conservatives ont été développées comme celle de Holm [4]. En 2010, Goeman et Solari [5] ont proposé un cadre général permettant de décrire la majorité de ces procédures via le *sequential rejection principle of familywise error control*. L’idée est de réajuster à chaque étape le nombre d’hypothèses à tester en fonction du nombre d’hypothèses déjà rejetées. Ainsi en augmentant le niveau de chaque test, le rejet d’hypothèses facilite le rejet des hypothèses restantes donc les procédures sont moins conservatives.

FDR

Malgré cette émergence de procédures moins conservatives, un critère moins restrictif que le FWER, plus adapté à certaines problématiques contemporaines, fut introduit en 1995 par Benjamini et Hochberg (BH) [6]. Il s’agit du *False Discovery Rate*, abrégé FDR. Ce critère se définit comme la moyenne de la proportion d’erreurs parmi les hypothèses rejetées, abrégée par FDP pour *False Discovery Proportion*,

$$\text{FDR} = \mathbb{E}[\text{FDP}], \quad \text{où } \text{FDP} = \frac{\text{nombre de faux positifs}}{\text{nombre de rejets}}.$$

Le prix à payer concernant l’étude du FDR est la difficulté mathématique supplémentaire due au caractère aléatoire du dénominateur du FDP. En pratique cette difficulté porte sur le type de dépendance liant les tests. Le Chapitre 3 présente en particulier quelques types de dépendance en lien avec ce manuscrit et le Chapitre 4 est consacré à l’étude du comportement du FDR de la procédure de Benjamini-Hochberg pour différentes structures de dépendance.

Nous noterons par ailleurs que même si le contrôle du FDR pour la procédure BH n’est établi que pour un certain type de dépendance (voir Chapitre 4), cette procédure est devenue la norme en pratique et constitue un des plus grands succès des statistiques théoriques en application de ces dernières années, comme en témoignent les actuelles 45881 citations de l’article introduisant la procédure de Benjamini et Hochberg [6].

Nous avons donc présenté deux critères permettant de quantifier la qualité d’une procédure de test multiple. Il s’en suit naturellement la question du choix du critère à contrôler.

CHAPITRE 1. INTRODUCTION (FRANÇAIS)

La section suivante répond à cette question en soulignant les différences entre le FWER et le FDR.

FWER versus FDR

D'une manière générale, un contrôle du FWER n'est pas plus souhaitable qu'un contrôle du FDR et inversement. Il s'agit essentiellement de choisir le critère le plus pertinent pour une problématique donnée.

Le contrôle du FWER assure avec une grande probabilité que l'ensemble des hypothèses rejetées ne contienne aucune erreur alors que le FDR évalue la proportion d'hypothèses nulles rejetées à tort parmi l'ensemble des hypothèses rejetées. Il est en ce sens moins restrictif, il s'adapte au nombre de rejets. Dans [7], Roquain propose une représentation visuelle de cette fondamentale différence entre FWER et FDR. Nous reportons cette illustration en la Figure 1.6. En bref, plusieurs hypothèses nulles sont testées simultanément soit en utilisant la méthode de Bonferroni (pour le contrôle du FWER) soit en utilisant la procédure de Benjamini et Hochberg (pour le contrôle du FDR) soit sans utiliser de correction multiple. Chaque rejet est alors matérialisé par un point noir sur la Figure 1.6 de sorte qu'une fausse découverte se situera à l'extérieur du cercle gris et inversement, un rejet correct se situera à l'intérieur du cercle gris. Le FWER assure avec grande probabilité qu'aucun point noir ne soit situé en dehors du cercle gris alors que le FDR assure que le nombre de point noir en dehors du cercle par rapport au nombre total de points noirs soit en moyenne plus petit qu'une certaine quantité (le niveau souhaité du test).

En pratique, les procédures contrôlant le FDR sont plus puissantes que celles contrôlant le FWER lorsque le nombre de fausses hypothèses nulles est grand et ce d'autant plus que l'intensité du signal est forte (illustré par la Figure 1.6, en bas). La contrepartie de ce gain de puissance est une augmentation du nombre de faux positifs.

En conclusion, nous retiendrons que le FWER permet un strict contrôle des hypothèses nulles rejetées à tort. Ce critère est ainsi très pertinent lorsqu'il s'agit de valider un résultat. Le test (1.2) fournit un exemple typique d'un tel cas pour lequel le test multiple sert à déterminer le traitement le plus adéquat d'un diabète de type 2. Le FDR autorise plus de faux positifs que le FWER. Cette augmentation n'est pas nécessairement nuisible et fait sens dans moult problématiques contemporaines. En particulier lorsqu'un test multiple est mené dans un but prospectif, contrôler le FDR s'avère pertinent. La section suivante fournit quelques exemples.

1.3. QUELQUES DOMAINES D'APPLICATIONS DES TESTS MULTIPLES

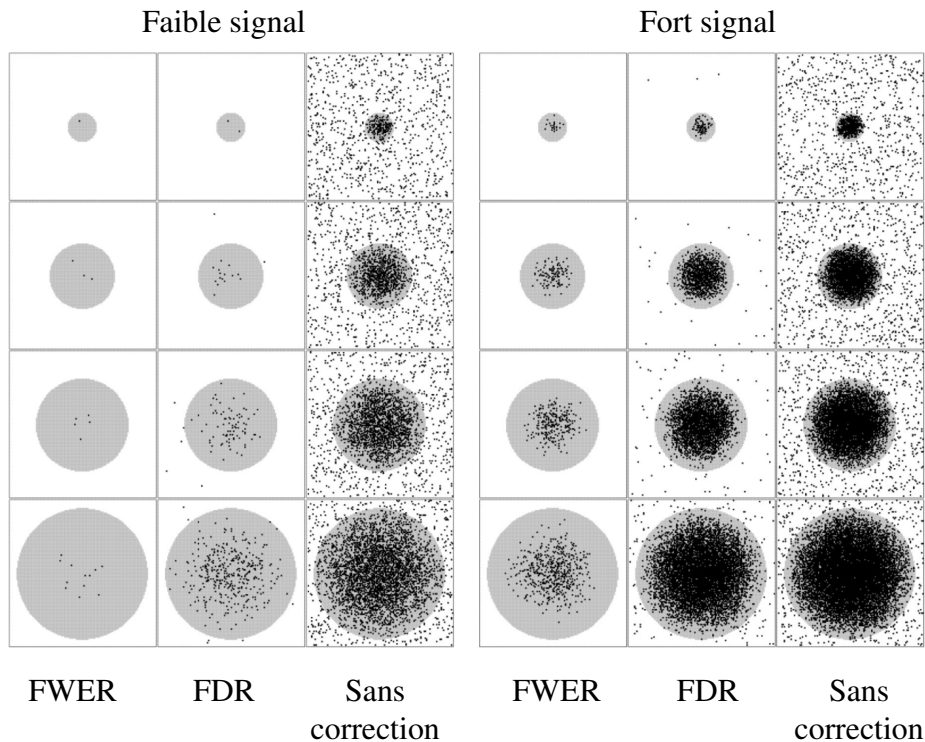


FIGURE 1.6 – Rejets obtenus en contrôlant le FWER (à gauche), le FDR (au milieu) et sans correction multiple (à droite). Les zones grisées (resp. blanches) correspondent à la localisation des rejets corrects (resp. incorrects). (Figure 1.2 de [7])

Nous noterons par ailleurs que lorsque toutes les hypothèses testées sont de vraies hypothèses nulles, le FDR et le FWER sont des critères équivalents. En effet, tous les potentiels rejets sont dans ce cas effectués à tort.

1.3 Quelques domaines d'applications des tests multiples

Au cours de ces dernières années, l'évolution des technologies a amené les statisticiens à traiter des données toujours plus complexes, à la fois en termes de nature des données et en termes de taille. Les chercheurs n'ont pas un paramètre à étudier mais des centaines de milliers de paramètres. Nous nous confrontons alors à un problème de multiplicité et si la nécessité d'une correction multiple a pu se poser, la nature des données amenées par les problématiques contemporaines a transformé ce questionnement en "comment prendre en compte la multiplicité des tests?". Nous illustrons ici l'utilisation faite des tests multiples dans quelques problématiques contemporaines. Loin d'être exhaustifs, nous ne mention-

CHAPITRE 1. INTRODUCTION (FRANÇAIS)

nons que les domaines les plus fréquemment concernés par la problématique des tests multiples. Précisément, nous donnons un aperçu des progrès techniques faits en génomique, neurosciences et climatologie puis nous transposons les conséquences de ces avancées technologiques au cadre mathématique à travers divers exemples tirés de la littérature.

Données issues de la génomique

L'avancée technologique significative dans le domaine de la génomique fut l'apparition des puces à ADN. De leurs balbutiements en 1991 à aujourd'hui, la technologie n'a cessé d'être améliorée. L'utilisation la plus connue des puces à ADN est la mesure de l'expression des gènes. Désormais, en une seule expérience, elles permettent de mesurer l'expression de centaines de milliers de gènes. Dans ce genre d'expérience, une fois ces mesures faites, un test multiple est alors typiquement utilisé pour déterminer les gènes différentiellement exprimés entre deux conditions. Dans ce cas, contrôler le FDR permet de fournir une liste raisonnable de gènes potentiellement significatifs quitte à commettre quelques erreurs, plutôt qu'une liste trop courte qui assurerait un contrôle strict de ces erreurs. Le lecteur intéressé pourra trouver une revue plus complète de l'utilisation des tests multiples en génomiques dans [8].

Exemple 1.1. (*Identification des gènes altérés en cas de déficience du gène Apo AI [9]*) Le gène Apo AI est connu pour jouer un rôle central dans la régulation du taux de cholestérol HDL. Le but de cette étude était d'identifier les gènes ayant une expression altérée entre des souris pour qui le gène Apo AI était totalement inactif et des souris ne présentant pas cette anomalie. D'un point de vue mathématique, les données résultant de cette expérience sont contenues dans une matrice de taille 6356 (nombres de gènes étudiés) \times 16 (nombres de souris).

Exemple 1.2. (*Etude comparative entre deux formes de leucémies [10]*) Il s'agissait dans cette étude, d'identifier les gènes différentiellement exprimés entre deux types de leucémie. Les données consistaient en une matrice de taille 6817 (nombres de gènes étudiés) \times 38 (nombre de sujets).

Comme le montrent les deux exemples précédents, les données issues des expériences génomiques amènent un nombre de variables (gènes étudiés) très élevés devant un nombre peu élevé de répétitions (les sujets de l'expérience). A cela s'ajoute la potentielle dépendance entre les gènes étudiés. Ces particularités font des données génomiques un véritable challenge mathématique.

Données issues des neurosciences

Un des domaines majeurs des neurosciences est l'imagerie par résonance magnétique (IRM). L'IRM est un moyen de mesurer l'activité cérébrale in vivo, de manière indirecte et

1.3. QUELQUES DOMAINES D'APPLICATIONS DES TESTS MULTIPLES

non invasive. Dans les années 70, le physicien britannique Peter Mansfield fut le premier à montrer comment les signaux de l'IRM pouvaient être analysés mathématiquement, rendant ainsi possible la conversion de ces signaux en une image exploitable. Mais ce n'est qu'une vingtaine d'années plus tard, que le chercheur japonais Seiji Ogawa permettra, grâce à ses travaux sur le magnétisme de l'hémoglobine, les premières images du cerveau en fonctionnement. Pour simplifier grossièrement, la détection de l'activité cérébrale est basée sur la mesure de l'effet BOLD : Blood Oxygen Level Dependant. Cet effet BOLD mesure des variations des propriétés du flux sanguin, lié à l'aimantation de l'hémoglobine contenue dans les globules rouges du sang. Une image volumique du cerveau est acquise toutes les 1 à 3 secondes. Pour chaque instant d'acquisition, un volume du cerveau correspond à l'acquisition de l'ordre de 100 000 voxels (= cube de 2 ou 4 mm de côté). Les acquisitions d'IRMf correspondent donc à l'enregistrement de séries temporelles pour chaque voxel couvrant ensemble le volume entier du cerveau.

Exemple 1.3. (IRMf d'activation [11]) Une expérience typique d'imagerie cérébrale consiste à déterminer les régions cérébrales activées lorsqu'un sujet effectue une tâche. Dans ce cas, un test multiple est utilisé pour tester chaque zone du cerveau. Pour chacune d'entre elles, nous testons si l'activation de la zone est significativement différente de zéro ou non. La Figure 1.7 est un exemple du type d'image statistique obtenue à la fin d'un tel test.

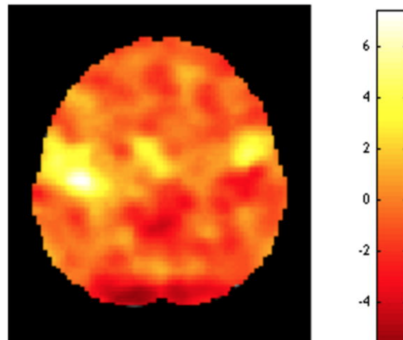


FIGURE 1.7 – Résultat d'un test multiple ayant servi à identifier les voxels significativement activés lors d'une tâche. Au total, 7505 voxels sont testés [11].

Là encore, compte tenu du coût et de la pénibilité de l'expérience pour le sujet, l'analyse statistique de ce type de données est rendue ardue par le grand nombre de tests (voxels ou régions cérébrales) devant le nombre de répétitions (longueur d'acquisition). Ces données sont également connues pour être fortement corrélées.

Dans ce domaine, la problématique des tests multiples a jusque tardivement été outrepassée par une partie de la communauté scientifique. Selon une étude portant sur 6 journaux

CHAPITRE 1. INTRODUCTION (FRANÇAIS)

majeurs de neuroimagerie, 25% à 30% des publications portant sur l'analyse de données d'IRMf n'utilisaient pas de seuils critiques corrigés en 2008 [12].

Données issues des sciences de l'atmosphère

L'atmosphère terrestre est une couche de gaz invisible qui entoure la terre. Son étude a été rendue possible par le développement des technologies spatiales, permettant la construction de satellites toujours plus puissants. Il est désormais possible de mesurer depuis l'espace la composition de l'atmosphère et ce à l'échelle de la planète. D'un point de vue mathématique, les données se présentent sous la forme d'une grille de points représentant une zone de l'atmosphère à une altitude donnée. Pour chaque point de la grille, l'enregistrement satellitaire correspond à une série temporelle de laquelle est déduit une statistique de test. Ce type de données est connu pour avoir une forte corrélation spatiale [13].

Exemple 1.4. (*Prévisions météorologiques [14]*). *Le SOI est un indice climatique qui mesure la différence de pression atmosphérique entre Tahiti et Darwin, en Australie. Dans cet exemple, il s'agissait de déterminer si cet indice pouvait être utilisé pour les prévisions météorologiques hivernales dans l'hémisphère nord. Pour cela, les tests statistiques permettaient d'identifier les corrélations statistiquement significatives entre les valeurs du SOI (pour les mois de juin/juillet/août) et les relevés de pression atmosphériques dans l'hémisphère nord (pour les mois de décembre/janvier/février). Ici, la zone étudiée correspondait à 936 points (voir Figure 1.8).*

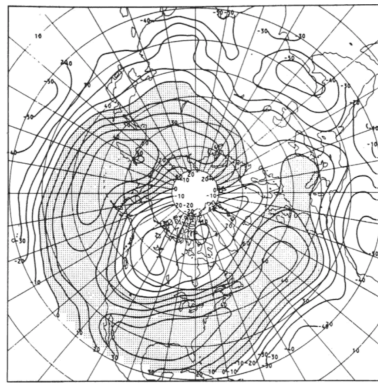


FIGURE 1.8 – Corrélations entre les valeurs de l'indice *SOI* en été et des relevés hivernaux de pression atmosphérique dans l'hémisphère nord (voir le texte) - Figure 5.6 de [14].

Il semblerait que ce domaine ne soit pas non plus épargné par la négligence du problème de multiplicité. Durant la première partie de l'année 2014, sur les 281 articles publiés dans *Journal of Climate*, moins de 3% des articles proposant une étude statistique impliquant des tests multiples utilisaient des seuils critiques corrigés [15].

1.4. RECONSTRUCTION DES RÉSEAUX DE CONNECTIVITÉ CÉRÉBRALE

Le domaine d'application qui motive cette thèse appartient au domaine des neurosciences et concerne l'analyse des données obtenues par imagerie par résonance magnétique fonctionnelle. La section suivante présente brièvement ce domaine sous l'angle du statisticien et situe clairement la place des tests multiples dans cette problématique.

1.4 Application à la reconstruction des réseaux de connectivité cérébrale en neurosciences

Au fil des années, les importants progrès technologiques ont permis une multiplication des études basées sur l'analyse de données obtenues par imagerie par résonance magnétique fonctionnelle (IRMf) (voir Figure 1.9). L'analyse de données d'IRMf rassemble de nom-

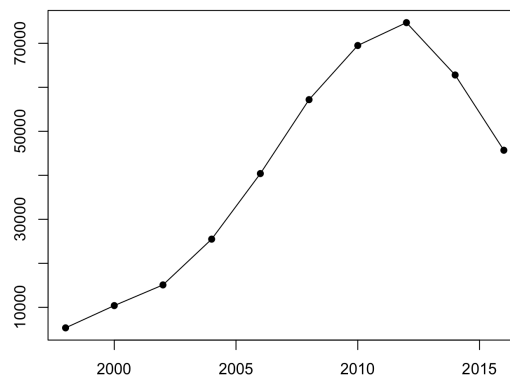


FIGURE 1.9 – Nombre d'articles par année référencés dans la base de données de Google scholar en utilisant les mots clés *fMRI study*.

breux protagonistes au profil divers : neuroscientifiques, psychologues, physiciens, statisticiens, entre autres. Typiquement, les données d'IRMf sont très bruitées et présentent une structure de dépendance "complexe". A cet égard, le rôle du statisticien est primordial dans la compréhension de la nature de ces données afin d'en tirer des résultats pertinents, faisant sens pour les neuroscientifiques.

Analyse statistique des données d'IRMf

D'un point de vue mathématique, l'issue d'une expérience d'IRMf consiste en l'acquisition de séries temporelles pour chaque voxel couvrant ensemble le volume entier du cerveau. Comme nous l'avons vu, le nombre de voxels nécessaires pour couvrir l'ensemble du cerveau est très élevé. Par conséquent, afin de faciliter l'analyse statistique, la taille des données est souvent réduite en regroupant certains voxels entre eux [16]. La mesure du

CHAPITRE 1. INTRODUCTION (FRANÇAIS)

signal correspondant à un regroupement est obtenue en agrégeant entre elles les mesures de chaque voxel (faisant partie du regroupement). Ce regroupement peut s'effectuer soit en fonction de critères purement anatomiques soit en fonction de critères fonctionnels soit en tenant compte de ces deux types de critères.

Les études basées sur l'IRMf sont généralement menées afin d'identifier soit les régions cérébrales qui échangent de l'information lorsqu'un sujet accomplit une tâche soit pour identifier les régions cérébrales activées lorsqu'une personne effectue une tâche ou adopte un certain comportement social (comme l'égoïsme, la jalousie, l'altruisme etc. . .). En 2008, les études en faveur de ce dernier objectif ont grandement été remise en question. La section suivante relate la controverse.

Polémique en neuroimagerie

A la fin des années 2008, en publiant l'article "Voodoo Correlations in Social Neuroscience" [16], les auteurs provoquent un houleux débat comme en témoignent les 1216 citations référencées par google scholar. Ils mettent en lumière les incohérences statistiques retrouvées dans de nombreuses analyses de données d'IRMf, en particulier lorsque ces études portent sur l'analyse des corrélations. Précisément, ils affirment que la plupart des corrélations (déduites d'une expérience d'IRMf) entre un comportement social et l'activation d'une région cérébrale sont fausses. Ils soutiennent que bon nombre de corrélations reportées dans ce type d'étude sont trop élevées pour être crédibles. En effet, la valeur de la corrélation entre deux mesures est limitée par la fiabilité de ces deux mesures. Pour une expérience d'IRMf, d'après [16], la fiabilité entre deux mesures n'est pas supposée dépasser 0.75. Ainsi, les auteurs de [16] ont effectué une revue de nombreux articles pour lesquels des mesures de corrélations obtenues par IRMf étaient reportées. Comme le montre la Figure 1.10, de nombreuses mesures de corrélations dépassent 0.75. Alors d'où vient ce problème ?

Un des points critiqués par les auteurs est que la méthode utilisée pour sélectionner les voxels qui doivent être testés n'est pas indépendante des tests effectués sur ces voxels. Ils affirment par ailleurs que les corrélations élevées obtenues en analysant la totalité du cerveau résultent uniquement du bruit. Nous noterons que ce point est contredit dans [17]. De plus, ils soutiennent que le problème de multiplicité est bien reconnu dans ce domaine et que dans la plupart des articles de neuroimagerie les auteurs utilisent des seuils critiques corrigés. Pourtant, comme nous l'avons vu, une étude [12] (également effectuée en 2008) souligne qu'une part non négligeable des auteurs continuent d'ignorer le problème de multiplicité.

Selon notre point de vue, l'utilisation "correcte" des tests multiples offre une solution aux problèmes soulevés dans [16]. En effet, même en prenant en compte le problème de multiplicité, nous devons utiliser une procédure de test multiple adaptée à la structure de dépendance des données et ce point est particulièrement épineux. A cet égard, le besoin

1.4. RECONSTRUCTION DES RÉSEAUX DE CONNECTIVITÉ CÉRÉBRALE

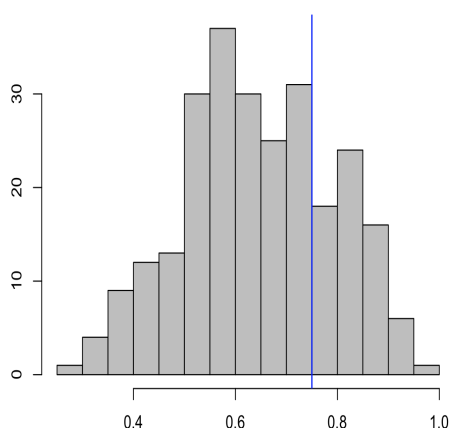


FIGURE 1.10 – Résultats établis dans [16] : Histogramme de valeurs de corrélations obtenues par IMRf issues des articles considérés par [16]. Ligne bleue : borne supérieure pour les valeurs des corrélations obtenues par IRMf imposée par la fiabilité des mesures.

de développement de procédures de test multiple aptes à analyser les données d'IRMf (*i.e.* adaptées à la nature des données sans être trop conservatives) constitue un remarquable et difficile enjeu pour le mathématicien.

Test multiple pour la reconstruction des réseaux de connectivité cérébrale

Il a été observé en neurosciences que la compréhension des processus cognitifs passe par l'identification de réseaux de régions cérébrales qui échangent de l'information. Ainsi, cette thèse est motivée par l'identification des connexions cérébrales fonctionnelles qui sont observables chez des sujets qui lors de l'acquisition sont au repos. Pour ce faire, d'un point de vue mathématique, nous voyons les régions cérébrales comme autant de sommets d'un graphe dont les arêtes correspondent aux paires de régions cérébrales échangeant de l'information, ces paires de régions étant identifiées par un test multiple.

Comme résumé par la Figure 1.11, cette thèse est donc liée à trois domaines bien distincts : celui de l'imagerie médicale (l'objet d'étude), celui des modèles graphiques (la modélisation mathématique utilisée) et celui des tests multiples (l'outil mathématique utilisé pour répondre à la problématique). Tout au long de cette thèse nous exploiterons des résultats déjà connus concernant les deux premiers domaines. Les contributions de celle-ci s'inscrivent dans le domaine des tests multiples. Précisément, en considérant la corrélation comme mesure de connectivité entre les paires de régions cérébrales, nous cherchons à détecter les paires de régions cérébrales significativement corrélées. L'enjeu est alors d'effectuer cette détection en assurant un contrôle des erreurs de type I.

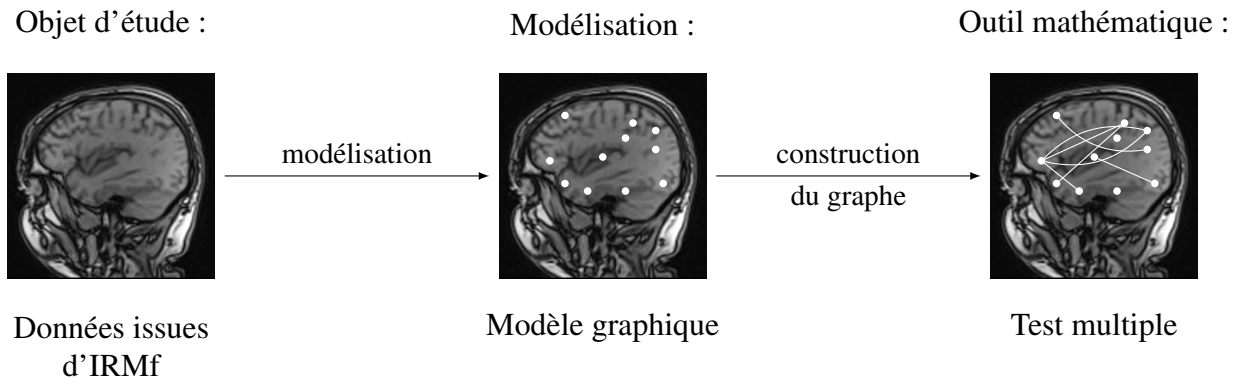


FIGURE 1.11 – Champs d'interaction concernés par cette thèse.

Nous étudierons le contrôle du FWER, définit ici par

$\text{FWER} = \mathbb{P}(\text{"il existe une paire de régions cérébrales déclarée significativement corrélées à tort"})$,

ainsi que le contrôle du FDR définit par la moyenne du FDP ici déterminé par,

$$\text{FDP} = \frac{\text{nombre de paires de régions cérébrales déclarées significativement corrélées à tort}}{\text{nombre de paires de régions cérébrales déclarées significativement corrélées}}.$$

La force du test multiple est donc de permettre un contrôle des arêtes incluses à tort (via le FWER ou le FDR) dans le graphe des réseaux de connectivité cérébrale que nous cherchons à construire. Nous nous heurtons à un problème de test multiple dans un cadre de forte dépendance et comme nous l'avons mentionné cette dépendance est l'obstacle majeur à l'analyse des données d'IRMf d'un point de vue mathématique. Ainsi, afin de faciliter cette dernière, les nouveaux résultats établis dans cette thèse se situent dans le cadre gaussien (profitant de ce fait de nombreux résultats théoriques). Ce manuscrit interroge donc le problème du contrôle du FWER et du FDR pour différentes structures de dépendance dans un cadre gaussien et ce dans une optique d'analyse de données d'IRMf.

Nous concluons cette introduction par la description du contenu de cette thèse.

Organisation du manuscrit

Le Chapitre 2 présente sommairement les modèles graphiques. Loin d'être exhaustifs, nous donnons au lecteur les clés mathématiques suffisantes pour justifier la démarche adoptée dans cette thèse : l'utilisation du test multiple pour la construction des réseaux de connectivité cérébrale quand la mesure de connectivité entre deux régions cérébrales est la corrélation entre celles-ci. Nous proposons par ailleurs une revue des techniques existantes de reconstruction de graphes afin de situer la place des tests multiples dans la littérature contemporaine (pour cette problématique).

1.4. RECONSTRUCTION DES RÉSEAUX DE CONNECTIVITÉ CÉRÉBRALE

Le Chapitre 3 formalise la problématique des tests multiples. Nous définissons en particulier les deux critères de décision étudiés dans cette thèse (FWER et FDR) ainsi que plusieurs types de structure de dépendance considérés dans la littérature et en lien avec ce manuscrit. Nous mentionnons quelques procédures permettant un contrôle du FWER et nous introduisons la procédure de Benjamini-Hochberg (BH). Nous reportons quelques résultats connus concernant le contrôle du FDR par cette procédure. Nous proposons également une nouvelle borne pour le FDR de BH pour une structure de dépendance positive plus faible que celles connues pour contrôler le FDR.

Le Chapitre 4 s'intéresse au FDR de BH pour différentes structures de dépendance, lorsque les statistiques de tests sont gaussiennes pour des tests (de moyenne) bilatéraux. En particulier, nous établissons théoriquement les premiers (à notre connaissance) résultats permettant d'améliorer significativement la borne proposée par Benjamini et Yekutieli pour des structures de dépendance non nécessairement positives. Dans un tel contexte, cette étude suggère par ailleurs que BH contrôle le FDR quelque soit la matrice de covariance.

Le Chapitre 5 concerne l'étude du contrôle des erreurs de type I lorsque les statistiques de tests sont asymptotiquement gaussiennes. Nous présentons quelques procédures de test multiple connues pour garantir un contrôle (asymptotique) du FWER en apportant par ailleurs une clarification théorique de certains résultats. Nous établissons également quelques premiers résultats théoriques concernant le contrôle (asymptotique) du FDR, non suffisants par ailleurs pour être utilisés dans le cadre des tests de corrélation.

Le Chapitre 6 propose d'étudier les performances des procédures de test multiple présentées au Chapitre 5, à la fois sur simulation et sur données réelles.

Chapter 1

Introduction (english)

This introduction is deliberately informal and presents (hopefully) the issue of a multiple hypothesis testing from an intuitive point of view. We also emphasize the connection between the multiple testing problem and some contemporary scientific challenges. We especially focus on the one that motivated this thesis, that is, the neuroimaging data analysis.

Contents

1.1	The multiple testing problem	31
1.2	Criteria and decisions	35
1.3	Multiple testing: some contemporary applications	38
1.4	Estimation of cerebral connectivity	42

1.1 The multiple testing problem

Single hypothesis testing

Let us begin with a short reminder of single hypothesis testing. Let H_0 be an hypothesis (called null hypothesis) supposed to be true *a priori*. Roughly speaking, hypothesis testing is a statistical inference aiming at deciding if this assumption is true or not from a real data set (the observed phenomena). When the null hypothesis is rejected, a test concludes that data support an alternative hypothesis, denoted by H_1 in general. Two types of errors can occur, namely: rejecting a true null hypothesis (type I error), or retaining a false null hypothesis (type II error). Traditionally, the most undesirable error is making a type I error and in this regards, statisticians generally conduct a type I error controlling test, that is type I error is bounded by some $\alpha \in [0, 1]$, called the level of the test. In other words, if the real data set is uninformative, the null hypothesis H_0 will be not rejected by a test of level α , with probability larger than $1 - \alpha$. Therefore, the hypothesis H_0 is favored and a test is

informative only when the null hypothesis is rejected.

Let us provide an example to illustrate this non equivalence between the two types of errors. Consider the case of an i.i.d sample where each variable follows a $\mathcal{N}(\mu, 1)$ where μ is the unknown parameter of interest. A first test is $H_0 : \mu < 0$ against $H_1 : \mu \geq 0$. The test which rejects H_0 when $n^{1/2}\bar{X}_n = n^{-1/2} \sum_{i=1}^n X_i$ exceeds $\bar{\Phi}^{-1}(\alpha)$, where $\bar{\Phi}(\cdot)$ denotes the upper-tail function of a standard Gaussian distribution, is a test of level α . By switching the null and the alternative, we obtain a second test, $H'_0 : \mu \geq 0$ against $H'_1 : \mu < 0$. The test which rejects H'_0 when $n^{1/2}\bar{X}_n$ is smaller than $-\bar{\Phi}^{-1}(\alpha)$ is a test of level α . Conclusions of these two tests are represented in Figure 1.1.

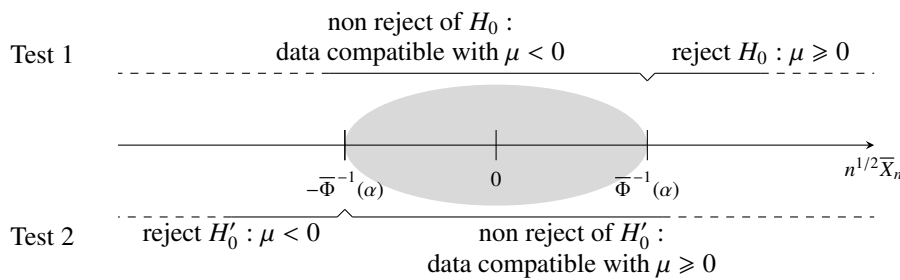


Figure 1.1 – Non equivalence between two types of errors: case of a test of the mean of *i.i.d* Gaussian variables.

In the shaded part, depending on the choice of the null hypothesis, both hypotheses " $\mu \geq 0$ " and " $\mu < 0$ " could be considered.

Multiple hypothesis testing: a first example

Multiple testing arises when more than one hypothesis is simultaneously involved in a statistical test. A multiple testing procedure can be seen as a process that associates the index set of the null hypotheses that it rejects to the index set of the null hypotheses to be tested. Although each individual test is realized in much in the same way as a single hypothesis test, the multiplicity of the tests implies different specificities. In particular, if the null hypotheses form a partition of the state space of the parameter of interest, no null hypothesis can be favored.

To illustrate the latter remark, we consider the case of a n -normal sample $X = (X_1, \dots, X_n)$ generated from a $\mathcal{N}(\mu, 1)$ distribution where the (unknown) parameter of interest is $\mu \in \mathbb{R}$, and consider the following test:

1.1. THE MULTIPLE TESTING PROBLEM

$$\begin{cases} H_0 : \text{"}\mu = 0\text{"} \\ H_{0^+} : \text{"}\mu > 0\text{"} \\ H_{0^-} : \text{"}\mu < 0\text{"} \end{cases} . \quad (1.1)$$

This test is a special case of the three-sided hypothesis testing considered in [1] and can be seen as a "multiple reformulation" of the tests of the previous section. The test which rejects H_0 when $n^{1/2}\bar{X}_n$ exceeds $\bar{\Phi}^{-1}(\alpha/2)$ or is smaller than $-\bar{\Phi}^{-1}(\alpha/2)$ is a test of level α . The test which rejects H_{0^+} when $n^{1/2}\bar{X}_n$ is smaller than $-\bar{\Phi}^{-1}(\alpha)$ is a test of level α and the test which rejects H_{0^-} when $n^{1/2}\bar{X}_n$ exceeds $\bar{\Phi}^{-1}(\alpha)$ is a test of level α . Since the three null hypotheses form a partition of \mathbb{R} (the state space of μ), only one null hypothesis can be incorrectly rejected. In this special case, the "global control" of the type I error is simply achieved by controlling the type I error of each individual test. In other words, testing simultaneously the three null hypotheses at level α is sufficient to obtain a global control of the type I error at level α . In the same way as Figure 1.1, Figure 1.2 draws the conclusions of the test (1.1).

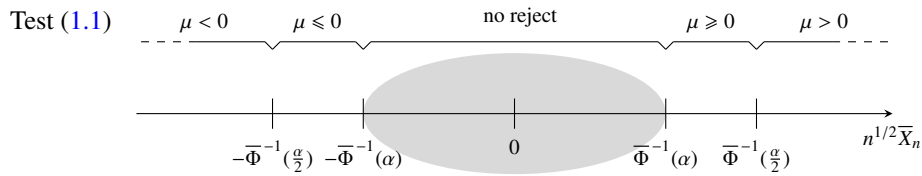


Figure 1.2 – Three-sided hypothesis test for means in the Gaussian setting (see text).

In Figure 1.2, the shaded part is a part in which no reject can occur. Since the three null hypotheses form a partition of \mathbb{R} , the only conclusion of the three-sided test (in the shaded part) is that the data are uninformative, no null hypothesis is favored.

The multiple testing (1.1) occurs in concrete situations, typically in the clinical trials. For instance, in [1], two treatments for treating type 2 diabetes are compared. Assume that the data correspond to the difference between the blood sugar level under insulin B and the blood sugar level under insulin A. Then the three-sided test can be reformulated as follows:

$$\begin{cases} H_0 : \text{"Insulin A and insulin B have the same efficiency"} \\ H_{0^+} : \text{"Insulin A is more efficient than insulin B"} \\ H_{0^-} : \text{"Insulin A is less efficient than insulin B"} \end{cases} . \quad (1.2)$$

In this case, the multiple testing is a way to prevent conclusions of a clinical trial from being influenced by the pharmaceutical industry.

In the previous example, the decisions of tests are obtained by bounding $n^{1/2}\bar{X}_n$ by some quantity defined in order to guarantee a control of type I error at an arbitrary level α . Consequently, if the level changes, another quantity has to be calculated to conclude the test. This is not convenient in practice. Moreover, the choice of α is quite arbitrary, so why considering $\alpha = 5\%$ and not $\alpha = 4\%$ or $\alpha = 6\%$? An interesting alternative is to consider the p -value. It is the largest α for which the null hypothesis can be rejected at level α . By nature, the p -value gives the decision of the test at all possible levels. A p -value is a random variable uniformly distributed on $[0, 1]$ under the null. The intuition behind is to measure the "probability" of observing the real dataset if the null is true. A "small" p -value provides evidence against H_0 and an "high" p -value indicates that there is no contradiction between the observed phenomena and the null.

Multiplicity issue

A false positive (also called a false discovery) occurs when more than one hypothesis are tested simultaneously and a null is incorrectly rejected. If the nulls do not form a partition of the state space of the interested parameter, the multiplicity issue arises. Control the level of each individual test is no longer sufficient to guarantee a control of the global error. Indeed, if all the nulls (potentially very numerous) are independently tested at the same level (and without multiple testing adjustment), the plurality of the errors will induce a loss of the global control and the conclusion of the test will be incorrect. For instance, let us consider a case with m independent hypotheses to test, and a significance level of 5%. What is the probability of observing at least one significant result just by chance? Figure 1.3 plots the probability of observing at least one false positive by chance.

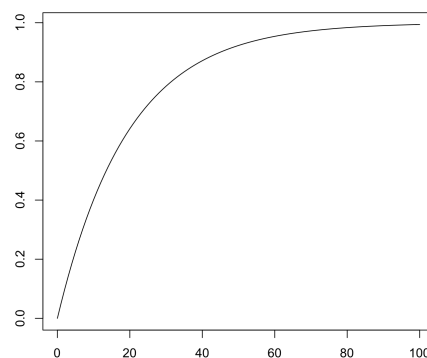


Figure 1.3 – Probability of observing at least one false positive by chance depending on the number of null hypotheses ($m \mapsto 1 - (1 - 0.05)^m$).

Indeed, we have

$$\begin{aligned}\mathbb{P}(\text{"get at least one false positive"}) &= 1 - \mathbb{P}(\text{"get no false positive"}) \\ &= 1 - (1 - 0.05)^m,\end{aligned}$$

where the independence is used for the last equality.

Figure 1.3 illustrates an effect strongly related to the repetition of tests, which can be summarized as "the more one looks for something, the more one finds it". Even if all of the tests are actually not significant, the more they rise in quantity, the higher the probability of getting a false discovery is. The neuroscience research field is concerned by this problem. More concretely, Figure 1.4 is a typical example of absurd multiple testing decision that can occur by ignoring the multiplicity issue. Indeed, as shown in [2], some cerebral regions of a dead salmon can be declared significantly activated by exploring enough cerebral regions.

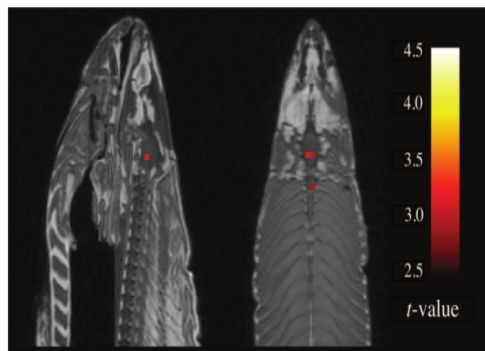


Figure 1.4 – Multiplicity issue: cerebral activity of a dead salmon [2].

1.2 Criteria and decisions

As seen above, the multiplicity problem is the main issue of a multiple testing procedure. How to guarantee that the nulls rejected by a multiple testing procedure are not rejected only by chance? How to measure the quality of a multiple testing procedure? In practice, the mathematician provides a probabilistic response and relies on two main (the most studied in the multiple testing literature) criteria: the Family-Wise Error Rate (FWER) and the False Discovery Rate (FDR). As we have seen, a single hypothesis test focuses on bounding the probability of type I error by the level of the test. In the same manner, a multiple testing procedure focuses on bounding either the FWER or the FDR by the level of the test. Thus the hypotheses rejected by a multiple testing procedure depend on the criterion that we aim at controlling. We now define this criteria more precisely.

FWER

The type I error can be naturally extended to the multiple testing framework via the FWER. Given a multiple testing procedure, the FWER is the probability of rejecting at least one true null hypothesis, that is, making one or more false discoveries. When several hypotheses are tested simultaneously, we have

$$\text{FWER} = \mathbb{P}(\text{"there exists at least one null hypothesis incorrectly rejected"}).$$

As we have seen, if the nulls do not form a partition of the state space of the interested parameter, control each individual test at level α is not sufficient to guarantee the control of the FWER at the same level α . For the latter purpose, we should adjust the level of each individual test. This process is called in the literature "adjusted/corrected threshold" or similarly "adjusted/corrected p -values". The classical method to control the FWER is the Bonferroni's method [3], which counteracts the problem of multiple comparisons by testing each individual hypothesis at a significance level of α/m , where α is the desired overall alpha level and m is the number of hypotheses. Although this method is very intuitive, it could be conservative if there are a large number of false hypotheses relative to the number of hypotheses being tested. Roughly speaking, it means that, by nature, the Bonferroni's method does not provide a "lot" of rejects. Otherwise, this method is very conservative if the test statistics are correlated. A first improvement has been obtained by Holm [4]. The Holm method starts with testing the most significant hypothesis by comparing its p -value to α/m , just as the Bonferroni method. If the hypothesis is rejected, then the p -value of the second most significant hypothesis is compared to $\alpha/(m-1)$, and so on, until the procedure comes to a stop. Necessarily, all hypotheses rejected by Bonferroni will also be rejected by Holm. However the Holm method could reject more hypotheses, that is, while still remaining the FWER control, the Holm method is more powerful. A lot of less conservative procedures have been developed over time. In 2010, Goeman and Solari [5] propose a general framework to describe most of these methods by using the sequential rejection principle of family-wise error control. From a high level, the gain of power is achieved by increasing the level of each remaining individual test depending on the hypotheses already rejected in previous steps.

FDR

A less conservative criterion than the FWER, called False Discovery Rate (FDR), has been introduced by Benjamini and Hochberg [6] in 1995. This criterion can be more suited for some contemporary problems. The FDR is defined by the mean of the False Discovery Proportion (FDP),

$$\text{FDR} = \mathbb{E}[\text{FDP}], \text{ where } \text{FDP} = \frac{\text{number of false discoveries}}{\text{number of discoveries}}.$$

Since the FDP is random, the study of FDR is more challenging from a mathematical point of view than the study of FWER. In practice, the difficulty relies on the type of dependence between the test statistics. Chapter 3 introduces some type of dependence assumptions linked to this thesis and Chapter 4 is devoted to the study of the FDR of the BH procedure with particularly structured dependencies.

Note that even if the FDR control of the BH procedure is only established in some particular dependency structures (see Chapter 4), this method is now a standard in practice. It is one of the greatest success of the application of the theoretical statistics in recent years as is proved by the current 45881 citations of the associated paper of Benjamini and Hochberg [6].

In order to quantify the quality of a multiple testing procedure, two criteria have been introduced. Thus, the natural question resulting from the latter is: how to choose a criterion? This is the topic of the following section.

FWER versus FDR

Generally, a FWER control is not better than a FDR control and conversely a FDR control is not better than a FWER control. Given a problem, we need to choose an appropriate criterion.

The control of the FWER guaranties that the event "there is no hypotheses wrongly rejected" occurs with high probability. The FDR estimates the proportion of false discoveries among the discoveries. In this respect, the FDR is less restrictive, it is adaptive with respect to the number of rejects. In [7], Roquain illustrates the philosophical difference between the FWER and the FDR. We report this illustration in Figure 1.5. Roughly, some null hypotheses are simultaneously tested using either the Bonferroni's procedure (for FWER control) or the BH procedure (for FDR control) or without multiple adjustment. Each reject is associated to a black dot in Figure 1.5 in such a way that a false discovery falls outside the gray disk and conversely a correct reject falls inside the gray disk. The FWER control ensures with high probability that no black dot falls outside the gray disk whereas the FDR control ensures that the number of black dots outside the disk out of the number of total black dots is, on average, less than the desired level of the test.

In practice, the FDR controlling procedures are more powerful than the FWER controlling procedures when the false null hypotheses are numerous and even more when the signal strength is strong (see Figure 1.5, bottom). The latter is at the price of an amount of false positives.

In conclusion, a strong control of false discoveries is possible with FWER control. This criterion is relevant when the multiple testing is the last scientific stage before a result is validated. For instance, the test (1.2) is typically used in such a case. The latter is the

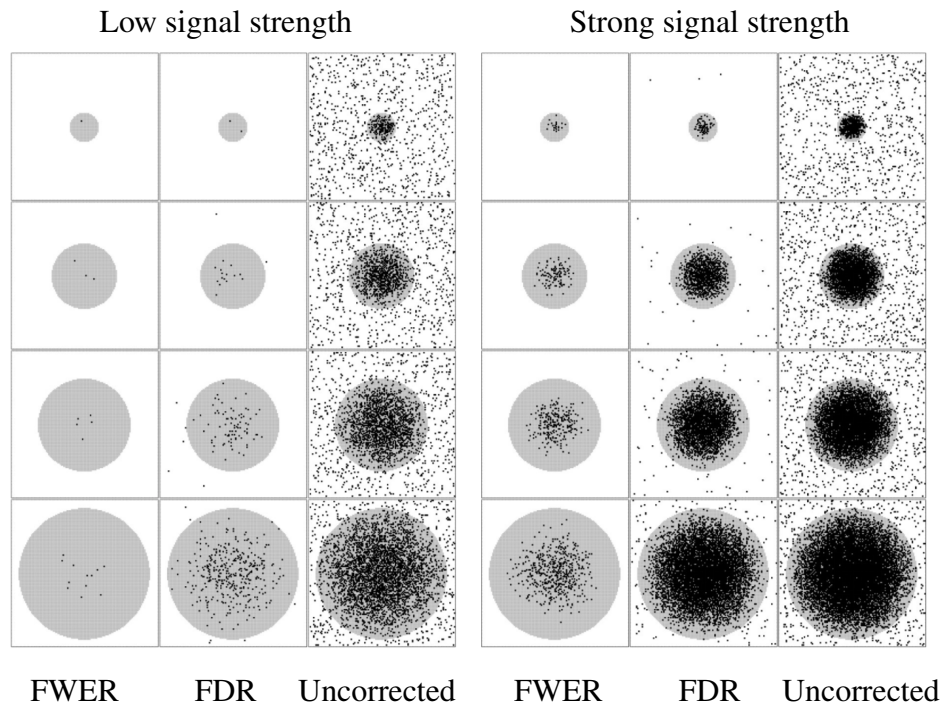


Figure 1.5 – Discoveries for FWER (left), FDR (middle) and without correction (right). The shaded parts (resp. white) are associated to the localization of correct (resp. incorrect) rejects. (Figure 1.2 de [7])

last scientific stage before a treatment for type 2 diabetes control is allowed on the market. FDR control allows more false discoveries than FWER control. The FDR is non necessary an inappropriate criterion and can be well-suited for some contemporary issue. In particular, when the analysis of an experiment resembles exploratory more than confirmatory research. The following section provides some exploratory experiments.

Finally, remark that when all the null hypotheses to be tested are the true nulls, FDR and FWER are equivalent criteria. Indeed, in this case, all potential rejects are incorrect rejects.

1.3 Multiple testing: some contemporary applications

In recent years, helped by decisive technological breakthroughs, the quality and quantity of data that must be analyzed by the statisticians have become more and more complex. The researchers have no longer one parameter to study but thousands or even hundreds of

1.3. MULTIPLE TESTING: SOME CONTEMPORARY APPLICATIONS

thousands parameters. Then the multiplicity arises and the question "is it really necessary to take into account the multiplicity issue?" has nowadays moved in "how to take into account the multiplicity issue?". Here, we provide a brief overview of the use of multiple testing in some contemporary research fields. This is hardly exhaustive, we focus on the most concerned applications with the multiplicity issue. More precisely, by using some examples from literature, we highlight the link between technological breakthroughs and mathematical analysis in genomics, neurosciences and atmospheric sciences.

Genomic data

In genomics research field, the DNA microarray is the most significant technological breakthrough. From its first steps in 1991 to nowadays, this technology is still on progress. The most popular use of the DNA microarray is the measure of gene expression. The latter allows the researchers to measure of hundreds of thousands gene expressions with only one experiment. In such a case, a multiple testing is typically used to identify the gene differentially expressed between two conditions and a FDR control is well-suited. Indeed, within this framework, the role of statisticians is more the result of futurology than of exact sciences. Providing a list of genes probably differentially expressed, containing few mistakes, is more relevant than providing a shorter list by ensuring a strong control of false positives. We refer the interested reader to [8] for a more complete review of the use of multiple testing in genomics.

Example 1.1. (*Identification of the genes with altered expression in Apo AI deficient mice*) The gene Apo AI is known to play a pivotal role in the HDL metabolism. This study aimed at identifying the genes with altered expression between Apo AI deficient mice and mice for which this gene is not deficient. From a mathematical point of view, the dataset is a 6356 (number of studied genes) \times 16 (number of mice) real-valued matrix.

Example 1.2. (*Comparison between two types of leukemia [10]*) This study aimed at identifying the genes differentially expressed between two types of leukemia. The dataset is a 6817 (number of studied genes) \times 38 (number of patients) real-valued matrix.

As is emphasized by the two previous examples, the dataset resulting from genomic experiments provide many variables (studied genes) with few repetitions (the patients). Furthermore, the genes are potentially dependent. It results from these specificities that the analysis of genomics data is really challenging for the mathematicians.

Neuroscientific data

Magnetic resonance imaging (MRI) plays a pivotal role in neuroscience research field. The MRI is a non invasive method to measure brain activity *in vivo*. In the 70s, the British physician Peter Mansfield achieved for the first time a mathematical analysis of the MRI

CHAPTER 1. INTRODUCTION (ENGLISH)

signals, which allowed to convert these signals into a usable image. Some twenty years later, thanks to his work on the magnetism of hemoglobin, the Japanese researcher Seiji Ogawa provided the first functional imaging of the brain. Roughly speaking, the detection of cerebral activity relies on the measure of the BOLD effect: Blood Oxygen Level Dependent. The BOLD effect reflects the amount of oxygen variations in blood as a function of brain activity. A brain functional image volume is acquired every 1 to 3 seconds during a scan. For each acquisition time, each functional image for the whole brain contains about 100,000 voxels (small cube-shaped regions). Thus, the output of functional MRI (fMRI) experiment consists in recording time series for each voxel in the whole brain.

Example 1.3. (*fMRI activation [11]*) Typically, an fMRI experiment aims at identifying the activated cerebral regions while somebody is doing a task. In such a case, for each cerebral region, the goal of a multiple testing decision is to infer from the data whether the activation of a cerebral region is significantly different from zero. Figure 1.6 provides an example of statistical image resulting from such a test.

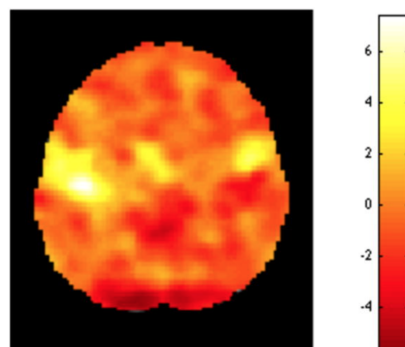


Figure 1.6 – Statistical image obtained via a multiple testing in order to identify the voxels significantly activated during a task. In this case, 7505 voxels are tested [11].

This type of experiment has a high financial cost and is uncomfortable for the patient. Thus, in the same way as genomics data, the mathematical challenge comes from the fact that there is numerous variables (voxels or cerebral regions) with few repetitions (acquisition time). Additionally, this type of dataset is known to be highly correlated.

In neuroscience research field, the multiplicity issue seems to have been taken into account late. According to an investigation of six major neuroimaging papers, 25% to 30% of publications related to fMRI data analysis override the problem of multiplicity in 2008 [12].

Atmospheric data

The atmosphere of Earth is a set of layers of gases surrounding the planet. Helped by more and more powerful satellites, it is now possible to measure from space, the composition of Earth's atmosphere around the globe. From a mathematical point of view, the dataset is a gridpoint associated to a part of the atmosphere at a fixed level. For each point of the grid, the satellite recording is associated to a time series from which a test statistic is deduced. This kind of dataset is known to be highly correlated in space [13].

Example 1.4. (*Weather forecasting [14]*) The SOI index is a climatic index which measures the difference of atmospheric pressure between Tahiti and Darwin, in Australia. This study aimed at determining if this index could be useful to weather forecasting in winter in northern hemisphere. In this respect, the multiple testing was used to identify the statistical significant correlations between the SOI index (for June/July/August) and the measures of the atmospheric pressure in northern hemisphere (for December/January/February). Here, the explored part corresponds to 936 points (see Figure 1.7).

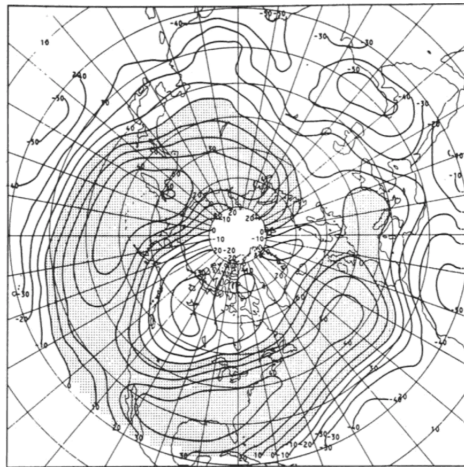


Figure 1.7 – Correlations between the SOI index in summer and values of atmospheric pressure in winter in northern hemisphere (see text) - Figure 5.6 from [14].

The multiplicity issue seems also to be neglected in a large number of geophysical studies. During the first half of 2014, less than 3% of articles including a multiple testing analysis out of the 281 published in Journal of Climate, used corrected thresholds [15].

This thesis is motivated by the analysis of the functional magnetic resonance imaging (fMRI). The following section briefly introduces the latter from a statistical point of view and situate the multiple testing with respect to our problem.

1.4 Application to estimation of cerebral connectivity in neurosciences

In recent years, helped by decisive technological breakthroughs, the number of neuroimaging studies based on functional Magnetic Resonance Imaging (fMRI) analysis has exploded (see Figure 1.8).

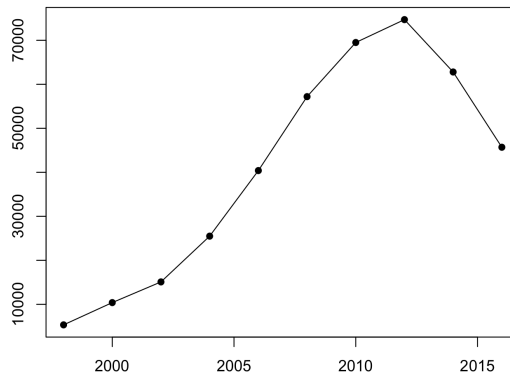


Figure 1.8 – Number of papers per year in the google scholar database using the keywords fMRI study.

The person involved in a such study may come from various backgrounds: neuroscience, psychology, physics and statistics, among others. An fMRI study provides very noisy data with a "complicated" covariance structure. In this regards, the role of statisticians is crucial to understand the nature of the data and obtain relevant results that can be interpreted by neuroscientists.

fMRI data analysis

From a mathematical point of view, the output of fMRI experiment consists in a time series for each voxel contained in the whole brain. As we have seen, the voxels contained in the whole brain are numerous. In order to do statistics, the size of data must be decreased by selecting a subset of voxels and aggregating measurements across them [16]. This selection can be based on purely anatomical or functional constraints, or a combination of the two. fMRI experiments are typically used either to identify cerebral connectivity or to identify which regions of a brain are activated while someone is doing a task or having a social behavior (e.g. selfishness, jealousy, altruism etc . . .). The latter instigated a major crisis in neuroscience research field in 2008. The following section relates the controversy.

Crisis in neuroimaging

With their paper "Voodoo Correlations in Social Neuroscience" [16] published at the end of 2008, the authors have given rise to a spirited debate (1216 citations on google scholar). They first pointed out statistical incoherences in the fMRI data analyses, especially when we are focused on correlation. They claimed that many studies which provide correlations between social behavior and measures of brain activation obtained by using fMRI should not be believed. They sustained that a lot of reported correlations are "impossibly high". Indeed, the correlation between two measures is upper bounded by the reliability of these two measures. For fMRI measures, according to [16], the reliability was not supposed to be greater than 0.75. Thus, the authors of [16] conducted a literature review of many articles for which correlations resulting on fMRI experiments were reported. As we can see in Figure 1.9, there is a lot of correlations greater than 0.75. So, what is the problem ?

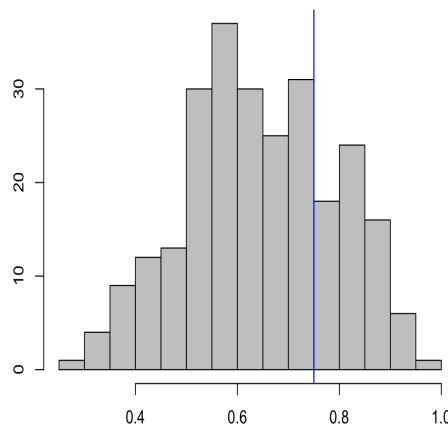


Figure 1.9 – Results of [16]: A histogram of the correlations resulting on fMRI experiments seen in the studies identified for analysis in [16]. Blue line: upper bound of reliabilities of fMRI and trait measures.

One of the practice condemned by the authors of this investigation is double-dipping: the method used to select which voxels should be tested is not independent of the tests performed on the resulting regions. They also claimed that large correlations from whole-brain regression analyses may just be due to noise. This assertion was contradicted in [17]. Moreover, the fMRI data analyses imply multiple comparisons, and they asserted that "adjustments of threshold are commonly employed". However, as we have seen previously, a study [12] (also conducted in 2008) showed that a lot of the authors still continued to ignore the multiplicity issue.

From our point of view, "properly" performed corrections for multiple testing allow

researchers to circumvent the problems raised by [16]. Even if one takes into account the multiplicity problem, one must apply a multiple testing procedure adapted to the covariance structure and this is the most challenging point. In this regards, the need for methods to "correctly" analyze fMRI data (*i.e.* adapted to the data without being too much conservative) gives rise to exciting new challenges for mathematicians.

Multiple testing for an estimation of cerebral connectivity

It comes from neurosciences that identifying the cerebral regions which exchange signals allows us to understand the underlying cognitive process. Thus this thesis is motivated by a resting state fMRI study, that is, an estimation of the regional interactions that occur when a subject is not performing an explicit task. In this regards, from a mathematical point of view, the cerebral regions are seen as the vertices of a graph in which an edge between two vertices means that the associated regions exchange signals. Additionally we aim at identifying the set of edges by a multiple testing.

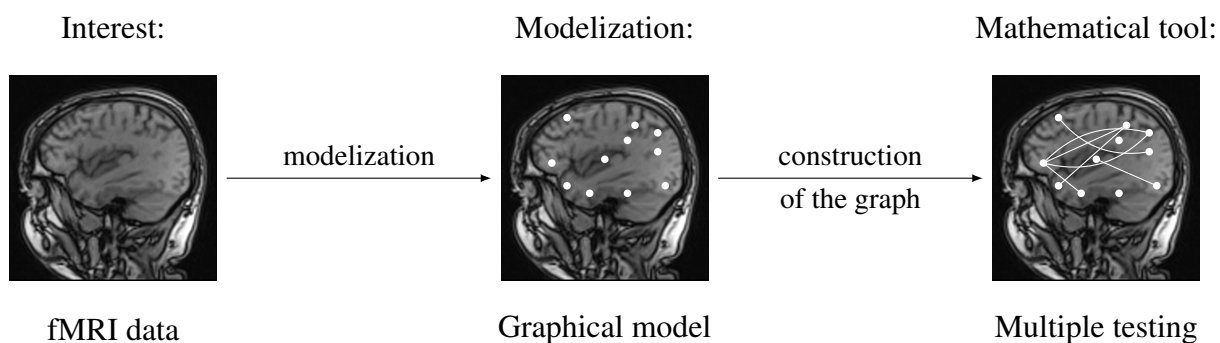


Figure 1.10 – The different scopes dealt with in this thesis.

Thus, this thesis relies on three distinct research fields: the medical imaging (the interest), the graphical models (the chosen modelling) and the multiple testing (the chosen mathematical tool to solve our problem). Throughout this thesis, we will use some known results of the first two research fields. The contributions of this thesis concern the multiple testing theory. More precisely, considering correlation as a connectivity measure between pairs of subset of cerebral regions, the aim is to identify pairs that are significantly correlated by using a multiple testing procedure. The issue of the latter is to detect the pairs while preserving a type I error control. We will focus either on the control of the FWER, here defined as

$$\text{FWER} = \mathbb{P}(\text{"there exists at least one pair of cerebral regions that are declared wrongly significant"}),$$

1.4. ESTIMATION OF CEREBRAL CONNECTIVITY

or on the FDR control, that is the mean of the FDP, here defined as

$$\text{FDP} = \frac{\text{number of pairs of cerebral regions that are declared wrongly significant}}{\text{number of pairs of cerebral regions that are declared significant}}.$$

Thanks to the multiple testing approach, it could be possible to control (via the FWER or the FDR) the edges incorrectly included in the graph of cerebral connectivity that we wish to build. This is a multiple testing problem under strong dependence and as previously mentioned this dependence will be the main difficulty from a mathematical point of view. In order to get strong theoretical results, the new results of this thesis hold in the Gaussian setting, which is bolstered by a rich literature. So, this manuscript deals with the problem of FDR and FWER control with particularly structured dependencies in order to analyze the fMRI data.

We conclude this introduction with the organization of this thesis.

Organization of this manuscript

Chapter 2 briefly introduces graphical models. It is hardly exhaustive but sufficient to understand the approach considered in this thesis: the use of multiple testing to estimate the cerebral connectivity when correlation is used as connectivity measure between two cerebral regions. Further, we propose a review of existing methods of graph inference in order to situate the multiple testing with respect to the contemporary literature (relatively to our problem).

Chapter 3 provides a rigorous formulation of the multiple testing research field. In particular, the two criteria (FWER and FDR) studied in this thesis are defined as well as some structured dependencies considered in the literature and connected with this manuscript. We introduce some FWER controlling procedures and the procedure of Benjamini and Hochberg (BH). We report some existing results for FDR control of the BH procedure. We also propose a new bound for the latter procedure that holds under a weaker positive dependence assumption than those known to obtain an FDR control.

Chapter 4 deals with the FDR control of the BH procedure with particular structured dependencies when the test statistics are normally distributed. In particular, we justify theoretically, for the first time to our knowledge, some results that significantly improve the bound of Benjamini and Yekutieli under non necessary positive dependence structures. Within this framework, our study seems to indicate that the FDR control of the BH procedure holds whatever the covariance matrix is.

CHAPTER 1. INTRODUCTION (ENGLISH)

Chapter 5 is devoted to the study of the type I errors control when the test statistics have an asymptotic Gaussian distribution. We review some multiple testing procedure known to provide an asymptotic FWER control and we clarify some proofs (it will be clearly indicate in such a case). We also provide some theoretical results for the asymptotic FDR control that are not yet sufficient for testing correlations.

Chapter 6 is dedicated to the numerical properties of the methods presented in Chapter 5 through the analysis of both simulated and real data.

Chapter 2

Graphical models

This chapter briefly introduces the graphical models in order to completely justify theoretically the considered modelling. We also establish the link between graph inference and multiple testing and we propose a multiple testing formulation of our problem.

Contents

2.1 Introduction	47
2.2 Well-posed problems	49
2.2.1 Preliminaries	49
2.2.2 Concentration graph	50
2.2.3 Covariance graph	53
2.3 How to infer the concentration or covariance graph?	54
2.3.1 A brief review	54
2.3.2 Multiple testing approaches	55

2.1 Introduction

The way towards an understanding of several complex phenomena (such as e.g. identification of functional relationships between genes, reconstruction of brain connectivity networks, or weather prediction) relies both on modelling (statistical analysis) and on experimentation (data collection). It is thus crucial that everyone involved in these different tasks, in order to communicate efficiently with each other, share a language they all understand. For the person who drives the experiments (e.g., in a medical environment), an appropriate model should be simple enough to be made sense of, as well as being able to outline the complexity of the observed phenomenon. For a statistician, an appropriate

CHAPTER 2. GRAPHICAL MODELS

model should be simple enough to study, while at the same time being as general and realistic as possible.

Graphical models stand as the perfect answer to these objectives. On the one hand, they are easily deciphered due to their graphical representation. On the other hand, they are powerful in depicting a complex reality, and they provide a simple statistical base (even more so in the Gaussian case considered here), bolstered by a rich literature.

As we have previously mentioned (see Figure 1.10), this work is motivated by the estimation of cerebral connectivity networks. Roughly speaking, within this framework, we assume that the variables (cerebral regions) indices are vertices of an unknown graph that we wish to build. An edge between two vertices will indicate that the two corresponding cerebral regions are "dependent". The two measures (as measure of cerebral connectivity for an fMRI experiment) traditionally considered are either the correlation between two variables or the partial correlation (also called conditional correlation in Gaussian setting) between two variables, which is the correlation between these two variables while controlling the effect of all the other ones. These two measures seem to be equally considered in the literature without any neuroscientific motivation behind the choice of the measure. For instance, in [18] the correlation is considered whereas the partial correlation is used in [19]. Thus, throughout this manuscript we consider these two measures.

Setting and aim

Throughout this chapter, let us consider the following statistical model,

$$\left(\mathbb{R}^{d^{\otimes n}}, \mathcal{B}(\mathbb{R}^d)^{\otimes n}, \{\mathcal{N}_d(0, \Gamma)\}^{\otimes n} \right),$$

where Γ is unknown and $\Gamma_{ii} = 1$ for all $i = 1, \dots, d$. From a technical point of view we should assume that Γ is invertible. Let $X \sim \mathcal{N}_d(0, \Gamma)$ and $\mathbb{X}^{(n)} = (X^{(1)}, \dots, X^{(n)})$ be an n -sample of X .

Following [20], we define a graph \mathcal{G} as a pair $\mathcal{G} = (V, E)$, where V is a finite set of vertices and the set of edges E is a subset of the set $V \times V$. The connection between a graph \mathcal{G} and the statistical model is made by identifying the vertices V of the graph \mathcal{G} with the variables X_1, \dots, X_d . Thus, the edge set E induces dependence properties.

Our goal is to infer the graph of marginal or conditional dependencies between X_1, \dots, X_d from the n -sample $\mathbb{X}^{(n)}$.

Note that it is a very hard problem in general and not assuming the Gaussian hypothesis is a very hard challenge among others. For instance, the high dimensional problems (particularly when $d > n$ or d/n increases while being smaller than one), the sparsity assumptions

in real applications or the control of false positive and false negative edges included in the graph.

2.2 Well-posed problems

Keep in mind that all of the following results hold only in the Gaussian setting with invertible covariance. For more general results related to graphical models, we refer the interested reader to the book of Lauritzen [20].

While many types of graphical models exist, we shall focus on undirected graphical models (also called a random Markov field). Indeed, the problem of the extraction of "the" graph of marginal (or conditional) dependencies from a n -sample is an ill-posed problem for other types of graphical models (see Chapter 7 of [21] for instance).

2.2.1 Preliminaries

Let $\mathcal{G} = (V, E)$ be a graph. An undirected edge, denoted $\{i, j\} \in E$, is an edge with both (i, j) and $(j, i) \in E$. The graph \mathcal{G} is an undirected graph if all edges $\{i, j\} \in E$ are undirected. In this case, for all $i, j \in V$, we have

$$i \stackrel{\mathcal{G}}{\sim} j \Leftrightarrow \text{there is an undirected edge between } i \text{ and } j \text{ in } \mathcal{G}.$$

Figure 2.1 provides an example of such a graph.

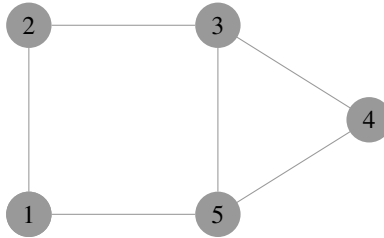


Figure 2.1 – An undirected graph $\mathcal{G} = (V, E)$ with $V = \{1, 2, 3, 4, 5\}$ and $E = \{\{1, 2\}, \{1, 5\}, \{2, 3\}, \{3, 4\}, \{3, 5\}, \{4, 5\}\}$.

Now, let us provide some basic definitions related to the graph theory that will be useful for the two following sections. Let $\mathcal{G} = (V, E)$ be an undirected graph. For all $i \in V$, the neighbors of i in the graph \mathcal{G} , denoted by $ne(i)$, are defined as

$$ne(i) := \{j : j \stackrel{\mathcal{G}}{\sim} i\} .,$$

The class of i , denoted by $cl(i)$, is defined as

$$cl(i) := ne(i) \cup \{i\} .$$

CHAPTER 2. GRAPHICAL MODELS

And the non neighbors of i in the graph \mathcal{G} , denoted by $nn(i)$, are defined as

$$nn(i) := V \setminus cl(i).$$

For instance, in the graph of Figure 2.1, the neighbors of 5 are $ne(5) = \{1, 3, 4\}$, the class of 5 is $cl(5) = \{1, 3, 4, 5\}$ and the non neighbors of 5 are $nn(5) = \{1, \dots, 5\} \setminus \{1, 3, 4, 5\} = \{2\}$. An undirected graph $\mathcal{G}' = (V', E')$ is a subgraph of \mathcal{G} , denoted by $\mathcal{G}' \subset \mathcal{G}$, if $V' \subset V$ and $E' \subset E$. Thus, a subgraph of \mathcal{G} is entirely determined by a subset of vertices of \mathcal{G} , there is no additional edge.

Armed with these preliminary definitions, we are now in position to formally introduce our problem of graph inference. As mentioned, we aim at inferring two types of graph: the graph of conditional dependencies and the graph of marginal dependencies between X_1, \dots, X_d . In the next two sections, following [22], we introduce different types of Markov properties for each type of graphs. Moreover we prove that the two problems are well-posed and amount to estimate respectively the locations of the nonzero entries of the precision matrix (also called concentration matrix) Γ^{-1} and the locations of the nonzero entries of the covariance matrix Γ .

2.2.2 Concentration graph

First, we focus on the conditional dependencies between X_1, \dots, X_d . Let \mathcal{G} be an undirected graph. The random vector X satisfies the Markov local property with respect to \mathcal{G} if for all $i \in \{1, \dots, d\}$,

$$X_i \perp\!\!\!\perp X_{nn(i)} \mid X_{ne(i)}. \quad (2.1)$$

In that case, X is a Gaussian graphical model with respect to \mathcal{G} . This yields an equivalence between the property “the distribution of X is a Gaussian graphical model with respect to \mathcal{G} ” and the specification of some conditional independences with respect to \mathcal{G} . For instance, if X is a Gaussian graphical with respect to the empty graph (graph with no edge) then X_1, \dots, X_d are independent variables. Conversely, if X is a Gaussian vector then X is a Gaussian graphical model with respect to the complete graph (each pair of vertices is linked by an edge). Moreover if \mathcal{G}' is a subgraph of \mathcal{G} and X is a Gaussian graphical model with respect to \mathcal{G}' then X is a Gaussian graphical model with respect to \mathcal{G} . Thus there exists a minimal graph (for the inclusion) such that X is a Gaussian graphical model with respect to \mathcal{G} . In the following, we define this graph and we prove that it is unique.

The conditional independence relationships between X_1, \dots, X_d are encoded in the precision matrix $K = \Gamma^{-1}$. Let \mathcal{G}_K be an undirected graph such that for all $i, j \in \{1, \dots, d\}$,

$$i \overset{\mathcal{G}_K}{\sim} j \Leftrightarrow K_{ij} \neq 0. \quad (2.2)$$

\mathcal{G}_K is called concentration graph and it satisfies the following theorem.

Theorem 2.1. Let $\mathcal{L}(X) = \mathcal{N}_d(0, \Gamma)$. For graph \mathcal{G}_K defined by (2.2), we have

1. X is a Gaussian graphical model with respect to \mathcal{G}_K .
2. \mathcal{G}_K is the minimal graph (for the inclusion) with respect to $\mathcal{L}(X)$.

In order to prove Theorem 2.1, we recall some additional results.

Proposition 2.1. Let U, W, Z be random variables having a density. Assume that all conditional densities exist, are well-defined and strictly positive. Then,

$$U \perp\!\!\!\perp W \mid Z \Leftrightarrow f(u \mid z) = f(u \mid z, w).$$

Proof.

$$\begin{aligned} U \perp\!\!\!\perp W \mid Z &\Leftrightarrow f(u, w \mid z) = f(u \mid z)f(w \mid z) \\ &\Leftrightarrow f(u \mid z) = \frac{f(u, w \mid z)}{f(w \mid z)} \\ &\Leftrightarrow f(u \mid z) = \frac{f(u, w, z)/f(z)}{f(w \mid z)} \\ &\Leftrightarrow f(u \mid z) = \frac{f(u \mid z, w)f(z, w)}{f(w \mid z)f(z)} = f(u \mid z, w). \end{aligned}$$

□

Proposition 2.2. Let U, V, W, Z be random variables having a density. Assume that all conditional densities exist, are well-defined and strictly positive. Then,

$$U \perp\!\!\!\perp (V, W) \mid Z \Rightarrow U \perp\!\!\!\perp V \mid (Z, W).$$

Proof. Assume that,

$$U \perp\!\!\!\perp (V, W) \mid Z. \tag{2.3}$$

Then,

$$\begin{aligned} f(u, v, z, w) &= f(u, v, w \mid z)f(z) \\ &= f(u \mid z)f(v, w \mid z)f(z) && \text{by using (2.3)} \\ &= f(u \mid z, w)f(v, w, z) && \text{by using Proposition 2.1.} \end{aligned}$$

□

Lemma 2.1. Let $I = \{1, \dots, k\}$ and $J = \{1, \dots, d\} \setminus I$. Let $Y = \begin{pmatrix} Y_I \\ Y_J \end{pmatrix} \sim \mathcal{N}_d(0, \Gamma)$. Assume that the precision matrix can be written as follows,

$$K = \begin{pmatrix} K_{II} & K_{IJ} \\ K_{JI} & K_{JJ} \end{pmatrix}$$

where $K_{II} \in \mathcal{M}_k(\mathbb{R})$, $K_{JJ} \in \mathcal{M}_{d-k}(\mathbb{R})$, $K_{IJ} \in \mathcal{M}_{k \times (d-k)}(\mathbb{R})$ and $K_{JI} \in \mathcal{M}_{(d-k) \times k}(\mathbb{R})$. Then $\mathcal{L}(Y_I \mid Y_J) = \mathcal{N}_k(-K_{II}^{-1}K_{IJ}Y_J, K_{II}^{-1})$.

CHAPTER 2. GRAPHICAL MODELS

Proof. Let h be the conditional density of Y_I with respect to Y_J . We have:

$$h(y_I|y_J) \propto \frac{1}{(2\pi)^{k/2}} \exp \left\{ -\frac{1}{2} y_I^T K_{II} y_I - y_I^T K_{IJ} y_J - \frac{1}{2} y_J^T (K_{JJ} - \Gamma_{JJ}^{-1}) y_J \right\}, \quad (2.4)$$

where Γ_{JJ} is the covariance matrix of Y_J and “ $y \propto z$ ” means that y is proportional to z . Using the block matrix inversion formula:

$$\begin{pmatrix} A & B \\ C & D \end{pmatrix}^{-1} = \begin{pmatrix} E^{-1} & -E^{-1}G \\ -FE^{-1} & D^{-1} + FE^{-1}G \end{pmatrix}, \text{ where } \begin{cases} E = A - BD^{-1}C \\ F = D^{-1}C \\ G = BD^{-1} \end{cases},$$

we obtain $\Gamma_{JJ}^{-1} = K_{JJ} - K_{JI}K_{II}^{-1}K_{IJ}$. The lemma follows by using this result in (2.4). \square

We are now in position to prove Theorem 2.1.

Proof. (Theorem 2.1)

1. Let $i \in \{1, \dots, d\}$, $I = \{i\} \cup nn(i)$ and $J = ne(i)$. The restriction of Γ^{-1} to I is given by $K_{II} = \begin{pmatrix} K_{ii} & 0 \\ 0 & K_{nn(i)nn(i)} \end{pmatrix}$. We deduce from Lemma 2.1 that the conditional density of X_I given X_J is Gaussian with covariance matrix $K_{II}^{-1} = \begin{pmatrix} K_{ii}^{-1} & 0 \\ 0 & K_{nn(i)nn(i)}^{-1} \end{pmatrix}$. Since the equivalence between independence and null covariance holds in Gaussian setting, we obtain that $\mathcal{L}(X) \sim \mathcal{G}_K$.
2. Let \mathcal{G} be an undirected graph such that $\mathcal{L}(X) \sim \mathcal{G}$. We want to prove that \mathcal{G}_K is minimal *i.e.* $\mathcal{G}_K \subseteq \mathcal{G}$. Let $(i,j) \in \{1, \dots, d\}^2 = V^2$ be such that $i \stackrel{\mathcal{G}}{\not\sim} j$. The previous result is achieved by proving the following contrapositive statement: if $i \stackrel{\mathcal{G}}{\sim} j$ then $K_{ij} = 0$. $\mathcal{L}(X) \sim \mathcal{G}$ means that $X_i \perp\!\!\!\perp X_{nn(i)} \mid X_{ne(i)}$ by using the definition. Since $i \stackrel{\mathcal{G}}{\sim} j$ then $j \in nn(i)$. From Proposition 2.2 it follows that,

$$X_i \perp\!\!\!\perp X_{nn(i)} \mid X_{ne(i)}, X_{nn(i)\setminus\{j\}}.$$

Since $ne(i) \cup (nn(i)\setminus\{j\}) = V \setminus \{i, j\}$, then $X_i \perp\!\!\!\perp X_{nn(i)} \mid X_{V \setminus \{i, j\}}$. Thus $X_i \perp\!\!\!\perp X_j \mid X_{V \setminus \{i, j\}}$ that is $K_{ij} \propto \text{cov}(X_i, X_j \mid X_{V \setminus \{i, j\}})$.

\square

Thus, the minimal graph \mathcal{G}_K of a standard non degenerated Gaussian vector is unique and entirely determined by the zeros of the precision matrix. Thus, the inference of the graph of conditional dependencies between X_1, \dots, X_d is a well-posed problem and consists of estimating the locations of the nonzero entries of K . With other words, it consists of inferring the concentration graph.

2.2.3 Covariance graph

Now, we focus on the marginal dependencies between X_1, \dots, X_d . Let \mathcal{G} be an undirected graph. The random vector X satisfies the marginal Markov local property with respect to \mathcal{G} if for all $i \in \{1, \dots, d\}$,

$$X_i \perp\!\!\!\perp X_{m(i)}. \quad (2.5)$$

In that case, X is called a covariance graph model [23] with respect to \mathcal{G} and the marginal dependence relationships between X_1, \dots, X_d are encoded in the covariance matrix Γ . Let \mathcal{G}_Γ be an undirected graph such that for all $i, j \in \{1, \dots, d\}$,

$$i \stackrel{\mathcal{G}_\Gamma}{\sim} j \Leftrightarrow \Gamma_{ij} \neq 0. \quad (2.6)$$

\mathcal{G}_Γ is called covariance graph. Since the equivalence between independence and null covariance holds in Gaussian setting, a similar result than Theorem 2.1 can be stated.

Theorem 2.2. *Let $X \sim N_d(0, \Gamma)$. For graph \mathcal{G}_Γ defined by (2.6), we have*

1. X is a covariance graph model with respect to \mathcal{G}_Γ .
2. \mathcal{G}_Γ is the minimal graph (for the inclusion) with respect to $\mathcal{L}(X)$.

Thus, the inference of the graph of marginal dependencies between X_1, \dots, X_d is a well-posed problem and consists of estimating the locations of the nonzero entries of Γ . With other words, it consists of inferring the covariance graph.

We end this section with an illustration in Figure 2.2 of the link between a graph and the zeros in the matrix that encodes the considered dependencies.



Figure 2.2 – Link between an undirected graph and the zeros in the matrix that encodes the considered dependencies.

If the graph of Figure 2.2 represents a covariance graph, we can for instance read the following relationships: $X_1 \perp\!\!\!\perp X_3$ and $X_5 \perp\!\!\!\perp X_2$. If this graph represents a concentration graph, X_1 is no longer independent of X_3 but we have $X_1 \perp\!\!\!\perp X_3 \mid \{X_2, X_5\}$ and X_5 is no longer independent of X_2 but $X_5 \perp\!\!\!\perp X_2 \mid \{X_1, X_3, X_4\}$.

We have thus checked that our problems are well-posed but actually how to solve them? This is the topic of the two following sections.

2.3 How to infer the concentration or covariance graph?

2.3.1 A brief review

Identify the zeros of the covariance or precision matrices is called covariance selection, as stated first in [24]. In this regards, from our point of view, three main approaches are used at the moment: *Graphical Lasso*, Bayesian graphical models and multiple testing for structured dependencies.

Driven by a wide range of contemporary applications, as seen in Chapter 1, statisticians must analyze high dimensional data. Unfortunately, all the well-known results (in the case of Gaussian variables) based on estimators obtained by maximum likelihood (see for example, [25]) are not true anymore when $n < d$. Indeed, the maximum likelihood estimator of the covariance matrix becomes singular and is no more invertible. Otherwise, in [26], the author shows that when $\frac{d}{n}$ increases while being smaller than one, the maximum likelihood provides a non consistent estimate, because its eigenvalues do not converge to those of the covariance matrix. This comes from theory of random matrices. Several methods have tried to reduce the number of parameters through assumption of sparsity for example or structured matrices. Note that, as mentioned in [27], since the covariance matrix and precision matrix encode different relationships, it is not possible to get at the same time sparse covariance and precision matrix, except for very particular cases ([28],[29]).

In graphical models, we are only interested in locating the zeros of the matrices (precision or covariance), so model selection may be more appropriate and less demanding. In the following of this section, we will present different methods for covariance selection. All of these methods have been developed under various structural assumptions to guarantee an accurate estimation of (covariance or precision) high-dimensional matrices.

Methods based on empirical covariance matrix and its inverse. Usually no sparsity assumption is made in these work. These include the banding methods in [30] and [31], thresholding rules in [32] for instance, shrinkage estimation in [33] and convex optimization techniques in [34]. These methods generally provide good performances for the covariance matrices, but it is hardly preserved for the precision matrices because of the inversion of the matrices.

Methods based on penalized likelihood estimation. This problem is formalized as a log-likelihood maximization with penalization (see for example [35] or [36]). This problem is difficult both in terms of finding a fast algorithm to solve the problem and in terms of deriving the statistical properties of the estimator.

The now classical method called *Graphical Lasso (Glasso)* [37] is proposing a very fast and simple algorithm to solve the log-likelihood maximization with sparsity constrained via adding a l_1 penalization term. The statistical properties of the estimator has been studied in detail in [38]. Several problems of this approach has been reported in [39]. First,

2.3. HOW TO INFER THE CONCENTRATION OR COVARIANCE GRAPH?

it is possible to construct an example where the estimator is not consistent [40]. Two equivalent conditions are needed for the *Glasso* estimators to be consistent: neighborhood stability conditions [41] or irrepresentability condition [38]. Alternative methods to the problem of neighborhood stability conditions have been proposed. These include for example scaled Lasso in [42], adaptive Lasso and non convex SCAD penalties in [43], non convex truncated Lasso in [44], additional nuclear norm penalty in [45], k-roots in [46] and cluster Lasso in [47]. Another problem induced by the *Glasso* approaches is that all the values are shrunk, not only the zeros [46].

Notice that similar methods using Lasso approaches have been developed to work with the covariance matrix [48].

Non likelihood based methods. Besides the previous methods based on a penalized log-likelihood method, several approaches have been proposed using different strategies, either close to the covariance selection of [24] or using empirical risks minimization [27]. These include neighborhood selection in [41], Dantzig selector in [49] or [50], CLIME in [51], scaled Lasso in [52] or l_1 penalised D -trace loss minimisation in [53].

Usually, in all the approaches above, the estimation of the covariance structure and the model selection occur at the same time and no control of false error edges identification is conducted. A new way that allows a control of false error edges identification relies on multiple testing procedures. The next section focuses on this approach.

2.3.2 Multiple testing approaches

There exists two types of hypothesis testing problems for covariance structures: global testing for the overall pattern of the covariance structures and simultaneous testing of a large collection of hypotheses for the local covariance structures such as pairwise correlations (see [54] for a general survey). As already mentioned, this thesis focuses on the latter. Within this framework, in multiple testing approaches, the covariance selection is divided into two different stages: first we get an estimation of the considered matrix (precision or covariance matrix) and then model selection is achieved through a multiple test based on the entries of this matrix.

More formally, let $\mathcal{G} = (V, E)$ be an undirected graph. Assuming that X is a Gaussian (or covariance) graphical model with respect to \mathcal{G} (unknown), we aim at obtaining an estimator $\widehat{\mathcal{G}} = (V, \widehat{E})$ of the minimal graph of \mathcal{G} from an n -sample $\mathbb{X}^{(n)}$ of X , that is inferring the concentration (or covariance) graph from $\mathbb{X}^{(n)}$. Denote $\mathcal{H} = \{\{i, j\}, 1 \leq i < j \leq d\}$ the index set of all possible (undirected) edges in \mathcal{G} and $m = d(d - 1)/2$ its cardinality.

CHAPTER 2. GRAPHICAL MODELS

Infer the concentration graph by multiple testing. Assume that X is a Gaussian graphical model with respect to \mathcal{G} . Thus, by definition, X satisfies the Markov local property. In fact, it is easy to see that X also satisfies the so-called pairwise Markov property: for all $i \in V$, for all $j \in V$ such that $j \in nn(i)$,

$$X_i \perp\!\!\!\perp X_j \mid ne(i).$$

Moreover, the partial correlation coefficients can be expressed using the precision matrix as follows.

Corollary 2.1. For all $i, j \in \{1, \dots, d\}$,

$$\text{cor}(X_i, X_j \mid X_k : k \neq i, j) = \frac{-K_{ij}}{\sqrt{K_{ii}K_{jj}}}.$$

Proof. We apply Lemma 2.1 with $I = \{i, j\}$ and $J = I^c$. We have:

$$K_{II}^{-1} = \begin{pmatrix} K_{ii} & K_{ij} \\ K_{ij} & K_{jj} \end{pmatrix}^{-1} = \frac{1}{K_{ii}K_{jj} - K_{ij}^2} \begin{pmatrix} K_{jj} & -K_{ij} \\ -K_{ij} & K_{ii} \end{pmatrix}.$$

By using the definition of partial correlation, we obtain

$$\text{cor}(X_i, X_j \mid X_k : k \neq i, j) = \frac{-K_{ij}}{K_{ii}K_{jj} - K_{ij}^2} \frac{\sqrt{K_{ii}K_{jj} - K_{ij}^2}}{\sqrt{K_{jj}}} \frac{\sqrt{K_{ii}K_{jj} - K_{ij}^2}}{\sqrt{K_{ii}}} = \frac{-K_{ij}}{\sqrt{K_{ii}K_{jj}}}.$$

□

Then using Theorem 2.1 we have that, for all $(i, j) \in V^2$,

$$K_{ij} = 0 \Leftrightarrow X_i \perp\!\!\!\perp X_j \mid X_{V \setminus \{i, j\}} \Leftrightarrow \rho_{\{i, j\}, V \setminus \{i, j\}} = 0,$$

where $\rho_{\{i, j\}, V \setminus \{i, j\}} = -K_{ij} / \sqrt{K_{ii}K_{jj}}$ is the partial correlation between X_i and X_j . Thus, inferring the concentration graph can be achieved through the following m tests:

$$H_{0,h} : \rho_{h, V \setminus \{h\}} = 0 \quad \text{against} \quad H_{1,h} : \rho_{h, V \setminus \{h\}} \neq 0, \quad h \in \mathcal{H}. \quad (2.7)$$

Infer the covariance graph by multiple testing. Assume that X is a covariance graph model. Thus, in particular, for all $i \in V$, for all $j \in V$ such that $j \in nn(i)$, we have

$$X_i \perp\!\!\!\perp X_j.$$

Moreover, for all $i, j \in V$,

$$\text{cor}(X_i, X_j) = \frac{\Gamma_{ij}}{\sqrt{\Gamma_{ii}\Gamma_{jj}}}.$$

2.3. HOW TO INFER THE CONCENTRATION OR COVARIANCE GRAPH?

Then using Theorem 2.2 we have that, for all $(i, j) \in V^2$,

$$\Gamma_{ij} = 0 \Leftrightarrow X_i \perp\!\!\!\perp X_j \Leftrightarrow \rho_{\{i,j\}} = 0,$$

where $\rho_{\{i,j\}} = \Gamma_{ij} / \sqrt{\Gamma_{ii}\Gamma_{jj}}$ is the (marginal) correlation between X_i and X_j . Thus, inferring the covariance graph can be achieved through the following m tests:

$$H_{0,h} : \rho_h = 0 \quad \text{against} \quad H_{1,h} : \rho_h \neq 0, \quad h \in \mathcal{H}. \quad (2.8)$$

One advantage of the multiple testing approaches is that the incorrect edges included in \widehat{G} can be controlled. We can state an equivalent formulation (in terms of graphs) of the control of the traditional errors considered (FWER and FDR) in multiple testing framework. Indeed, there is an edge between i and j in \widehat{G} if and only if the corresponding null hypothesis $H_{0,h}$ is rejected. Since for all $h \in \mathcal{H}$, $h \in E$ if $H_{0,h}$ is false and $H_{1,h}$ is true, another way of stating FWER control is:

$$\begin{aligned} & \mathbb{P}(\text{there exists } h \in \mathcal{H} \text{ such that } H_{0,h} \text{ is rejected and true}) \leq \alpha \\ \Leftrightarrow & \mathbb{P}(\text{there exists } h \in \mathcal{H} \text{ such that } h \text{ in } \widehat{E} \text{ and not in } E) \leq \alpha \\ \Leftrightarrow & \mathbb{P}(\text{there exists } i, j \text{ such that there exists an edge between } i \text{ and } j \text{ in } \widehat{G} \text{ and not in } G) \leq \alpha \\ \Leftrightarrow & \mathbb{P}(\widehat{G} \not\subseteq G) \leq \alpha \\ \Leftrightarrow & \mathbb{P}(\widehat{G} \subset G) \geq 1 - \alpha. \end{aligned}$$

In other words, a FWER controlling procedure identifies the set of the edges that are not included in G with probability $1 - \alpha$. The latter infers a subgraph of G rather than a graph with more edges. We can analogously state the asymptotic FWER control in terms of graphs. A multiple testing procedure provides an asymptotic control of the FWER at level α if:

$$\limsup_{n \rightarrow +\infty} \mathbb{P}(\widehat{G} \subseteq G) \geq 1 - \alpha.$$

For instance, as we will see in details in Chapter 5, Drton and Perlman [55] proposed an asymptotic FWER controlling procedure for the estimation of a Gaussian graphical model. Their method is valid for $m < n$ but for $m > n$, the sample partial correlation coefficients are not well defined.

We can also state a formulation of the FDR control in terms of graphs. Controlling the FDR at level α allows us to select a graph \widehat{G} such that the proportion of incorrect edges among all the edges of \widehat{G} is smaller than or equal to α in expectation, that is,

$$\mathbb{E} \left[\frac{\text{number of edges incorrectly included in } \widehat{G}}{\text{number of edges included in } \widehat{G}} \right] \leq \alpha.$$

CHAPTER 2. GRAPHICAL MODELS

In addition, a multiple testing procedure provides an asymptotic control of the FDR at level α if:

$$\limsup_{n \rightarrow +\infty} \mathbb{E} \left[\frac{\text{number of edges incorrectly included in } \widehat{G}}{\text{number of edges included in } \widehat{G}} \right] \leq \alpha.$$

For instance, Cai and Liu [56] establish a large-scale multiple testing procedure for testing correlations that controls the FDR at the price of sparsity assumptions. Their assumptions are related to the signal and weaker than usually. Roughly speaking, their method requires "small" correlation coefficients rather than null coefficients. Their results are proved when the number of tests tends to the infinity. Other procedures have been recently proposed, for example in [57], [58] or [59].

Advantages and drawbacks of graph inference procedures of the two previous sections are summed up by Table 2.1. This table is a brief table and roughly aggregates a lot of methods. For each aggregation, we give possible advantage(s) or disadvantage(s) of one of these methods, without exhaustiveness.

Covariance selection		
Methods	- -	++
Covariance matrix and take the inverse (shrinkage, thresholding, convex optimization ...)	- bad performances for the precision matrices - no control of false edges identification	- no sparsity assumption - good performances for the covariance matrices
penalization added to likelihood estimation (<i>Glasso</i> , scaled Lasso, k-roots, cluster Lasso, non convex truncated Lasso ...)	- the estimator is not always consistent - Shrink all the values not only the zeros - no control of false edges identification	computational cost
Non likelihood estimation (neighborhood selection, Dantzig selector, CLIME, scaled Lasso ...)	- computational cost - no control of false edges identification	no Gaussian assumption
Multiple testing	technically challenging for dependent tests	- control of false error edges identification - take into account some previous knowledge about the graph

Table 2.1 – Advantages and drawbacks of graph inference procedures defined in Section 2.3.1 and Section 2.3.2.

2.3. HOW TO INFER THE CONCENTRATION OR COVARIANCE GRAPH?

So, this chapter has theoretically justified the considered modelling and introduced a multiple testing formulation of our problem via the problem of testing correlation (tests (2.8)) or partial correlation (tests (2.7)). Then, the natural following steps are: the construction of test statistics and the construction of a multiple testing procedure.

In what follows, for the sake of clarity, the index of edges is denoted by ij instead of $\{i, j\}$, for all $h \in \mathcal{H}$. Throughout this manuscript, we assume that $d < n$. In this case, the maximum likelihood estimator of correlation, denoted by $\widehat{\rho}_{n,\cdot} = (\widehat{\rho}_{n,h})_{h \in \mathcal{H}}$, is well-defined and is expressed as

$$\widehat{\rho}_{n,h} = \frac{\frac{1}{n} \sum_{l=1}^n X_i^{(l)} X_j^{(l)} - \bar{X}_i \bar{X}_j}{\widehat{\sigma}(X_i) \widehat{\sigma}(X_j)}, \quad (2.9)$$

where $\bar{X}_i = \frac{1}{n} \sum_{l=1}^n X_i^{(l)}$ and $\widehat{\sigma}(X_i) = \sqrt{\frac{1}{n} \sum_{l=1}^n (X_i^{(l)} - \bar{X}_i)^2}$. Then, the correlation test (2.8) can be based on the following test statistic:

$$T_{n,h}(X^{(n)}) = \sqrt{n} \widehat{\rho}_{n,h}.$$

Moreover, Aitkin [60] established the asymptotic normality of the vector of empirical correlations.

Proposition 2.3 (Aitkin [60]). *The vector of empirical correlations $\widehat{\rho}_{n,\cdot} = (\widehat{\rho}_{n,h})_{h \in \mathcal{H}}$ is asymptotically normal,*

$$\sqrt{n}(\widehat{\rho}_{n,\cdot} - \rho_{\cdot}) \xrightarrow[n \rightarrow +\infty]{\mathcal{L}} \mathcal{N}_m(0, \Omega(\Gamma)),$$

with

$$\Omega(\Gamma) = (\omega_{ij,kl})_{(ij,kl) \in \mathcal{H}^2} \text{ given by} \quad (2.10)$$

$$\omega_{ij,ij} = (1 - \rho_{ij}^2)^2,$$

$$\omega_{ij,il} = -\frac{1}{2} \rho_{ij} \rho_{il} (1 - \rho_{ij}^2 - \rho_{il}^2 - \rho_{jl}^2) + \rho_{jl} (1 - \rho_{ij}^2 - \rho_{il}^2),$$

$$\omega_{ij,kl} = \frac{1}{2} \rho_{ij} \rho_{kl} (\rho_{ik}^2 + \rho_{il}^2 + \rho_{jk}^2 + \rho_{jl}^2) + \rho_{ik} \rho_{jl} + \rho_{il} \rho_{jk} - \rho_{ik} \rho_{jk} \rho_{kl} - \rho_{ij} \rho_{ik} \rho_{il} - \rho_{ij} \rho_{jk} \rho_{jl} - \rho_{il} \rho_{jl} \rho_{kl}.$$

Similar result of Proposition 2.3 holds for the vector of the empirical partial correlations $\widehat{\rho}_{h,V \setminus \{h\}}$ if we replace all ρ_{ij} by $\rho_{ij,V \setminus \{ij\}}$ in the equation (2.10), where the sample size n should also be replaced by $n_d = n - d + 1$ [25].

Consequently, the multiple testing problems (2.8) and (2.7) can be viewed as multiple testing problems for means when the covariance matrix is unknown in the asymptotic Gaussian setting. In this regards, we need to construct a relevant (in terms of error control) multiple testing procedure. As a first step, Chapter 4 investigates the FDR of the BH procedure

CHAPTER 2. GRAPHICAL MODELS

when the covariance is known in Gaussian setting. Then, Chapter 5 is devoted to the construction of multiple testing procedures that (asymptotically) control the FWER or FDR in asymptotic Gaussian setting. Before moving to the problem of construction of a multiple testing procedure, we should formally introduce the multiple testing research field. This is the topic of the next chapter.

We end up with a summary of the organization of this manuscript.

2.3. HOW TO INFER THE CONCENTRATION OR COVARIANCE GRAPH?

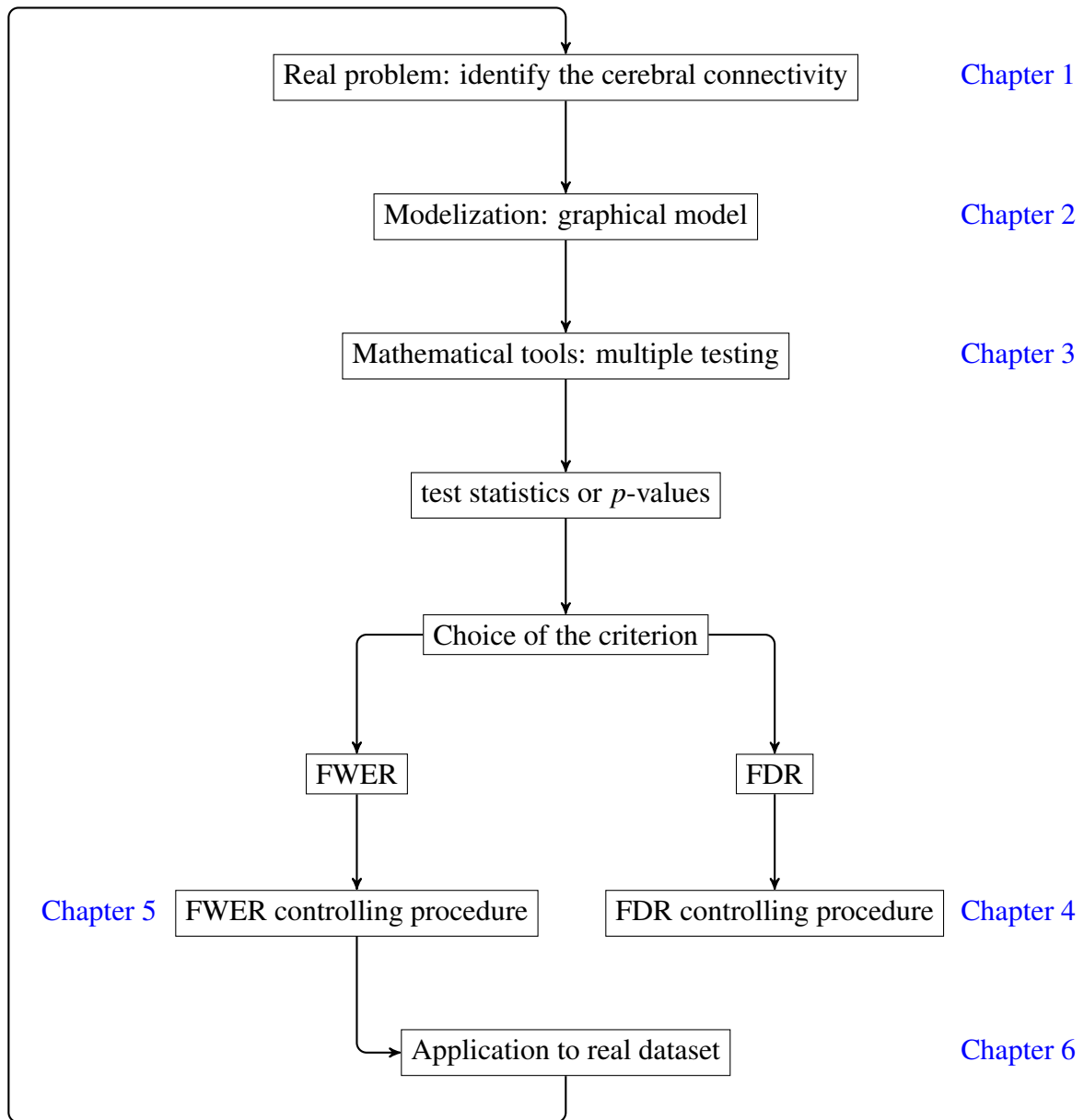


Figure 2.3 – Organization of the manuscript.

Chapter 3

Multiple testing framework

This chapter presents the mathematical background (which will be used throughout this manuscript) of multiple testing theory. We formally introduce the two main criteria of quality considered in the literature, that is the FWER and the FDR. Moreover, several assumptions concerning the dependency structure are discussed, for which we report some results about FWER control and FDR control. Our contributions in this chapter are: some corrections of proofs relative to the least favorable configurations for FDR control of the Benjamini-Hochberg (BH) procedure, and we propose a new upper-bound for the FDR of BH procedure valid under weaker positive dependence assumption than positive properties traditionally considered in the literature.

Contents

3.1	Setting	64
3.1.1	Model	64
3.1.2	Criteria and decisions	65
3.2	Dependence assumptions	66
3.2.1	MTP ₂ : multivariate total positivity of order 2	66
3.2.2	PRD: positive regression dependence	67
3.2.3	PA: positive association	68
3.2.4	PLOD: positive lower orthant dependence	69
3.3	Classical FWER control	71
3.4	Classical FDR control	72
3.4.1	The BH procedure	73
3.4.2	Bounds of FDR of BH procedure: known results	73
3.5	A new result: relaxation of the PRDS condition	79

3.1 Setting

3.1.1 Model

Let $(\mathfrak{X}, \mathcal{X}, \mathcal{P})$ be a statistical model. We observe X generated from $P \in \mathcal{P}$. For all fixed $m \geq 2$, let $(\mathcal{M}_{0,i})_{1 \leq i \leq m}$ be a sequence of subsets of \mathcal{P} . We consider the problem of testing the null hypotheses $H_{0,i} : "P \in \mathcal{M}_{0,i}"$ against the alternative hypotheses $H_{1,i} : "P \in \mathcal{M}_{0,i}^c"$ simultaneously for all $i \in \{1, \dots, m\}$. For all $P \in \mathcal{P}$, we denote by $\mathcal{H}_0(P) = \{i \in \{1, \dots, m\}, P \in \mathcal{M}_{0,i}\}$ the set of indexes i for which P satisfies $H_{0,i}$, that is the index set of the true null hypotheses. We also denote by $m_0(P) = |\mathcal{H}_0(P)|$ its cardinality. Similarly, for all $P \in \mathcal{P}$, let $\mathcal{H}_1(P) = \{1, \dots, m\} \setminus \mathcal{H}_0(P)$ be the set of indexes i for which P does not satisfy $H_{0,i}$ (the false null hypothesis) and we denote by $m_1(P) = m - m_0(P)$ its cardinality. We aim at identifying the true/false null hypotheses based on the knowledge of X . In this regards, a multiple testing procedure relies on the sequence of p -values $\mathbf{p}(X) = (p_i(X))_{1 \leq i \leq m} \in [0, 1]^m$ resulting from each m individual tests. From now on, we write \mathbf{p} instead of $\mathbf{p}(X)$ for short (even if it actually depends on the observations). Throughout this thesis, assume that each p -value corresponding to a true null hypothesis must be stochastically lower-bounded by a uniform variable, namely

$$\forall P \in \mathcal{P}, \forall i \in \mathcal{H}_0(P), \forall u \in [0, 1], \mathbb{P}(p_i \leq u) \leq u. \quad (3.1)$$

This assumption is usually used in the literature and with words, it can be seen as a generalization of the construction of a single test to the case of multiple null hypotheses. Indeed, the event "the null hypothesis is rejected at level α " occurs with probability lower than or equal to α for a single hypothesis test at level α .

The probably most common illustration of multiple testing is given by the Gaussian statistical model for which we aim at identifying the non null means. Let X be a Gaussian vector with mean $\mu \in \mathbb{R}^m$ and covariance matrix Σ , which is assumed to satisfy $\Sigma_{ii} = 1$. For all $i \in \{1, \dots, m\}$, two multiple testing problems are traditionally considered:

- one-sided case: $H_{0,i} : "\mu_i \leq 0"$ against $H_{1,i} : "\mu_i > 0"$, $1 \leq i \leq m$,
- two-sided case: $H_{0,i} : "\mu_i = 0"$ against $H_{1,i} : "\mu_i \neq 0"$, $1 \leq i \leq m$.

In the one-sided case (resp. two-sided), the associated p -values are defined as $p_i = \overline{\Phi}(X_i)$ (resp. $p_i = 2\overline{\Phi}(|X_i|)$), $1 \leq i \leq m$. We can merely check that these p -values satisfy the property (3.1).

Formally speaking, a multiple testing procedure is a measurable function of the observations $R : [0, 1]^m \rightarrow \mathcal{P}(\{1, \dots, m\})$, which returns a set $R(\mathbf{p}) \equiv \mathcal{R}$ corresponding to the index set of null hypotheses that are rejected. This set \mathcal{R} is given by

$$\mathcal{R} = \{1 \leq i \leq m : p_i(X) \leq \widehat{t}\},$$

where $\widehat{t} \in [0, 1]$ is a random variable that possibly depends on the family of the p -values.

3.1.2 Criteria and decisions

Accounting for the multiplicity of individual tests can be achieved by controlling an appropriate error rate. We present here the two main criteria considered in the literature.

Family-Wise Error Rate (FWER)

The notion of type I error can be naturally extended to the multiple testing framework by the Family-Wise Error Rate (FWER). The FWER depends on the rejected set \mathcal{R} and on the (unknown) distribution P of the observations. The FWER corresponds to the probability of rejecting at least one true null hypothesis, namely

$$\forall P \in \mathcal{P}, \text{FWER}(\mathcal{R}, P) = \mathbb{P}(|\mathcal{R} \cap \mathcal{H}_0(P)| \geq 1). \quad (3.2)$$

Section 3.3 provides some examples of FWER controlling procedures.

False Discovery Rate (FDR)

By denoting Q the proportion of false discoveries, for all $P \in \mathcal{P}$, the FDR is defined as follows

$$\text{FDR}(\mathcal{R}, P) = \mathbb{E}[Q], \text{ where } Q = \frac{|\mathcal{R} \cap \mathcal{H}_0(P)|}{|\mathcal{R}| \vee 1}. \quad (3.3)$$

" $|\mathcal{R}| \vee 1$ " is the maximum between $|\mathcal{R}|$ and 1 in order to guarantee the well-definition of Q when no hypothesis is rejected ($|\mathcal{R}| = 0$). Throughout this manuscript when the set \mathcal{R} will be identify without ambiguity, \mathcal{R} will be dropped in the notation for short and we will write $\text{FDR}(P)$ instead of $\text{FDR}(\mathcal{R}, P)$. In the same way, if P depends on only one parameter θ , we will write $\text{FDR}(\theta)$ and m_0 (resp. m_1) instead of $m_0(\theta)$ (resp. $m_1(\theta)$).

Despite there is a philosophical difference between FWER control and FDR control, these two criteria are not disconnected from each other. Since $0 \leq Q \leq 1$,

$$\forall P \in \mathcal{P}, \text{FDR}(\mathcal{R}, P) = \mathbb{E}[Q] \leq \mathbb{P}(Q > 0) = \text{FWER}(\mathcal{R}, P).$$

Therefore a FWER controlling procedure is also a FDR controlling procedure. This obviously does not hold conversely. Since the FDR is upper bounded by the FWER, FDR control is easier to guarantee than FWER control (at the same level). Thus the FDR controlling procedures reject generally more hypotheses than the FWER controlling procedures. Otherwise note that if all null hypotheses are true, Q is a binomial random variable and FDR and FWER are equivalent criteria.

3.2 Dependence assumptions

The statistical study of the FDR is very challenging from a mathematical point of view because the cardinality of the rejected set \mathcal{R} is random. To get rigorous controlling results, given a multiple testing procedure, some assumptions (rarely corresponding to realistic situations) relying on the distribution P of X , are traditionally considered. The strongest but very convenient in practice assumption is the independence, namely

$(p_i(X), i \in \mathcal{H}_0(P))$ is a family of mutually independent variables,

and $(p_i(X), i \in \mathcal{H}_0(P))$ is independent of $(p_i(X), i \in \mathcal{H}_1(P))$.

Under this assumption, a lot of both FWER and FDR controlling procedures have been developed. When the independence assumption is no longer valid, the accurate controlling results proposed in the literature generally hold under positive dependence assumption. Here, we present some different assumptions of the general notion of positive dependence and we precise the relation between these assumptions. For a more complete review, we refer the interested reader to [61] for example, for bivariate distribution.

3.2.1 MTP_2 : multivariate total positivity of order 2

The notion of positive dependence, called Multivariate Total Positivity of order 2 (MTP_2), has been first introduced by T.K. Sarkar in 1969 [62]. If the joint density f of the p -values satisfies the following property

$$\forall y, z \in \mathbb{R}^m, f(y)f(z) \leq f(\min(y, z))f(\max(y, z)),$$

then the p -values are MTP_2 , where the minimum (resp. maximum) between y and z should be considered component-wise. Let us report some examples for which the distribution considered is MTP_2 .

Example 3.1. *All families of independent random variables have an MTP_2 joint density.*

Example 3.2. [62] *Let us consider the problem of simultaneously testing means in the one-sided Gaussian setting. For all invertible matrix Σ , if*

$$P \in \{\mathcal{N}_m(\mu, \Sigma) : \text{for all } i \neq j, -(\Sigma^{-1})_{ij} \geq 0\}$$

then the MTP_2 property is satisfied for the p -values.

Example 3.3. [63] *Let us consider the problem of simultaneously testing means in the two-sided Gaussian setting. For all invertible matrix Σ , for all diagonal matrix D such that $i \in \{1, \dots, m\}, D_{ii} \in \{-1, 1\}$, if*

$$P \in \{\mathcal{N}_m(0, \Sigma) : \text{for all } i \neq j, \exists D, (-D\Sigma^{-1}D)_{ij} \geq 0\}$$

then the MTP_2 property is satisfied for the p -values.

For instance, a special case of interest is the centered positive equi-correlated case,

$$P \in \{\mathcal{N}_m(0, \Sigma(\rho)), \rho \geq 0\},$$

where $\Sigma(\rho)$ is the matrix such that $\Sigma_{ii}(\rho) = 1$ and $\Sigma_{ij}(\rho) = \rho$ for $i \neq j$. Indeed, $\Sigma^{-1}(\rho)$ can be written as $cM \in \mathcal{M}(\mathbb{R}^m)$ where

$$\begin{cases} c = (1 - \rho)^{-1} \frac{1+(m-1)\rho-\rho}{1+(m-1)\rho} \\ M_{i,i} = 1 \\ M_{i,j} = -\frac{\rho}{1-\rho+(m-1)\rho} \text{ pour tout } i \neq j \end{cases} .$$

Thus, by choosing $D = I_m$, we can check that for all $i \neq j$, $(-D\Sigma^{-1}D)_{ij} \geq 0$.

3.2.2 PRD: positive regression dependence

The notion of positive dependence called Positive Regression Dependence (PRD) was first introduced by Lehmann [64] in the bivariate case. Let us begin by defining the notion of increasing set. An increasing set $D \in [0, 1]^m$ is characterized by the following property

For all $y, z \in [0, 1]^m$ such that for all $i \in \{1, \dots, m\}$, $y_i \leq z_i$ and $y \in D$, then $z \in D$.

\mathbf{p} is PRD if for all increasing set D , for all $\{p_1, \dots, p_i\} \subset \mathbf{p}$, $\mathbb{P}(\mathbf{p} \in D \mid p_1 = v_1, \dots, p_i = v_i)$ is increasing in (v_1, \dots, v_i) .

By conditioning by only one variable and when the monotonicity condition is only relied on one subset of \mathbf{p} , a weaker notion of positive dependency can be defined, which is called Positive Regression Dependency on each one from a Subset (PRDS). The family \mathbf{p} is said PRDS on $S \in \{1, \dots, m\}$ if for all increasing set $D \in [0, 1]^m$, the function $u \mapsto \mathbb{P}(\mathbf{p} \in D \mid p_i = u)$ is nondecreasing for all $i \in S$. Thus, by definition, if \mathbf{p} is PRD then \mathbf{p} is PRDS on all subsets of $\{1, \dots, m\}$. In addition, if \mathbf{p} is a family of independent random variables then \mathbf{p} is PRDS on all subsets of $\{1, \dots, m\}$.

Otherwise, a weaker dependence property than PRDS, which is called weak PRDS, is usually considered. The only difference between these two types of dependence is that the weak PRDS property requires that the function $\mathbb{P}(\mathbf{p} \in D \mid p_i \leq u)$ (instead of $\mathbb{P}(\mathbf{p} \in D \mid p_i = u)$) is nondecreasing in u .

Proposition 3.1. *For all $P \in \mathcal{P}$, if $(p_i(X))_{1 \leq i \leq m}$ is PRDS on $\mathcal{H}_0(P)$ then $(p_i(X))_{1 \leq i \leq m}$ is also weak PDRS on $\mathcal{H}_0(P)$.*

A proof of this result is proposed in the proof of Proposition 3.6 of [65].

CHAPTER 3. MULTIPLE TESTING FRAMEWORK

Karlin and Rinott [66] showed that the MTP_2 property implies the PRDS property. Thus if \mathbf{p} is MTP_2 then \mathbf{p} is in particular PRDS on $\mathcal{H}_0(P)$ and thus also weak PRDS. Therefore the previous examples 3.1 and 3.2 provide also PRDS distributions. Note that even if the PRDS property is weaker than the MTP_2 property, the MTP_2 property is often easier to establish in practice.

Example 3.4. [67] *Let us consider the problem of simultaneously testing means in the one-sided Gaussian setting. If*

$$P \in \{\mathcal{N}_m(\mu, \Sigma) : \text{for all } i \in \mathcal{H}_0, \text{ and for all } j \neq i, \Sigma_{ij} \geq 0\},$$

then the p -values are PRDS on $\mathcal{H}_0(P)$.

We finally provide two counter-examples for which the PRDS property does not hold.

Counter-example 3.1. *Yekutieli [68] shows that the dependence induced by the absolute values of a Gaussian vector is not positive. In other words, for the problem of simultaneously testing means in the two-sided Gaussian setting, the tests statistic are not positively dependent. Let $X_1 \sim \mathcal{N}(0, 1)$ and $X_2 = X_1 + 2$. Then $(X_1, X_2) \sim \mathcal{N}_2\left(\begin{pmatrix} 0 \\ 2 \end{pmatrix}, \begin{pmatrix} 1 & \rho \\ \rho & 1 \end{pmatrix}\right)$ where $\rho = 1$ and*

$$\mathbb{P}(|X_2| > 1 \mid |X_1| = t) = \begin{cases} 1 & \text{for } t < 1 \text{ or } t > 3 \\ 1/2 & \text{for } 1 \leq t \leq 3 \end{cases}.$$

Counter-example 3.2. [68] *For all $i \neq j$ let $X_{i,j} = (Z_i - Z_j)/\sqrt{2}$ where Z_1, \dots, Z_m are independent and identically distributed (i.i.d) from $\mathcal{N}(0, 1)$. Then $\{|X_{3,2}|, |X_{3,1}|, |X_{2,1}|\}$ is not PRDS. Indeed, let $D = \{(X_{3,2}, X_{3,1}) : |X_{3,2}| > 1, |X_{3,1}| > 1\}$ be a nondecreasing set in $|X_{3,2}|, |X_{3,1}|$. We can merely check that $\mathcal{L}(X_{3,2} \mid X_{2,1} = t) = \mathcal{N}(-t/2, 3/4)$. Since $X_{3,1} = X_{2,1} + X_{3,2}$, for $0 < t < 2$, D can be written as $\{X_{3,2} > 1\} \cup \{X_{3,2} < -1 - t\}$ and it follows that*

$$\mathbb{P}(D \mid |X_{2,1}| = t) = 2\Phi\left(\frac{-1 - t/2}{\sqrt{3/4}}\right),$$

is decreasing in t .

3.2.3 PA: positive association

The notion of Positive Association (PA) was first introduced by Esary and al [69] in 1967. The random variables X_1, \dots, X_m are PA if for all nondecreasing functions f and g , the following property is satisfied,

$$\text{cov}(f(X_1, \dots, X_m), g(X_1, \dots, X_m)) \geq 0,$$

as soon as the quantities $\mathbb{E}[f(X)]$, $\mathbb{E}[g(X)]$ and $\mathbb{E}[f(X)g(X)]$ are well-defined.

Let us provide some natural properties, which are useful to understand the relation between the PA and the next type of dependence presented in this section.

Proposition 3.2. *All nondecreasing functions of positively associated random variables are positively associated.*

Proposition 3.3. *If X_1, \dots, X_m are binary random variables that are positively associated then $1 - X_1, \dots, 1 - X_m$ are binary random variables that also are positively associated.*

By using Proposition 3.3, we are able to prove the following theorem:

Theorem 3.1. *If X_1, \dots, X_m are binary random variables that are positively associated, then*

$$\mathbb{P}(X_1 = 1, \dots, X_m = 1) \geq \mathbb{P}(X_1 = 1) \dots \mathbb{P}(X_m = 1) \quad (3.4)$$

and

$$\mathbb{P}(X_1 = 0, \dots, X_m = 0) \geq \mathbb{P}(X_1 = 0) \dots \mathbb{P}(X_m = 0). \quad (3.5)$$

Proof. Let $\gamma(X) = X_1$ and $\delta(X) = X_2 \dots X_m$ be two nondecreasing functions of X . By using the definition of positive association, we obtain that $\mathbb{E}[X_1 \dots X_m] \geq \mathbb{E}[X_1] \mathbb{E}[X_2 \dots X_m]$. Repeating the application of the same argument enables us to write $\mathbb{E}[X_1 \dots X_m] \geq \mathbb{E}[X_1] \dots \mathbb{E}[X_m]$. Since X is a binary random variable, $\mathbb{E}[X] = \mathbb{P}(X = 1)$, (3.4) holds. Finally, (3.5) is obtained by using (3.4) and Proposition 3.3. \square

We finally provide an example of statistics positively associated.

Example 3.5. [67] *For the problem of simultaneously testing means in the one-sided Gaussian setting, if $P \in \{\mathcal{N}_m(\mu, \Sigma) : \text{for all } i \neq j, \Sigma_{ij} \geq 0\}$ then the p -values are PA.*

3.2.4 PLOD: positive lower orthant dependence

The type of positive dependence called Positive Lower Orthant Dependence (PLOD) is a generalization of positive quadrant dependence introduced by Lehmann [64] in the bivariate case to the multivariate case.

For all constant $b \in \mathbb{R}^m$, for all $Z \in \mathcal{R}^m$, the notation " $Z \leq b$ " should be considered component-wise. Z is said PLOD if for all constant $b \in \mathbb{R}^m$, the following inequality holds,

$$\mathbb{P}(Z \leq b) \geq \prod_{i=1}^m \mathbb{P}(Z_i \leq b_i). \quad (3.6)$$

Theorem 3.2. [70] *If X is PA then X is PLOD.*

CHAPTER 3. MULTIPLE TESTING FRAMEWORK

Proof. Let $b \in \mathbb{R}^m$. For all $i \in \{1, \dots, m\}$, let Y_i be the binary random variable defined as: $Y_i = 1$ if $X_i > b_i$, $Y_i = 0$ if $X_i \leq b_i$. By using Proposition 3.2, we have that Y_1, \dots, Y_m are positively associated and the result follows from Theorem 3.1. \square

Another remarkable example (since the result requires few assumptions) of PLOD statistics is given by the Šidák's inequality (1967) [71]:

Theorem 3.3. (*Šidák's inequality*) *If $X \sim \mathcal{N}_m(0, \Sigma)$ and Σ is invertible then $|X|$ is PLOD e.g. for all positive constant $b \in \mathbb{R}^m$, the inequality (3.6) is satisfied for $Z = |X|$.*

Note that Theorem 3.3 holds for all invertible Σ , the only "real" requirement is to be centered ($\mu = 0$, that is being under the full null in a multiple testing framework). This theorem especially still holds even if there exists some indexes i, j for which $\Sigma_{ij} < 0$. This result has been established a few years ago (1958) by Dunn [72] in the cases either $m = 2$ or $m = 3$ or by using additional assumptions on Σ .

The Šidák's inequality relies on a theorem of Anderson (Corollary 2 in [73]), which generalizes to the dimension m the following intuitive result:

Let $f : \mathbb{R} \rightarrow \mathbb{R}$ be a density function such that

1. For all $x \in \mathbb{R}$, $f(x) = f(-x)$,
2. For all $x \in \mathbb{R}$, for all $0 \leq k \leq 1$, $f(kx) \geq f(x)$.

Then for all $l \geq 0$, $\max \left\{ \int_I f(x) dx, I \subset \mathbb{R}, |I| = l \right\}$ is attained for $I = [-l/2, l/2]$ e.g. $\max \left\{ \mathbb{P}(X \in I), I \subset \mathbb{R}, |I| = l \right\}$ is attained for $I = [-l/2, l/2]$.

Figure 3.1 illustrates this result when f is the density function of the centered Gaussian distribution.

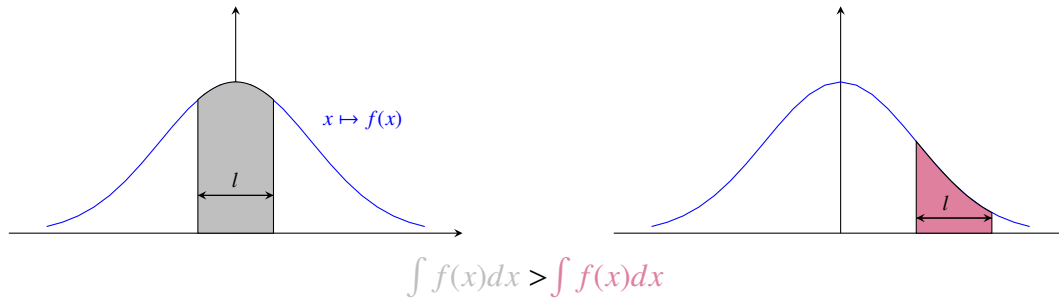


Figure 3.1 – Given an interval of fixed size l , $\int f(x) dx$ is maximal when this interval is centered at the origin when f is the density function of the centered Gaussian distribution.

The generalization of this result to the dimension m is stated as follows:

Theorem 3.4. (*Anderson*). *Let $X \in \mathbb{R}^m$ be a random vector having a density f satisfying*

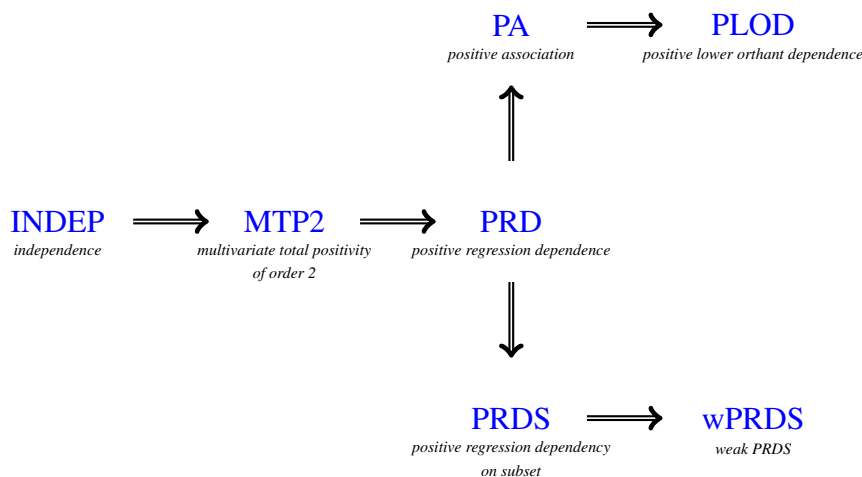
1. for all $x \in \mathbb{R}^m$, $f(x) = f(-x)$,
2. $K_u = \{x \in \mathbb{R}^m : f(x) \geq u\}$ is convex for all $u \geq 0$.

Let E be a convex set and symmetric about the origin. Let $Y \in \mathbb{R}^m$ be a determinist vector. Then for all $0 \leq k \leq 1$,

$$\mathbb{P}(X + kY \in E) \geq \mathbb{P}(X + Y \in E). \tag{3.7}$$

The proof of Šidák’s inequality relies on Theorem 3.4. We refer the interested reader to the supplementary results, see proof A.1.1.

We now conclude this section by summering the relations between the different notions of positive dependence presented in this section:



Note that the two types of dependence PA and PRDS are not related to each other. The examples 3.4 and 3.5 show that the PRDS property does not imply the PA property. Otherwise, Lehmann [64] provides several counter-examples in the bivariate case, which prove that the conversely is not true anymore.

3.3 Classical FWER control

Many procedures that control the FWER were build over time. We here choose to focus on two of them: the Bonferroni’s procedure, which is the most intuitive correction and the Šidák’s procedure, which is in connection with this manuscript.

CHAPTER 3. MULTIPLE TESTING FRAMEWORK

As mentioned in Chapter 1, the Bonferroni's correction ($\widehat{t}_{bonf} = \alpha/m$) is the first method used to counteract the problem of multiple comparisons by controlling the FWER. Indeed:

$$\forall P \in \mathcal{P}, \text{FWER}(\mathcal{R}, P) = \mathbb{P}\left(\bigcup_{i \in \mathcal{H}_0(P)} \{p_i \leq \widehat{t}_{bonf}\}\right) \leq \sum_{i \in \mathcal{H}_0(P)} \widehat{t}_{bonf} \leq \frac{\alpha m_0(P)}{m}. (\leq \alpha) \quad (3.8)$$

Note that the multiple testing procedure based on the Bonferroni's correction is valid whatever the structure of dependence of p -values is. When all the tested null hypotheses are independent, the FWER of Bonferroni's method is equal to $1 - (1 - \alpha/m)^m \simeq \alpha$ when α is small. Thus under independence, the Bonferroni's procedure is almost sharp. However, when there is a lot of false hypotheses regarding to the number of tested null hypotheses, this procedure is extremely conservative. Indeed, the bound 3.8 satisfies $\frac{\alpha m_0}{m} = \alpha - \frac{\alpha m_1}{m} \ll \alpha$ when m_1/m is large.

We now report another example (in connection with this manuscript) of FWER controlling procedure. When the p -values are PLOD, a FWER controlling procedure can be derived by the Šidák's inequality: for all $i \in \{1, \dots, m\}$, the procedure that rejects H_{0i} when $p_i \leq 1 - (1 - \alpha)^{1/m} = \widehat{t}_{sidak}$, controls the FWER at level α . Indeed, for all P such that \mathbf{p} is PLOD, we have

$$\begin{aligned} \text{FWER}(\mathcal{R}, P) &= 1 - \mathbb{P}\left(\bigcap_{i \in \mathcal{H}_0(P)} \{p_i > \widehat{t}_{sidak}\}\right) \\ &= 1 - \mathbb{P}\left(\bigcap_{i \in \mathcal{H}_0(P)} \{p_i > 1 - (1 - \alpha)^{1/m}\}\right). \end{aligned} \quad (3.9)$$

By using the Šidák's inequality (see Theorem 3.3), (3.9) is upper bounded by

$$\begin{aligned} 1 - \prod_{i \in \mathcal{H}_0(P)} \mathbb{P}(p_i > 1 - (1 - \alpha)^{1/m}) &\leq 1 - \prod_{i \in \mathcal{H}_0(P)} 1 - (1 - (1 - \alpha)^{1/m}) \\ &\leq 1 - (1 - \alpha)^{m_0/m} \\ &\leq \alpha. \end{aligned}$$

Note that since $\widehat{t}_{sidak} \geq \widehat{t}_{bonf}$, the Šidák's procedure is more powerful than the Bonferroni's procedure.

3.4 Classical FDR control

The most famous procedure that potentially controls the FDR is the Benjamini-Hochberg (BH) procedure [6]. We here focus on this approach.

3.4.1 The BH procedure

By denoting $(p_{(i)})_{1 \leq i \leq m}$ the family of the ordered p -values, the index set of rejected null hypotheses by the BH procedure is given by

$$\mathcal{R}_{BH} = \{i \in \{1, \dots, m\}, p_i \leq \alpha \widehat{k}/m\}, \quad (3.10)$$

where $\widehat{k} = \max\{k \in \{0, \dots, m\} : p_{(k)} \leq \alpha k/m\}$ with the convention $p_{(0)} = 0$. Figure 3.2 provides an example in the case of five simultaneous tests.

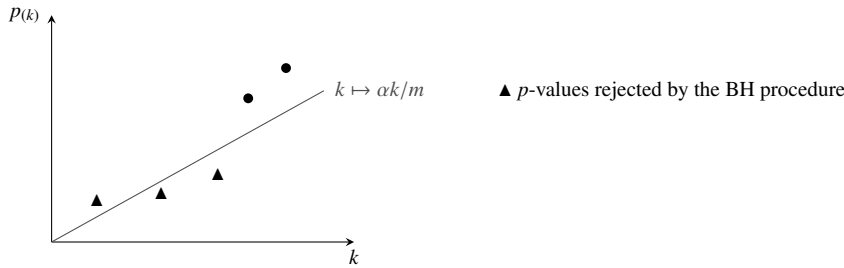


Figure 3.2 – For $m = 5$ tests, the BH procedure rejects 3 hypotheses.

In order to obtain the index set of rejected hypotheses by the BH procedure, we compare the ordered p -values $p_{(1)} \leq \dots \leq p_{(m)}$ with the critical constants of BH procedure $t_k = \alpha k/m$, $k = 1, \dots, m$. So, the index set of the rejected hypotheses relies on the largest ordered p -value that is upper bounded by its critical threshold. This process classifies the BH procedure as a step-up procedure. More generally, considering any sequence of nondecreasing constants $(\tau_k)_{1 \leq k \leq m}$, the index set of rejected hypotheses by the step-up procedure based on the critical constants τ_k , $k = 1, \dots, m$, denoted by $SU(\tau)$, is given by

$$\mathcal{R}_{SU(\tau)} = \{i \in \{1, \dots, m\}, p_i \leq \tau_{\widehat{k}}\}, \quad (3.11)$$

where $\widehat{k} = \max\{k \in \{0, \dots, m\} : p_{(k)} \leq \tau_k\}$ with the convention $p_{(0)} = 0$.

3.4.2 Bounds of FDR of BH procedure: known results

FDR control for BH procedure

The BH procedure has been developed in 1995 in order to guarantee an FDR control when the test statistics are independent. In 2001, Benjamini and Yekutieli proved that BH procedure still provides FDR control when the test statistics are PRDS.

CHAPTER 3. MULTIPLE TESTING FRAMEWORK

Theorem 3.5. *Under the weak PRDS assumption, the BH procedure provides an FDR control at level $\alpha m_0/m$ [67], namely if $P \in \mathcal{P}$ is such that the p -values are weak PRDS on $\mathcal{H}_0(P)$, then*

$$FDR(P) \leq m_0(P)\alpha/m. \quad (3.12)$$

Note that in the independent case, the inequality (3.12) is an equality when the p -values are uniform. Even if the independence assumption (resp. PRDS assumption) is rarely satisfied in realistic situations, it provides a remarkable theoretical result: the control of the FDR.

Under the PRDS assumption, we report a shorter proof than the original proof of Benjamini and Yekutieli, which is mentioned for instance in [74].

Proof. For all $P \in \mathcal{P}$, we have

$$\begin{aligned} FDR(P) &= \sum_{k=1}^m \frac{1}{k} \mathbb{E}[|\mathcal{R} \cap \mathcal{H}_0(P)| \cdot \mathbb{1}_{|\mathcal{R}|=k}] \\ &= \sum_{k=1}^m \frac{1}{k} \mathbb{E} \left[\sum_{i \in \mathcal{H}_0(P)} \mathbb{1}_{p_i \leq \frac{\alpha k}{m}} \cdot \mathbb{1}_{|\mathcal{R}|=k} \right] \\ &\leq \frac{\alpha}{m} \sum_{i \in \mathcal{H}_0(P)} \sum_{k=1}^m \mathbb{P}(\widehat{k} = k \mid p_i \leq \frac{\alpha k}{m}). \end{aligned}$$

Furthermore, the set $\{\mathbf{p} \in [0, 1]^m \text{ tel que } \widehat{k}(\mathbf{p}) < k - 1\}$ is a nondecreasing set. Thus by using the PRDS assumption, the function $u \mapsto \mathbb{P}(\widehat{k} \leq k - 1 \mid p_i = u)$ is a nondecreasing function. Thus the function $u \mapsto \mathbb{P}(\widehat{k} \leq k - 1 \mid p_i \leq u)$ is also a nondecreasing function. This yields

$$\begin{aligned} \sum_{k=1}^m \mathbb{P}(\widehat{k} = k \mid p_i \leq \frac{\alpha k}{m}) &= \sum_{k=1}^m \left[\mathbb{P}(\widehat{k} \leq k \mid p_i \leq \frac{\alpha k}{m}) - \mathbb{P}(\widehat{k} \leq k - 1 \mid p_i \leq \frac{\alpha k}{m}) \right] \\ &\leq \sum_{k=1}^m \left[\mathbb{P}(\widehat{k} \leq k \mid p_i \leq \frac{\alpha k}{m}) - \mathbb{P}(\widehat{k} \leq k - 1 \mid p_i \leq \frac{\alpha(k-1)}{m}) \right] \\ &\leq \mathbb{P}(\widehat{k} \leq m \mid p_i \leq \alpha) \\ &\leq 1. \end{aligned}$$

Finally,

$$FDR(P) \leq \frac{\alpha}{m} \sum_{i \in \mathcal{H}_0(P)} \sum_{k=1}^m \mathbb{P}(\widehat{k} = k \mid p_i \leq \frac{\alpha k}{m}) \leq \alpha m_0(P)/m.$$

□

On least favorable configurations for FDR control for the BH procedure

Without any assumption on the dependence structure of the p -values, Benjamini and Yekutieli provided an upper bound for the FDR of BH procedure [67]. Moreover Guo and Rao [75] proved that this bound is sharp by constructing a very artificial probability distribution of p -values. We here report the details of this construction and we propose an alternative proof for the sharpness of the bound of Benjamini and Yekutieli. Otherwise, we will see in Chapter 4 that this bound can be improved for more realistic dependence structures.

Theorem 3.6. *Whatever the dependence structure of the p -values is, for all $P \in \mathcal{P}$, we have [67]:*

$$\text{FDR}(P) \leq \frac{m_0(P)}{m} \alpha \sum_{i=1}^m \frac{1}{i}. \quad (3.13)$$

Moreover, for all $m_0 \in \{0, \dots, m\}$, if $\alpha \left[\frac{m_1(P)}{m} + \frac{m_0(P)}{m} \sum_{j=1}^{m_0} \frac{1}{j} \right] \leq 1$, and if for all distribution ν on $[0, 1]^m$ with m_0 uniform marginals, there exists $P^* \in \mathcal{P}$ with $|\mathcal{H}_0(P^*)| = m_0$ and $(p_i(X))_i \stackrel{X \sim P^*}{\sim} \nu$, then there exists a joint distribution of p -values for which the bound (3.13) is attained [75].

Note that the bound (3.13) is not sufficient to guarantee the FDR control at level α as soon as m is higher than or equal to 2.

Proof. Let us begin by proving the inequality (3.13). In this regards, we consider the convention $\frac{0}{0} = 0$. For $P \in \mathcal{P}$, we have

$$\begin{aligned} \text{FDR}(P) &= \sum_{i \in \mathcal{H}_0(P)} \sum_{k=1}^m \frac{1}{k} \mathbb{P} \left(\widehat{k} \leq k, p_i \leq \frac{\alpha k}{m} \right) - \frac{1}{k} \mathbb{P} \left(\widehat{k} \leq k-1, p_i \leq \frac{\alpha k}{m} \right) \\ &\leq \sum_{i \in \mathcal{H}_0(P)} \sum_{k=1}^m \frac{1}{k} \mathbb{P} \left(\widehat{k} \leq k, p_i \leq \frac{\alpha k}{m} \right) - \frac{1}{k} \mathbb{P} \left(\widehat{k} \leq k-1, p_i \leq \frac{\alpha(k-1)}{m} \right) \\ &\leq \sum_{i \in \mathcal{H}_0(P)} \sum_{k=1}^m \frac{1}{k} \mathbb{P} \left(\widehat{k} \leq k, p_i \leq \frac{\alpha k}{m} \right) - \frac{1}{k-1} \mathbb{P} \left(\widehat{k} \leq k-1, p_i \leq \frac{\alpha(k-1)}{m} \right) \\ &\quad + \sum_{i \in \mathcal{H}_0(P)} \sum_{k=2}^m \left(\frac{1}{k-1} - \frac{1}{k} \right) \mathbb{P} \left(\widehat{k} \leq k-1, p_i \leq \frac{\alpha(k-1)}{m} \right) \\ &\leq \sum_{i \in \mathcal{H}_0(P)} \frac{1}{m} \mathbb{P}(p_i \leq \alpha) + \sum_{i \in \mathcal{H}_0(P)} \sum_{k=2}^m \frac{1}{k(k-1)} \mathbb{P} \left(p_i \leq \frac{\alpha(k-1)}{m} \right) \\ &\leq \frac{\alpha m_0(P)}{m} + \frac{\alpha m_0(P)}{m} \left(\frac{1}{2} + \dots + \frac{1}{m} \right). \end{aligned}$$

CHAPTER 3. MULTIPLE TESTING FRAMEWORK

Now, let us to construct a probability distribution P^* that attains this bounds. Let $(U_i)_{1 \leq i \leq m}$ be a family of independent random variables such that $U_i \sim \mathcal{U}\left[\frac{i-1}{m}\alpha, \frac{i}{m}\alpha\right]$, $1 \leq i \leq m$ and $U_{m+1} \sim \mathcal{U}[\alpha, 1]$. Let N a random variable taking values in $\{1, \dots, m+1\}$ and generated from the following law:

$$\mathbb{P}(N = n) = \begin{cases} \frac{m_0}{m} \frac{1}{n} \alpha & \text{if } 1 \leq n \leq m_0 \\ \alpha/m & \text{if } m_0 + 1 \leq n \leq m \\ 1 - \left[\frac{m_1}{m} + \frac{m_0}{m} \sum_{j=1}^{m_0} \frac{1}{j} \right] \alpha & \text{if } n = m + 1 \end{cases}.$$

Depending on the value of N , we associate one p -value to each index $1, 2, \dots, m$, by the following way:

- Case $N = n \in I_0 = \{1, \dots, m_0\}$: pick n indexes $\{i_1, \dots, i_n\} := I_{0,n}$ randomly and without replacement in I_0 . For all $j \in \{1, \dots, n\}$, the index i_j is associated to $p_{i_j} = U_n$. For all $s \in \{1, \dots, m\} \setminus I_{0,n} = I \setminus I_{0,n}$, the index s is associated to $p_s = U_{m+1}$.
- Case $N = n \in I_1 = \{m_0 + 1, \dots, m\}$: choose the indexes i_1, \dots, i_{m_0} in I_0 and pick randomly and without replacement the $(n - m_0)$ remaining indexes in I_1 . We denote by $I_{1,n} := \{i_1, \dots, i_{m_0}, i_{m_0+1}, \dots, i_n\}$ the index set of picked indexes. For all $j \in \{1, \dots, n\}$, the index i_j is associated to $p_{i_j} = U_n$. For all $s \in I \setminus I_{1,n}$, the index s is associated to $p_s = U_{m+1}$.
- Case $N = m + 1$: for all $i \in I_0$, the index i is associated to $p_i = U_{m+1}$. For all $j \in I_1$, the index j is associated to $p_j = U_1$.

Now, we show that this construction induces the following property

$$\forall i \in I_0, \forall x \in]0, 1[, \mathbb{P}(p_i \leq x) = x.$$

Let $i \in I_0$. For all $x \in]0, 1[$,

$$\begin{aligned} \mathbb{P}(p_i \leq x) &= \sum_{n=1}^{m_0} \mathbb{P}(p_i \leq x \mid N = n) \mathbb{P}(N = n) + \sum_{n=m_0+1}^m \mathbb{P}(p_i \leq x \mid N = n) \mathbb{P}(N = n) \\ &+ \mathbb{P}(p_i \leq x \mid N = m + 1) \mathbb{P}(N = m + 1). \end{aligned} \quad (3.14)$$

If $n \in I_0$, we have

$$\begin{aligned} \mathbb{P}(p_i \leq x \mid N = n) &= \mathbb{P}(p_i \leq x \mid N = n, i \in I_{0,n}) \mathbb{P}(i \in I_{0,n} \mid N = n) \\ &+ \mathbb{P}(p_i \leq x \mid N = n, i \in I_0 \setminus I_{0,n}) \mathbb{P}(i \in I_0 \setminus I_{0,n} \mid N = n) \\ &= \mathbb{P}(U_n \leq x) \frac{n}{m_0} + \mathbb{P}(U_{m+1} \leq x) \frac{m_0 - n}{m_0}. \end{aligned}$$

If $n \in I_1$, we have

$$\begin{aligned}\mathbb{P}(p_i \leq x \mid N = n) &= \mathbb{P}(p_i \leq x \mid N = n, i \in \{i_1, \dots, i_{m_0}\})\mathbb{P}(i \in \{i_1, \dots, i_{m_0}\} \mid N = n) \\ &= \mathbb{P}(U_n \leq x).\end{aligned}$$

Finally if $n = m + 1$, we have

$$\begin{aligned}\mathbb{P}(p_i \leq x \mid N = m + 1) &= \mathbb{P}(p_i \leq x \mid N = m + 1, i \in I_0)\mathbb{P}(i \in I_0 \mid N = m + 1) \\ &= \mathbb{P}(U_{m+1} \leq x).\end{aligned}$$

Thus (3.14) can be written as

$$\begin{aligned}\mathbb{P}(p_i \leq x) &= \sum_{n=1}^{m_0} \mathbb{P}(U_n \leq x) \frac{\alpha}{m} + \sum_{n=1}^{m_0} \mathbb{P}(U_{m+1} \leq x) \frac{\alpha}{m} \frac{m_0 - n}{n} \\ &\quad + \sum_{n=m_0+1}^m \mathbb{P}(U_n \leq x) \frac{\alpha}{m} \\ &\quad + \mathbb{P}(U_{m+1} \leq x) \left[1 - \frac{m - m_0}{m} \alpha - \frac{m_0}{m} \alpha \sum_{j=1}^{m_0} \frac{1}{j} \right].\end{aligned}\tag{3.15}$$

For all $i \in I_0$, we establish that the constructed p -values are uniform by considering the three cases,

1. $x \in [(\tilde{n} - 1)\alpha/m, \tilde{n}\alpha/m[\subset [0, m_0\alpha/m[$,
2. $x \in [(\tilde{n} - 1)\alpha/m, \tilde{n}\alpha/m[\subset [m_0\alpha/m, \alpha[$,
3. $x \in [\alpha, 1[$.

We have:

1.

$$\begin{aligned}\mathbb{P}(p_i \leq x) &= \sum_{n=1}^{\tilde{n}-1} \mathbb{P}(U_n \leq x) \frac{\alpha}{m} + \mathbb{P}(U_{\tilde{n}} \leq x) \frac{\alpha}{m} \\ &= (\tilde{n} - 1) \frac{\alpha}{m} + \frac{\alpha}{m} \frac{x - (\tilde{n} - 1)\alpha/m}{\alpha/m} \\ &= x\end{aligned}$$

2.

$$\begin{aligned}\mathbb{P}(p_i \leq x) &= \frac{m_0}{m} \alpha + \sum_{n=m_0+1}^{\tilde{n}-1} \mathbb{P}(U_n \leq x) \frac{\alpha}{m} + \mathbb{P}(U_{\tilde{n}} \leq x) \frac{\alpha}{m} \\ &= \frac{m_0}{m} \alpha + \frac{\alpha}{m} (\tilde{n} - m_0 - 1) + x - \frac{\tilde{n}\alpha}{m} + \frac{\alpha}{m} \\ &= x\end{aligned}$$

3.

$$\begin{aligned}
 \mathbb{P}(p_i \leq x) &= \frac{m_0}{m}\alpha + \frac{\alpha}{m}\mathbb{P}(U_{m+1} \leq x) \sum_{n=1}^{m_0} \frac{m_0 - n}{n} + \frac{m - m_0}{m}\alpha \\
 &\quad + \mathbb{P}(U_{m+1} \leq x) \left[1 - \frac{m - m_0}{m}\alpha - \frac{m_0}{m}\alpha \sum_{j=1}^{m_0} \frac{1}{j} \right] \\
 &= \alpha + \frac{\alpha}{m}\mathbb{P}(U_{m+1} \leq x) \left[m_0 \sum_{n=1}^{m_0} \frac{1}{n} - m_0 \right] \\
 &\quad + \mathbb{P}(U_{m+1} \leq x) \left[1 - \frac{m - m_0}{m}\alpha - \frac{m_0}{m}\alpha \sum_{j=1}^{m_0} \frac{1}{j} \right] \\
 &= \alpha + (1 - \alpha)\mathbb{P}(U_{m+1} \leq x) \\
 &= \alpha + (1 - \alpha) \frac{x - \alpha}{1 - \alpha} \\
 &= x.
 \end{aligned}$$

Moreover, we can merely check that this construction induces the following value of the FDP:

$$\text{FDP}(P^*) = \begin{cases} 1 & \text{si } 1 \leq n \leq m_0 \\ m_0/n & \text{si } m_0 + 1 \leq n \leq m. \\ 0 & \text{si } n = m + 1 \end{cases}$$

This yields

$$\begin{aligned}
 \text{FDR}(P^*) &= \mathbb{E}[\text{FDP}(P^*)] \\
 &= \mathbb{E} \left[\sum_{n=1}^{m+1} \mathbb{E}[\text{FDP}(P^*) | N = n] \mathbb{1}\{N = n\} \right] \\
 &= \sum_{n=1}^{m_0} \mathbb{E} \left[\underbrace{\mathbb{E}[\text{FDP}(P^*) | N = n]}_{=1} \mathbb{1}\{N = n\} \right] + \sum_{n=m_0+1}^m \mathbb{E} \left[\underbrace{\mathbb{E}[\text{FDP}(P^*) | N = n]}_{=m_0/n} \mathbb{1}\{N = n\} \right] \\
 &\quad + \mathbb{E} \left[\underbrace{\mathbb{E}[\text{FDP}(P^*) | N = m + 1]}_{=0} \mathbb{1}\{N = m + 1\} \right] \\
 &= \sum_{n=1}^{m_0} \mathbb{P}(N = n) + \sum_{n=m_0+1}^m \frac{m_0}{n} \mathbb{P}(N = n) \\
 &= \sum_{n=1}^{m_0} \frac{m_0}{m} \frac{1}{n} \alpha + \sum_{n=m_0+1}^m \frac{m_0}{n} \frac{\alpha}{m},
 \end{aligned}$$

and the desired bound is proved, $\text{FDR}(P^*) = \frac{\alpha m_0}{m} \sum_{n=1}^m \frac{1}{n}$. \square

3.5. A NEW RESULT: RELAXATION OF THE PRDS CONDITION

Under a weaker condition than the PRDS assumption, Guo and Rao [76] proved that the critical constants of BH procedure are somewhat optimal.

Theorem 3.7. *For all distribution $P \in \mathcal{P}$, among all step-up procedures $SU(\tau)$ having a nondecreasing sequence of critical values $(\tau_k)_{1 \leq k \leq m}$, which controls the FDR at level $\alpha m_0(P)/m$, that is*

$$FDR(\mathcal{R}_{SU(\tau)}, P) \leq \frac{\alpha m_0(P)}{m},$$

and such that

$$\forall i \in \mathcal{H}_0(P), \sum_{k=1}^m \mathbb{P}(|\mathcal{R}_{SU(\tau)}| = k \mid p_i \leq \tau_k) \leq 1, \quad (3.16)$$

the critical constants of the BH procedure are optimal, namely

$$\forall P \in \mathcal{P}, \forall SU(\tau) \text{ such that } (\tau, \mathbf{p}) \text{ satisfies (3.16), we have } \forall k \in \{1, \dots, m\}, \tau_k \leq t_k.$$

Note that this notion of optimality is weaker than the one usually considered because assumption (3.16) concerns both the law of the p -values and the sequence of critical constants $(\tau_k)_{1 \leq k \leq m}$. Our contribution is the correction of some steps in the proof of Theorem 3.7. Here, we provide the sketch of proof and we refer the interested reader to the supplementary results (see A.1.3) for the whole proof and the details of corrections.

Proof. (Sketch of proof of Theorem 3.7)

Let $(\tau_k)_{1 \leq k \leq m}$ be a nondecreasing sequence of critical constants and $k_1 = \max \left\{ \arg \max_{k \in \{1, \dots, m\}} \left\{ \frac{\tau_k}{k} \right\} \right\}$.

This proof is divided into three steps.

First, we show that for all $P \in \mathcal{P}$, for all procedure $SU(\tau)$ such that (τ, \mathbf{p}) satisfies (3.16), then $FDR(\mathcal{R}_{SU(\tau)}, P) \leq m_0(P)\tau_{k_1}/k_1$ and by using the same construction than in Theorem 3.6, we show that there exists a probability distribution P^* for which the joint distribution of the p -values attains this bound.

Second, we prove that the joint distribution of these p -values satisfies the assumption (3.16) but is not PRDS on $\{1, \dots, m_0\}$.

We finally obtain the desired result (the optimality of BH thresholds) by using the definition of k_1 . □

3.5 A new result: relaxation of the PRDS condition

From now on, given a probability distribution $P \in \mathcal{P}$, assume that the family of p -values is PRDS on one subset of $\mathcal{H}_0(P)$. In this case, the only bound that is valid for the FDR is the bound of Benjamini and Yekutieli. The idea of the following theorem is to exploit the gap between the very general bound of Benjamini and Yekutieli and the bound valid under PRDS assumption. Thus by taking into account the control of the FDR on the positively

CHAPTER 3. MULTIPLE TESTING FRAMEWORK

correlated subset of $\mathcal{H}_0(P)$, we aim at improving the bound of Benjamini and Yekutieli in such a case. The relaxation of the PRDS assumption is particularly motivated by the neuroscientific application presented in Chapter 1. Indeed, it seems that the correlations between the different areas of the brain are mostly positive [77].

Theorem 3.8. *For all $P \in \mathcal{P}$, let $I_+^0(P) \subset \{i \in \mathcal{H}_0(P), (p_j)_{1 \leq j \leq m} \text{ is PRDS on } i\}$ and $I_-^0(P) = \mathcal{H}_0(P) \setminus I_+^0(P)$. Then, we have*

$$FDR(P) \leq \frac{\alpha}{m} \left(|I_+^0(P)| + |I_-^0(P)| \sum_{k=1}^m \frac{1}{k} \right). \quad (3.17)$$

Proof. For all $P \in \mathcal{P}$,

$$\begin{aligned} FDR(P) &\leq \frac{\alpha}{m} \sum_{i \in \mathcal{H}_0(P)} \sum_{k=1}^m \mathbb{P}(\widehat{k} = k \mid p_i \leq \alpha k/m) \\ &\leq \frac{\alpha}{m} \sum_{i \in I_+^0(P)} \sum_{k=1}^m \mathbb{P}(\widehat{k} = k \mid p_i \leq \alpha k/m) + \frac{\alpha}{m} \sum_{i \in I_-^0(P)} \sum_{k=1}^m \mathbb{P}(\widehat{k} = k \mid p_i \leq \alpha k/m) \\ FDR(P) &\leq \frac{\alpha}{m} \left(|I_+^0(P)| + |I_-^0(P)| \sum_{k=1}^m \frac{1}{k} \right) \end{aligned}$$

where the last inequality is obtained by using Theorem 3.5 in the first member of the expression and the bound of Benjamini and Yekutieli for the second. \square

Note that when $|I_+^0(P)| = 0$, the inequality (3.17) is equivalent to the bound of Benjamini and Yekutieli. Otherwise, when $|I_-^0(P)| = 0$ we obtain the control of the FDR for the positively dependent test statistics.

Corollary 3.1. *Let $\gamma_m = \sum_{j=1}^m 1/j$. For all $P \in \mathcal{P}$, if $|I_-^0(P)| \leq (m - m_0(P))/(\gamma_m - 1)$ then $FDR(P) \leq \alpha$.*

Proof. The result follows from that for all $P \in \mathcal{P}$,

$$\begin{aligned} |I_-^0(P)| \leq (m - m_0(P))/(\gamma_m - 1) &\Leftrightarrow (\gamma_m - 1)|I_-^0(P)| \leq m - m_0(P) \\ &\Leftrightarrow |I_+^0(P)| + |I_-^0(P)| + |I_-^0(P)|\gamma_m - |I_-^0(P)| \leq m \\ &\Leftrightarrow |I_+^0(P)| + |I_-^0(P)|\gamma_m \leq m. \end{aligned}$$

\square

Example 3.6. *If $P = N_m(\mu, \Sigma)$ and for all $i = 1, \dots, m$, we consider the problem of testing the null hypotheses $H_{0,i} : \mu_i = 0$ against the alternative hypotheses $H_{1,i} : \mu_i > 0$. Then $I_+^0 = \{i \in \mathcal{H}_0, \Sigma_i \geq 0\}$ satisfies the assumption of Theorem 3.8.*

3.5. A NEW RESULT: RELAXATION OF THE PRDS CONDITION

For instance, considering the matrix $\Sigma = \begin{pmatrix} 1 & -\rho & 0 & \dots & 0 \\ -\rho & 1 & 0 & & \vdots \\ 0 & 0 & \ddots & \ddots & \vdots \\ \vdots & & \ddots & \ddots & 0 \\ 0 & \dots & \dots & 0 & 1 \end{pmatrix}$ with $\rho > 0$, we have

$I_+^0 = \{i \in \{1, \dots, m\} : \mu_i = 0 \text{ and } \forall j \neq i, \Sigma_{ij} \geq 0\}$, $I_-^0 = \{i \in \{1, \dots, m\} : \mu_i = 0 \text{ and } \exists j \neq i, \Sigma_{ij} < 0\}$ and $|I_-^0| = 2$. By using Corollary 3.1, the BH procedure provides an FDR control at level α if $m_1 \geq 2(\gamma_m - 1) \sim 2 \log m$. While this condition seems to be reasonable, it may be hard to check in practice.

Chapter 4

On BH procedure in the two-sided Gaussian setting

This chapter investigates the quality of the BH procedure in the two-sided Gaussian setting. We are interested in studying bounds on FDR of BH valid under some particularly structured dependencies. Our contributions are: an extension of the scope of the bound established by Reiner-Benaim both theoretically and numerically, and a formalization of the results of Cohen [78]. Moreover, we will prove that a bound proposed by Cohen is almost sharp.

Contents

4.1	Statistical setting	83
4.2	Background	84
4.3	Known results and new formalized proofs	85
4.3.1	Result under the full null	85
4.3.2	Results for perfect correlation	86
4.4	New results	89
4.4.1	On sharpness of (CohenBound)	89
4.4.2	Extending scope of (ReinerBound)	91
4.4.3	Numerical study of (ReinerBound) with exact calculation	92
4.5	Summary and conjectures	95

4.1 Statistical setting

Let $m \geq 2$. Let us observe

$$X_i = \mu_i + Y_i, \quad 1 \leq i \leq m,$$

CHAPTER 4. ON BH PROCEDURE IN THE TWO-SIDED GAUSSIAN SETTING

where $\mu \in \mathbb{R}^m$ is the parameter of interest and Y is an m -dimensional centered Gaussian vector with covariance matrix Σ invertible with $\Sigma_{ii} = 1$ for all $i \in \{1, \dots, m\}$. We consider the problem of testing

$$H_{0,i} : \mu_i = 0 \text{ against } H_{1,i} : \mu_i \neq 0,$$

simultaneously for all $i \in \{1, \dots, m\}$. Introduce the associated p -values as

$$p_i = 2\bar{\Phi}(|X_i|), 1 \leq i \leq m,$$

where $\bar{\Phi}(z) = \mathbb{P}(Z \geq z)$ for $Z \sim \mathcal{N}(0, 1)$. We also denote by $\mathcal{H}_0(\mu)$ the index set of the true null hypotheses, that is $\mathcal{H}_0(\mu) = \{i \in \{1, \dots, m\} : \mu_i = 0\}$ and we denote by $m_0(\mu) = |\mathcal{H}_0(\mu)|$ its cardinality. Similarly, the index set of false hypotheses is defined as $\mathcal{H}_1(\mu) = \{i \in \{1, \dots, m\} : \mu_i \neq 0\}$ and its cardinality is denoted by $m_1(\mu) = |\mathcal{H}_1(\mu)|$. Throughout this chapter, to shorten notation, μ is sometimes dropped in the notation. The ordered p -values are denoted by

$$p_{(0)} \leq p_{(1)} \leq \dots \leq p_{(m)},$$

with the convention $p_{(0)} = 0$. Let us recall that the FDR of BH procedure is defined by the following quantity

$$\text{FDR}(\mu, \Sigma) = \mathbb{E} \left[\frac{\sum_{i \in \mathcal{H}_0} \mathbb{1}\{p_i \leq \widehat{\alpha}k/m\}}{\widehat{k} \vee 1} \right]$$

where $\widehat{k} = \max\{k \in \{0, \dots, m\} : p_{(k)} \leq \alpha k/m\}$ and we write " $\widehat{k} \vee 1$ " for the maximum between \widehat{k} and 1, in order to guarantee the well-definition of the FDR when $\widehat{k} = 0$. Throughout this chapter, we denote by $\tau_k, i = 1, \dots, m$, the following sequence of non-increasing constants

$$\tau_k = \bar{\Phi}^{-1}(\alpha k/2m), \text{ for } 1 \leq k \leq m. \quad (4.1)$$

We are interested in studying bounds on $\text{FDR}(\mu, \Sigma)$ valid under some assumptions on μ and Σ .

4.2 Background

As seen in Chapter 3 (see Theorem 3.6), a first bound is given by Benjamini and Yekutieli [67]:

$$\forall \mu, \Sigma, \quad \text{FDR}(\mu, \Sigma) \leq \alpha \frac{m_0}{m} \gamma_m, \quad \gamma_m = \sum_{i=1}^m 1/i. \quad (\text{BYBound})$$

Furthermore, Guo and Rao [75] proved that this bound is sharp (if $\alpha \leq \gamma_m^{-1}$) when the set of p -value distribution is larger than the one considered in this section. However, in the restricted two-sided Gaussian case, the bound (BYBound) is likely to be too conservative.

4.3. KNOWN RESULTS AND NEW FORMALIZED PROOFS

We might conjecture that a bound smaller than α is valid in the two-sided Gaussian case. For instance, when $m_0 = 1$, the bound ([BYBound](#)) is always smaller than α . But what about the other situations?

Reiner-Benaim [[79](#)] proposed the following bound:

$$\alpha \frac{m_0}{m} \left[1 + \frac{1}{2} \left(1 - \frac{m_0}{m} \right) \right] \leq \alpha, \quad (\text{ReinerBound})$$

and has proved it for $m = 2$ and Σ being the matrix with all entries equal to 1.

Next, Cohen [[78](#)] suggested that ([ReinerBound](#)) might be violated when the alternative means are not equal and proposed the larger bound

$$\alpha \frac{m_0}{m} \left[1 + \mathbb{1}_{\{m_1 > 0\}} \frac{1}{2} \sum_{\ell=m_0+1}^m \frac{1}{\ell} \right] \leq \alpha. \quad (\text{CohenBound})$$

He established that the bound indeed holds for any m, μ and Σ being the matrix with all entries equal to 1. He also suggested the following more liberal bound

$$\alpha \frac{m_0}{m} \left[1 + \mathbb{1}_{\{m_1 > 0\}} \sum_{\ell=m_0+1}^m \frac{1}{\ell} \right] \leq \alpha \quad (\text{CohenBound2})$$

and proved that it holds under a similar constraint, but only concerning the submatrix $(\Sigma_{i,j})_{(i,j) \in \mathcal{H}_0^2}$. Note that this bound coincides with ([BYBound](#)) when $m_0 = 1$.

As seen in Chapter 3 (see Counter-exemple [3.1](#)), the PRDS property does no longer hold in the two-sided Gaussian framework. Thus, the results of this chapter provide an interesting direction towards showing that the PRDS assumption is a sufficient but not necessary condition for the FDR control of the BH procedure.

4.3 Known results and new formalized proofs

4.3.1 Result under the full null

When $\mu = 0$, the p -values are given simply by

$$2\overline{\Phi}(|Y_i|), \text{ where } Y_i \sim \mathcal{N}(0, 1), 1 \leq i \leq m.$$

Then, Theorem 3.1 of [[63](#)], combined with [[67](#)] provides the following result.

CHAPTER 4. ON BH PROCEDURE IN THE TWO-SIDED GAUSSIAN SETTING

Theorem 4.1. *Assume Σ is such that there exists a diagonal matrix D such that the element of the diagonal are valued in $\{-1, +1\}$ and the off-diagonal elements of $-D\Sigma^{-1}D$ are all non-negative, then*

$$FDR(0, \Sigma) \leq m_0\alpha/m = \alpha. \quad (4.2)$$

Note that the assumption on Σ is equivalent to assume that the p -values are MTP_2 in our framework (see Example 3.3 in Chapter 3). Several examples of such matrix Σ are provided in Section 4 of [63]. For instance, the identity matrix, the ρ -equicorrelated case are included, but also some other cases like 1 everywhere except in two blocks equal to ρ for both ρ negative and positive.

Recently, Royen proves the Gaussian correlation conjecture, see [80]. This might indicate that (4.2) is valid for any Σ . But the rejection sets involved in

$$FDR(0, \Sigma) = \mathbb{P}(\exists k \in \{0, \dots, m\} : p_{(k)} \leq \alpha k/m)$$

are not convex, so the argument does not apply directly.

Note also that Theorem 4.1 entails that for $m = 2$, for all μ and Σ ,

$$FDR(\mu, \Sigma) \leq \alpha.$$

Indeed, it establishes the control for $m_0(\mu) = 2$. The control is trivial when $m_0(\mu) = 0$ (since no incorrect reject can occur). Next, the bound (BYBound) is $\alpha 1/2(1 + 1/2) = 3\alpha/4 \leq \alpha$ when $m_0(\mu) = 1$.

4.3.2 Results for perfect correlation

Cohen [78] stated the following results. However, the proofs were not completely formalized. We provide (hopefully) complete proofs below.

Proposition 4.1. *Let Σ being the covariance matrix with all entries equal to 1, then for all $\mu \in \mathbb{R}^m$,*

$$FDR(\mu, \Sigma) \leq \alpha \frac{m_0}{m} \left[1 + \mathbb{1}_{\{m_1 > 0\}} \frac{1}{2} \sum_{\ell=m_0+1}^m \frac{1}{\ell} \right] \leq \alpha. \quad (4.3)$$

Proof. By the equicorrelated structure, $Y_i = Y_j$ for all i, j . Let us denote by Z this common value. Denote $\mathcal{H}_1^- = \{i \in \{1, \dots, m\} : \mu_i < 0\}$, $\mathcal{H}_1^+ = \{i \in \{1, \dots, m\} : \mu_i > 0\}$ and m_1^-, m_1^+ their cardinal, respectively. Finally, FDP denotes the false discovery proportion of BH procedure.

By definition, the BH procedure rejects the hypotheses corresponding to the indexes $\{i \in \{1, \dots, m\} : |Y_i + \mu_i| \geq \tau_{\widehat{k}}\}$, where $\widehat{k} = \max\{k \in \{0, \dots, m\} : \sum_{i=1}^m \mathbb{1}_{\{|Y_i + \mu_i| \geq \tau_k}\} \geq k\}$, where the τ_k 's are defined by (4.1).

4.3. KNOWN RESULTS AND NEW FORMALIZED PROOFS

Let us first focus on the event where $\{Z > 0, \text{FDP} > 0\}$. We have in that case for all k ,

$$\sum_{i=1}^m \mathbb{1}_{\{|Y_i + \mu_i| \geq \tau_k\}} = \sum_{i \in \mathcal{H}_0} \mathbb{1}_{\{Z \geq \tau_k\}} + \sum_{i \in \mathcal{H}_1^+} \mathbb{1}_{\{Z + \mu_i \geq \tau_k\}} + \sum_{i \in \mathcal{H}_1^-} \mathbb{1}_{\{|Z + \mu_i| \geq \tau_k\}}.$$

Also, since $\text{FDP} > 0$, there exists $i \in \mathcal{H}_0$ such that $|X_i| \geq \tau_{\widehat{k}}$, that is, $Z \geq \tau_{\widehat{k}}$, which in turn implies $\widehat{k} = \sum_{i=1}^m \mathbb{1}_{\{|Y_i + \mu_i| \geq \tau_{\widehat{k}}\}} \geq m_0 + m_1^+$. Also, on that event, we have $\text{FDP} = \frac{m_0}{\widehat{k}}$. This yields

$$\begin{aligned} \mathbb{E} [\text{FDP} \mathbb{1}_{\{Z > 0\}}] &\leq \sum_{\ell=0}^{m_1^-} \frac{m_0}{m_0 + m_1^+ + \ell} \mathbb{P} \left(Z \geq \tau_{m_0 + m_1^+ + \ell}, \widehat{k} = m_0 + m_1^+ + \ell \right) \\ &= \sum_{\ell=0}^{m_1^-} \frac{m_0}{m_0 + m_1^+ + \ell} \mathbb{P} \left(Z \geq \tau_{m_0 + m_1^+ + \ell}, \widehat{k} \geq m_0 + m_1^+ + \ell \right) \\ &\quad - \sum_{\ell=0}^{m_1^-} \frac{m_0}{m_0 + m_1^+ + \ell} \mathbb{P} \left(Z \geq \tau_{m_0 + m_1^+ + \ell}, \widehat{k} \geq m_0 + m_1^+ + \ell + 1 \right). \end{aligned}$$

Therefore, by re-indexing the second sum, we obtain

$$\begin{aligned} \mathbb{E} [\text{FDP} \mathbb{1}_{\{Z > 0\}}] &\leq \frac{m_0}{m_0 + m_1^+} \mathbb{P} \left(Z \geq \tau_{m_0 + m_1^+} \right) \\ &\quad + \sum_{\ell=1}^{m_1^-} \frac{m_0}{m_0 + m_1^+ + \ell} \mathbb{P} \left(Z \geq \tau_{m_0 + m_1^+ + \ell}, \widehat{k} \geq m_0 + m_1^+ + \ell \right) \\ &\quad - \sum_{\ell=1}^{m_1^-} \frac{m_0}{m_0 + m_1^+ + \ell - 1} \mathbb{P} \left(Z \geq \tau_{m_0 + m_1^+ + \ell - 1}, \widehat{k} \geq m_0 + m_1^+ + \ell \right). \end{aligned}$$

Now, since $\frac{m_0}{m_0 + m_1^+ + \ell - 1} \geq \frac{m_0}{m_0 + m_1^+ + \ell}$, the latter is smaller than or equal to

$$\begin{aligned} &\alpha \frac{m_0}{2m} + \sum_{\ell=1}^{m_1^-} \frac{m_0}{m_0 + m_1^+ + \ell} \left[\mathbb{P} \left(Z \geq \tau_{m_0 + m_1^+ + \ell}, \widehat{k} \geq m_0 + m_1^+ + \ell \right) - \mathbb{P} \left(Z \geq \tau_{m_0 + m_1^+ + \ell - 1}, \widehat{k} \geq m_0 + m_1^+ + \ell \right) \right] \\ &= \alpha \frac{m_0}{2m} + \sum_{\ell=1}^{m_1^-} \frac{m_0}{m_0 + m_1^+ + \ell} \left[\mathbb{P} \left(\tau_{m_0 + m_1^+ + \ell} \leq Z < \tau_{m_0 + m_1^+ + \ell - 1}, \widehat{k} \geq m_0 + m_1^+ + \ell \right) \right] \\ &\leq \alpha \frac{m_0}{2m} \left[1 + \sum_{\ell=1}^{m_1^-} \frac{1}{m_0 + m_1^+ + \ell} \right]. \end{aligned}$$

This proves

$$\mathbb{E} [\text{FDP} \mathbb{1}_{\{Z > 0\}}] \leq \alpha \frac{m_0}{2m} \left[1 + \sum_{\ell=1}^{m_1^-} \frac{1}{m_0 + m_1^+ + \ell} \right].$$

CHAPTER 4. ON BH PROCEDURE IN THE TWO-SIDED GAUSSIAN SETTING

Now, by the symmetry of the distribution of Z , we also have

$$\mathbb{E}[\text{FDP} \mathbb{1}_{\{Z < 0\}}] \leq \alpha \frac{m_0}{2m} \left[1 + \sum_{\ell=1}^{m_1^+} \frac{1}{m_0 + m_1^- + \ell} \right],$$

which yields

$$\begin{aligned} \text{FDR} &\leq \frac{\alpha m_0}{m} \left[1 + \frac{1}{2} \left(\sum_{\ell=1}^{m_1^-} \frac{1}{m_0 + m_1^+ + \ell} + \sum_{\ell=1}^{m_1^+} \frac{1}{m_0 + m_1^- + \ell} \right) \right] \\ &\leq \frac{\alpha m_0}{m} \left[1 + \frac{1}{2} \left(\sum_{\ell=1}^{m_1^-} \frac{1}{m_0 + \ell} + \sum_{\ell=1}^{m_1^+} \frac{1}{m_0 + m_1^- + \ell} \right) \right]. \end{aligned}$$

□

Under a light relaxation of the perfect correlation assumption, the FDR control still remains and the FDR is upper-bounded by the slightly increased bound ([CohenBound2](#)).

Proposition 4.2. *Let $\mu \in \mathbb{R}^m$ and Σ be any covariance matrix with diagonal entries equal to 1 and such that the submatrix $(\Sigma_{i,j})_{(i,j) \in \mathcal{H}_0^2}$ has all entries equal to 1. Then we have*

$$\text{FDR}(\mu, \Sigma) \leq \alpha \frac{m_0}{m} \left[1 + \mathbb{1}_{\{m_1 > 0\}} \sum_{\ell=m_0+1}^m \frac{1}{\ell} \right] \leq \alpha. \quad (4.4)$$

This result implies that $\text{FDR}(\mu, \Sigma) \leq \alpha$ holds for any m and Σ provided that μ is such that $m_0(\mu) = 1$. This might be of interest in situations where null hypotheses are not compatible (if one is true, all the others are necessarily false).

Proof. Denote by Z the common value of Y_i , $i \in \mathcal{H}_0$. Also, the same reasoning as before entails that the event $\{\text{FDP} > 0\}$ implies $\widehat{k} \geq m_0$ with $\text{FDP} = \frac{m_0}{\widehat{k}}$. This yields

$$\begin{aligned} \text{FDR}(\mu, \Sigma) &\leq \sum_{\ell=0}^{m_1} \frac{m_0}{m_0 + \ell} \mathbb{P}(|Z| \geq \tau_{m_0+\ell}, \widehat{k} = m_0 + \ell) \\ &= \sum_{\ell=0}^{m_1} \frac{m_0}{m_0 + \ell} \mathbb{P}(|Z| \geq \tau_{m_0+\ell}, \widehat{k} \geq m_0 + \ell) - \sum_{\ell=0}^{m_1-1} \frac{m_0}{m_0 + \ell} \mathbb{P}(|Z| \geq \tau_{m_0+\ell}, \widehat{k} \geq m_0 + \ell + 1) \\ &= \mathbb{P}(|Z| \geq \tau_{m_0}) \\ &\quad + \sum_{\ell=1}^{m_1} \left(\frac{m_0}{m_0 + \ell} \mathbb{P}(|Z| \geq \tau_{m_0+\ell}, \widehat{k} \geq m_0 + \ell) - \frac{m_0}{m_0 + \ell - 1} \mathbb{P}(|Z| \geq \tau_{m_0+\ell-1}, \widehat{k} \geq m_0 + \ell) \right). \end{aligned} \quad (4.5)$$

Hence, we obtain

$$\begin{aligned} \text{FDR}(\mu, \Sigma) &\leq \alpha m_0/m + \sum_{\ell=1}^{m_1} \frac{m_0}{m_0 + \ell} \left(\mathbb{P}(|Z| \geq \tau_{m_0+\ell}, \widehat{k} \geq m_0 + \ell) - \mathbb{P}(|Z| \geq \tau_{m_0+\ell-1}, \widehat{k} \geq m_0 + \ell) \right) \\ &\leq \alpha m_0/m + \sum_{\ell=1}^{m_1} \frac{m_0}{m_0 + \ell} \alpha/m, \end{aligned}$$

which implies (4.4). \square

These two previous results do not really use the Gaussian structure. Proposition 4.1 would be extended to a non-Gaussian two-sided location model. Proposition 4.2 only uses that all the p -values under the null are equal (use neither Gaussian nor two-sided).

4.4 New results

4.4.1 On sharpness of (CohenBound)

The following proposition shows that (CohenBound) is essentially sharp under perfect correlation.

Proposition 4.3. *Let $m \geq 2$, Σ being the $m \times m$ matrix with all entries equal to 1, and $m_0 \in \{1, \dots, m-1\}$ such that $\tau_{m_0} \leq \tau_{m-1} + 2\tau_m$ is satisfied. Then, by taking $\mu^0 \in \mathbb{R}^m$ with $m_1 = m - m_0$ non-zero coordinates such that*

$$\mu_i^0 = \tau_{m-i+1} + \tau_{m-i}, \quad 1 \leq i \leq m_1,$$

we have

$$\text{FDR}(\mu^0, \Sigma) \geq \alpha \frac{m_0}{m} \left[1 + \mathbb{1}_{\{m_1 > 0\}} \frac{1}{2} \sum_{\ell=m_0+1}^m \frac{1}{\ell} \right] - \Phi\left(-3\overline{\Phi}^{-1}(\alpha/2)\right). \quad (4.6)$$

The latter can be used to show that (ReinerBound) is not valid in general. For instance, for $m = 3$, $m_0 = 1$, $\alpha = 0.05$, (ReinerBound) is $(4/9)\alpha \leq 0.0223$. Meanwhile, (CohenBound) is $(17/36)\alpha \geq 0.0235$ and the remaining term is negligible: $\Phi\left(-3\overline{\Phi}^{-1}(\alpha/2)\right) \leq 3 \times 10^{-9}$. The assumption of Proposition 4.3 is satisfied in this particular example: $\tau_{m_0} \leq 2.4 \leq 6 \leq \tau_{m-1} + 2\tau_m$.

Proof. We use the notation of the proof of Proposition 4.1. The idea is to build $\mu \in \mathbb{R}^m$ with $\mu_i = 0$ for $1 \leq i \leq m_0$ and where the non-zero values of the μ_i 's are increasing and chosen so that, for $1 \leq \ell \leq m_1$,

$$\{\widehat{k} = m_0\} \supset \{-\tau_{m_0} - \mu_1 < Z < -\tau_{m_0}\} \quad \text{and} \quad \{\widehat{k} = m_0 + \ell\} \supset \{-\tau_{m_0+\ell-1} < Z < -\tau_{m_0+\ell}\}.$$

CHAPTER 4. ON BH PROCEDURE IN THE TWO-SIDED GAUSSIAN SETTING

Case $\ell = m_1$ Let us consider the first event for which $-\tau_{m-1} < Z < -\tau_m$. Then, we show that all the test statistics are larger than or equal to τ_m . First, $|Z| = -Z > \tau_m$. Second, for all $1 \leq i \leq m_1$, $Z + \mu_i \geq Z + \mu_1 \geq -\tau_{m-1} + \mu_1 = \tau_m$, by choosing

$$\mu_1 = \tau_m + \tau_{m-1}.$$

Case $\ell = m_1 - 1$ Now, consider the event for which $-\tau_{m-2} < Z < -\tau_{m-1}$. We have $|Z| = -Z > \tau_{m-1}$ and for all $2 \leq i \leq m_1$, $Z + \mu_i \geq Z + \mu_2 \geq -\tau_{m-2} + \mu_2 = \tau_{m-1}$, by choosing

$$\mu_2 = \tau_{m-1} + \tau_{m-2}.$$

We also have

$$\begin{aligned} Z + \mu_1 &< -\tau_{m-1} + \mu_1 = \tau_m \\ -Z - \mu_1 &< \tau_{m-2} - \mu_1 = \tau_{m-2} - \tau_{m-1} - \tau_m \leq \tau_m. \end{aligned}$$

because $\tau_{m-2} \leq \tau_{m_0} \leq 2\tau_m + \tau_{m-1}$ by assumption. This entails that the BH procedure rejects exactly $m - 1$ nulls.

Cases $2 \leq \ell \leq m_1 - 2$ we can deal with these cases in a similar way.

Case $\ell = 1$ The corresponding event is $\{-\tau_{m_0} < Z < -\tau_{m_0+1}\}$. In that case, $|Z| = -Z \geq \tau_{m_0+1}$ and we have

$$Z + \mu_{m_1} \geq -\tau_{m_0} + \mu_{m_1} = \tau_{m_0+1},$$

by letting $\mu_{m_1} = \tau_{m_0} + \tau_{m_0+1}$, so the number of rejections of BH procedure is higher than or equal to $m_0 + 1$. Let us prove that it actually does not make more than $m_0 + 1$ rejections. Let $k \geq m_0 + 2$. For all $i \leq m_1 - 1$, we have

$$Z + \mu_i < -\tau_{m_0+1} + \mu_i = -\tau_{m_0+1} + \tau_{m-i+1} + \tau_{m-i} \leq \tau_{m-i+1} \leq \tau_k,$$

for all i such that $k \leq m - i + 1$, that is, all $1 \leq i \leq m - k + 1$ (note that this bound is smaller than or equal to $m_1 - 1$). Furthermore, for all i ,

$$-Z - \mu_i < \tau_{m_0} - \mu_1 = \tau_{m_0} - \tau_m - \tau_{m-1} \leq \tau_m,$$

because $\tau_{m_0} \leq \tau_{m-1} + 2\tau_m$ by assumption. This implies that the number of test statistics larger than τ_k is smaller than $m - (m - k + 1) = k - 1$, for any $k \geq m_0 + 2$. Hence the BH procedure makes exactly $m_0 + 1$ rejections.

Case $\ell = 0$ Now consider the case $-\tau_{m_0+1} - \mu_1 < Z < -\tau_{m_0}$. Then we have $|Z| = -Z > \tau_{m_0}$ and for all $i \in \{1, \dots, m_1\}$,

$$Z + \mu_i \leq -\tau_{m_0} + \mu_{m_1} = \tau_{m_0+1}$$

and $-Z - \mu_i \leq -Z - \mu_1 < \tau_{m_0+1}$. This means that the BH procedure makes m_0 rejections.

Combining the events Now, remember that the event $\{Z > 0, \text{FDP} > 0\} = \{Z > \tau_m, \text{FDP} = m_0/m\}$, while $\{Z < 0, \text{FDP} > 0\} = \{Z < -\tau_m, \text{FDP} \geq m_0/m\}$. Hence, we have

$$\begin{aligned} \text{FDR} &= \mathbb{E}[\text{FDP} \mathbf{1}_{\{Z > \tau_m\}}] + \sum_{i=m_0+1}^m \mathbb{E}[\text{FDP} \mathbf{1}_{\{-\tau_{i-1} \leq Z < -\tau_i\}}] + \mathbb{E}[\text{FDP} \mathbf{1}_{\{Z < -\tau_{m_0}\}}] \\ &\geq \frac{m_0\alpha}{2m} + \sum_{i=m_0+1}^m \frac{m_0}{i} \mathbb{P}(-\tau_{i-1} \leq Z < -\tau_i) + \mathbb{P}[-\tau_{m_0+1} - \mu_1 < Z < -\tau_{m_0}] \\ &= \frac{m_0\alpha}{m} + \frac{m_0\alpha}{2} \sum_{i=m_0+1}^m \frac{1}{i} - \mathbb{P}[Z < -\tau_{m_0+1} - \mu_1]. \end{aligned}$$

Hence, the result follows from

$$-\tau_{m_0+1} - \mu_1 \leq -3\tau_m = -3\bar{\Phi}^{-1}(\alpha/2).$$

□

4.4.2 Extending scope of (ReinerBound)

Here, we extend the scope of (ReinerBound) that was only proved for $m = 2$ and Σ being the matrix with all entries equal to 1, see [79]. Let us recall that this bound is valid only when all means under the alternative are equal.

Proposition 4.4. *Let $\mu \in \mathbb{R}^m$ and Σ be the covariance matrix such that the submatrices $(\Sigma_{i,j})_{(i,j) \in \mathcal{H}_0^2}$ and $(\Sigma_{i,j})_{(i,j) \in \mathcal{H}_1^2}$ have all entries equal to 1, and $(\Sigma_{i,j})_{(i,j) \in \mathcal{H}_0 \times \mathcal{H}_1}$ has all entries equal to $\rho \in [-1, 1]$. Assume moreover that $\forall i \in \mathcal{H}_1, \mu_i = \delta$, for some $\delta \neq 0$. Then we have*

$$\text{FDR}(\mu, \Sigma) \leq \alpha \frac{m_0}{m} \left[1 + \frac{1}{2} \left(1 - \frac{m_0}{m} \right) \right] \leq \alpha. \quad (4.7)$$

Proof. Let us assume $\mathcal{H}_0 = \{1, \dots, m_0\}$ so that $(Y_1, \dots, Y_{m_0}, Y_{m_0+1}, \dots, Y_m)$ can be realized as $(Z, \dots, Z, W, \dots, W)$, where (W, Z) is a Gaussian vector with $\text{Var}(W) = \text{Var}(Z) = 1$. Now, since the alternative p -values are all equal, the only values \widehat{k} making the FDP positive are m_0 and m . This implies

$$\begin{aligned} \text{FDR}(\mu, \Sigma) &= \mathbb{P}(|Z| \geq \tau_{m_0}, \widehat{k} = m_0) + \frac{m_0}{m} \mathbb{P}(|Z| \geq \tau_m, \widehat{k} = m) \\ &= \alpha \frac{m_0}{m} + \mathbb{P}(|Z| \geq \tau_{m_0}, \widehat{k} = m_0) - \frac{m_0}{m} \mathbb{P}(|Z| \geq \tau_m, \widehat{k} = m_0) \\ &= \alpha \frac{m_0}{m} + \mathbb{P}(|Z| \geq \tau_{m_0}, |W + \mu| < \tau_m) - \frac{m_0}{m} \mathbb{P}(|Z| \geq \tau_m, |W + \mu| < \tau_m) \\ &= \alpha \frac{m_0}{m} + \mathbb{P}(|Z| \geq \tau_{m_0}, |W + \mu| \leq \tau_m) - \frac{m_0}{m} \mathbb{P}(|Z| \geq \tau_m, |W + \mu| \leq \tau_m). \end{aligned}$$

CHAPTER 4. ON BH PROCEDURE IN THE TWO-SIDED GAUSSIAN SETTING

Now note that $I = \{w \in \mathbb{R} : |w + \mu| \leq \tau_m\}$ is a closed interval of \mathbb{R} , thus I can be put under the form $A \cup B$, for $A = [-a, a]$, $a > 0$, B being either included in \mathbb{R}^- or \mathbb{R}^+ and $A \cap B = \emptyset$. This gives

$$\begin{aligned} \text{FDR}(\mu, \Sigma) &= \alpha \frac{m_0}{m} + \mathbb{P}(|Z| \geq \tau_{m_0}, |W| \leq a) - \frac{m_0}{m} \mathbb{P}(|Z| \geq \tau_m, |W| \leq a) \\ &\quad + \mathbb{P}(|Z| \geq \tau_{m_0}, W \in B) - \frac{m_0}{m} \mathbb{P}(|Z| \geq \tau_m, W \in B). \end{aligned}$$

Now since $(|Z|, |W|)$ is MTP_2 , it is PRDS and we have

$$\mathbb{P}(|W| \leq a \mid |Z| \geq \tau_{m_0}) \leq \mathbb{P}(|W| \leq a \mid |Z| \geq \tau_m).$$

As a result,

$$\begin{aligned} \text{FDR}(\mu, \Sigma) &\leq \alpha \frac{m_0}{m} + \mathbb{P}(|Z| \geq \tau_{m_0}, W \in B) - \frac{m_0}{m} \mathbb{P}(|Z| \geq \tau_m, W \in B) \\ &= \alpha \frac{m_0}{m} + \left(1 - \frac{m_0}{m}\right) \mathbb{P}(|Z| \geq \tau_{m_0}, W \in B). \end{aligned}$$

Now, since B is either included in \mathbb{R}^- or \mathbb{R}^+ and by symmetry of (W, Z) , we have

$$\text{FDR}(\mu, \Sigma) \leq \alpha \frac{m_0}{m} + \frac{1}{2} \left(1 - \frac{m_0}{m}\right) \mathbb{P}(|Z| \geq \tau_{m_0}).$$

This proves the result. □

Note that the proof of Proposition 4.4 relies explicitly on the two-sided Gaussian structure. This might be a good sign.

4.4.3 Numerical study of (ReinerBound) with exact calculation

Let us consider the case where $\mu \in \mathbb{R}^m$ and Σ is the ρ -equicorrelated matrix. Assume moreover that for all $i \in \mathcal{H}_1$, $\mu_i = \delta$, for some $\delta \neq 0$, and $\rho \geq 0$.

Roquain and Villers [81] obtained the exact formula for computing the FDR in the one-sided case. Here, adapting their result to the two-sided case enables us to then numerically investigate (ReinerBound) for some $\rho \neq 1$.

Let us begin by rewriting the observations $X_k, k = 1 \dots, m$. Indeed, the dependence structure mentioned at this section leads that for all $k \in \{1, \dots, m\}$, for all $\rho \geq 0$

$$X_k = \sqrt{\rho} \bar{\Phi}^{-1}(U) + \sqrt{1 - \rho} \bar{\Phi}^{-1}(U_k) + \mu_k,$$

where U, U_k are independent and identically random variables generated from an uniform distribution on $[0, 1]$. Using this decomposition, conditionally on $U = u$, the p -values

are a sequence of independent random variables having marginal laws given by, for $k \in \{1, \dots, m\}$,

$$P_k \sim \begin{cases} F_0(t_k, u) = \bar{\Phi}\left(\frac{\bar{\Phi}^{-1}(\frac{t_k}{2}) - \sqrt{\rho}\bar{\Phi}^{-1}(u)}{\sqrt{1-\rho}}\right) + \Phi\left(\frac{-\bar{\Phi}^{-1}(\frac{t_k}{2}) - \sqrt{\rho}\bar{\Phi}^{-1}(u)}{\sqrt{1-\rho}}\right) & \text{if } k \in \mathcal{H}_0 \\ F_1(t_k, u) = \bar{\Phi}\left(\frac{\bar{\Phi}^{-1}(\frac{t_k}{2}) - \sqrt{\rho}\bar{\Phi}^{-1}(u) - \mu_k}{\sqrt{1-\rho}}\right) + \Phi\left(\frac{-\bar{\Phi}^{-1}(\frac{t_k}{2}) - \sqrt{\rho}\bar{\Phi}^{-1}(u) - \mu_k}{\sqrt{1-\rho}}\right) & \text{if } k \in \mathcal{H}_1 \end{cases},$$

where $t_k = \frac{\alpha k}{m}$, $1 \leq k \leq m$ are the critical constants of the BH procedure. We also denote by $\mathbf{t} = (t_k)_{1 \leq k \leq m}$ this sequence.

In [81], Roquain and Villers provided the exact formula for computing the FDR for all $m \geq 2$ when the p -values under the true null hypotheses are uniformly generated. Here, we generalize their result to a non uniform distribution. Let \mathcal{F} be a set of continuous cumulative distribution functions on $[0, 1]$ to $[0, 1]$. For all $F, G \in \mathcal{F}$, for all $0 \leq k_0 \leq k$, let Ψ be the auxiliary function defined as follows

$$\Psi_{k, k_0, F, G}(\mathbf{t}) = \mathbb{P}(Z_{(1)} \leq t_1, \dots, Z_{(k)} \leq t_k),$$

where $(Z_i)_{1 \leq i \leq k}$ is a sequence of independent random variables such that for all $1 \leq i \leq k_0$, $Z_i \sim G$, for all $k_0 + 1 \leq i \leq k$, $Z_i \sim F$, with the convention $\Psi_{0,0,F,G}(\cdot) = 1$.

Thus, using the Ψ function, we can establish the exact formula for computing the FDR in the two-sided case. In particular, Theorem 3 of [82] is rewritten as follows.

Theorem 4.2. *For all $m \geq 2$, $m_0 \in \{1, \dots, m\}$, $0 \leq k \leq m$, $i \leq m_0$, $k - i \leq m - m_0$, $\rho \geq 0$, we have*

$$\begin{aligned} FDR(\mu, \Sigma) &= \sum_{k=1}^{m-1} \sum_{i \in \mathcal{H}_0} \frac{i}{k} \binom{m_0}{i} \binom{m - m_0}{k - i} \int_0^1 (F_0(t_k, u))^i (F_1(t_k, u))^{k-i} \\ &\quad \times \Psi_{m-k, m_0-i, \bar{F}_1, \bar{F}_0}(1 - t_m, \dots, 1 - t_k + 1) du + \frac{m_0}{m} \int_0^1 \Psi_{m, m_0, F_1, F_0}(t_1, \dots, t_m) du, \end{aligned}$$

where $\bar{F}_\varepsilon(t, u) = 1 - F_\varepsilon(1 - t, u)$ for all $t \in [0, 1]$ and $\varepsilon \in \{0, 1\}$.

Note that the true null hypotheses are associated to the first m_0 indexes.

In practice, we use the following recursion formula to compute $\Psi_{k, k_0, F, G}$, which is an adaptation of Proposition 1 of [82] to the two-sided case.

CHAPTER 4. ON BH PROCEDURE IN THE TWO-SIDED GAUSSIAN SETTING

Proposition 4.5. For all $F, G \in \mathcal{F}$, for all $0 \leq k_0 \leq k$, we have

$$\begin{aligned} \Psi_{k,k_0,F,G}(\mathbf{t}) &= G(t_k)^{k_0} F(t_k)^{k-k_0} - \sum_{\substack{0 \leq j_0 \leq j \leq k-2 \\ j_0 \leq k_0 \\ j-j_0 \leq k-k_0}} \binom{k_0}{j_0} \binom{k-k_0}{j-j_0} (G(t_k) - G(t_{j+1}))^{k_0-j_0} \\ &\quad \times (F(t_k) - F(t_{j+1}))^{k-k_0-j+j_0} \Psi_{j,j_0,F,G}(t_1, \dots, t_j). \end{aligned}$$

Figure 4.1 represents the FDR of the BH procedure as function of μ for some $\rho \neq 1$ where three hypotheses are tested, respectively four. This figure shows that (ReinerBound) still holds for $m = 3$ and $m = 4$ even if $\rho \neq 1$. More precisely, in the four cases represented in Figure 4.1, we have:

Case $m = 3, \mathbf{m}_0 = 1$ (ReinerBound) = $\alpha \frac{1}{3} \left[1 + \frac{1}{2} \left(1 - \frac{1}{3} \right) \right] \simeq 0.09$.

Case $m = 3, \mathbf{m}_0 = 2$ (ReinerBound) = $\alpha \frac{2}{3} \left[1 + \frac{1}{2} \left(1 - \frac{2}{3} \right) \right] \simeq 0.16$.

Case $m = 4, \mathbf{m}_0 = 1$ (ReinerBound) = $\alpha \frac{1}{4} \left[1 + \frac{1}{2} \left(1 - \frac{1}{4} \right) \right] \simeq 0.07$.

Case $m = 4, \mathbf{m}_0 = 3$ (ReinerBound) = $\alpha \frac{3}{4} \left[1 + \frac{1}{2} \left(1 - \frac{3}{4} \right) \right] \simeq 0.17$.

Therefore according to Figure 4.1, (ReinerBound) might not be sharp.

We obtain similar results when the number of tests is increasing ($m \leq 20$). (ReinerBound) might still hold and this work provides a numerical proof of (ReinerBound) for some cases when $\rho \neq 1$ (in the ρ -equicorrelated structure).

We conclude this section by summering the bounds mentioned in this chapter that are valid for the FDR of the BH procedure.

Two-sided equi-correlated Gaussian setting

Type of dependence	Indep PRDS	$\forall i, j \in \mathcal{H}_1,$ $\mu_i = \mu_j = \mu$ $* m = 2$ or $* \forall (i, j) \in \mathcal{H}_0^2,$ $\forall (i, j) \in \mathcal{H}_1^2, \Sigma_{ij} = 1;$ $\forall (i, j) \in \mathcal{H}_0 \times \mathcal{H}_1,$ $\Sigma_{ij} = \rho$	$\forall (i, j) \in \mathcal{H}^2,$ $\Sigma_{ij} = 1$	$\forall (i, j) \in \mathcal{H}_0^2,$ $\Sigma_{ij} = 1$ $\forall (i, j) \in \mathcal{H}^2 \setminus \mathcal{H}_0^2,$ $\Sigma_{ij} = \rho$	All types of dependence
Bounds	$\frac{\alpha m_0}{m} \leq$	(ReinerBound)	\leq (CohenBound)	\leq (CohenBound2)	\leq (BYBound)

FDR control

4.5. SUMMARY AND CONJECTURES

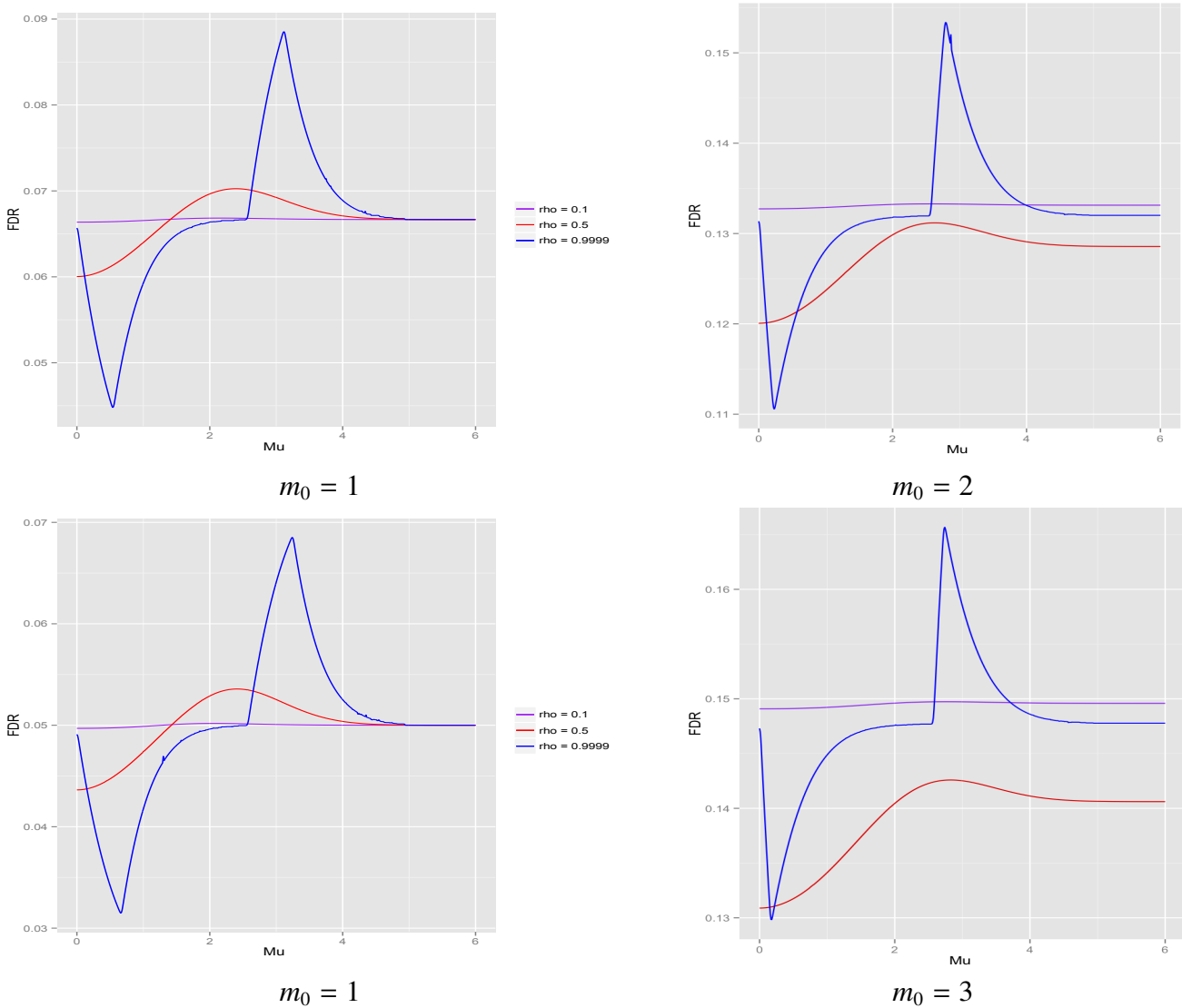


Figure 4.1 – $FDR(\mu, \rho)$ as function of μ for different values of ρ , $m = 3$ (on the top) and $m = 4$ (on the bottom), $\alpha = 0.2$.

4.5 Summary and conjectures

In this chapter, we have theoretically justified the FDR control of the BH procedure in particular configurations that induce non-positive dependence structure.

Considering the two-sided problem of testing means in a Gaussian setting, a simulation study suggests that [ReinerBound](#) is still valid for a structure dependence in which the correlations decrease exponentially. More generally, we conjecture the following result.

CHAPTER 4. ON BH PROCEDURE IN THE TWO-SIDED GAUSSIAN SETTING

Conjecture 4.1. *In the statistical model $(\mathbb{R}^m, \mathcal{B}(\mathbb{R}^m), \{\mathcal{N}_m(\mu, \Sigma); \mu \in \mathbb{R}^m, \Sigma \in \mathcal{M}_m(\mathbb{R})\})$, we consider two-sided hypothesis tests for means. Then the BH procedure controls the FDR at level α , namely,*

$$\forall \mu \in \mathbb{R}^m, \forall \Sigma \in \mathcal{M}_m(\mathbb{R}), FDR(\mu, \Sigma) \leq \alpha.$$

This conjecture implies another one as stated by several researchers (e.g. [83] or [84]):

Conjecture 4.2. *Let $X \sim \mathcal{N}_m(0, \Sigma)$. In the two-sided case, for all matrix Σ , the following inequality is satisfied,*

$$\mathbb{P}\left(\exists k \in \{1, \dots, m\}, p_k \leq \frac{\alpha k}{m}\right) \leq \alpha.$$

The previous inequality is known as Simes inequality. In the special case of positive dependence (called MTP_2 , see Chapter 3 for details), it has been proved by Sarkar [85].

Chapter 5

Asymptotic results

This chapter investigates the problem of the FWER or FDR control for a multiple testing procedure when the test statistics have an asymptotic Gaussian distribution. We present four existing asymptotic FWER controlling procedures that are particularly well-suited for testing correlation. Our contribution in this chapter, is to clarify or complete existing results in the literature and it is clearly mentioned in the text. We also provide some first results regarding FDR control.

Contents

5.1	Asymptotic Gaussian setting	97
5.2	Asymptotic control of the FWER	99
5.2.1	Subset pivotality fails	99
5.2.2	Single step methods	100
5.2.3	Step-down methods	106
5.2.4	Application to correlation tests	109
5.3	Towards an asymptotic control of the FDR	110
5.3.1	On BH procedure in the one-sided asymptotic Gaussian setting	110
5.3.2	Correlation tests	113

5.1 Asymptotic Gaussian setting

Let us consider the statistical model $(\mathfrak{X}^{\otimes n}, \mathcal{X}^{\otimes n}, P^{\otimes n})$. Let $\mathbb{X}^{(n)} = (X^{(1)}, \dots, X^{(n)})$ be the data generated from some unknown probability distribution P which is assumed to belong to a family of probability distributions \mathcal{P} .

Let $m \geq 2$ be some fixed number of null hypotheses to be tested. Let us consider the case

CHAPTER 5. ASYMPTOTIC RESULTS

in which interest focuses on a parameter vector

$$\theta(P) = (\theta_1(P), \dots, \theta_m(P)).$$

Observing $\mathbb{X}^{(n)}$, we aim at testing either the following two-sided null hypotheses, for all $i \in \{1, \dots, m\}$

$$H_{0,i} : "\theta_i = 0" \text{ against } H_{1,i} : "\theta_i \neq 0", \quad (5.1)$$

or the following one-sided null hypotheses,

$$H_{0,i} : "\theta_i = 0" \text{ against } H_{1,i} : "\theta_i > 0". \quad (5.2)$$

For all $P \in \mathcal{P}$, let $\mathcal{H}_0(P) = \{i \in \{1, \dots, m\} : \theta_i = 0\}$ be the index set of true null hypotheses, that is, the index set of all i such that $H_{0,i}$ is satisfied for P . Denote $m_0(P) = |\mathcal{H}_0(P)|$ its cardinality. We consider the following test statistics, for all $i \in \{1, \dots, m\}$

$$T_{n,i}(\mathbb{X}^{(n)}) = \sqrt{n} \widehat{\theta}_{n,i}, \quad (5.3)$$

where $\widehat{\theta}_{n,\cdot} = (\widehat{\theta}_{n,i})_{1 \leq i \leq m}$ is an estimate of $\theta(P)$ computed using the data $\mathbb{X}^{(n)}$. Throughout this chapter, we assume that $\widehat{\theta}_{n,\cdot}$ has an asymptotic Gaussian distribution, namely, for all $P \in \mathcal{P}$

$$\sqrt{n} (\widehat{\theta}_{n,\cdot} - \theta(P)) \xrightarrow[n \rightarrow +\infty]{\mathcal{L}} \mathcal{N}_m(0, \Sigma), \quad (5.4)$$

where we assume that Σ is invertible and $\Sigma_{ii} = 1$.

Let $(p_{n,i}(\mathbb{X}^{(n)}))_{1 \leq i \leq m}$ be a family of p -values resulting from each m individual test. Since under $H_{0,i}$, $T_{n,i} \xrightarrow[n \rightarrow +\infty]{\mathcal{L}} \mathcal{N}_m(0, 1)$, the asymptotic Gaussian assumption (5.4) gives rise to the p -value process (asymptotically valid) for the two-sided testing problem (5.1)

$$\forall i \in \{1, \dots, m\}, p_{n,i}(\mathbb{X}^{(n)}) = 2 \left[1 - \Phi \left(\left| T_{n,i}(\mathbb{X}^{(n)}) \right| \right) \right], \quad (5.5)$$

and for the one-sided testing problem (5.2), we have

$$\forall i \in \{1, \dots, m\}, p_{n,i}(\mathbb{X}^{(n)}) = 1 - \Phi \left(T_{n,i}(\mathbb{X}^{(n)}) \right). \quad (5.6)$$

Since we consider an asymptotic p -value process, we can only get asymptotic results in terms of control of the errors. A multiple testing procedure \mathcal{R} is said to asymptotically control the FWER at level α if for all $P \in \mathcal{P}$,

$$\limsup_{n \rightarrow +\infty} \text{FWER}(\mathcal{R}, P) \leq \alpha. \quad (5.7)$$

In the same way, a multiple testing procedure \mathcal{R} is said to asymptotically control the FDR at level α if for all $P \in \mathcal{P}$,

$$\limsup_{n \rightarrow +\infty} \text{FDR}(\mathcal{R}, P) \leq \alpha. \quad (5.8)$$

In what follows, for the sake of clarity, the test statistic is denoted by $T_{n,i}$ instead of $T_{n,i}(\mathbb{X}^{(n)})$ and the p -values are denoted by $p_{n,i}$ instead of $p_{n,i}(\mathbb{X}^{(n)})$ even if these quantities depend on the observations $\mathbb{X}^{(n)}$.

5.2 Asymptotic control of the FWER

Throughout this section we consider the two-sided testing problem (5.1).

5.2.1 Subset pivotality fails

One classical requirement to construct a procedure that controls asymptotically the FWER is the "asymptotic subset pivotality" (aSP) condition defined by Westfall and Young [86]. This condition is the existence of least favorable distribution for all configurations of true hypotheses. First, let us define this condition in a more general framework in which the asymptotic Gaussian hypothesis (5.4) does not necessarily hold. Assume that, for all $P \in \mathcal{P}$, $(T_{n,i})_{i \in \mathcal{H}_0(P)} \xrightarrow{\mathcal{L}} Z \in \mathbb{R}^{\mathcal{H}_0(P)}$. Let $Q(P)$ be the distribution of Z . Then, formally speaking, (aSP) can take the following form,

$$\exists P_0 \in \mathcal{P}, \forall P \in \mathcal{P}, Q(P) = Q(P_0). \quad (\text{aSP})$$

In a few words, the subset pivotality condition requires that the joint distribution of any subvector of p -values is unaffected by the truth or falsehood of the remaining hypotheses. Therefore, under this condition, thresholding is done under the full null hypothesis. A procedure, which asymptotically controls the FWER at level α , can be obtained using the (aSP) condition. Define the rejection set,

$$\mathcal{R}_\alpha^{SP} = \{i \in \{1, \dots, m\} : |T_{n,i}| > t_{n,\alpha}^{SP}\},$$

where $t_{n,\alpha}^{SP}$ is the $(1 - \alpha)$ -quantile of $\mathcal{L}_{X \sim P_0}(\max\{|T_{n,i}|, 1 \leq i \leq m\})$. Then, for all $P \in \mathcal{P}$,

$$\begin{aligned} \text{FWER}(\mathcal{R}_\alpha^{SP}, P) &= \mathbb{P}_{X \sim P}(\exists i \in \mathcal{H}_0(P) : |T_{n,i}| > t_{n,\alpha}^{SP}) \\ &= \mathbb{P}_{X \sim P}(\max\{|T_{n,i}|, i \in \mathcal{H}_0(P)\} > t_{n,\alpha}^{SP}). \end{aligned}$$

Finally, using the (aSP) condition, and since $\max\{|T_{n,i}|, i \in \mathcal{H}_0(P)\} \leq \max\{|T_{n,i}|, 1 \leq i \leq m\}$,

$$\begin{aligned} \limsup_{n \rightarrow +\infty} \text{FWER}(\mathcal{R}_\alpha^{SP}, P) &\leq \mathbb{P}_{X \sim P_0}(\max\{|T_{n,i}|, 1 \leq i \leq m\} > t_{n,\alpha}^{SP}) \\ &\leq \alpha. \end{aligned}$$

Hence the procedure \mathcal{R}_α^{SP} controls asymptotically the FWER at level α .

Unfortunately, checking the (aSP) condition is difficult in general. In particular, the asymptotic subset pivotality fails for correlation tests. As mentioned in [60], consider the

CHAPTER 5. ASYMPTOTIC RESULTS

case of a Gaussian vector $X \sim \mathcal{N}_3(0, \Gamma(u))$ where the covariance matrix Γ depends on a parameter u , that is,

$$\Gamma(u) = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & u \\ 0 & u & 1 \end{pmatrix}, u \in [-1, 1] \setminus \{0\},$$

so that $\mathcal{H}_0(P) = \{\{1, 2\}, \{1, 3\}\}$. If the asymptotic subset pivotality holds, by using a result of Aitkin [60] (see Chapter 2, Proposition 2.3), there exists $\Gamma_0 \in \mathcal{M}_3^{cov}(\mathbb{R})$ such that for all $u \in [-1, 1] \setminus \{0\}$, $\mathcal{N}_3(0, \Omega(\Gamma(u))) = \mathcal{N}_3(0, \Omega(\Gamma_0))$. That is, there exists $\Gamma_0 \in \mathcal{M}_3^{cov}(\mathbb{R})$ such that,

$$\forall u \in [-1, 1] \setminus \{0\}, \begin{pmatrix} 1 & u & 0 \\ u & 1 & 0 \\ 0 & & (1 - u^2)^2 \end{pmatrix} = \Omega(\Gamma_0).$$

This gives a contradiction because the correlation under \mathcal{H}_0 depends on the signal under \mathcal{H}_1 , the index set of false null hypotheses. Therefore, the procedures of Westfall and Young fail to provide an asymptotically (since the asymptotic normality of the vector of correlation is used) FWER control for correlation tests.

The next section presents four methods which provide an asymptotic control of the FWER for the two-sided testing problem (5.1) without using the subset pivotality assumption.

5.2.2 Single step methods

Bonferroni

The Bonferroni procedure [3] is the most classical example of FWER control.

Method 5.1 (Bonferroni). *The Bonferroni multiple testing procedure is defined by*

$$\mathcal{R}_\alpha^{bonf} = \left\{ i \in \{1, \dots, m\} : p_{n,i} \leq \frac{\alpha}{m} \right\}. \quad (5.9)$$

Theorem 5.1. *For the two-sided testing problem (5.1) based on the asymptotic p -value process (5.5), the method $\mathcal{R}_\alpha^{bonf}$ provides an asymptotic control of the FWER at level α , that is, for all $P \in \mathcal{P}$,*

$$\lim_{n \rightarrow +\infty} \text{FWER}(\mathcal{R}_\alpha^{bonf}, P) \leq \alpha. \quad (5.10)$$

Proof. For all $P \in \mathcal{P}$,

$$\text{FWER}(\mathcal{R}_\alpha^{bonf}, P) = \mathbb{P} \left(\bigcup_{i \in \mathcal{H}_0(P)} \{p_{n,i} \leq \alpha/m\} \right).$$

5.2. ASYMPTOTIC CONTROL OF THE FWER

Thus,

$$\begin{aligned} \lim_{n \rightarrow +\infty} \text{FWER}(\mathcal{R}_\alpha^{\text{bonf}}, P) &\leq \sum_{i \in \mathcal{H}_0(P)} \alpha/m \\ &\leq \frac{\alpha m_0(P)}{m} \\ &\leq \alpha. \end{aligned}$$

□

This control does not require any assumptions about dependence among the p -values but under strong dependence the Bonferroni correction is extremely conservative.

Šidák

As mentioned by Westfall and Young [86], an asymptotic FWER controlling procedure can be derived by Šidák's inequality [71].

Theorem 5.2 (Šidák's inequality). *Let X be a random vector having an m -multivariate normal distribution with zero mean values and invertible covariance matrix. Then X satisfies the following inequality, for every positive constant $b \in \mathbb{R}_+^m$,*

$$\mathbb{P}(|X_1| \leq b_1, \dots, |X_m| \leq b_m) \geq \prod_{i=1}^m \mathbb{P}(|X_i| \leq b_i). \quad (5.11)$$

For the specific case of correlation testing, Drton and Perlmann [55] used this inequality to construct a procedure that asymptotically controls the FWER for the problem (5.1).

Method 5.2 (Šidák). *Let $c_\alpha^s = \Phi^{-1}\left(\frac{1}{2}(1 - \alpha)^{1/m} + \frac{1}{2}\right) > 0$. The Šidák's multiple testing procedure is defined by*

$$\mathcal{R}_\alpha^s = \{i \in \{1, \dots, m\} : |T_{n,i}| > c_\alpha^s\}. \quad (5.12)$$

Theorem 5.3. *For the two-sided testing problem (5.1) for which the asymptotic Gaussian assumption (5.4) holds, the method \mathcal{R}_α^s provides an asymptotic control of the FWER at level α , namely, for all invertible Σ ,*

$$\lim_{n \rightarrow +\infty} \text{FWER}(\mathcal{R}_\alpha^s, \Sigma) \leq \alpha. \quad (5.13)$$

Our contribution in this section is to clarify and complete the proof of this theorem. It requires the following previous result.

Proposition 5.1. *For all $x > 0$, for all $P \in \mathcal{P}$,*

$$\lim_{n \rightarrow +\infty} \left\{ \mathbb{P} \left(\sqrt{n} \left| \widehat{\theta}_{n,i} - \theta_i(P) \right| \leq x, \text{ for all } i \in \{1, \dots, m\} \right) \right\} \geq (2\Phi(x) - 1)^m. \quad (5.14)$$

CHAPTER 5. ASYMPTOTIC RESULTS

Proof. Let Z be an m -dimensional random vector with multivariate Gaussian distribution $\mathcal{N}_m(0, \Sigma)$. Let $B = \{z \in \mathbb{R}^m : \|z\|_\infty \leq x\}$. Denote ∂B the frontier of the set B . Since for all $x > 0$,

$$\begin{aligned} \mathbb{P}(Z \in \partial B) &= \mathbb{P}(\|Z\|_\infty = x) \\ &= \mathbb{P}\left(\sup_{1 \leq j \leq m} |Z_j| = x\right) \\ &= \mathbb{P}(\exists j \in \{1, \dots, m\} : |Z_j| = x) \\ &\leq \sum_{j=1}^m \mathbb{P}(|Z_j| = x) \\ &\leq 0, \end{aligned}$$

then, by using Portmanteau's lemma [87], it follows that for all $P \in \mathcal{P}$,

$$\lim_{n \rightarrow +\infty} \mathbb{P}\left(\sqrt{n} \|\widehat{\theta}_{n,\cdot} - \theta(P)\|_\infty \leq x\right) = \mathbb{P}(\|Z\|_\infty \leq x).$$

Finally, applying (5.11), we get, for all $P \in \mathcal{P}$,

$$\lim_{n \rightarrow +\infty} \mathbb{P}\left(\sqrt{n} \|\widehat{\theta}_{n,\cdot} - \theta(P)\|_\infty \leq x\right) \geq \left[\mathbb{P}(|Z_1| \leq x)\right]^m,$$

and (5.14) holds. \square

We now are in position to prove Theorem 5.3.

Proof of Theorem 5.3. Using Proposition 5.1, we can state that for all invertible Σ , for all $P \in \mathcal{P}$,

$$\begin{aligned} \text{FWER}(\mathcal{R}_\alpha^s, \Sigma) &= \mathbb{P}(\exists i \in \mathcal{H}_0(P), |T_{n,i}| > c_\alpha^s) \\ &= 1 - \mathbb{P}(\forall i \in \mathcal{H}_0(P), |T_{n,i}| \leq c_\alpha^s) \\ &= 1 - \mathbb{P}(\|(T_{n,i})_{i \in \mathcal{H}_0(P)}\|_\infty \leq c_\alpha^s) \\ &= 1 - \mathbb{P}\left(\sqrt{n} \left|\widehat{\theta}_{n,i} - \theta_i(P)\right| \leq c_\alpha^s, \forall i \in \mathcal{H}_0(P)\right) \end{aligned}$$

Since $\|(T_{n,i})_{i \in \mathcal{H}_0}\|_\infty \leq \|(T_{n,i})_{1 \leq i \leq m}\|_\infty$, we finally get that for all invertible Σ , for all $P \in \mathcal{P}$,

$$\text{FWER}(\mathcal{R}_\alpha^s, \Sigma) \leq 1 - \mathbb{P}\left(\sqrt{n} \left|\widehat{\theta}_{n,i} - \theta_i(P)\right| \leq c_\alpha^s, \forall i \in \{1, \dots, m\}\right),$$

and using (5.14),

$$\begin{aligned} \lim_{n \rightarrow +\infty} \text{FWER}(\mathcal{R}_\alpha^s, \Sigma) &\leq 1 - (2\Phi(c_\alpha^s) - 1)^m \\ &\leq \alpha. \end{aligned}$$

\square

5.2. ASYMPTOTIC CONTROL OF THE FWER

The procedure is valid for any structured dependencies, as soon as the inequality (5.11) holds, which is true in Gaussian setting. Note that this inequality provides the same way of constructing the critical values as if subset pivotality holds.

The Šidák's procedure is less conservative than the Bonferroni procedure. This comparison is illustrated in table 2.2 of Westfall and Young [86]. Moreover, using Taylor-series expansion, they show that the difference between the two adjustments becomes smaller with smaller α and becomes larger with larger m^{-1} .

Non parametric bootstrap

Romano and Wolf [88] propose an asymptotic FWER controlling procedure which only requires a monotonic assumption on the family of thresholds. This allows us to be free of subset pivotality assumption (which does not hold for testing correlations).

Method 5.3 (Romano-Wolf). *For all $C \subset \{1, \dots, m\}$, let $t_{n,\alpha}(\Sigma, C)$ be the $(1 - \alpha)$ -quantile of $\mathcal{L}(\|\mathcal{N}(0, \Sigma)|_C\|_\infty)$, where $\mathcal{N}(0, \Sigma)|_C$ is the restriction of the Gaussian distribution $\mathcal{N}_m(0, \Sigma)$ on C , namely $\mathcal{N}(0, (\Sigma)_{i,i' \in C})$. The Romano-Wolf's multiple testing procedure is defined by*

$$\mathcal{R}_\alpha^{\text{rw}} = \{i \in \{1, \dots, m\} : |T_{n,i}| > \widehat{t}_{n,\alpha}(\Sigma, C_m)\}, \quad (5.15)$$

where $C_m = \{1, \dots, m\}$ and $\widehat{t}_{n,\alpha}(\Sigma, C_m)$ is computed using bootstrap resamples of $\mathbb{X}^{(n)}$ where a bootstrap resample from $\mathbb{X}^{(n)}$ is denoted by $\mathbb{X}^{(n)*}$ and defined as an n iid sample from the empirical distribution of $\mathbb{X}^{(n)}$.

Note that this method closely relies on its ability to approximate the joint distribution of the test statistics.

Theorem 5.4. *Assume that for any metric d metrizing the weak convergence on $\mathbb{R}^{m_0(P)}$,*

$$d\left(\mathcal{L}\left(\left(T_{n,i}(\mathbb{X}^{(n)*})\right)_{i \in \mathcal{H}_0(P)} \mid \mathbb{X}^{(n)*}\right), \mathcal{L}\left(\left(T_{n,i}(\mathbb{X}^{(n)})\right)_{i \in \mathcal{H}_0(P)}\right)\right) \xrightarrow{\mathbb{P}} 0, \quad (5.16)$$

where $\mathcal{L}\left(\left(T_{n,i}(\mathbb{X}^{(n)*})\right)_{i \in \mathcal{H}_0(P)} \mid \mathbb{X}^{(n)*}\right)$ denotes the conditional distribution of $\left(T_{n,i}(\mathbb{X}^{(n)*})\right)_{i \in \mathcal{H}_0(P)}$ given $\mathbb{X}^{(n)*}$. Then, for the two-sided testing problem (5.1) for which the asymptotic Gaussian assumption (5.4) holds, the method $\mathcal{R}_\alpha^{\text{rw}}$ provides an asymptotic control of the FWER at level α , that is, for all Σ invertible,

$$\limsup_{n \rightarrow +\infty} \text{FWER}(\mathcal{R}_\alpha^{\text{rw}}, \Sigma) \leq \alpha. \quad (5.17)$$

Proof. First, since under P , the matrix Σ is assumed to be invertible, the joint distribution of $\left(T_{n,i}(\mathbb{X}^{(n)})\right)_{i \in \mathcal{H}_0(P)}$ has a nondegenerate limit law. Then the result follows from Theorem 7 of [88]. \square

Parametric bootstrap

Drton and Perlman [22] mention a parametric bootstrap method for testing (5.1) when interest focuses on partial correlation coefficients. The only difference between this method and the Romano-Wolf's method is that an estimation of the matrix Σ is used. We denote by $\widehat{\Sigma}$ such an estimator. We report this method in our framework.

Method 5.4 ($\max T_\infty$). Let $t_{n,\alpha}(\widehat{\Sigma})$ be the $(1 - \alpha)$ -quantile of $\mathcal{L}(\|\mathcal{N}_m(0, \widehat{\Sigma})\|_\infty)$. The $\max T_\infty$'s multiple testing procedure is defined by

$$\mathcal{R}_\alpha^{\max T_\infty} = \{i \in \{1, \dots, m\} : |T_{n,i}| > t_{n,\alpha}(\widehat{\Sigma})\}, \quad (5.18)$$

where $t_{n,\alpha}(\widehat{\Sigma})$ is computed using (simulated) samples of $\mathcal{N}_m(0, \widehat{\Sigma})$.

In this section, our contribution is to prove the following theorem.

Theorem 5.5. Assume that $\widehat{\Sigma}$ is a consistent estimator of Σ , that is,

$$\widehat{\Sigma} \xrightarrow{\mathbb{P}} \Sigma. \quad (5.19)$$

Then, for the two-sided testing problem (5.1) for which the asymptotic Gaussian assumption (5.4) holds, the method $\mathcal{R}_\alpha^{\max T_\infty}$ provides an asymptotic control of the FWER at level α , namely, for all invertible Σ ,

$$\limsup_{n \rightarrow +\infty} \text{FWER}(\mathcal{R}_\alpha^{\max T_\infty}, \Sigma) \leq \alpha. \quad (5.20)$$

Proof. For all $P \in \mathcal{P}$, we denote by $\mathcal{N}(0, \Sigma)|_{\mathcal{H}_0(P)}$ the restriction of the Gaussian distribution $\mathcal{N}_m(0, \Sigma)$ on $\mathcal{H}_0(P)$, namely $\mathcal{N}(0, (\Sigma)_{i,i' \in \mathcal{H}_0(P)})$. Let $P \in \mathcal{P}$.

First, since $(T_{n,i})_{i \in \mathcal{H}_0(P)} \xrightarrow[n \rightarrow +\infty]{\mathcal{L}} \mathcal{N}(0, \Sigma)|_{\mathcal{H}_0(P)}$ and $x \in \mathbb{R}^{\mathcal{H}_0} \mapsto \|x\|_\infty$ is continuous, by the continuous mapping theorem,

$$\|(T_{n,i})_{i \in \mathcal{H}_0(P)}\|_\infty \xrightarrow[n \rightarrow +\infty]{\mathcal{L}} \|\mathcal{N}(0, \Sigma)|_{\mathcal{H}_0(P)}\|_\infty. \quad (5.21)$$

Second, let us show that

$$t_{n,\alpha}(\widehat{\Sigma}) \xrightarrow{\mathbb{P}} t_{n,\alpha}(\Sigma). \quad (5.22)$$

We introduce for all $x \in \mathbb{R}$, the cumulative distribution function φ_x , defined by

$$\varphi_x : M \in \mathcal{M}_{\mathcal{H} \times \mathcal{H}}^{\text{cov}}(\mathbb{R}) \mapsto \mathbb{P}(\|M^{1/2}\xi\|_\infty \leq x),$$

where $\xi \sim \mathcal{N}_m(0, I_m)$, namely,

$$\varphi_x(\widehat{\Sigma}) \xrightarrow{\mathbb{P}} \varphi_x(\Sigma). \quad (5.23)$$

5.2. ASYMPTOTIC CONTROL OF THE FWER

Suppose that φ_x is strictly increasing. Using the assumption (5.19) and the continuous mapping theorem, it is sufficient to prove the continuity of φ_x . Let (M_n) be a sequence of matrices such that $(M_n) \xrightarrow[n \rightarrow +\infty]{} M$. Then $M_n^{1/2} \xrightarrow[n \rightarrow +\infty]{} M^{1/2}$ [89] and $M_n^{1/2}\xi \xrightarrow[n \rightarrow +\infty]{\mathcal{L}} M^{1/2}\xi$. Hence, the continuous mapping theorem involves $\|M_n^{1/2}\xi\|_\infty \xrightarrow[n \rightarrow +\infty]{\mathcal{L}} \|M^{1/2}\xi\|_\infty$. In addition, applying Portmanteau's lemma [87], we obtain that $\varphi_x(M_n) \xrightarrow[n \rightarrow +\infty]{} \varphi_x(M)$, that is, φ_x is continuous. Therefore, for all $\varepsilon > 0$,

$$\varphi_{t_{n,\alpha}(\Sigma)-\varepsilon}(\widehat{\Sigma}) \xrightarrow{\mathbb{P}} \varphi_{t_{n,\alpha}(\Sigma)-\varepsilon}(\Sigma). \quad (5.24)$$

For all $\varepsilon > 0$,

$$\begin{aligned} \mathbb{P}(t_{n,\alpha}(\widehat{\Sigma}) \leq t_{n,\alpha}(\Sigma) - \varepsilon) &= \mathbb{P}(\varphi_{t_{n,\alpha}(\Sigma)-\varepsilon}(\widehat{\Sigma}) \geq 1 - \alpha) \\ &= \mathbb{P}(\varphi_{t_{n,\alpha}(\Sigma)-\varepsilon}(\widehat{\Sigma}) - \varphi_{t_{n,\alpha}(\Sigma)-\varepsilon}(\Sigma) \geq 1 - \alpha - \varphi_{t_{n,\alpha}(\Sigma)-\varepsilon}(\Sigma)). \end{aligned}$$

Since for all $\varepsilon > 0$, $\varphi_{t_{n,\alpha}(\Sigma)-\varepsilon}(\Sigma) = 1 - \alpha - \delta$, for $\delta > 0$ because φ is supposed to be strictly increasing, the right-hand-side is bounded by

$$\mathbb{P}(\varphi_{t_{n,\alpha}(\Sigma)-\varepsilon}(\widehat{\Sigma}) - \varphi_{t_{n,\alpha}(\Sigma)-\varepsilon}(\Sigma) \geq \delta) \leq \mathbb{P}(|\varphi_{t_{n,\alpha}(\Sigma)-\varepsilon}(\widehat{\Sigma}) - \varphi_{t_{n,\alpha}(\Sigma)-\varepsilon}(\Sigma)| \geq \delta).$$

Then, convergence (5.22) follows from (5.24).

We now finally state the asymptotic FWER control for the procedure $\mathcal{R}^{\max T_\infty}$. By definition,

$$\text{FWER}(\mathcal{R}_\alpha^{\max T_\infty}, \Sigma) = \mathbb{P}(\|(T_{n,i})_{i \in \mathcal{H}_0(P)}\|_\infty \geq t_{n,\alpha}(\widehat{\Sigma})).$$

For all $\delta > 0$, this quantity is bounded by

$$\mathbb{P}(\|(T_{n,i})_{i \in \mathcal{H}_0(P)}\|_\infty > t_{n,\alpha}(\Sigma) - \delta) + \mathbb{P}(t_{n,\alpha}(\widehat{\Sigma}) \leq t_{n,\alpha}(\Sigma) - \delta).$$

Then, using (5.21) and (5.22), we have that

$$\limsup_{n \rightarrow +\infty} \text{FWER}(\mathcal{R}_\alpha^{\max T_\infty}, \Sigma) \leq \mathbb{P}(\|\mathcal{N}(0, \Sigma)|_{\mathcal{H}_0(P)}\|_\infty > t_{n,\alpha}(\Sigma) - \delta).$$

The asymptotic control is obtained by letting $\delta \rightarrow 0$. □

This result is available for any test statistic with asymptotic normal distribution. This is thus well-suited for applications where the distribution P may be unknown. Moreover, the quality of $\max T_\infty$ procedure relies on the quality of estimation of Σ .

5.2.3 Step-down methods

Single-step methods can be too conservative and improved by step-wise methods, while still preserving FWER control. The idea behind this refinement is that the critical values become smaller with each step for the remaining hypotheses, then step-down versions increase possibly the number of discoveries and the power of single-step versions can thus be improved. Let R_C be a rejected index set and $A_C = \{1, \dots, m\} \setminus R_C$ for some arbitrary $C \subset \{1, \dots, m\}$.

Step-down Algorithm.

- Step 0 : $C_0 = \{1, \dots, m\}$;
 - Step j ($j \geq 1$) : $C_j = A_{C_{j-1}}$.
- If $C_j = C_{j-1}$ stop and $\widehat{C} = C_j$. Otherwise go to step $j + 1$.

If a single step method provides an asymptotic control of the FWER, the following proposition provides sufficient conditions under which its step-down version preserves this control.

Proposition 5.2. *Let $(R_C)_{C \subset \mathcal{H}}$ be a family of multiple testing procedure such that:*

- For all $P \in \mathcal{P}$,

$$\limsup_{n \rightarrow +\infty} FWER(R_{\mathcal{H}_0(P)}, P) \leq \alpha, \quad (5.25)$$

- For all $C \subset \{1, \dots, m\}$, $C \mapsto A_C$ is increasing in C , that is,

$$\forall C, C' \subset \{1, \dots, m\}, C \subset C' \Rightarrow A_C \subset A_{C'}. \quad (5.26)$$

Then, for all $P \in \mathcal{P}$,

$$\limsup_{n \rightarrow +\infty} FWER(R_{\widehat{C}}, P) \leq \alpha.$$

Proof. Let $P \in \mathcal{P}$. There exists an event \mathcal{E} , such that

$$\mathbb{P}(\mathcal{E}) \xrightarrow[n \rightarrow +\infty]{} 1 - \alpha, \quad (5.27)$$

and on \mathcal{E} , $\mathcal{H}_0(P) \subset A_{\mathcal{H}_0(P)}$, by (5.25).

For all $j \geq 0$, $\mathcal{H}_0(P) \subset C_j$ implies that $A_{\mathcal{H}_0(P)} \subset A_{C_j} = C_{j+1}$ by (5.26). Thus on \mathcal{E} , we have $\mathcal{H}_0(P) \subset C_{j+1}$. Since $\mathcal{H}_0(P) \subset C_0 = \{1, \dots, m\}$, on \mathcal{E} , we have $\mathcal{H}_0(P) \subset C_j$ for all $j \geq 0$. It results that on \mathcal{E} , $\mathcal{H}_0(P) \subset A_{\widehat{C}}$.

Finally, using the convergence (5.27) and the equality $\mathbb{P}(\mathcal{H}_0(P) \subset A_{\widehat{C}}) = 1 - FWER(R_{\widehat{C}}, P)$, Proposition 5.2 is proved. \square

Bonferroni step-down

Method 5.5 (Bonferroni step-down). *The Bonferroni step-down multiple testing procedure is defined by, for all $C \subset \{1, \dots, m\}$,*

$$R_C^{bonf} = \left\{ i \in \{1, \dots, m\} : p_{n,i} \leq \frac{\alpha}{|C|} \right\}, \quad (5.28)$$

and the result is obtained as $\mathcal{R}_\alpha^{bonf, sd} = \left\{ i \in \{1, \dots, m\} : p_{n,i} \leq \frac{\alpha}{|C|} \right\}$.

Note that the Bonferroni step-down procedure provides the Holm procedure.

Corollary 5.1. *For the two-sided testing problem (5.1) for which the asymptotic Gaussian assumption (5.4) holds, the method $\mathcal{R}_\alpha^{bonf, sd}$ is an asymptotic FWER controlling procedure at level α , that is, for all $P \in \mathcal{P}$,*

$$\lim_{n \rightarrow +\infty} \text{FWER}(\mathcal{R}_\alpha^{bonf, sd}, P) \leq \alpha. \quad (5.29)$$

Proof. $\mathcal{R}_\alpha^{bonf, sd}$ satisfies (5.25) by Theorem 5.1 and for all $C \subset C'$, $\frac{\alpha}{|C|} \geq \frac{\alpha}{|C'|}$, that is, $\mathcal{R}_\alpha^{bonf, sd}$ satisfies (5.26). \square

Šidák step-down

Method 5.6 (Šidák step-down). *The Šidák step-down multiple testing procedure is defined by, for all $C \subset \{1, \dots, m\}$,*

$$R_C^{s, sd} = \left\{ i \in \{1, \dots, m\} : p_{n,i} \leq 1 - (1 - \alpha)^{1/|C|} \right\}, \quad (5.30)$$

and the result is $\mathcal{R}_\alpha^{s, sd} = \left\{ i \in \{1, \dots, m\} : p_{n,i} \leq 1 - (1 - \alpha)^{1/|C|} \right\}$.

Corollary 5.2. *For the two-sided testing problem (5.1) for which the asymptotic Gaussian assumption (5.4) holds, the method $\mathcal{R}_\alpha^{s, sd}$ is an asymptotic FWER controlling procedure at level α , namely, for all invertible Σ ,*

$$\text{FWER}(\mathcal{R}_\alpha^{s, sd}, \Sigma) \leq \alpha. \quad (5.31)$$

Proof. $\mathcal{R}_\alpha^{s, sd}$ satisfies (5.25) by Theorem 5.3 and for all $C \subset C'$, $1 - (1 - \alpha)^{1/|C|} \geq 1 - (1 - \alpha)^{1/|C'|}$, that is, $\mathcal{R}_\alpha^{s, sd}$ satisfies (5.26). \square

Non parametric bootstrap step-down

Method 5.7 (Romano-Wolf step-down). For all $C \subset \{1, \dots, m\}$, let $t_{n,\alpha}(\Sigma, C)$ be the $(1 - \alpha)$ -quantile of $\mathcal{L}(\|\mathcal{N}(0, \Sigma)|_C\|_\infty)$. The Romano-Wolf's multiple testing procedure is defined by, for all $C \subset \{1, \dots, m\}$,

$$R_C^{rw, sd} = \left\{ i \in \{1, \dots, m\} : |T_{n,i}| \geq \widehat{t}_{n,\alpha}(\Sigma, C) \right\}, \quad (5.32)$$

and the result is $\mathcal{R}_\alpha^{rw, sd} = \left\{ i \in \{1, \dots, m\} : |T_{n,i}| \geq \widehat{t}_{n,\alpha}(\widehat{\Sigma}, \widehat{C}) \right\}$.

Corollary 5.3. Assume that assumption (5.16) holds. Then, for the two-sided testing problem (5.1) for which the asymptotic Gaussian assumption (5.4) holds, the method $\mathcal{R}^{rw, sd}$ provides an asymptotic control of the FWER at level α , namely, for all invertible Σ ,

$$\limsup_{n \rightarrow +\infty} \text{FWER}(\mathcal{R}^{rw, sd}, \Sigma) \leq \alpha. \quad (5.33)$$

Proof. Inequality (5.25) holds by Theorem 5.4 and it is easy to see that the application $C \mapsto t_{n,\alpha}(\Sigma, C)$ is nondecreasing, that is

$$\forall C, C' \subset \{1, \dots, m\} \text{ such that } C \subset C', \text{ we have } t_{n,\alpha}(\Sigma, C) \leq t_{n,\alpha}(\Sigma, C').$$

□

Parametric bootstrap step-down

For all $C \subset \{1, \dots, m\}$, let $t_{n,\alpha}(\Sigma, C)$ be the $(1 - \alpha)$ -quantile of $\mathcal{L}(\|\mathcal{N}(0, \Sigma)|_C\|_\infty)$.

Method 5.8 ($\max T_\infty$ step-down). The $\max T_\infty^{sd}$'s multiple testing procedure is defined by, for all $C \subset \{1, \dots, m\}$,

$$R_C^{\max T_\infty, sd} = \left\{ i \in \{1, \dots, m\} : |T_{n,i}| \geq t_{n,\alpha}(\widehat{\Sigma}, C) \right\}, \quad (5.34)$$

and the result is $\mathcal{R}_\alpha^{\max T_\infty, sd} = \left\{ i \in \{1, \dots, m\} : |T_{n,i}| \geq t_{n,\alpha}(\widehat{\Sigma}, \widehat{C}) \right\}$.

Corollary 5.4. Assume that $\widehat{\Sigma} \xrightarrow{\mathbb{P}} \Sigma$. Then, for the two-sided testing problem (5.1) for which the asymptotic Gaussian assumption (5.4) holds, the method $\mathcal{R}^{\max T_\infty, sd}$ provides an asymptotic control of the FWER at level α , namely, for all invertible Σ ,

$$\limsup_{n \rightarrow +\infty} \text{FWER}(\mathcal{R}^{\max T_\infty, sd}, \Sigma) \leq \alpha. \quad (5.35)$$

Proof. $\mathcal{R}_\alpha^{\max T_\infty, sd}$ satisfies (5.25) by Theorem 5.5 and the application $C \mapsto t_{n,\alpha}(\Sigma, C)$ is increasing in C . □

5.2.4 Application to correlation tests

From now on, we consider the problem of correlation tests in the Gaussian case. We fix $P = \mathcal{N}_d(0, \Gamma)$, where $\Gamma = (\rho_{ij})_{1 \leq i, j \leq d}$ and $\rho_{ii} = 1$ for all $i = 1, \dots, d$. Let us recall that $\mathcal{H} = \{\{i, j\}, 1 \leq i < j \leq d\}$ denotes the index set of pairwise correlation coefficients and $m = d(d-1)/2$ its cardinality.

Tests based on correlation coefficients

We aim at testing, for all $h \in \mathcal{H}$,

$$H_{0,h} : \rho_h = 0 \quad \text{against} \quad H_{1,h} : \rho_h \neq 0. \quad (5.36)$$

As seen in Chapter 2, the test can be based on the following test statistics: for all $h \in \mathcal{H}$,

$$T_{n,h}(\mathbb{X}^{(n)}) = \sqrt{n} \widehat{\rho}_{n,h}, \quad \text{where} \quad \widehat{\rho}_{n,h} = \frac{\frac{1}{n} \sum_{l=1}^n X_i^{(l)} X_j^{(l)} - \bar{X}_i \bar{X}_j}{\widehat{\sigma}(X_i) \widehat{\sigma}(X_j)}, \quad (5.37)$$

$\bar{X}_i = \frac{1}{n} \sum_{l=1}^n X_i^{(l)}$ and $\widehat{\sigma}(X_i) = \sqrt{\frac{1}{n} \sum_{l=1}^n (X_i^{(l)} - \bar{X}_i)^2}$. The non parametric bootstrap approach is valid for the analysis of sample correlations (see Example 9 of [88]). Moreover, since $\widehat{\rho}_{n,\cdot}$ has an asymptotic Gaussian distribution, all of the previous methods are well-suited for the correlation tests (5.36). These should be applied with $\Sigma = \Omega(\Gamma)$ and $\widehat{\Sigma} = \Omega(\widehat{\Gamma})$, where $\Omega(\widehat{\Gamma})$ is the matrix, which provides the asymptotic correlation of the correlation coefficients, given by equation (2.10) by replacing ρ_h with $\widehat{\rho}_{n,h}$, for all $h \in \mathcal{H}$.

Note that the $\max T_\infty$ procedure is applicable with any estimation $\widehat{\Gamma}$ of Γ satisfying $\widehat{\Gamma} \xrightarrow{\mathbb{P}} \Gamma$. A natural choice is the empirical estimator $\widehat{\Gamma} = (\widehat{\rho}_{i,i'})_{i,i' \in \{1, \dots, m\}}$. The matrix $(\widehat{\Gamma} - \Gamma)$ can be rewritten as $\frac{1}{\sqrt{n}} \left(\sqrt{n} (\widehat{\Gamma} - \Gamma) \right)$. Since, using Proposition 2.3, $\sqrt{n} (\widehat{\Gamma} - \Gamma) \xrightarrow[n \rightarrow +\infty]{\mathcal{L}} \mathcal{N}_m(0, \Omega(\Gamma))$ and $\frac{1}{\sqrt{n}} \xrightarrow[n \rightarrow +\infty]{\mathbb{P}} 0$, Slutsky's theorem ([90], Theorem 11.4) implies that $\widehat{\Gamma} - \Gamma \xrightarrow{\mathbb{P}} 0$, that is, $\widehat{\Gamma} \xrightarrow{\mathbb{P}} \Gamma$.

The test statistics (5.37) are used by Romano and Wolf for correlation tests (see Section 5.2 of [88]).

Another test statistics usually considered in the literature is the Fisher's transformation. Denote F the latter. F is defined as $F : x \mapsto \frac{1}{2} \log \frac{1+x}{1-x}$ and then the multiple testing (5.36) can also be based on the following test statistics: for all $h \in \{1, \dots, m\}$,

$$\widetilde{T}_{n,h}(\mathbb{X}^{(n)}) = \sqrt{n-3} F(\widehat{\rho}_{n,h}). \quad (5.38)$$

Indeed, using the delta method [91], it is easy to check that

$$\sqrt{n-3} (F(\widehat{\rho}_{n,h}) - F(\rho_h)) \xrightarrow[n \rightarrow +\infty]{} \mathcal{N}(0, 1).$$

CHAPTER 5. ASYMPTOTIC RESULTS

This approximation is accurate for moderate or large values of n [25].

In Chapter 6, we will study, for each test statistic, the numerical properties of the fourth previous multiple testing procedures for the correlation tests through the analysis of both simulated and real data.

Tests based on partial correlation coefficients

Let $V = \{1, \dots, d\}$. We aim at testing, for all $h \in \mathcal{H}$,

$$H_{0,h} : \rho_{h,V \setminus \{h\}} = 0 \quad \text{against} \quad H_{1,h} : \rho_{h,V \setminus \{h\}} \neq 0. \quad (5.39)$$

Then, the test is usually based on the following test statistic:

$$T_{n,h}^*(\mathbb{X}^{(n)}) = \sqrt{n_d} F(\widehat{\rho}_{h,V \setminus \{h\}}), \quad (5.40)$$

where $n_d = n - d + 1$.

The same result of Proposition 2.3 holds for the vector of the empirical partial correlations $\widehat{\rho}_{h,V \setminus \{h\}}$ if we replace all ρ_{ij} by $\rho_{ij,V \setminus \{i,j\}}$ in the equation (2.10), where the sample size n should also be replaced by n_d [25]. Moreover, by using the delta method [91], we have that $\sqrt{n_d}[F(\widehat{\rho}_{h,V \setminus \{h\}}) - F(\rho_{h,V \setminus \{h\}})]_{h \in \mathcal{H}}$ is an asymptotic Gaussian centered vector,

$$\sqrt{n_d}(F(\widehat{\rho}_{h,V \setminus \{h\}}) - F(\rho_{h,V \setminus \{h\}})) \xrightarrow[n \rightarrow +\infty]{} \mathcal{N}(0, 1),$$

that is, the asymptotic Gaussian assumption (5.4) holds for correlation tests based on partial correlation coefficients. Thus, all the previous methods are also well-suited for the partial correlation tests (5.39). The test statistic (5.40) have been considered in [22] for instance.

5.3 Towards an asymptotic control of the FDR

In [92], Romano *et al* propose both subsampling and bootstrap procedures that control the false discovery rate (FDR) under dependence, especially for the two-sided testing problem (5.1). Unfortunately, their methods require that all the p -values under the alternative are equal to zero. Otherwise, as already mentioned, some asymptotic FDR controlling procedures have been obtained by considering sparsity assumptions. Here, we no longer want to suppose such assumptions.

5.3.1 On BH procedure in the one-sided asymptotic Gaussian setting

From now on, consider the one-sided testing problem (5.2). This section provides a formal answer in the one-sided case to the informal question "how does multiple testing correction

5.3. TOWARDS AN ASYMPTOTIC CONTROL OF THE FDR

work in the case of asymptotic Gaussian statistical tests?". In particular, we will establish a sufficient condition for the limiting distribution of the statistical tests to obtain an asymptotic FDR control by using the BH procedure.

Throughout this section, for technical purposes, we consider a local alternative, that is, assume that the Gaussian assumption (5.4) is satisfied for $\theta(P) = c/\sqrt{n}$ where $c \in \mathbb{R}^m$ is a fixed constant. This gives rise to the p -value process

$$\forall i \in \{1, \dots, m\}, p_{n,i} = 1 - \Phi(Y_i + c) \text{ where } Y \sim \mathcal{N}_m(c, \Sigma) := P_\infty.$$

The "local alternative" assumption is actually reasonable since it eliminates the trivial configuration where all p -values under the alternative are asymptotically equal to zero.

In this section, we will study the FDR of BH procedure when n tends to infinity and the number of tests m is fixed.

Asymptotic FDR of the BH procedure

Now, we prove that the limit of the FDR of the BH procedure is equal to the FDR of the limiting experience. Assume that

$$\forall P \in \mathcal{P}, \sqrt{n}(\widehat{\theta}_{n,\cdot} - \theta(P)) \xrightarrow{\mathcal{L}} P_\infty. \quad (5.41)$$

Theorem 5.1. *Under (5.41), for all $P \in \mathcal{P}$*

$$\lim_{n \rightarrow +\infty} \text{FDR}(P) = \text{FDR}(P_\infty).$$

Proof. Let $f : \mathbb{R} \rightarrow [0, 1], x \mapsto 1 - \Phi(x+c)$ and $Y^{(n)} = \sqrt{n}(\widehat{\theta}_{n,\cdot} - \theta(P))$. For all $k \in \{1, \dots, m\}$, for all $i \in \mathcal{H}_0(P)$, let us define the set $\mathcal{B}_{k,i}$ as follows

$$\begin{aligned} \mathcal{B}_{k,i} &= \{y \in \mathbb{R}^m, \widehat{k} \leq k-1, f(y_i) \leq \alpha(k-1)/m\} \\ &= \{y \in \mathbb{R}^m, f(y)_{(j)} > \alpha j/m \text{ for } j = k, \dots, m, f(y_i) \leq \alpha(k-1)/m\}. \end{aligned} \quad (5.42)$$

Denote η the Lebesgue's measure and $\partial\mathcal{B}_{k,i}$ the frontier of the set $\mathcal{B}_{k,i}$. Let us begin by showing that for all $k \in \{1, \dots, m\}$, for all $i \in \mathcal{H}_0(P)$, $\eta(\partial\mathcal{B}_{k,i}) = 0$. For all $k \in \{1, \dots, m\}$,

CHAPTER 5. ASYMPTOTIC RESULTS

for all $i \in \mathcal{H}_0(P)$, we have

$$\begin{aligned}
 \eta(\partial \mathcal{B}_{k,i}) &= \eta\left(\partial\left[\left(\bigcap_{j=k}^m \{y : f(y)_{(j)} > \alpha j/m\}\right) \cap \{y : f(y)_i \leq \alpha(k-1)/m\}\right]\right) \\
 &\leq \eta\left(\left(\bigcup_{j=k}^m \partial\{y : f(y)_{(j)} > \alpha j/m\}\right) \cup \partial\{y : f(y)_i \leq \alpha(k-1)/m\}\right) \\
 &\leq \sum_{j=k}^m \eta(\{y : f(y)_{(j)} = \alpha j/m\}) + \eta(\{y : f(y)_i = \alpha(k-1)/m\}) \\
 &\leq \sum_{l=1}^m \sum_{j=k}^m \eta(\{y : f(y)_l = \alpha j/m\}) + \eta(\{y : f(y)_i = \alpha(k-1)/m\}) \\
 &= 0.
 \end{aligned}$$

Thus for all $k \in \{1, \dots, m\}$, for all $i \in \mathcal{H}_0(P)$, $\mathbb{P}(Y \in \partial \mathcal{B}_{k,i}) = 0$. Using the Portmanteau's lemma, it follows

$$\lim_{n \rightarrow +\infty} \mathbb{P}(Y^{(n)} \in \mathcal{B}_{k,i}) = \mathbb{P}(Y \in \mathcal{B}_{k,i}).$$

If $\mathcal{B}'_{k,i} = \{y \in \mathbb{R}^m, \widehat{k} \leq k-1, f(y)_i \leq \alpha k/m\}$, in the same manner, we would see that

$$\lim_{n \rightarrow +\infty} \mathbb{P}(Y^{(n)} \in \mathcal{B}'_{k,i}) = \mathbb{P}(Y \in \mathcal{B}'_{k,i}).$$

Therefore, we obtain

$$\begin{aligned}
 \text{FDR}(P) &= \sum_{i \in \mathcal{H}_0(P)} \sum_{k=1}^m \frac{1}{k} \left[\mathbb{P}(\widehat{k} \leq k, f(Y_i^{(n)}) \leq \alpha k/m) - \mathbb{P}(\widehat{k} \leq k-1, f(Y_i^{(n)}) \leq \alpha k/m) \right] \\
 &= \sum_{i \in \mathcal{H}_0(P)} \sum_{k=1}^m \frac{1}{k} \left[\mathbb{P}(Y^{(n)} \in \mathcal{B}_{k+1,i}) - \mathbb{P}(Y^{(n)} \in \mathcal{B}'_{k,i}) \right].
 \end{aligned}$$

Then, by taking the limit in the previous expression, it finally follows that

$$\begin{aligned}
 \lim_{n \rightarrow +\infty} \text{FDR}(P) &= \sum_{i \in \mathcal{H}_0(P)} \sum_{k=1}^m \frac{1}{k} \left[\mathbb{P}(Y \in \mathcal{B}_{k+1,i}) - \mathbb{P}(Y \in \mathcal{B}'_{k,i}) \right] \\
 &= \text{FDR}(P_\infty).
 \end{aligned}$$

□

5.3. TOWARDS AN ASYMPTOTIC CONTROL OF THE FDR

Asymptotic FDR control

Theorem 5.1 gives a sufficient condition for FDR control under weaker assumptions than usually at the price of an asymptotic control (instead of an exact control). Indeed, this theorem is valid under asymptotic Gaussian assumption whatever the distribution P of $X^{(j)}$ is. Thus it can be useful even when the distribution of the observations is unknown. More precisely, Theorem 5.1 indicates that to control asymptotically the FDR of the BH procedure at level α , it is sufficient to control the limiting FDR at level α . In particular, Theorem 5.1 combined with Theorem 3.5 provides the following result.

Theorem 5.2. *Under (5.41), if P_∞ is PRDS on $\{1 \leq i \leq m : c_i = 0\}$, then for all $P \in \mathcal{P}$, we have*

$$\limsup_{n \rightarrow +\infty} \text{FDR}(P) \leq \alpha.$$

As seen in Chapter 3, in the one-sided Gaussian testing framework, the PRDS assumption is satisfied whenever $\Sigma_{ij} \geq 0$, for all $i \neq j$, see [67].

5.3.2 Correlation tests

Theorem 5.2 shows that to asymptotically control the FDR of BH procedure, it is sufficient that the asymptotic distribution of the test statistic should be PRDS on \mathcal{H}_0 . Unfortunately, this result is not helpful for testing correlations in two-sided Gaussian setting. Indeed, resulting from a simulation study, we can state that the matrix $\Omega(\Gamma)$ of Proposition 2.3 does not provide a positive structure dependence in general. More precisely, 100,000 covariance d -matrices Γ are randomly generated using the R-package *ClusterGeneration* [93] and we evaluate the occurrence of MTP_2 distributions on these 100,000 matrices (see Table 5.1).

Distribution of:	$ X $	$(\rho_h)_{h \in \mathcal{H}}$	
		without MTP_2 constraint on $ X $	with MTP_2 constraint on $ X $
$d = 3$	20.26%	13.27%	65.53%
$d = 4$	0.09%	0%	0%
$d = 5$	0%	0%	×

Table 5.1 – Occurrence of MTP_2 distributions on 100,000 simulations.

The evaluation of the proportion of MTP_2 distributions is based on Theorem 3.1 and Theorem 3.1' of [63], which give necessary and sufficient condition on the matrix Γ (resp. $\Omega(\Gamma)$) for $|X|$ (resp. $(|\rho_h|)_{h \in \mathcal{H}}$) having an MTP_2 distribution. The proportion of the vector of the absolute value of correlation coefficients of X , that is, $(|\rho_h|)_{h \in \mathcal{H}}$, having an MTP_2 distribution is evaluated in two different cases: when $|X|$ does not have an MTP_2 distribution

CHAPTER 5. ASYMPTOTIC RESULTS

(see the third column of Table 5.1) and when $|X|$ has an MTP_2 distribution (see the fourth column of Table 5.1).

As shown in Table 5.1, the MTP_2 constraint seems to be a very restrictive constraint (see the second column of Table 5.1). Moreover, even if the absolute value of the observed vector X has an MTP_2 distribution, the transformation Ω does not preserve the positive dependence structure (see the last column of Table 5.1) for the correlation coefficients.

However, if the conjecture 4.1 is proved then Theorem 5.1 guarantees that the BH procedure provides an asymptotic FDR control for testing correlation coefficients.

Chapter 6

Application to neuroscientific framework

This chapter investigates the numerical properties of the four multiple testing procedures defined in Section 5.2 of Chapter 5, through the analysis of both simulated and real data. Section 6.1 examines the performance of these multiple testing procedures by simulations. A real data analysis is discussed in Section 6.2.

Contents

6.1	Simulation study	115
6.1.1	Statistical setting	115
6.1.2	FWER control	117
6.1.3	Power with fixed alternative	120
6.1.4	Power study for the chain graph	123
6.2	Application to estimation of cerebral connectivity	126
6.2.1	Description of the data and methodological context	126
6.2.2	Results	127
6.2.3	Discussion on power	130

6.1 Simulation study

6.1.1 Statistical setting

We generate n independent random vectors $X^{(1)}, \dots, X^{(n)}$ from a 10-dimensional multivariate normal distribution $\mathcal{N}_{10}(0, \Gamma)$. Four covariance matrix models are considered. The latter are represented in Figure 6.1.

CHAPTER 6. APPLICATION TO NEUROSCIENTIFIC FRAMEWORK

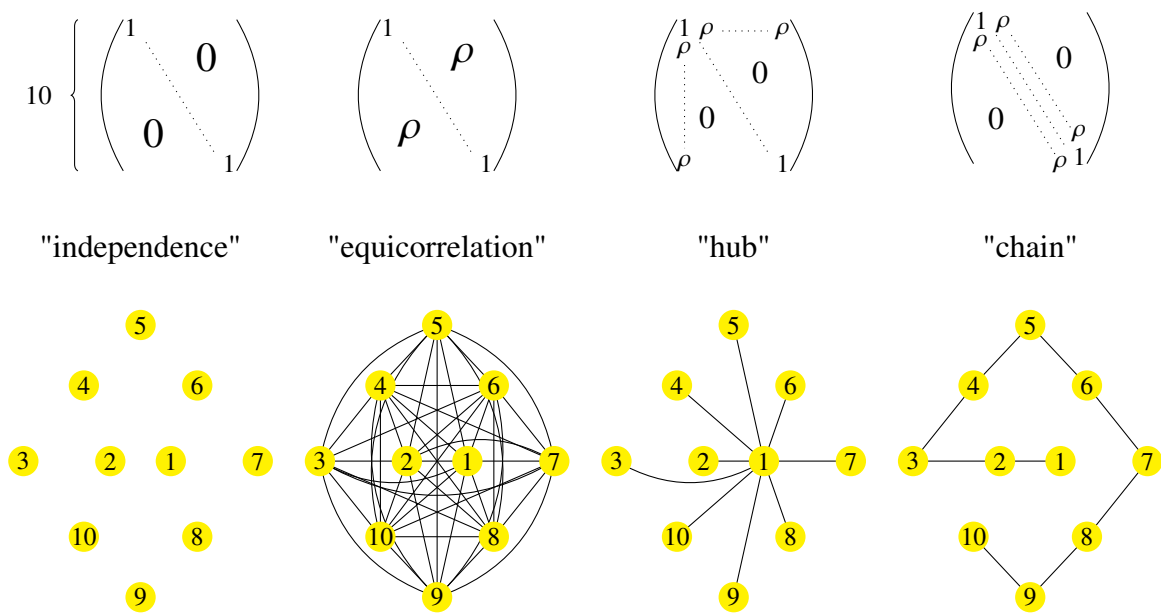


Figure 6.1 – The four covariance matrix models considered for simulation study with their graphical representations.

For each covariance matrix model, we plot the matrix $\Omega(\Gamma)$ (matrix of Proposition 2.3 with the true value of parameter ρ). Results are displayed in Figure 6.2.

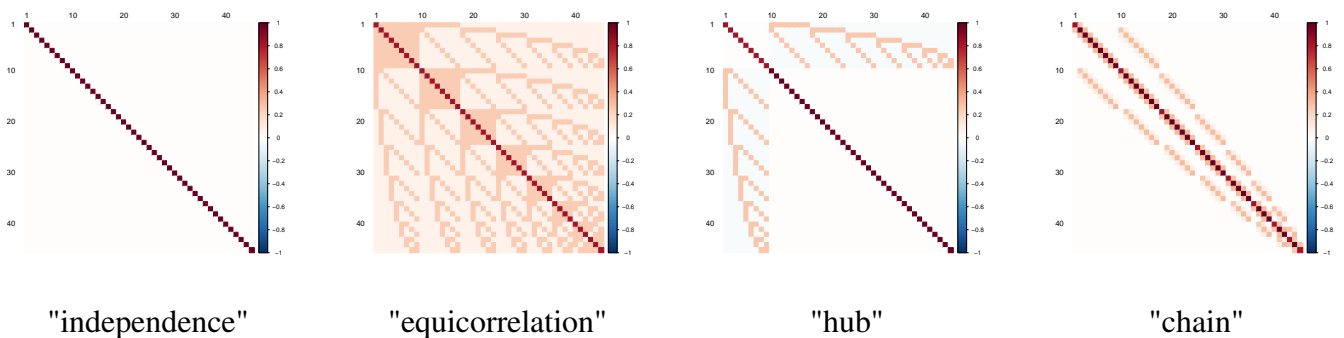


Figure 6.2 – The matrix $\Omega(\Gamma)$ (matrix of Proposition 2.3) with the true value of parameter $\rho = 0.3$.

Since we generate 10-dimensional vectors, there are $10 \times 9 / 2 = 45$ pairwise correlations to test. Throughout this chapter, we consider the four multiple testing procedures defined in Section 5.2 for the two-sided testing problem (2.8), that is we aim at testing pairwise

correlation coefficients: for all $h \in \mathcal{H}$,

$$H_{0,h} : \rho_h = 0 \quad \text{against} \quad H_{1,h} : \rho_h \neq 0,$$

based on the test statistics $T_{n,h}(\mathbb{X}^{(n)}) = \sqrt{n} \widehat{\rho}_{n,h}$. Table 6.1 recalls the methods of Section 5.2. Moreover, we also consider the oracle $\max T_\infty$ procedure in our simulation study, which is the $\max T_\infty$ procedure in which the matrix $\Omega(\Gamma)$ is computed with the true parameter ρ .

Method	Rejection set of single step version	Asymptotic FWER control	
		single step	step-down
Bonferroni	$\{h \in \mathcal{H} : T_{n,h} > \Phi^{-1}\left(1 - \frac{\alpha}{2m}\right)\}$	Theorem 5.1	Corollary 5.1
Šidák	$\{h \in \mathcal{H} : T_{n,h} > \Phi^{-1}\left(\frac{1}{2}(1 - \alpha)^{1/m} + \frac{1}{2}\right)\}$	Theorem 5.3	Corollary 5.2
Bootstrap RW	$\{h \in \mathcal{H} : T_{n,h} > \widehat{t}_{n,\alpha}(\Omega(\Gamma), \mathcal{H})\}$, where $\widehat{t}_{n,\alpha}(\Omega(\Gamma), \mathcal{H})$ is the $(1 - \alpha)$ -quantile of $\mathcal{L}(\ \mathcal{N}_m(0, \Omega(\Gamma))\ _\infty)$ computed using bootstrap resamples of $\{X^{(1)}, \dots, X^{(n)}\}$.	Theorem 5.4	Corollary 5.3
$\max T_\infty$	$\{h \in \mathcal{H} : T_{n,h} > t_{n,\alpha}(\Omega(\widehat{\Gamma}))\}$ where $t_{n,\alpha}(\Omega(\widehat{\Gamma}))$ is the $(1 - \alpha)$ -quantile of $\mathcal{L}(\ \mathcal{N}_m(0, \Omega(\widehat{\Gamma}))\ _\infty)$ computed using simulated samples of $\mathcal{N}_m(0, \Omega(\widehat{\Gamma}))$.	Theorem 5.5	Corollary 5.4
oracle $\max T_\infty$	$\{h \in \mathcal{H} : T_{n,h} > t_{n,\alpha}(\Omega(\Gamma))\}$ where $t_{n,\alpha}(\Omega(\Gamma))$ is the $(1 - \alpha)$ -quantile of $\mathcal{L}(\ \mathcal{N}_m(0, \Omega(\Gamma))\ _\infty)$ computed using simulated samples of $\mathcal{N}_m(0, \Omega(\Gamma))$ with the true value of Γ .	Theorem 5.5	Corollary 5.4

Table 6.1 – Multiple testing procedures considered in this chapter.

The performance criteria here are the FWER and the power.

6.1.2 FWER control

First, we investigate the FWER control with respect to n that varies in $\{25, 50, 75, \dots, 500\}$. The nominal FWER level is 0.1 and it is calculated based on $N = 100,000$ replicates. For the Romano-Wolf's procedure and the $\max T_\infty$ procedure, the number of the bootstrap

CHAPTER 6. APPLICATION TO NEUROSCIENTIFIC FRAMEWORK

resamples is equal to 1,000. Figure 6.3 displays the FWER of the single step version of the procedures.

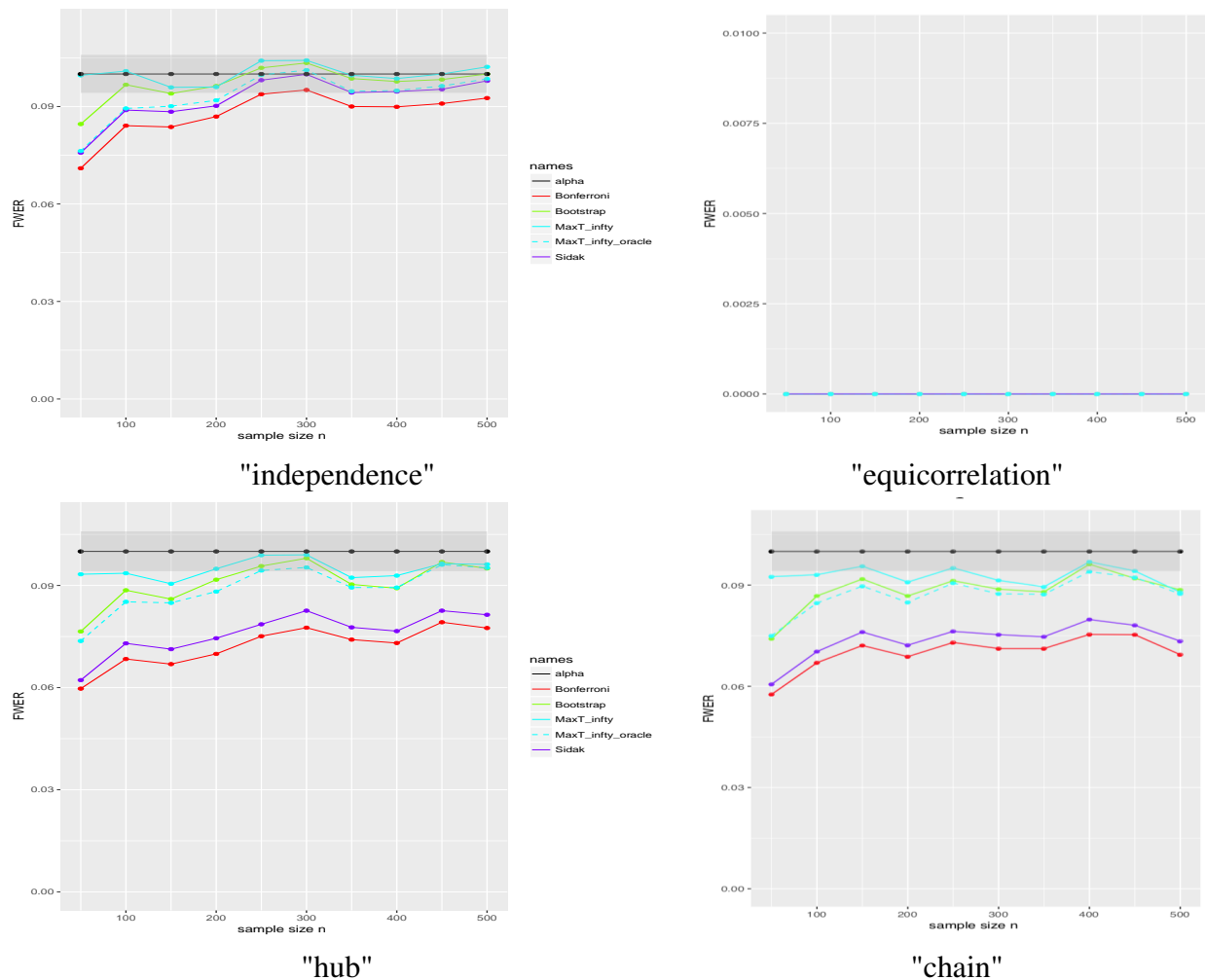


Figure 6.3 – FWER of the multiple testing procedures defined in Section 5.2 based on $N = 100,000$ replicates. The number of the bootstrap resamples is equal to 1,000 and $\alpha = 0.1$. The shaded zone around the α -line corresponds to the confidence interval of a Binomial distribution with parameters N and α , with confidence 95%.

As shown in Figure 6.3, since there is no true null hypothesis in the equicorrelation case, the FWER is equal to zero. Moreover, as theoretically established in Chapter 5, all the methods control the FWER for the four covariance matrix models.

As expected, the Bonferroni's method is the most conservative, all the more so when the proportion of false null hypotheses is far from 1, that is for the dependence structures hub and chain respectively. Figure 6.3 also shows that the Šidák's procedure allows a little more false discoveries than the Bonferroni's procedure. Let us recall that the Šidák's procedure does not require additional theoretical assumption with respect to the Bonferroni's procedure, so the Šidák's procedure appears as an interesting alternative to Bonferroni's procedure in the two-sided Gaussian setting when the dependence structure is unknown.

The Romano-Wolf's bootstrap procedure and $\max T_\infty$ procedure seem to provide similar performances as soon as n is large enough ($n > 100$), even if the $\max T_\infty$ procedure seems to be a little less conservative procedure.

Finally the oracle $\max T_\infty$ procedure improves performance with respect to non oracle procedure, which is natural.

We also apply step-down procedures to the same simulated dataset. Resulting FWER is displayed on Figure 6.4. It illustrates the fact that the step-down version of a multiple testing procedure is less conservative than its single step version. The control is still acquired for most methods. Only the $\max T_\infty$ seems to provide a FWER larger than α for the hub structure, even if the difference is low. This is probably due to the quality of the estimation of the covariance matrix Γ plugged into (5.18).

For the hub dependence structure, both the step-down Bonferroni's procedure and the step-down Šidák's procedure are almost sharp as soon as n is large enough ($n > 250$), which is no longer true for their single step versions.

For the chain dependence structure, all the methods are more accurate and the step-down Romano-Wolf's bootstrap procedure and the step-down $\max T_\infty$ procedure seem to be almost sharp for $n > 100$.

Given a multiple testing procedure, once the control of FWER is ensured, its power is critical for real data applications. In this regards, the behavior of the power of procedures with respect to the dataset properties is important because it provides guidance to the quality of recordings required for extracting information. The two following sections are devoted to the study of the power.

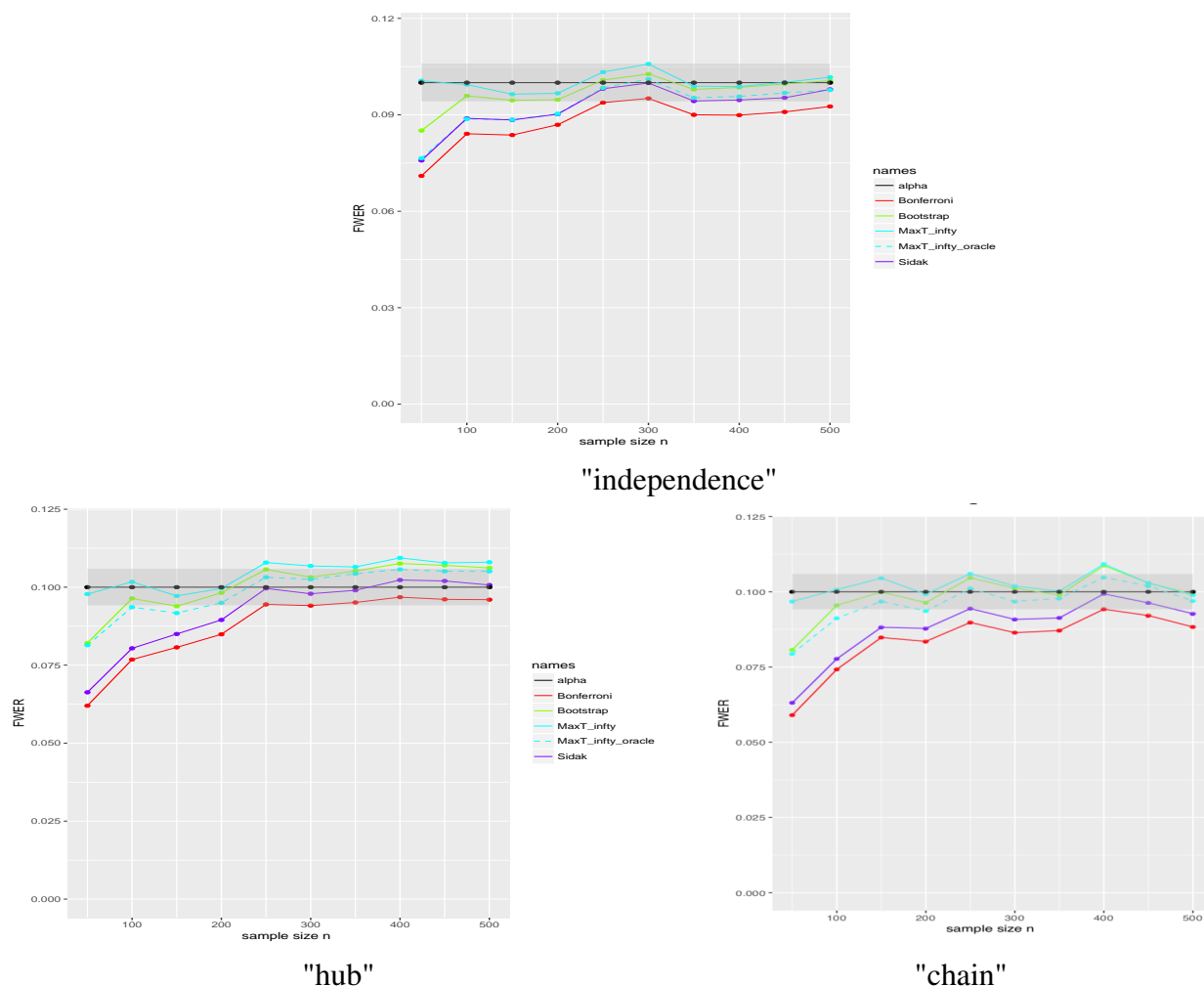


Figure 6.4 – FWER of the step-down versions of the multiple testing procedures defined in Section 5.2 based on $N = 100,000$ replicates. The number of the bootstrap resamples is equal to 1,000 and $\alpha = 0.1$. The shaded zone around the α -line corresponds to the confidence interval of a Binomial distribution with parameters N and α , with confidence 95%.

6.1.3 Power with fixed alternative

The power here is defined as the proportion of null hypotheses correctly rejected and it is calculated based on $N = 100,000$ replicates. The nominal FWER level is 0.1. For the Romano-Wolf's procedure and the $\max T_\infty$ procedure, the number of the bootstrap resamples is equal to 1,000.

Figure 6.5 displays the power of each procedure with respect to the sample size n . Since

6.1. SIMULATION STUDY

there is no false hypothesis in the independent case, the power is equal to zero.

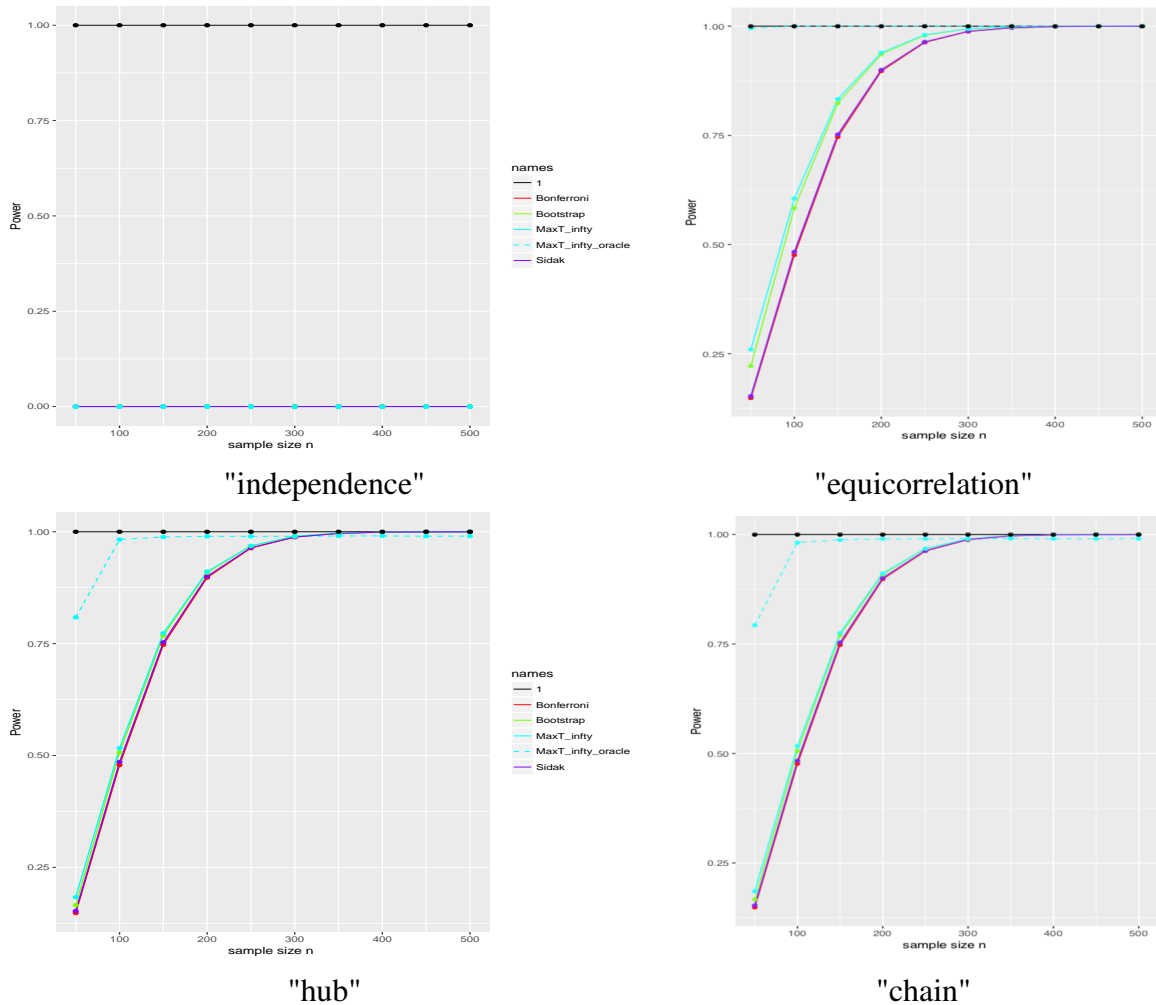


Figure 6.5 – Power of the multiple testing procedures defined in Section 5.2 based on 100,000 replicates. The number of the bootstrap resamples is equal to 1,000 and $\alpha = 0.1$.

First, oracle $\max T_\infty$ procedure gives an optimal power of 1, as soon as the number n of observations is high enough to ensure identifiability.

All procedures have a power tending to 1 when n goes to infinity. Furthermore, the $\max T_\infty$ and the Romano-Wolf's bootstrap procedure are more powerful than the Bonferroni's procedure and the Šidák's procedure, especially for the equicorrelation dependence structure in which all rejected null hypotheses are correctly rejected since there is no true null hypothesis.

We also compare the difference of power between the single step version and the step-

CHAPTER 6. APPLICATION TO NEUROSCIENTIFIC FRAMEWORK

down version for each procedure. The results are reported in Figure 6.6.

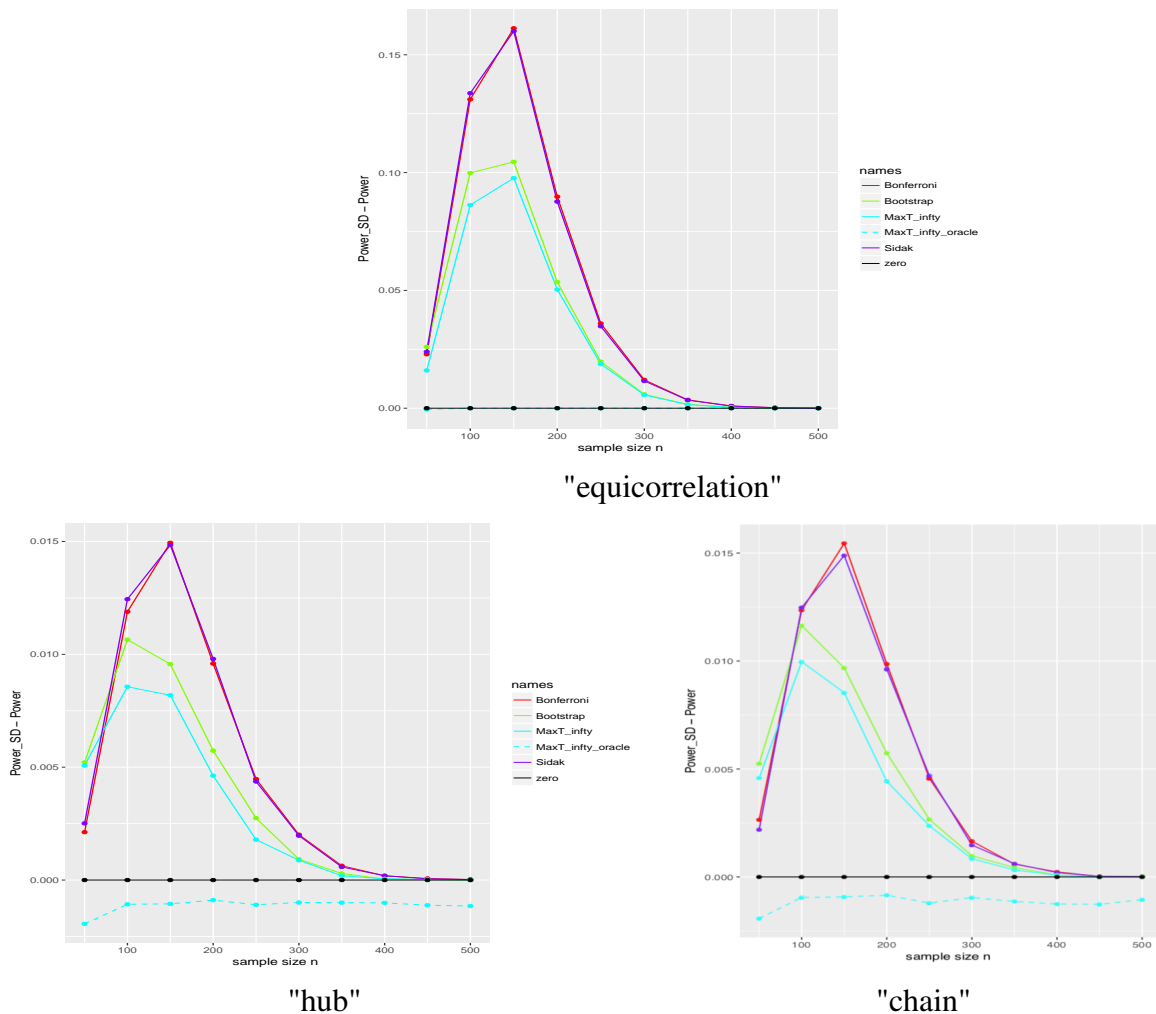


Figure 6.6 – Difference of power between the single step and the step-down version of the multiple testing procedures defined in Section 5.2 based on 100,000 replicates. The number of the bootstrap resamples is equal to 1,000 and $\alpha = 0.1$.

It illustrates the effective gain of power obtained with step-down procedures. When n is "small" ($n < 100$), the gain is less important since the power of single test procedure is inherently very low. For larger values of n , the efficiency decreases with the number of observations is high since the power of single test procedure is also high.

Note that the order of magnitude is 0.1 for the equicorrelation dependence structure (which corresponds roughly to 4 additional null hypotheses rejected) and 0.01 for hub graph and chain graph (equivalent to 1 additional null hypothesis rejected).

6.1.4 Power study for the chain graph

A more deeper study of the power is now realized. We consider the chain graph, and evaluate the power characteristics when the parameter ρ and the number of nodes in the graph vary.

First, we can expect that the power increases as the strength of the signal (values of ρ) increases. This is indeed what can be observed in Figure 6.7. Remark that we restrict the study to ρ smaller than 0.5 since for ρ higher than 0.5 the chain matrix is no more a covariance matrix.

In real data applications, we must be aware of the number of observations needed to ensure a sufficiently high power. For example to reject 75% of false null hypotheses, with $\rho = 0.2$, correlations must be estimated with more than 350 observations in the sample. Indeed as seen in Figure 6.7, the first curve, which is above the crossing point between the axis $y = 0.75$ and the axis $x = 0.2$, corresponds to the curve of $n_{\text{obs}} = 350$. As observed in Section 6.1.3, step-down procedures enable an increase of power. Similar procedures were applied at the chain dependence structure to study the power behavior with respect to n and ρ . Results, in term of gain of power, are displayed in Figure 6.8.

First, step-down procedure for oracle method does not increase much the power. Next, for others multiple testing methods, two regimes can be observed:

- $n \leq 150$: It is needed that ρ is sufficiently high for single test procedure to discriminate null and alternative hypotheses. Thus, step-down procedures are efficient only if ρ is higher than a given level (depending on n).
- $n > 150$: Identifiability does not intervene anymore. The gain is then higher when ρ is low. Indeed for high values of ρ , single testing procedures, and consequently multiple testing procedures, have a high power.

CHAPTER 6. APPLICATION TO NEUROSCIENTIFIC FRAMEWORK

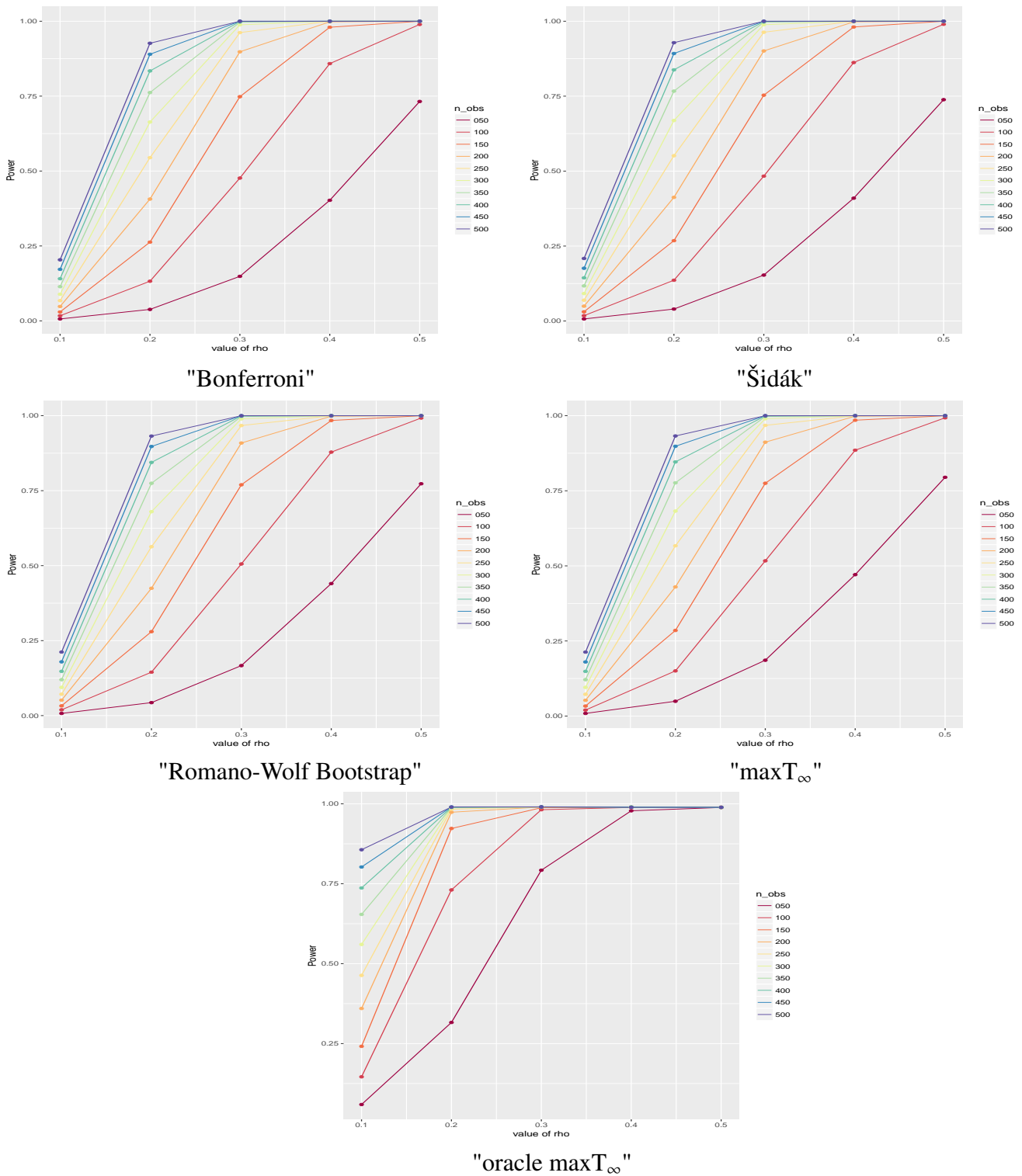


Figure 6.7 – Power of the multiple testing procedures defined in Section 5.2 based on 100,000 replicates of chain graph. The number of the bootstrap resamples is equal to 1,000 and $\alpha = 0.1$.

6.1. SIMULATION STUDY

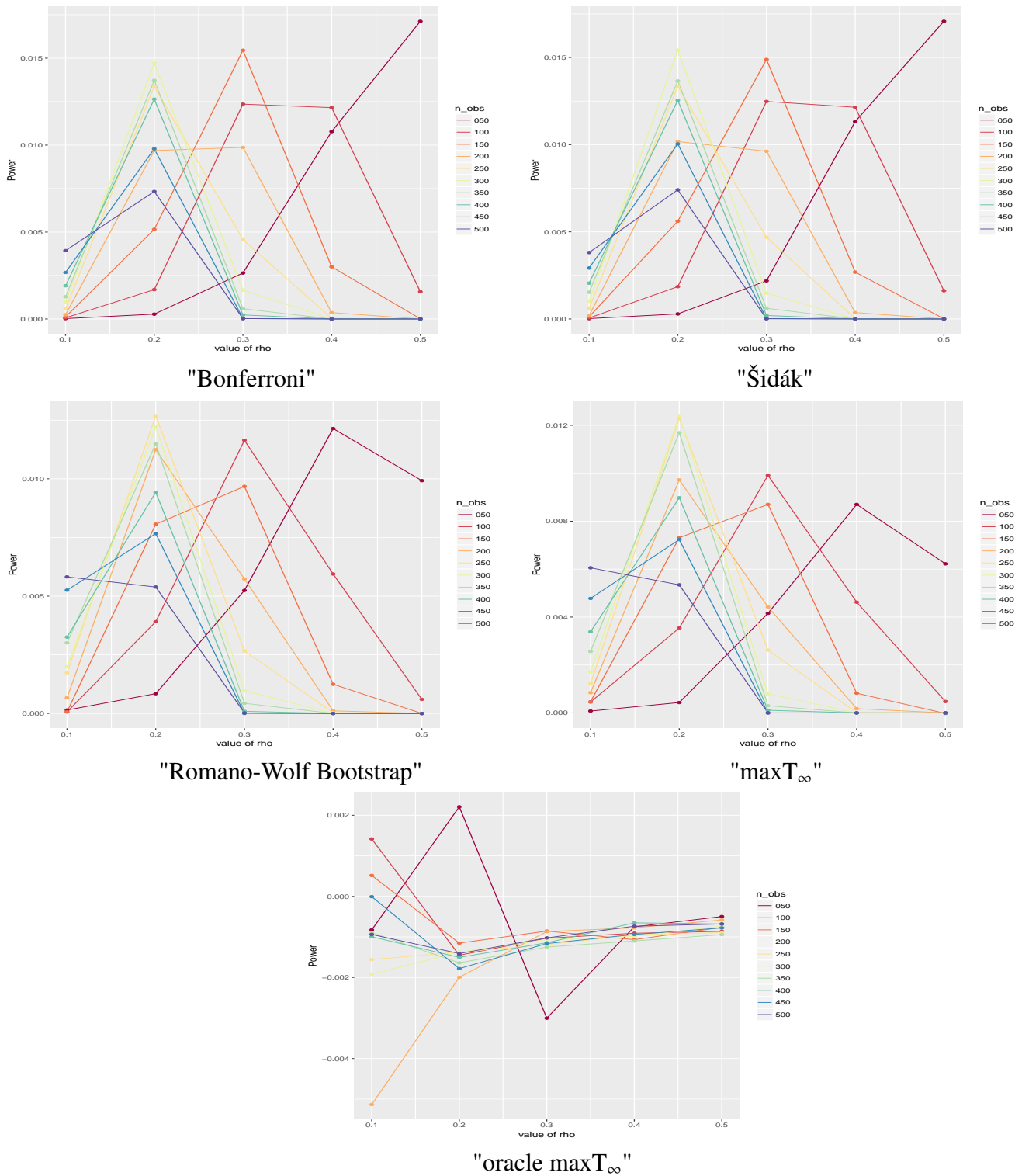


Figure 6.8 – Gain of power of the step-down multiple testing procedures with respect to single step procedures, based on 100,000 replicates of chain graph. The number of the bootstrap resamples is equal to 1,000 and $\alpha = 0.1$.

6.2 Application to estimation of cerebral connectivity

Consider the two-sided testing problem (2.7). In order to apply the multiple testing procedures defined in Section 5.2 on real dataset, we acquired functional Magnetic Resonance images (fMRI) of both dead and alive rats.

6.2.1 Description of the data and methodological context

Data

Rats are used for acquiring data on the functioning of the brain using fMRI. This provides us a set of time series with spatial localization in the brain of the rats. Four rats are scanned dead (rats 1 to 4), and seven rats are scanned alive under anaesthetic (rats 5 to 11). The duration of the scanning is 30 minutes with a time repetition of 0.5 second so that 3,600 time points are available at the end of experience. After preprocessing as explained in [94], we extracted 51 time series covering the whole brain of rats.

Multiple testing approach

Since the whole brain of rats are aggregated into 51 cerebral regions, we have to deal with $50 \times 51/2 = 1275$ tests for each of the 11 rats.

The time series resulting from fMRI experiment are nonstationary with long memory properties, which is not convenient from a mathematical point of view. To avoid such properties, the correlation coefficients are estimated in the wavelet transform domain and then the statistical tests are based on wavelet correlation coefficients [95, 96].

First we are decomposing each time series using a wavelet basis and then for each wavelet scale we are studying all the possible pairs of correlations. From the literature in neuroscience, it is convenient and adequate to focus on low frequencies because we obtain the best signal noise ratio. Here we will focus on wavelet scale 5 corresponding to the frequency interval [0.03 ; 0.06] Hz.

The wavelet correlation coefficients are typically supposed to have a Gaussian distribution [97]. Thus, the wavelet correlation coefficients satisfy the asymptotic Gaussian assumption (5.4). Note that in practice, we observe that the the wavelet correlation coefficients actually satisfy (5.4) and this assumption is usually made in the neuroscientific literature. Hence, all procedures of Section 5.2 are well-suited for the two-sided testing problem (2.7) and the results of these procedures on the real dataset are discussed in the following section.

6.2.2 Results

Figure 6.9 displays the proportion of significant correlations obtained by Bonferroni, Šidák, Romano-Wolf and $\max T_\infty$ procedures. The nominal FWER level is 0.05. As stated previously, only scale 5 of wavelet decomposition is considered, corresponding to the frequency interval [0.03 ; 0.06] Hz.

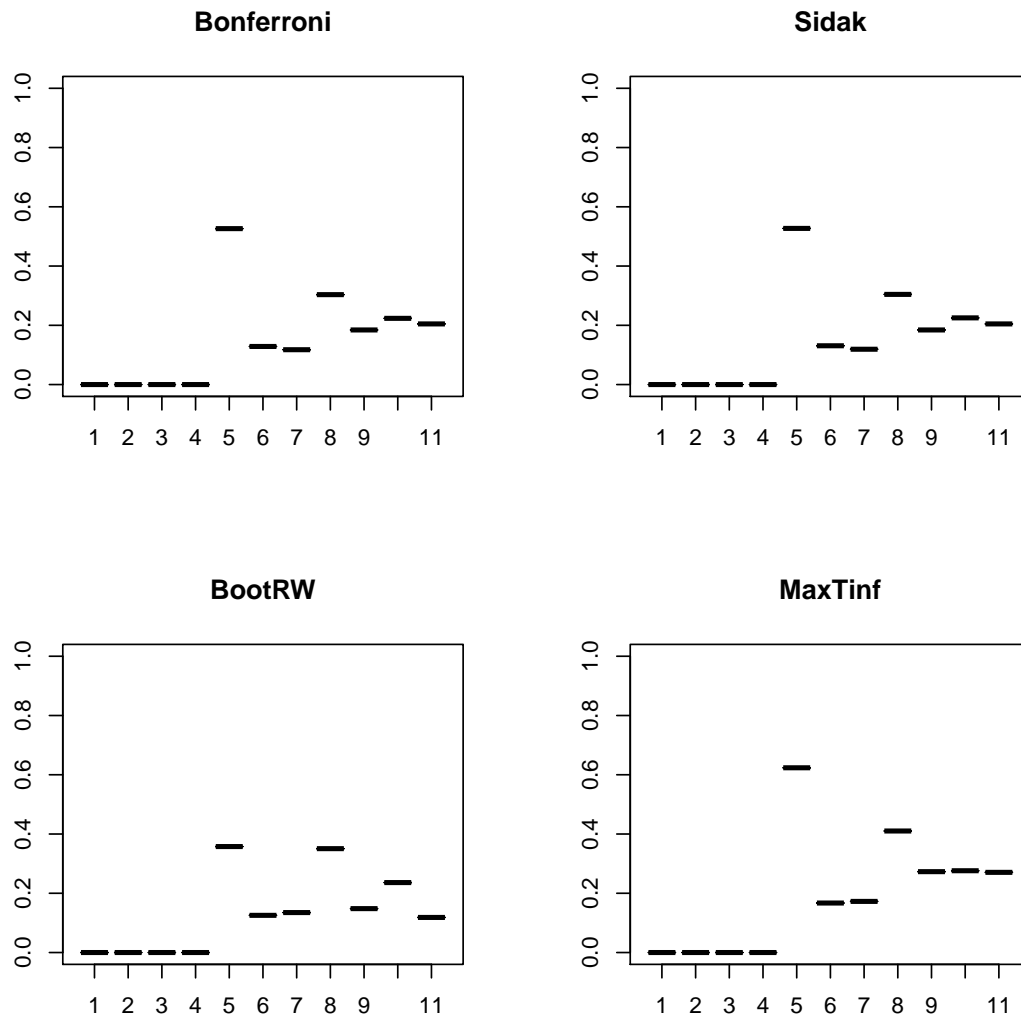


Figure 6.9 – Proportion of rejected null hypotheses obtained by multiple testing procedures defined in Section 5.2 for the rats 1 to 11 (in x -axis). 1275 tests are done for each rat.

CHAPTER 6. APPLICATION TO NEUROSCIENTIFIC FRAMEWORK

It is not possible to evaluate the accuracy of procedures on real dataset because of the lack of ground truth. However, since there is actually no cerebral activity in a brain of a dead rat, all null hypotheses to be tested are true nulls and the number of rejected hypotheses must be equal to zero in this case. Indeed, Figure 6.9 shows that the number of significant correlations is zero or near to zero for dead rats, whatever the method. The highest values are less than 5% (see Table 6.2), which is in coherence with the fact that we expect to be under the full null hypothesis.

For alive rats, the proportion of rejected null hypotheses is always higher than 5%. Note that there is a high variability with respect to rats, but all methods give the same pattern. The highest number of detections is obtained for the rats 5 and 8 and for the others, the number of detections is between 15% and 25%

Otherwise, two examples of estimated graphs for alive rats are given at Figure 6.10.

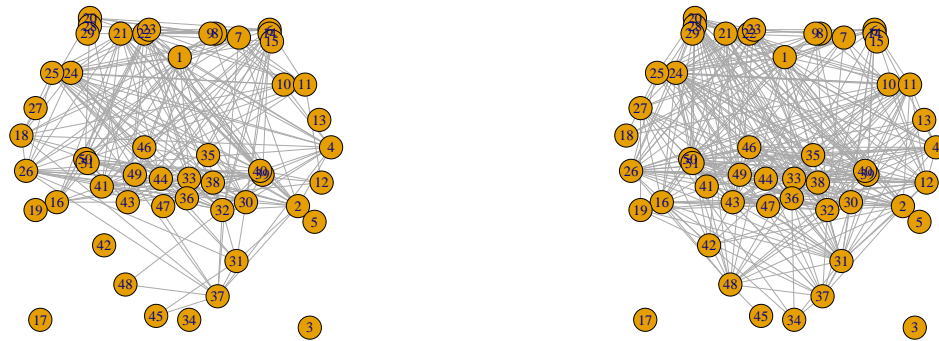


Figure 6.10 – Two estimated graphs for two alive rats, obtained by $\max T_\infty$ multiple testing, at scale 5.

As mentioned, since we do not know the true null hypotheses, we can not evaluate the accuracy of the procedures. Nevertheless, apply a bootstrap procedure on time series allow us to estimate the number of activated cerebral connections. This is discussed in the following section.

Power study for alive rats

We applied a bootstrap procedure on time series using moving windows in time (with overlapping). Results are given in Table 6.2. We observe an important difference between alive and dead rats. Dead rats have a mean proportion of rejected null hypotheses equal to

6.2. APPLICATION TO ESTIMATION OF CEREBRAL CONNECTIVITY

zero or really near to zero. The highest mean of rejected hypotheses is indeed 5 for 1275 tests. On the contrary for alive rats we observe always more than 10% of null hypotheses rejected.

Unexpectedly, bootstrap procedure of Romano-Wolf leads to a lower number of rejected null hypotheses, with respect to Bonferroni or Šidák procedures. No such behavior were observed on simulated data.

Dead rat	1	2	3	4
Bonferroni	6e-04 (5e-04)	1e-04 (2e-04)	0 (0)	0 (0)
Šidák	6e-04 (5e-04)	1e-04 (2e-04)	0 (0)	0 (0)
Bootstrap RW	5e-04 (5e-04)	1e-04 (2e-04)	0 (0)	0 (0)
maxT _∞	8e-04 (5e-04)	1e-04 (2e-04)	0 (0)	0 (0)

Alive rat	5	6	7	8
Bonferroni	0.41 (0.08)	0.10 (0.03)	0.11 (0.02)	0.24 (0.06)
Šidák	0.41 (0.08)	0.10 (0.03)	0.11 (0.03)	0.24 (0.06)
Bootstrap RW	0.18 (0.10)	0.10 (0.03)	0.11 (0.03)	0.26 (0.07)
maxT _∞	0.50 (0.08)	0.14 (0.03)	0.14 (0.03)	0.33 (0.08)

Alive rat	9	10	11
Bonferroni	0.22 (0.04)	0.22 (0.03)	0.14 (0.04)
Šidák	0.22 (0.04)	0.23 (0.03)	0.14 (0.04)
Bootstrap RW	0.19 (0.04)	0.25 (0.03)	0.11 (0.03)
maxT _∞	0.30 (0.05)	0.28 (0.03)	0.19 (0.05)

Table 6.2 – Estimation of the proportion of rejected null hypotheses obtained by bootstrap with overlapping sliding windows. The value given is the mean proportion of rejected null hypothesis. The value between parenthesis is the standard deviation.

Step-down procedures were also considered. Results are reported in Table 6.3. The latter shows that, as expected, the number of rejected null hypotheses increase. Again, the number of rejections is closed to zero for dead rats, and presents much variability for alive rats. Note that the minimal number of significant edges for an alive rat is obtained for rat 6, with a mean of approximately 140 significant edges.

CHAPTER 6. APPLICATION TO NEUROSCIENTIFIC FRAMEWORK

Dead rat	1	2	3	4
Bonferroni SD	6e-04 (5e-04)	1e-04 (2e-04)	0 (0)	0 (0)
Šidák SD	6e-04 (5e-04)	1e-04 (2e-04)	0 (0)	0 (0)
Bootstrap RW SD	5e-04 (4e-04)	0 (1e-04)	0 (0)	0 (0)
maxT _∞ SD	9e-04 (6e-04)	1e-04 (3e-04)	0 (1e-04)	0 (0)

Alive rat	5	6	7	8
Bonferroni SD	0.45 (0.09)	0.11 (0.03)	0.11 (0.03)	0.26 (0.07)
Šidák SD	0.45 (0.09)	0.11 (0.03)	0.11 (0.03)	0.26 (0.07)
Bootstrap RW SD	0.20 (0.11)	0.11 (0.03)	0.12 (0.04)	0.28 (0.07)
maxT _∞ SD	0.52 (0.09)	0.14 (0.03)	0.15 (0.04)	0.35 (0.08)

Alive rat	9	10	11
Bonferroni SD	0.24 (0.05)	0.24 (0.03)	0.15 (0.04)
Šidák SD	0.24 (0.05)	0.24 (0.03)	0.15 (0.04)
Bootstrap RW SD	0.21 (0.05)	0.27 (0.03)	0.12 (0.03)
maxT _∞ SD	0.31 (0.06)	0.29 (0.03)	0.2 (0.05)

Table 6.3 – Estimation of the proportion of rejected null hypotheses obtained by bootstrap with overlapping sliding windows, applying step-down procedures. The value given is the mean proportion of rejected null hypothesis. The value between parenthesis is the standard deviation.

6.2.3 Discussion on power

For a large part, the neuroscientific community does not take into account the power of the procedures that are used in their analyses. In this regards, the classical approach consists of inferring a graph with a fixed number of edges whatever the power of the procedure used is (in [98] for instance). This approach is typically used in order to compare different brains. However, it makes no sense and if we want to select more edges than a procedure could be detect, the additional edges included in the inferred graph will only result from noise. For instance, according to Table 6.2, with respect to the single step procedures used in this chapter we should not select a graph with more than 127 edges for rat 6 (10% of 1275 tests, the percentage of detections obtained).

Figure 6.11 provides another example of the issue occasioned by the power of a procedure. Looking at anatomical data from the scan of a patient who shows a deficiency on

6.2. APPLICATION TO ESTIMATION OF CEREBRAL CONNECTIVITY

the left hemisphere, we expect to see a lack of brain connection in that hemisphere. In that figure, we depict below the anatomical data the position of (respectively, from left to right) the 100, 200 and 400 edges showing the strongest correlations signals. Observe that the expected lack of correlation is far from plain when looking at 400 edges. This example illustrates that care has to be taken in order to "correctly" threshold a multiple testing procedure, that is, in order not to consider too many non significant edges (false discoveries). This points out the need of taking into account the power of a procedure.

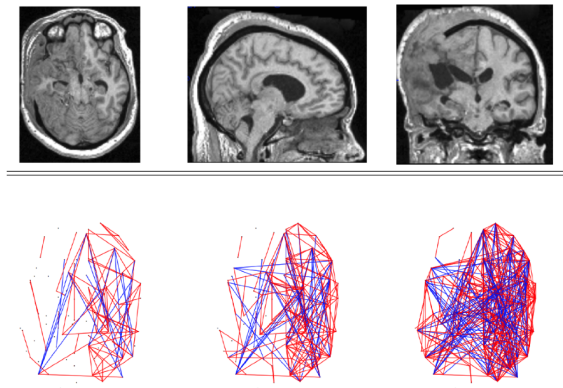


Figure 6.11 – Problem of identifying cerebral connectivity networks: the critical point of the power in real data application [99] (see text). The length of red edges is smaller than 85mm and the length of blue edges is higher than 85mm with respect to the Euclidean distance.

Conclusion and perspectives

Solving an applied mathematical problem requires an appropriate model that should be simple enough to study, while at the same time being as realistic as possible. The latter point is critical in practice and the theoretical assumptions, needed to get accurate results, rarely correspond to realistic situations. First, we usually establish the results under strong theoretical assumptions and then we try to extend the scope of these results in order to solve the real problem. The new results contained in this thesis contribute to in this ever-expanding pattern.

We investigated the problem of both FWER and FDR control for fMRI data analysis and throughout this manuscript the dependence assumptions have been relaxed step by step.

Chapter 2 briefly introduces graphical models and theoretically justifies the problem of graph inference by a multiple testing procedure. However, a deeper study of the properties of Gaussian graphical models could be relevant, especially to take into account previous knowledge on the graph that we wish to build. For instance, a recent reference is [100].

In Chapter 3, we establish a new bound for the FDR of BH procedure by relaxing the PRDS assumption on \mathcal{H}_0 . It results that the FDR is controlled by the BH procedure when "few" coefficients of the covariance matrix are negative in Gaussian setting.

When the test statistics are normally distributed, Chapter 4 completely formalizes the FDR control of the BH procedure with particular structured dependencies in which the properties of positive dependence do not hold. We also show that (CohenBound) is essentially sharp under perfect correlation. In order to deal with real problem, the scope of these results must be extended. Some possible directions directly in line with this chapter are:

- Generalize (CohenBound) without perfect correlation between the null and the alternative (by generalizing the proof of Proposition 4.4 for instance);
- Prove (CohenBound2) for $\mu \in \mathbb{R}^m$ and Σ a covariance matrix with diagonal entries equal to 1, such that the submatrix $(\Sigma_{ij})_{(i,j) \in \mathcal{H}_0^c}$ is equicorrelated;
- Establish that the perfect correlation is somewhat a least favorable configuration.

Conclusion and perspectives

In Chapter 5, we present some existing FWER controlling procedures in asymptotic Gaussian setting. It theoretically clarifies this control, when the literature were incomplete to our knowledge. We also show that to asymptotically control the FDR of the BH procedure, it is sufficient that the asymptotic distribution of the test statistic is PRDS on \mathcal{H}_0 in one-sided Gaussian setting. This results from the fact that to control asymptotically the FDR of the BH procedure, it is sufficient to control the limiting FDR. The latter could be extended to a non Gaussian limit. For instance, in order that the proof of Theorem 5.1 remains valid, it is sufficient that each p -value should be a continuous transformation of the statistical test and should admit a density with respect to the Lebesgue's measure. In this regard, we could formally introduce the notion of asymptotic positive dependence.

Chapter 6 illustrates the theoretical properties of the FWER controlling procedures defined in Chapter 5 on both simulated and real data set. The FWER control has been checked and the different procedures have been compared. The $\max T_\infty$ procedure appears as the most powerful procedure. A remarkable fact is that the full null hypothesis has been verified on real data set (for all dead rats). The multiple testing approach is often criticized in the literature because of its impossibility to satisfy the null hypothesis in practice (*e.g.* [101]). Our study allows us to contradict this assertion. Furthermore, as mentioned, the wavelet correlation coefficients are typically supposed asymptotically normally distributed and under some assumptions, their limit covariance matrix has been more precisely established in [102]. It would be interesting to adapt their results to our work for getting more accurate results on real data set. The covariance matrix $\Omega(\Gamma)$ would be replaced by a more efficient estimator based on the wavelet properties.

Finally, this work provides a rigorous framework for the problem of graph inference in neuroscience. The neuroscientific community has so far considered the FWER criterion in their analyses. In this regards, some existing FWER controlling procedures are reviewed and theoretically clarified. This allows us to provide a real data analysis based on accurate theoretical results, which is rarely proposed in the neuroscientific literature.

Otherwise, the exploratory nature of the research questions in fMRI data analysis makes the FDR criterion an interesting criterion in this field. So, this thesis also investigates the FDR control, especially for non positive dependence structure and without supposing classical assumptions of sparsity. In general, the study of FDR control is technically demanding and the proof of Conjecture 4.1 is still an open problem and constitutes a very exciting mathematical challenge. It would be interesting to numerically explore this conjecture thanks to a deeper simulation study.

Bibliography

- [1] Jelle J Goeman, Aldo Solari, and Theo Stijnen. Three-sided hypothesis testing: Simultaneous testing of superiority, equivalence and inferiority. *Statistics in medicine*, 29(20):2117–2125, 2010. [15](#), [16](#), [33](#)
- [2] Craig M Bennett, Abigail A Baird, Michael B Miller, and George L Wolford. Neural correlates of interspecies perspective taking in the post-mortem atlantic salmon: an argument for proper multiple comparisons correction. *Journal of Serendipitous and Unexpected Results*, 1:1–5, 2011. [17](#), [18](#), [35](#)
- [3] Carlo E Bonferroni. *Il calcolo delle assicurazioni su gruppi di teste*. Tipografia del Senato, 1935. [19](#), [36](#), [100](#)
- [4] Sture Holm. A simple sequentially rejective multiple test procedure. *Scandinavian journal of statistics*, pages 65–70, 1979. [19](#), [36](#)
- [5] Jelle J Goeman and Aldo Solari. The sequential rejection principle of familywise error control. *The Annals of Statistics*, pages 3782–3810, 2010. [19](#), [36](#)
- [6] Yoav Benjamini and Yosef Hochberg. Controlling the false discovery rate: a practical and powerful approach to multiple testing. *Journal of the Royal Statistical Society. Series B (Methodological)*, pages 289–300, 1995. [19](#), [36](#), [37](#), [72](#)
- [7] Etienne Roquain. *Contributions to multiple testing theory for high-dimensional data*. PhD thesis, Université Pierre et Marie Curie, 2015. [20](#), [21](#), [37](#), [38](#)
- [8] Jelle J Goeman and Aldo Solari. Multiple hypothesis testing in genomics. *Statistics in medicine*, 33(11):1946–1978, 2014. [22](#), [39](#)
- [9] Matthew J Callow, Sandrine Dudoit, Elaine L Gong, Terence P Speed, and Edward M Rubin. Microarray expression profiling identifies genes with altered expression in HDL-deficient mice. *Genome research*, 10(12):2022–2029, 2000. [22](#)
- [10] Todd R Golub, Donna K Slonim, Pablo Tamayo, Christine Huard, Michelle Gaasenbeek, Jill P Mesirov, Hilary Coller, Mignon L Loh, James R Downing, Mark A

BIBLIOGRAPHY

- Caligiuri, et al. Molecular classification of cancer: class discovery and class prediction by gene expression monitoring. *science*, 286(5439):531–537, 1999. [22](#), [39](#)
- [11] Martin A Lindquist and Amanda Mejia. Zen and the art of multiple comparisons. *Psychosomatic medicine*, 77(2):114, 2015. [23](#), [40](#)
- [12] Craig M Bennett, George L Wolford, and Michael B Miller. The principled control of false positives in neuroimaging. *Social cognitive and affective neuroscience*, 4(4):417–422, 2009. [24](#), [26](#), [40](#), [43](#)
- [13] Robert E Livezey and WY Chen. Statistical field significance and its determination by Monte Carlo techniques. *Monthly Weather Review*, 111(1):46–59, 1983. [24](#), [41](#)
- [14] Daniel S Wilks. *Statistical methods in the atmospheric sciences*, volume 100. Academic press, 2011. [24](#), [41](#)
- [15] Daniel S Wilks. “the stippling shows statistically significant grid points”: How research results are routinely overstated and overinterpreted, and what to do about it. *Bulletin of the American Meteorological Society*, 97(12):2263–2273, 2016. [24](#), [41](#)
- [16] Edward Vul, Christine Harris, Piotr Winkielman, and Harold Pashler. Puzzlingly high correlations in fMRI studies of emotion, personality, and social cognition. *Perspectives on psychological science*, 4(3):274–290, 2009. [25](#), [26](#), [27](#), [42](#), [43](#), [44](#)
- [17] Matthew D Lieberman, Elliot T Berkman, and Tor D Wager. Correlations in social neuroscience aren’t voodoo: Commentary on vul et al.(2009). *Perspectives on Psychological Science*, 4(3):299–307, 2009. [26](#), [43](#)
- [18] Felix Carbonell, Keith J Worsley, Nelson J Trujillo-Barreto, and M Vega-Hernandez. The geometry of time-varying cross-correlation random fields. *Computational Statistics & Data Analysis*, 53(9):3291–3304, 2009. [48](#)
- [19] Lei Nie, Xian Yang, Paul M Matthews, Zhi-Wei Xu, and Yi-Ke Guo. Inferring functional connectivity in fmri using minimum partial correlation. *International Journal of Automation and Computing*, 14(4):371–385, 2017. [48](#)
- [20] Steffen L Lauritzen. *Graphical models*, volume 17. Clarendon Press, 1996. [48](#), [49](#)
- [21] Christophe Giraud. *Introduction to high-dimensional statistics*, volume 138. CRC Press, 2014. [49](#)
- [22] Mathias Drton and Michael D Perlman. Multiple testing and error control in Gaussian graphical model selection. *Statistical Science*, pages 430–449, 2007. [50](#), [104](#), [110](#)

-
- [23] David Roxbee Cox and Nanny Wermuth. *Multivariate dependencies: Models, analysis and interpretation*. Chapman and Hall/CRC, 2014. [53](#)
- [24] Arthur P Dempster. Covariance selection. *Biometrics*, pages 157–175, 1972. [54](#), [55](#)
- [25] T.W. Anderson. *An Introduction to Multivariate Statistical Analysis*. John Wiley & Sons, third edition edition, 2003. [54](#), [59](#), [110](#)
- [26] Iain M Johnstone. On the distribution of the largest eigenvalue in principal components analysis. *Annals of statistics*, pages 295–327, 2001. [54](#)
- [27] Jianqing Fan, Yuan Liao, and Han Liu. An overview of the estimation of large covariance and precision matrices. *The Econometrics Journal*, 19(1), 2016. [54](#), [55](#)
- [28] Larry Wasserman, Mladen Kolar, Alessandro Rinaldo, et al. Berry-esseen bounds for estimating undirected graphs. *Electronic Journal of Statistics*, 8(1):1188–1224, 2014. [54](#)
- [29] Somayeh Sojoudi. Equivalence of graphical lasso and thresholding for sparse graphs. *Journal of Machine Learning Research*, 17(115):1–21, 2016. [54](#)
- [30] Peter J Bickel and Elizaveta Levina. Regularized estimation of large covariance matrices. *The Annals of Statistics*, pages 199–227, 2008. [54](#)
- [31] Wei Biao Wu and Mohsen Pourahmadi. Banding sample autocovariance matrices of stationary processes. *Statistica Sinica*, pages 1755–1768, 2009. [54](#)
- [32] Peter J Bickel, Elizaveta Levina, et al. Covariance regularization by thresholding. *The Annals of Statistics*, 36(6):2577–2604, 2008. [54](#)
- [33] Olivier Ledoit, Michael Wolf, et al. Nonlinear shrinkage estimation of large-dimensional covariance matrices. *The Annals of Statistics*, 40(2):1024–1060, 2012. [54](#)
- [34] Onureena Banerjee, Laurent El Ghaoui, Alexandre d’Aspremont, and Georges Natsoulis. Convex optimization techniques for fitting sparse Gaussian graphical models. In *Proceedings of the 23rd international conference on Machine learning*, pages 89–96. ACM, 2006. [54](#)
- [35] Jianhua Z Huang, Naiping Liu, Mohsen Pourahmadi, and Linxu Liu. Covariance matrix selection and estimation via penalised normal likelihood. *Biometrika*, 93(1):85–98, 2006. [54](#)
- [36] Alexandre d’Aspremont, Onureena Banerjee, and Laurent El Ghaoui. First-order methods for sparse covariance selection. *SIAM Journal on Matrix Analysis and Applications*, 30(1):56–66, 2008. [54](#)

BIBLIOGRAPHY

- [37] Jerome Friedman, Trevor Hastie, and Robert Tibshirani. Sparse inverse covariance estimation with the graphical lasso. *Biostatistics*, 9(3):432–441, 2008. [54](#)
- [38] Pradeep Ravikumar, Martin J Wainwright, Garvesh Raskutti, Bin Yu, et al. High-dimensional covariance estimation by minimizing ℓ_1 -penalized log-determinant divergence. *Electronic Journal of Statistics*, 5:935–980, 2011. [54](#), [55](#)
- [39] Shuheng Zhou, Sara van de Geer, and Peter Bühlmann. Adaptive lasso for high dimensional regression and gaussian graphical modelling. *arXiv preprint arXiv:0903.2515*, 2009. [54](#)
- [40] Nicolai Meinshausen. A note on the lasso for Gaussian graphical model selection. *Statistics & probability letters*, 78(7):880–884, 2008. [55](#)
- [41] Nicolai Meinshausen and Peter Bühlmann. High-dimensional graphs and variable selection with the Lasso. *Annals of Statistics*, 34(3):1436–1462, June 2006. [55](#)
- [42] Tingni Sun and Cun-Hui Zhang. Sparse matrix inversion with scaled lasso. *The Journal of Machine Learning Research*, 14(1):3385–3418, 2013. [55](#)
- [43] Jianqing Fan, Yang Feng, and Yichao Wu. Network exploration via the adaptive LASSO and SCAD penalties. *The annals of applied statistics*, 3(2):521, 2009. [55](#)
- [44] Yunzhang Zhu, Xiaotong Shen, and Wei Pan. Structural pursuit over multiple undirected graphs. *Journal of the American Statistical Association*, 109(508):1683–1696, 2014. [55](#)
- [45] Ashwini Maurya. A joint convex penalty for inverse covariance matrix estimation. *Computational Statistics & Data Analysis*, 75:15–27, 2014. [55](#)
- [46] Vahe Avagyan, Andrés M Alonso, and Francisco J Nogales. Improving the Graphical Lasso Estimation for the Precision Matrix Through Roots of the Sample Covariance Matrix. *Journal of Computational and Graphical Statistics*, pages 1–8, 2017. [55](#)
- [47] Kean Ming Tan, Daniela Witten, and Ali Shojaie. The cluster graphical lasso for improved estimation of gaussian graphical models. *Computational statistics & data analysis*, 85:23–36, 2015. [55](#)
- [48] Jacob Bien and Robert J Tibshirani. Sparse estimation of a covariance matrix. *Biometrika*, 98(4):807–820, 2011. [55](#)
- [49] Emmanuel Candes and Terence Tao. The Dantzig selector: Statistical estimation when p is much larger than n . *The Annals of Statistics*, 35(6):2313–2351, 2007. [55](#)

-
- [50] Ming Yuan. High dimensional inverse covariance matrix estimation via linear programming. *Journal of Machine Learning Research*, 11(Aug):2261–2286, 2010. [55](#)
- [51] Tony Cai, Weidong Liu, and Xi Luo. A constrained ℓ_1 minimization approach to sparse precision matrix estimation. *Journal of the American Statistical Association*, 106(494):594–607, 2011. [55](#)
- [52] Tingni Sun and Cun-Hui Zhang. Scaled sparse linear regression. *Biometrika*, 99(4):879–898, 2012. [55](#)
- [53] Teng Zhang and Hui Zou. Sparse precision matrix estimation via lasso penalized D-trace loss. *Biometrika*, 101(1):103–120, 2014. [55](#)
- [54] T Tony Cai. Global testing and large-scale multiple testing for high-dimensional covariance structures. *Annual Review of Statistics and Its Application*, 4:423–446, 2017. [55](#)
- [55] Mathias Drton and Michael D Perlman. Model selection for Gaussian concentration graphs. *Biometrika*, 91(3):591–602, 2004. [57](#), [101](#)
- [56] T Tony Cai and Weidong Liu. Large-scale multiple testing of correlations. *Journal of the American Statistical Association*, 111(513):229–240, 2016. [58](#)
- [57] Tony Cai, Weidong Liu, and Yin Xia. Two-sample covariance matrix testing and support recovery in high-dimensional and sparse settings. *Journal of the American Statistical Association*, 108(501):265–277, 2013. [58](#)
- [58] Weidong Liu. Gaussian graphical model estimation with false discovery rate control. *The Annals of Statistics*, pages 2948–2978, 2013. [58](#)
- [59] Jie Liu, Chunming Zhang, David Page, et al. Multiple testing under dependence via graphical models. *The Annals of Applied Statistics*, 10(3):1699–1724, 2016. [58](#)
- [60] Murray A Aitkin. Some tests for correlation matrices. *Biometrika*, pages 443–446, 1969. [59](#), [99](#), [100](#)
- [61] George Kimeldorf and Allan R Sampson. A framework for positive dependence. *Annals of the Institute of Statistical Mathematics*, 41(1):31–45, 1989. [66](#)
- [62] T.K Sarkar. *Some lower bounds of reliability*. PhD thesis, ProQuest LLC, Ann Arbor, MI. Stanford University, 1969. [66](#)
- [63] Samuel Karlin and Yosef Rinott. Total positivity properties of absolute value multinormal variables with applications to confidence interval estimates and related probabilistic inequalities. *The Annals of Statistics*, pages 1035–1049, 1981. [66](#), [85](#), [86](#), [113](#)

BIBLIOGRAPHY

- [64] Erich Leo Lehmann. Some concepts of dependence. *The Annals of Mathematical Statistics*, pages 1137–1153, 1966. [67](#), [69](#), [71](#)
- [65] Gilles Blanchard and Etienne Roquain. Two simple sufficient conditions for fdr control. *Electronic journal of Statistics*, 2:963–992, 2008. [67](#)
- [66] Samuel Karlin and Yosef Rinott. Classes of orderings of measures and related correlation inequalities. I. Multivariate totally positive distributions. *Journal of Multivariate Analysis*, 10(4):467–498, 1980. [68](#)
- [67] Yoav Benjamini and Daniel Yekutieli. The control of the false discovery rate in multiple testing under dependency. *Annals of statistics*, pages 1165–1188, 2001. [68](#), [69](#), [74](#), [75](#), [84](#), [85](#), [113](#)
- [68] Daniel Yekutieli. False discovery rate control for non-positively regression dependent test statistics. *Journal of Statistical Planning and Inference*, 138(2):405–415, 2008. [68](#)
- [69] James D Esary, Frank Proschan, David W Walkup, et al. Association of random variables, with applications. *The Annals of Mathematical Statistics*, 38(5):1466–1474, 1967. [68](#)
- [70] RL Dykstra, John E Hewett, and WA Thompson Jr. Events which are almost independent. *The Annals of Statistics*, pages 674–681, 1973. [69](#)
- [71] Zbyněk Šidák. Rectangular confidence regions for the means of multivariate normal distributions. *Journal of the American Statistical Association*, 62(318):626–633, 1967. [70](#), [101](#)
- [72] Olive Jean Dunn. Estimation of the means of dependent variables. *The Annals of Mathematical Statistics*, pages 1095–1111, 1958. [70](#)
- [73] Theodore W Anderson. The integral of a symmetric unimodal function over a symmetric convex set and some probability inequalities. *Proceedings of the American Mathematical Society*, 6(2):170–176, 1955. [70](#)
- [74] Gilles Blanchard and Etienne Roquain. Adaptive false discovery rate control under independence and dependence. *Journal of Machine Learning Research*, 10(Dec):2837–2871, 2009. [74](#)
- [75] Wenge Guo and M Bhaskara Rao. On control of the false discovery rate under no assumption of dependency. *Journal of Statistical Planning and Inference*, 138(10):3176–3188, 2008. [75](#), [84](#)

- [76] Wenge Guo and M Bhaskara Rao. On optimality of the Benjamini-Hochberg procedure for the false discovery rate. *Statistics & Probability Letters*, 78(14):2024–2030, 2008. [79](#), [153](#)
- [77] Christopher R Genovese, Nicole A Lazar, and Thomas Nichols. Thresholding of statistical maps in functional neuroimaging using the false discovery rate. *Neuroimage*, 15(4):870–878, 2002. [80](#)
- [78] Cohen Rami. Finding maximum fdr, at two tailed hypotheses testing in a case that the test statistics are normally distributed. internal communication. [83](#), [85](#), [86](#)
- [79] Anat Reiner-Benaim. FDR control by the BH procedure for two-sided correlated tests with implications to gene expression data analysis. *Biometrical Journal*, 49(1):107–126, 2007. [85](#), [91](#)
- [80] Thomas Royen. A simple proof of the Gaussian correlation conjecture extended to multivariate Gamma distributions. *arXiv preprint arXiv:1408.1028*, 2014. [86](#)
- [81] Etienne Roquain and Fanny Villers. Exact calculations for false discovery proportion with application to least favorable configurations. *The Annals of Statistics*, pages 584–612, 2011. [7](#), [92](#), [93](#)
- [82] Gilles Blanchard, Thorsten Dickhaus, Etienne Roquain, and Fanny Villers. On least favorable configurations for step-up-down tests. *arXiv preprint arXiv:1108.5262*, 2011. [93](#)
- [83] Helmut Finner, M Roters, and K Strassburger. On the simes test under dependence. *Statistical Papers*, 58(3):775–789, 2017. [96](#)
- [84] Taras Bodnar and Thorsten Dickhaus. On the simes inequality in elliptical models. *Annals of the Institute of Statistical Mathematics*, 69(1):215–230, 2017. [96](#)
- [85] Sanat K Sarkar. Some probability inequalities for ordered MTP_2 random variables: a proof of the simes conjecture. *Annals of Statistics*, pages 494–504, 1998. [96](#)
- [86] Peter H Westfall and S Stanley Young. *Resampling-based multiple testing: Examples and methods for p-value adjustment*, volume 279. John Wiley & Sons, 1993. [99](#), [101](#), [103](#)
- [87] Patrick Billingsley. *Convergence of probability measures*. John Wiley & Sons, 2013. [102](#), [105](#)
- [88] Joseph P Romano and Michael Wolf. Exact and approximate stepdown methods for multiple hypothesis testing. *Journal of the American Statistical Association*, 100(469):94–108, 2005. [103](#), [109](#)

BIBLIOGRAPHY

- [89] Roger A Horn and Charles R Johnson. *Matrix analysis*. Cambridge university press, 2012. [105](#)
- [90] Allan Gut. *Probability: a graduate course*, volume 75. Springer Science & Business Media, 2012. [109](#)
- [91] Larry Wasserman. *All of statistics: a concise course in statistical inference*. Springer Science & Business Media, 2013. [109](#), [110](#)
- [92] Joseph P Romano, Azeem M Shaikh, and Michael Wolf. Control of the false discovery rate under dependence using the bootstrap and subsampling. *Test*, 17(3):417, 2008. [110](#)
- [93] Harry Joe. Generating random correlation matrices based on partial correlations. *Journal of Multivariate Analysis*, 97(10):2177–2189, 2006. [113](#)
- [94] Christopher P. Pawela, Bharat B. Biswal, Younghoon R. Cho, Dennis S. Kao, Ruyang Li, Seth R. Jones, Marie L. Schulte, Hani S. Matloub, Anthony G. Hudetz, and James S. Hyde. Resting-state functional connectivity of the rat brain. *Magnetic Resonance in Medicine*, 59(5):1021–1029, 2008. [126](#)
- [95] S. Achard and I. Gannaz. Multivariate wavelet whittle estimation in long-range dependence. *Journal of Time Series Analysis*, 37(4):476–512, 2016. arXiv preprint arXiv:1412.0391. [126](#)
- [96] E. Moulines, F. Roueff, and M. Taqqu. On the spectral density of the wavelet coefficients of long memory time series with application to the log-regression estimation of the memory parameter. *Journal of Time Series Analysis*, 28:155–187, 2007. [126](#)
- [97] B. Whitcher, P. Guttorp, and D. B. Percival. Wavelet analysis of covariance with application to atmospheric time series. *Journal of Geophysical Research*, 105(D11)(14):941–962, 2000. [126](#)
- [98] Martijn P van den Heuvel, Siemon C de Lange, Andrew Zalesky, Caio Seguin, BT Thomas Yeo, and Ruben Schmidt. Proportional thresholding in resting-state fMRI functional connectivity networks and consequences for patient-control connectome studies: Issues and recommendations. *Neuroimage*, 152:437–449, 2017. [130](#)
- [99] Sophie Achard, Chantal Delon-Martin, Petra E Vértes, Félix Renard, Maleka Schenck, Francis Schneider, Christian Heinrich, Stéphane Kremer, and Edward T Bullmore. Hubs of brain functional networks are radically reorganized in comatose patients. *Proceedings of the National Academy of Sciences*, 109(50):20608–20613, 2012. [131](#)

- [100] Shaun Fallat, Steffen Lauritzen, Kayvan Sadeghi, Caroline Uhler, Nanny Wermuth, Piotr Zwiernik, et al. Total positivity in markov structures. *The Annals of Statistics*, 45(3):1152–1184, 2017. [133](#)
- [101] Martin A Lindquist and Andrew Gelman. Correlations and multiple comparisons in functional imaging: a statistical perspective (commentary on Vul et al., 2009). *Perspectives on Psychological Science*, 4(3):310–313, 2009. [134](#)
- [102] Brandon Whitcher, Peter Gutterp, and Donald Percival. Mathematical background for wavelet estimators of cross-covariance and cross-correlation. 1999. [134](#)
- [103] Vitali D Milman and Gideon Schechtman. *Asymptotic Theory of Finite Dimensional Normed Spaces: Isoperimetric Inequalities in Riemannian Manifolds*, volume 1200. Springer, 2009. [145](#)
- [104] Robb J Muirhead. *Aspects of multivariate statistical theory*, volume 197. John Wiley & Sons, 2009. [149](#)

BIBLIOGRAPHY

Appendix A

Supplementary results

A.1 Proofs of Chapter 3

A.1.1 Proof of Theorem 3.4

Consider the measure space $(\mathbb{R}^m, \mathcal{B}(\mathbb{R}^m), \lambda)$, where λ corresponds to the Lebesgue's measure. The proof of Theorem 3.4 requires the theorem of Brunn-Minkowski. Let us begin by introduce the latter.

Denote $V\{A\} = \lambda(A)$ for all $A \subseteq \mathbb{R}^m$. Let K_0 and K_1 be two compact sets of \mathbb{R}^m . The set $K_0 + K_1$ is defined as follows

$$K_0 + K_1 = \{x \in \mathbb{R}^m : \exists (k_0, k_1) \in K_0 \times K_1, x = k_0 + k_1\}.$$

Theorem. (*Brunn-Minkowski*) Let K_0 and K_1 be two non-empty compact sets of \mathbb{R}^m . Then, we have

$$V\{K_0 + K_1\}^{1/m} \geq V\{K_0\}^{1/m} + V\{K_1\}^{1/m} \quad (\text{A.1})$$

Proof. (Theorem of Brunn-Minkowski)

The sketch of the proof is given by [103] (see Appendix III.1). Given two m -tuples of real numbers, the rectangular cuboid $P(a, b)$ is defined as

$$P(a, b) = \{x \in \mathbb{R}^m : \forall i \in \llbracket 1, m \rrbracket, a_i \leq x_i \leq b_i\}.$$

First, we prove that (A.1) holds when K_0 and K_1 are finite union of rectangular cuboid with non-empty and disjoint interior. We define K_0 and K_1 as

$$K_0 = \bigcup_{i=1}^{m_0} P(a^{(i)}, b^{(i)}), \quad K_1 = \bigcup_{i=1}^{m_1} P(c^{(i)}, d^{(i)}).$$

We proceed by induction on $m = m_0 + m_1$.

APPENDIX A. SUPPLEMENTARY RESULTS

Case $m = 2$ We have $K_0 = P(a, b)$, $K_1 = P(c, d)$ and $V\{K_0\}^{1/m} = \prod_{i=1}^m l_{0i}^{1/m}$ where $l_{0i} = b_i - a_i$, $V\{K_1\}^{1/m} = \prod_{i=1}^m l_{1i}^{1/m}$ where $l_{1i} = d_i - c_i$, $V\{K_0 + K_1\}^{1/m} = \prod_{i=1}^m (l_{0i} + l_{1i})^{1/m}$. We conclude by using the inequality of arithmetic and geometric means:

$$\begin{aligned} \frac{V\{K_0\}^{1/m} + V\{K_1\}^{1/m}}{V\{K_0 + K_1\}^{1/m}} &= \prod_{i=1}^m \left(\frac{l_{0i}}{l_{0i} + l_{1i}} \right)^{1/m} + \prod_{i=1}^m \left(\frac{l_{1i}}{l_{0i} + l_{1i}} \right)^{1/m} \\ &\leq \frac{1}{m} \sum_{i=1}^m \frac{l_{0i}}{l_{0i} + l_{1i}} + \frac{1}{m} \sum_{i=1}^m \frac{l_{1i}}{l_{0i} + l_{1i}} = 1. \end{aligned}$$

Cases $m > 2$ Assume that (A.1) holds for all K'_0 and K'_1 , finite union of rectangular cuboid with non-empty and disjoint interior, such that $m'_0 + m'_1 < m$. Let $m_0 \geq 2$. A 3-tuple $(k, t, u) \in \llbracket 1, m \rrbracket \times \mathbb{R} \times \mathbb{R}$ is constructed such that one of the cuboid contained in K_0 is contained in both $\{x_k \geq t\}$ and $\{x_k \leq t\}$ and such that we have

$$\frac{V\{K_0 \cap \{x_k \leq t\}\}}{V\{K_0 \cap \{x_k \geq t\}\}} = \frac{V\{K_1 \cap \{x_k \leq u\}\}}{V\{K_1 \cap \{x_k \geq u\}\}}.$$

Since the interiors of $P(a^{(1)}, b^{(1)})$ and $P(a^{(2)}, b^{(2)})$ are disjoint, there exists $k \in \llbracket 1, m \rrbracket$ such that $]a_k^{(1)}, b_k^{(1)}[\cap]a_k^{(2)}, b_k^{(2)}[= \emptyset$. Thus we can choose $t = \min\{b_k^{(1)}, b_k^{(2)}\}$.

Otherwise, it results from the intermediate value theorem that the function $z \mapsto V\{K_1 \cap \{x_k \leq z\}\}$ is surjective on $[0, V\{K_1\}]$. Thus there exists a real number u such that

$$V\{K_1 \cap \{x_k \leq u\}\} = \frac{V\{K_1\}V\{K_0 \cap \{x_k \leq t\}\}}{V\{K_0\}} \in [0, V\{K_1\}],$$

and we can choose such a u .

Let $K_0^- = K_0 \cap \{x_k \leq t\}$, $K_0^+ = K_0 \cap \{x_k \geq t\}$, $K_1^- = K_1 \cap \{x_k \leq u\}$ and $K_1^+ = K_1 \cap \{x_k \geq u\}$. We have

$$V\{K_0 + K_1\} \geq V\{K_0^- + K_1^-\} + V\{K_0^+ + K_1^+\}.$$

Moreover, the interiors of the cuboid contained in $K_0^- \cup K_1^-$ and $K_0^+ \cup K_1^+$ are non-empty and disjoint, and by the nature of k , their quantity is strictly less than m . It follows from induction hypothesis and the definition of u that

$$\begin{aligned} V\{K_0 + K_1\} &\geq \left(V\{K_0^-\}^{1/m} + V\{K_1^-\}^{1/m} \right)^m + \left(V\{K_0^+\}^{1/m} + V\{K_1^+\}^{1/m} \right)^m \\ &= V\{K_0^-\} \left(1 + \frac{V\{K_1^-\}^{1/m}}{V\{K_0^-\}^{1/m}} \right)^m + V\{K_0^+\} \left(1 + \frac{V\{K_1^+\}^{1/m}}{V\{K_0^+\}^{1/m}} \right)^m \\ &= \left(V\{K_0^-\}^{1/m} + V\{K_1^-\}^{1/m} \right)^m. \end{aligned}$$

A finite union of rectangular cuboid can be written as union of rectangular cuboid with disjoint interiors. Thus, the inequality is still valid in such a case.

We now extend the latter to two non-empty bounded open sets U_0 and U_1 that are respectively countable union of rectangular cuboid $(P_n)_{n \in \mathbb{N}}$ and $(P'_n)_{n \in \mathbb{N}}$. Then $V\{U_0\} = \lim_{n \rightarrow +\infty} V\{\bigcup_{0 \leq p \leq n} P_p\}$ et $V\{U_1\} = \lim_{n \rightarrow +\infty} V\{\bigcup_{0 \leq p \leq n} P'_p\}$. It follows from the previous result that

$$\begin{aligned} \mathcal{V}(n) &= V\left\{\bigcup_{0 \leq p \leq n} P_p\right\}^{1/m} + V\left\{\bigcup_{0 \leq p \leq n} P'_p\right\}^{1/m} \leq \left(V\left\{\bigcup_{0 \leq p \leq n} P_p\right\} + V\left\{\bigcup_{0 \leq p \leq n} P'_p\right\}\right)^{1/m} \\ &\leq V\{U_0 + U_1\}^{1/m}. \end{aligned}$$

We conclude by taking the limit of $\mathcal{V}(n)$.

Let K_0 and K_1 be two non-empty compact sets. Define U_n and V_n as

$$U_n = \left\{x \in \mathbb{R}^m : d(x, K_0) < \frac{1}{2^n}\right\}, V_n = \left\{x \in \mathbb{R}^m : d(x, K_1) < \frac{1}{2^n}\right\}.$$

For all $n \in \mathbb{N}$, $V\{K_0\}^{1/m} + V\{K_1\}^{1/m} \leq V\{U_n\}^{1/m} + V\{V_n\}^{1/m} \leq V\{U_n + V_n\}^{1/m}$, where the last inequality results from the previous inequality. Thus, $V\{K_0\}^{1/m} + V\{K_1\}^{1/m} \leq \inf_{n \in \mathbb{N}} V\{U_n + V_n\}^{1/m} = V\left\{\bigcap_{n \in \mathbb{N}} (U_n + V_n)\right\}^{1/m}$. The conclusion follows from $K_0 + K_1 = \bigcap_{n \in \mathbb{N}} (U_n + V_n)$. We prove that $\bigcap_{n \in \mathbb{N}} (U_n + V_n) \subseteq K_0 + K_1$, the converse inclusion is trivial. Let $a \in \bigcap_{n \in \mathbb{N}} (U_n + V_n)$. For all $n \in \mathbb{N}$, there exists $(x_n, y_n) \in U_n \times V_n$ such that $a = x_n + y_n$. Since $(x_n)_{n \in \mathbb{N}}$ is bounded, the latter has a convergent subsequence $(x_{n_k})_{k \in \mathbb{N}}$. Denoting by x its limit, the limit of the sequence $(y_{n_k})_{k \in \mathbb{N}}$ is $y = a - x$. Since for all $k \in \mathbb{N}$, $d(x_{n_k}, K_0) \leq 1/2^{n_k}$ and $d(y_{n_k}, K_1) \leq 1/2^{n_k}$, we have $d(x, K_0) = 0$ and $x \in K_0$ because K_0 is closed. Similarly, $y \in K_1$ and finally $a = x + y \in K_0 + K_1$. \square

The inequality (A.1) can be rewritten as: let K_0 and K_1 two non-empty compact sets of \mathbb{R}^m . Then, for all $t \in [0, 1]$, we have

$$V\{tK_0 + (1-t)K_1\}^{1/m} \geq tV\{K_0\}^{1/m} + (1-t)V\{K_1\}^{1/m}.$$

We are now in position to report the proof of the Anderson's theorem, which relies on the Brunn-Minkowski's theorem.

Proof. (Theorem 3.4)

Let E be a convex set symmetric with respect to the origin and $Y \in \mathbb{R}^m$ a deterministic vector. First, we prove that for all $u \geq 0$, we have

$$H(u) = V\{(E + kY) \cap K_u\} \geq V\{(E + Y) \cap K_u\} = H^*(u).$$

Denote $\alpha[(E + Y) \cap K_u] + (1-\alpha)[(E - Y) \cap K_u]$ the set of all linear combinations $\alpha z + (1-\alpha)w$ where $z \in (E + Y) \cap K_u$, $w \in (E - Y) \cap K_u$ and $\alpha \in [0, 1]$. Define $\alpha = (1+k)/2$ so that

$$\alpha Y + (1-\alpha)(-Y) = kY. \tag{A.2}$$

APPENDIX A. SUPPLEMENTARY RESULTS

It results from the convexity of K_u that $K_u \supseteq \alpha K_u + (1-\alpha)K_u$. Using (A.2) and the convexity of E we obtain that $E + kY \supseteq \alpha E + (1-\alpha)E + kY = \alpha(E + Y) + (1-\alpha)(E - Y)$. This yields $(E + kY) \cap K_u \supseteq \alpha[(E + Y) \cap K_u] + (1-\alpha)[(E - Y) \cap K_u]$ and thus,

$$V\{(E + kY) \cap K_u\} \geq V\{\alpha[(E + Y) \cap K_u] + (1-\alpha)[(E - Y) \cap K_u]\}.$$

Since $(E + Y) \cap K_u$ is the symmetric set with respect to the origin of $(E - Y) \cap K_u$, we have $V\{(E + Y) \cap K_u\} = V\{(E - Y) \cap K_u\}$. For all $u > 0$, $V\{K_u\} = \int \mathbb{1}\{f(x) \geq u\} dx \leq \frac{1}{u} \int f(x) dx$, then K_u is bounded and the result is achieved by the Brunn-Minkowski's theorem,

$$\begin{aligned} V\{(E + kY) \cap K_u\} &\geq \left(\alpha V\{(E + Y) \cap K_u\}^{1/m} + (1-\alpha)V\{(E - Y) \cap K_u\}^{1/m}\right)^m \\ &= \left(\alpha V\{(E + Y) \cap K_u\}^{1/m} + (1-\alpha)V\{(E + Y) \cap K_u\}^{1/m}\right)^m \\ &= V\{(E + Y) \cap K_u\}. \end{aligned}$$

For all $0 \leq u_1 \leq u_2$, we have $H(u_1) - H(u_2) = V\{(E + kY) \cap \{x \in \mathbb{R}^m : u_1 \leq f(x) \leq u_2\}\}$ and using the definition of the Stieltjes integral, it follows that

$$\begin{aligned} \int_{E+kY} f(x) dx - \int_{E+Y} f(x) dx &= - \int_0^\infty u dH(u) + \int_0^\infty u dH^*(u) \\ &= \int_0^\infty u [dH^*(u) - dH(u)]. \end{aligned}$$

By integration by parts we obtain

$$\begin{aligned} \int_a^b u [dH^*(u) - dH(u)] &= \left[u(H^*(u) - H(u)) \right]_a^b - \int_a^b [H^*(u) - H(u)] du \\ &= b(H^*(b) - H(b)) - a(H^*(a) - H(a)) + \int_a^b [H(u) - H^*(u)] du. \end{aligned}$$

Furthermore, $\lim_{b \rightarrow +\infty} bH(b) = 0 = \lim_{b \rightarrow +\infty} bH^*(b)$ and for all $a \geq 0$, $-a(H^*(a) - H(a)) \geq 0$ and $\int_a^b [H(u) - H^*(u)] du \geq 0$. Thus, we finally have that $\int_0^\infty u [dH^*(u) - dH(u)] \geq 0$. \square

A.1.2 Proof of Theorem 3.3

Proof. (Theorem 3.3)

Let $X = (X_1, \dots, X_m) \in \mathbb{R}^m$ be a centered Gaussian vector having an arbitrary covariance matrix Σ and (c_2, \dots, c_m) a sequence of positive constants. Assume that Σ is invertible. First, we prove that

$$\mathbb{P}(|X_1| \leq c_1, \dots, |X_m| \leq c_m) \geq \mathbb{P}(|X_1| \leq c_1) \mathbb{P}(|X_2| \leq c_2, \dots, |X_m| \leq c_m). \quad (\text{A.3})$$

Inequality (A.3) is achieved by proving that the function F defined as

$$F(c_1) = \int_{-c_1}^{c_1} \int_{-c_2}^{c_2} \dots \int_{-c_m}^{c_m} f(x_1, \dots, x_m) dx_1 dx_2 \dots dx_m \\ - \int_{-c_1}^{c_1} \int_{-c_2}^{c_2} \dots \int_{-c_m}^{c_m} f(x_1) f(x_2, \dots, x_m) dx_1 dx_2 \dots dx_m,$$

is positive for all $c_1 \geq 0$. Using the symmetry of the Gaussian distribution, we can rewrite F as

$$F(c_1) = 2 \int_0^{c_1} f(x_1) \int_{-c_2}^{c_2} \dots \int_{-c_m}^{c_m} f(x_2, \dots, x_m | x_1) - f(x_2, \dots, x_m) dx_2 \dots dx_m dx_1.$$

Thus

$$\frac{\partial F(c_1)}{\partial c_1} = 2f(c_1) \int_{-c_2}^{c_2} \dots \int_{-c_m}^{c_m} f(x_2, \dots, x_m | c_1) - f(x_2, \dots, x_m) dx_2 \dots dx_m.$$

Now, using Theorem 3.4, we prove that the function G defined as

$$G(c_1) = \int_{-c_2}^{c_2} \dots \int_{-c_m}^{c_m} f(x_2, \dots, x_m | c_1) - f(x_2, \dots, x_m) dx_2 \dots dx_m,$$

is a nonincreasing function. Define H as

$$H(c_1) = \int_{-c_2}^{c_2} \dots \int_{-c_m}^{c_m} f(x_2, \dots, x_m | c_1) dx_2 \dots dx_m.$$

$f(x_2, \dots, x_m | c_1)$ corresponds to a Gaussian density function with $\rho_{12}\sigma_2\sigma^{-1}c_1, \dots, \rho_{1m}\sigma_m\sigma^{-1}c_1$ means for which the variances and the correlation coefficients are independent of c_1 (see Theorem 1.2.11 in [104]). Thus, the density of $f(x_2, \dots, x_m | c_1)$ can be obtained by "shifting" the density of $f(x_2, \dots, x_m | 0)$ so that

$$H(c_1) = \int_{E'} f(x_2, \dots, x_m | 0) dx_2 \dots dx_m,$$

where $E' = [-c_2 - \rho_{12}\sigma_2\sigma^{-1}c_1, c_2 - \rho_{12}\sigma_2\sigma^{-1}c_1] \times \dots \times [-c_m - \rho_{1m}\sigma_m\sigma^{-1}c_1, c_m - \rho_{1m}\sigma_m\sigma^{-1}c_1]$. It results from Theorem 3.4 (used with $E = [-c_2, c_2] \times \dots \times [-c_m, c_m]$, $Y = (-\rho_{12}\sigma_2\sigma^{-1}c_1, \dots, -\rho_{1m}\sigma_m\sigma^{-1}c_1)$ and $f(x) = f(x_2, \dots, x_m | 0)$) that H is a nondecreasing function and thus G is also a nondecreasing function. Moreover, since $\lim_{c_1 \rightarrow +\infty} H(c_1) = 0$, we have $\lim_{c_1 \rightarrow +\infty} G(c_1) < 0$ and then two cases can occur:

Case a For all $c_1 \geq 0$, $G(c_1) \leq 0$.

APPENDIX A. SUPPLEMENTARY RESULTS

Case b There exists $c^* > 0$ such that $G(c_1) > 0$ for $0 \leq c_1 \leq c^*$ and $G(c_1) \leq 0$ for $c_1 > c^*$.

Note that

$$F(0) = 0 \quad \text{and} \quad \lim_{c_1 \rightarrow +\infty} F(c_1) = 0, \quad (\text{A.4})$$

and since $f(c_1) \geq 0$, $\partial F(c_1)/\partial c_1$ has same sign of G . Thus

Case a For all $c_1 \geq 0$, F is nonincreasing. (A.4) yields that for all $c_1 \geq 0$, $F(c_1) = 0$ and then we obtain the case of equality in (A.3).

Case b There exists a constant $c^* > 0$ such that $F(c_1)$ is nondecreasing for $0 \leq c_1 < c^*$ and nonincreasing on $]c^*, +\infty[$. Thus, $F(c_1) \geq 0$ for all $c_1 \geq 0$. Finally, for all $c_1 \geq 0$, we have

$$\int_{-c_1}^{c_1} \int_{-c_2}^{c_2} \dots \int_{-c_m}^{c_m} f(x_1, \dots, x_m) dx_1 dx_2 \dots dx_m \geq \int_{-c_1}^{c_1} \int_{-c_2}^{c_2} \dots \int_{-c_m}^{c_m} f(x_1) f(x_2, \dots, x_m) dx_1 dx_2 \dots dx_m,$$

and Šidák's inequality is obtained by induction. \square

A.1.3 Proof of Theorem 3.7

Let $(\tau_k)_{1 \leq k \leq m}$ be a nondecreasing sequence of critical values and $k_1 = \max \left\{ \arg \max_{k \in \{1, \dots, m\}} \left\{ \frac{\tau_k}{k} \right\} \right\}$.

We have divided the proof of Theorem 3.7 into a sequence of intermediate results. Let us begin by proving the following theorem:

Theorem A.1. *For all $P \in \mathcal{P}$, for all procedures $SU(\tau)$ such that (τ, \mathbf{p}) satisfies the hypothesis (3.16), we have*

$$FDR(\mathcal{R}_{SU(\tau)}, P) \leq \frac{\tau_{k_1} m_0(P)}{k_1}. \quad (\text{A.5})$$

Moreover, for all $m_0 \in \{0, \dots, m\}$, if for all distribution ν on $[0, 1]^m$ with m_0 uniform marginals, there exists $P^ \in \mathcal{P}$ with $|\mathcal{H}_0(P^*)| = m_0$ and $(p_i(X))_i \underset{X \sim P^*}{\sim} \nu$, then there exists a joint distribution of p -values for which the bound (A.5) is attained.*

Proof. (Theorem A.1)

For all $P \in \mathcal{P}$, for all procedures $SU(\tau)$ such that (τ, \mathbf{p}) satisfies the hypothesis (3.16), the

FDR can be written as

$$\begin{aligned}
 \text{FDR}(\mathcal{R}_{S U(\tau)}, P) &= \sum_{k=1}^m \frac{1}{k} \mathbb{E} \left[|\mathcal{R}_{S U(\tau)} \cap \mathcal{H}_0(P)| \cdot \mathbb{1}_{|\mathcal{R}_{S U(\tau)}|=k} \right] \\
 &= \sum_{k=1}^m \frac{1}{k} \mathbb{E} \left[\sum_{i \in \mathcal{H}_0(P)} \mathbb{1}_{p_i \leq \tau_k} \cdot \mathbb{1}_{|\mathcal{R}_{S U(\tau)}|=k} \right] \\
 &= \sum_{i \in \mathcal{H}_0(P)} \sum_{k=1}^m \frac{1}{k} \mathbb{P}(|\mathcal{R}_{S U(\tau)}| = k \mid p_i \leq \tau_k) \mathbb{P}(p_i \leq \tau_k).
 \end{aligned}$$

It results from (3.1) and the definition of k_1 that

$$\text{FDR}(\mathcal{R}_{S U(\tau)}, P) \leq \frac{\tau_{k_1}}{k_1} \sum_{i \in \mathcal{H}_0(P)} \sum_{k=1}^m \mathbb{P}(|\mathcal{R}_{S U(\tau)}| = k \mid p_i \leq \tau_k). \quad (\text{A.6})$$

Finally (A.5) follows from the hypothesis (3.16).

Now, let us construct a probability distribution P^* that attains this bounds. Let U_1, \dots, U_{m+1} be a family of independent random variables such that $U_i \sim \mathcal{U}[\tau_{i-1}, \tau_i]$, $i = 1 \dots, m$ and $U_{m+1} \sim \mathcal{U}[\tau_m, 1]$. Let N a random variable taking values in $\{1, \dots, k_1\} \cup \{m+1\}$ and generated from the following law:

$$\mathbb{P}(N = n) = \begin{cases} \pi_n & n = 1, \dots, k_1 \\ 1 - \sum_{i=1}^{k_1} \pi_i & n = m+1 \end{cases}, \text{ where } \pi_n = \begin{cases} m_0(\tau_n - \tau_{n-1})/k_1 & \text{if } k_1 \leq m_0 \\ \tau_n - \tau_{n-1} & \text{otherwise} \end{cases}.$$

Depending of the value of N , we associate one p -value to each index $1, 2, \dots, m$ by the following way:

Case $N = n \in \{1, \dots, k_1\} \subset I_0 = \{1, \dots, m_0\}$ Pick n indexes $\{i_1, \dots, i_n\} := I_{0,n}$ randomly and without replacement in I_0 . For all $j \in \{1, \dots, n\}$, the index i_j is associated to $p_{i_j} = U_n$. For all $s \in \{1, \dots, m\} \setminus I_{0,n} = I \setminus I_{0,n}$, the index s is associated to $p_s = U_{m+1}$.

Case $N = n \in I_1 = \{m_0 + 1, \dots, m\}$ ($k_1 > m_0$) Choose the indexes i_1, \dots, i_{m_0} in I_0 and pick randomly and without replacement the $(n - m_0)$ remaining indexes in I_1 . Denote $I_{1,n} := \{i_1, \dots, i_{m_0}, i_{m_0+1}, \dots, i_n\}$ the index set of picked indexes. For all $j \in \{1, \dots, n\}$, the index i_j is associated to $p_{i_j} = U_n$. For all $s \in I \setminus I_{1,n}$, the index s is associated to $p_s = U_{m+1}$.

APPENDIX A. SUPPLEMENTARY RESULTS

Case $N = m + 1$ For all $i \in I_0$, the index i is associated to $p_i = U_{m+1}$. For all $j \in I_1$, the index j is associated to $p_j = 0$.

When there exists $i \in I_0$ and $k \in \{1, \dots, m\}$ such that $p_i \leq \tau_k$, Figure A.1 and Figure A.2 illustrate this construction. Otherwise, we deduce for each case, the number of rejects, denoted by R , and the associated value of the FDP.

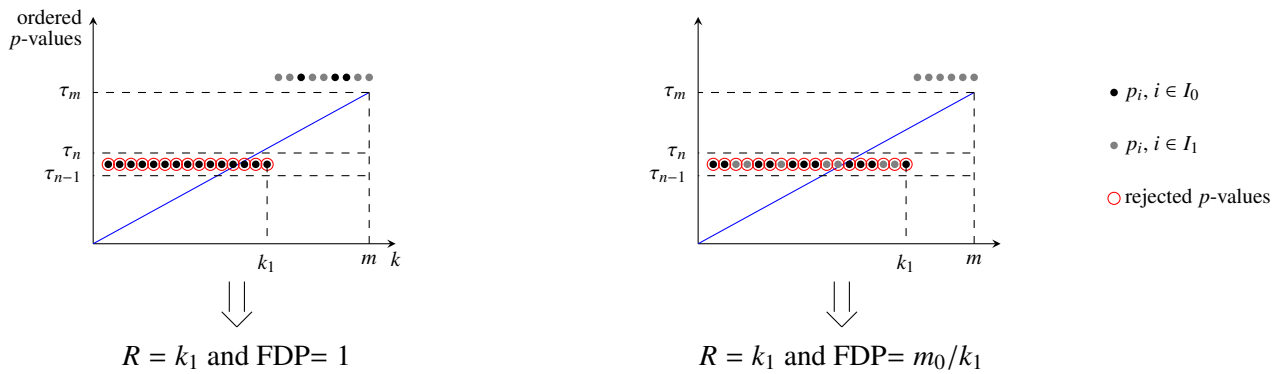


Figure A.1 – $n \in I_0$ (left), $n \in I_1$ (right).

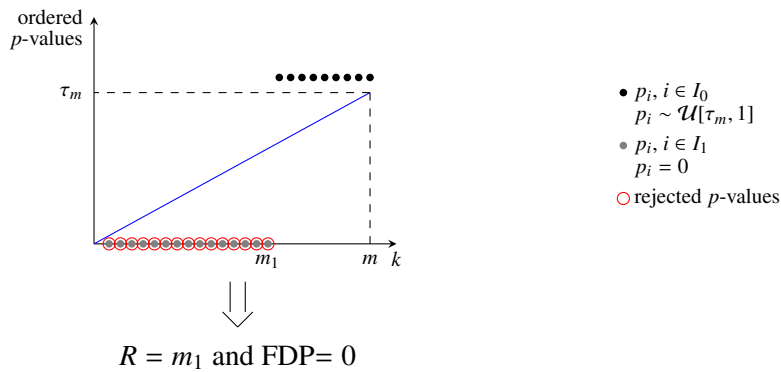


Figure A.2 – $n = m + 1$.

This yields

$$\begin{aligned}
\text{FDR}(P^*) &= \sum_{n=1}^{k_1} \mathbb{E} \left[\underbrace{\mathbb{E}[\text{FDP}(P^*) | N = n]}_{=1} \mathbb{1}\{N = n\} \right] + \mathbb{E} \left[\underbrace{\mathbb{E}[\text{FDP}(P^*) | N = m + 1]}_{=0} \mathbb{1}\{N = m + 1\} \right] \\
&= \begin{cases} \sum_{n=1}^{k_1} \mathbb{E} \left[\underbrace{\mathbb{E}[\text{FDP}(P^*) | N = n]}_{=1} \mathbb{1}\{N = n\} \right] & \text{if } k_1 \leq m_0 \\ \sum_{n=1}^{k_1} \mathbb{E} \left[\underbrace{\mathbb{E}[\text{FDP}(P^*) | N = n]}_{=m_0/k_1} \mathbb{1}\{N = n\} \right] & \text{if } k_1 > m_0 \end{cases} \\
&= \begin{cases} \sum_{n=1}^{k_1} \mathbb{P}(N = n) & \text{if } k_1 \leq m_0 \\ (m_0/k_1) \sum_{n=1}^{k_1} \mathbb{P}(N = n) & \text{if } k_1 > m_0 \end{cases} \\
&= m_0 \tau_{k_1} / k_1.
\end{aligned}$$

□

Now, we provide a second intermediate result useful for the proof of Theorem 3.7.

Proposition A.1. *The p -values induced by the probability distribution P^* :*

1. *are upper-bounded by a uniform variable, that is, satisfy (3.1),*
2. *satisfy the hypothesis (3.16),*
3. *are not PRDS on $\{1, \dots, m_0\}$.*

In [76], the authors stated that the p -values induced by P^* are exactly uniform. We prove that the latter only can be upper-bounded by a uniform variable. Moreover we correct the proof of the third assertion.

Proof. (Proposition A.1)

1. Let $i \in \{1, \dots, m_0\} := I_0$. We have,

$$\begin{aligned}
\mathbb{P}(p_i \leq x) &= \sum_{n=1}^{k_1} \mathbb{P}(p_i \leq x | N = n) \mathbb{P}(N = n) + \mathbb{P}(p_i \leq x | N = m + 1) \mathbb{P}(N = m + 1) \\
&= \sum_{n=1}^{k_1} \mathbb{P}(p_i \leq x | N = n) \mathbb{P}(N = n) + \mathbb{P}(U_{m+1} \leq x) \mathbb{P}(N = m + 1) \quad (\text{A.7})
\end{aligned}$$

APPENDIX A. SUPPLEMENTARY RESULTS

Let $J := \{1, \dots, k_1\}$. Let us begin by the case $k_1 \leq m_0$. We have:

$$\begin{aligned} \mathbb{P}(p_i \leq x \mid N = n) &= \mathbb{P}(p_i \leq x \mid i \in J) \mathbb{P}(i \in J \mid N = n) \\ &\quad + \mathbb{P}(p_i \leq x \mid i \notin J) \mathbb{P}(i \notin J \mid N = n) \\ &= \mathbb{P}(U_n \leq x) \frac{k_1}{m_0} + \mathbb{P}(U_{m+1} \leq x) \left(1 - \frac{k_1}{m_0}\right). \end{aligned}$$

Thus for all $x \in]0, 1[$, (A.7) can be expressed as

$$\mathbb{P}(p_i \leq x) = \sum_{n=1}^{k_1} \left[\mathbb{P}(U_n \leq x) \frac{k_1}{m_0} + \mathbb{P}(U_{m+1} \leq x) \left(1 - \frac{k_1}{m_0}\right) \right] \mathbb{P}(N = n) + \mathbb{P}(U_{m+1} \leq x) \mathbb{P}(N = m+1). \quad (\text{A.8})$$

If $x \in [\tau_{\tilde{n}-1}, \tau_{\tilde{n}}] \subseteq [0, \tau_m[$, (A.8) is equal to

$$\begin{aligned} \sum_{n=1}^{k_1} \mathbb{P}(U_n \leq x) \frac{k_1}{m_0} \mathbb{P}(N = n) &= \sum_{n=1}^{\tilde{n}-1} \mathbb{P}(U_n \leq x) \frac{k_1}{m_0} \mathbb{P}(N = n) + \mathbb{P}(U_{\tilde{n}} \leq x) \frac{k_1}{m_0} \mathbb{P}(N = \tilde{n}) \\ &= \sum_{n=1}^{\tilde{n}-1} (\tau_n - \tau_{n-1}) + (\tau_{\tilde{n}} - \tau_{\tilde{n}-1}) \frac{x - \tau_{\tilde{n}-1}}{\tau_{\tilde{n}} - \tau_{\tilde{n}-1}} \\ &= x. \end{aligned}$$

If $x \in [\tau_{\tilde{n}-1}, \tau_{\tilde{n}}] \subseteq [\tau_m, 1[$, (A.8) is equal to

$$\begin{aligned} &\sum_{n=1}^{k_1} \mathbb{P}(U_n \leq x) \frac{k_1}{m_0} \mathbb{P}(N = n) + \mathbb{P}(U_{m+1} \leq x) \sum_{n=1}^{k_1} \mathbb{P}(N = n) - \frac{k_1}{m_0} \mathbb{P}(U_{m+1} \leq x) \sum_{n=1}^{k_1} \mathbb{P}(N = n) \\ &\quad + \mathbb{P}(U_{m+1} \leq x) \mathbb{P}(N = m+1) \\ &= \sum_{n=1}^{k_1} (\tau_n - \tau_{n-1}) + \frac{m_0}{k_1} \mathbb{P}(U_{m+1} \leq x) \sum_{n=1}^{k_1} (\tau_n - \tau_{n-1}) - \frac{k_1}{m_0} \mathbb{P}(U_{m+1} \leq x) \sum_{n=1}^{k_1} \frac{m_0}{k_1} (\tau_n - \tau_{n-1}) \\ &\quad + \mathbb{P}(U_{m+1} \leq x) - \frac{m_0}{k_1} \mathbb{P}(U_{m+1} \leq x) \sum_{n=1}^{k_1} (\tau_n - \tau_{n-1}) \\ &= \tau_{k_1} - \tau_{k_1} \mathbb{P}(U_{m+1} \leq x) + \mathbb{P}(U_{m+1} \leq x) \\ &= \tau_{k_1} + (1 - \tau_{k_1}) \frac{x - \tau_m}{1 - \tau_m} \\ &\leq x. \end{aligned}$$

Indeed, for all $x \in [\tau_m, 1[$, since the sequence $(\tau_i)_{i \in I}$ is nondecreasing, we have

$$\begin{aligned} \underbrace{(\tau_m - \tau_{k_1})}_{\geq 0} \underbrace{(x - 1)}_{\leq 0} \leq 0 &\Leftrightarrow \frac{\tau_{k_1} - \tau_{k_1} \tau_m + x - \tau_m - x \tau_{k_1} + \tau_{k_1} \tau_m - x + x \tau_m}{1 - \tau_m} \leq 0 \\ &\Leftrightarrow \tau_{k_1} + \frac{x - \tau_m}{1 - \tau_m} - \tau_{k_1} \frac{x - \tau_m}{1 - \tau_m} - x \leq 0 \\ &\Leftrightarrow \tau_{k_1} + (1 - \tau_{k_1}) \frac{x - \tau_m}{1 - \tau_m} \leq x. \end{aligned}$$

We emphasize that the uniformity of the p -values induced by P^* holds if and only if $k_1 = m$ but it is not the case in general.

Next we move to the case $k_1 > m_0$. We have:

$$\begin{aligned} \mathbb{P}(p_i \leq x \mid N = n) &= \mathbb{P}(p_i \leq x \mid i \in J) \mathbb{P}(i \in J \mid N = n) \\ &\quad + \mathbb{P}(p_i \leq x \mid i \notin J) \mathbb{P}(i \notin J \mid N = n) \\ &= \mathbb{P}(p_i \leq x \mid i \in \{1, \dots, m_0\}) \mathbb{P}(i \in \{1, \dots, m_0\} \mid N = n) \\ &= \mathbb{P}(U_n \leq x). \end{aligned}$$

Thus for all $x \in]0, 1[$, (A.7) can be expressed as

$$\mathbb{P}(p_i \leq x) = \sum_{n=1}^{k_1} \mathbb{P}(U_n \leq x) \mathbb{P}(N = n) + \mathbb{P}(U_{m+1} \leq x) \mathbb{P}(N = m + 1). \quad (\text{A.9})$$

If $x \in [\tau_{\bar{n}-1}, \tau_{\bar{n}}] \subseteq [0, \tau_m[$, (A.9) is equal to

$$\begin{aligned} \sum_{n=1}^{k_1} \mathbb{P}(U_n \leq x) \mathbb{P}(N = n) &= \sum_{n=1}^{\bar{n}-1} \mathbb{P}(U_n \leq x) (\tau_n - \tau_{n-1}) + \mathbb{P}(U_{\bar{n}} \leq x) (\tau_{\bar{n}} - \tau_{\bar{n}-1}) \\ &= x. \end{aligned}$$

If $x \in [\tau_{\bar{n}-1}, \tau_{\bar{n}}] \subseteq [\tau_m, 1[$, (A.9) is equal to

$$\begin{aligned} \sum_{n=1}^{k_1} \mathbb{P}(U_n \leq x) \mathbb{P}(N = n) + \left(1 - \sum_{i=1}^{k_1} \pi_i\right) \frac{x - \tau_m}{1 - \tau_m} &= \sum_{n=1}^{k_1} (\tau_n - \tau_{n-1}) + \frac{x - \tau_m}{1 - \tau_m} - \frac{x - \tau_m}{1 - \tau_m} \sum_{i=1}^{k_1} (\tau_i - \tau_{i-1}) \\ &= \tau_{k_1} + (1 - \tau_{k_1}) \frac{x - \tau_m}{1 - \tau_m} \\ &\leq x. \end{aligned}$$

2. As we can see in Figure A.1 and Figure A.2 the p -values induced by P^* are such that

$$\forall i \in I_0, [p_i \leq \alpha_k] \Rightarrow [N \neq m + 1 \text{ et } R = k_1].$$

Thus, $\sum_{k=1}^m \mathbb{P}(R = k \mid p_i \leq \tau_k) = \mathbb{P}(R = k_1 \mid p_i \leq \tau_{k_1}) = 1$.

APPENDIX A. SUPPLEMENTARY RESULTS

3. For all $i \in I_0$ we have,

$$\mathbb{P}(R \leq k_1 \mid p_i > \tau_m) < 1 \quad \text{et} \quad \mathbb{P}(R \leq k_1 \mid p_i \leq \tau_{k_1}) = 1,$$

then $x \mapsto \mathbb{P}(R \leq k_1 \mid p_i = x)$ is a nonincreasing function. Indeed, denoting $f(x) = \mathbb{P}(R \leq k_1 \mid p_i = x)$ for $i \in I_0$, if f were nondecreasing, we would obtain

$$\begin{aligned} 1 = \mathbb{P}(R \leq k_1 \mid p_i \leq \tau_{k_1}) &= \int_0^{\tau_{k_1}} f(x) d\mathcal{L}(p_i \mid p_i \leq \tau_{k_1}) \leq f(\tau_{k_1}) \\ &\leq f(\tau_m) \leq \int_{\tau_m}^1 f(x) d\mathcal{L}(p_i \mid p_i > \tau_m) < 1, \end{aligned}$$

absurd.

□

We are now in position to prove Theorem 3.7.

Proof. (Theorem 3.7)

It results from Theorem A.1 and Proposition A.1 that for all $P \in \mathcal{P}$, for all procedure $SU(\tau)$ such that (τ, \mathbf{p}) satisfies (3.16), $\text{FDR}(\mathcal{R}_{SU(\tau)}, P) \leq m_0(P)\alpha/m$ if $m_0(P)\tau_{k_1}/k_1 \leq m_0(P)\alpha/m$ that is (by definition of k_1) if for all $k \in \{1, \dots, m\}$, $\tau_k \leq \alpha k/m$. Conversely, if $\tau_k \leq \alpha k/m$ for all $k \in \{1, \dots, m\}$ then $\text{FDR}(\mathcal{R}_{SU(\tau)}, P) \leq m_0(P)\tau_{k_1}/k_1 \leq m_0(P)\alpha/m$. □

