

VITESSES DE SÉPARATION PAR FAMILLE OU FAMILY-WISE ERROR RATES (FWSR) POUR LES TESTS MULTIPLES

Magalie Fromont ¹ & Matthieu Lerasle ² & Patricia Reynaud-Bouret ³

¹ IRMAR, Université Rennes 2, Campus de Villejean, 35043 Rennes Cedex,
magalie.fromont@univ-rennes2.fr

² Laboratoire Jean Alexandre Dieudonné, Université Nice Sophia-Antipolis, 06100 Nice,
mlerasle@unice.fr

³ Laboratoire Jean Alexandre Dieudonné, Université Nice Sophia-Antipolis, 06100 Nice,
reynaudb@unice.fr

Résumé. Partant d'un parallèle entre les tests multiples et les tests adaptatifs (au sens du minimax) d'une hypothèse nulle, construits selon un principe général d'agrégation, nous définissons de nouveaux critères d'évaluation liés à l'erreur de seconde espèce pour les tests multiples. La *vitesse de séparation par famille faible* ou *weak Family-Wise Separation Rate* (*wFWSR*) et la *vitesse de séparation par famille* ou *Family-Wise Separation Rate* (*FWSR*) généralisent ainsi la notion de vitesse de séparation uniforme pour les tests d'une hypothèse nulle, constituant le point de départ d'une théorie minimax pour les tests multiples. L'étude d'un problème de test multiple simple dans un cadre de régression gaussienne vient illustrer l'intérêt de cette théorie, qui permet de formaliser la supériorité souvent constatée en pratique de certaines procédures de tests multiples de type step-down, comme les procédures min- p , par rapport aux procédures de type single-step, comme celle de Bonferroni.

Mots-clés. Test multiple, Family-Wise Error Rate, procédure step-down, vitesse de séparation, test minimax, test adaptatif.

Abstract. Starting from a parallel between multiple tests and (minimax) adaptive tests of a single null hypothesis, based on a general aggregation principle, we define new second kind error-related evaluation criteria for multiple tests. The *weak Family-Wise Separation Rate* (*wFWSR*) and the *Family-Wise Separation Rate* (*FWSR*) thus generalize the notion of uniform separation rate for the tests of a single null hypothesis, laying the foundations of a minimax theory for multiple tests. The study of a simple multiple testing problem in a Gaussian regression framework illustrates the interest of this theory, which formalizes the superiority, often observed in practice, of some step-down multiple testing procedures, such as the min- p ones, with respect to single-step multiple testing procedures, such as Bonferroni's one.

Keywords. Multiple testing, Family-Wise Error Rate, step-down procedure, separation rate, minimax test, adaptive test.

1 Introduction

Généralisant le risque de première espèce pour les tests classiques d’une hypothèse nulle, de nombreux critères liés à l’erreur de première espèce ont été introduits pour les tests multiples. Parmi ces critères figurent notamment le traditionnel *Family-Wise Error Rate* (FWER), défini comme la probabilité d’avoir au moins un faux positif (une hypothèse nulle vraie mais rejetée néanmoins), et le très populaire *False Discovery Rate* (FDR), défini comme le taux moyen de faux positifs. Différentes procédures de tests multiples ont été construites avec comme principal objectif de contrôler l’un et/ou l’autre de ces critères, comme la procédure de Bonferroni, celle de Holm, ou les procédures de type min- p , décrites dans l’ouvrage de Dudoit et Van Der Laan (2007), contrôlant le FWER, donc le FDR.

Si les performances de ces procédures en termes d’erreur de deuxième espèce sont souvent étudiées en pratique, très peu de critères de comparaison ou d’optimalité liés à l’erreur de deuxième espèce ont été définis et analysés en théorie. À notre connaissance, seuls les articles de Lehmann, Romano, et Shaffer (2005), et de Romano, Shaikh et Wolf (2011) fournissent des résultats d’optimalité de type *maximin*.

Nous proposons ici de nouveaux critères liés à l’erreur de deuxième espèce permettant d’évaluer les procédures de tests multiples dont le FWER est contrôlé par un niveau α dans $]0, 1[$ fixé. Ces critères viennent généraliser la notion de vitesse de séparation uniforme introduite par Baraud (2002), donnant une version non asymptotique de la théorie minimax pour les tests non paramétriques d’une hypothèse nulle initialement introduite par Ingster (1993). Partant de l’observation que la plupart des tests adaptatifs au sens du minimax sont construits sur un principe général d’agrégation, pouvant être relié aux procédures de Bonferroni, de Holm, ou de type min- p , nous définissons les notions de *vitesse de séparation par famille faible* ou *weak Family-Wise Separation Rate* (wFWSR) et de *vitesse de séparation par famille* ou *Family-Wise Separation Rate* (FWSR), puis celle de *vitesse de séparation par famille minimax* ou *minimax Family-Wise Separation Rate* (mFWSR), conduisant à différents critères d’optimalité au sens du minimax.

Considérant un problème de test multiple simple dans un cadre de régression, nous montrons que dans certains cas, une procédure de type min- p est optimale au sens du minimax, alors que celle de Bonferroni est clairement sous-optimale, ce qui justifie les résultats de nombreuses études pratiques.

2 Tests multiples et tests agrégés

Soit X une variable aléatoire observée, à valeurs dans un espace mesurable $(\mathbb{X}, \mathcal{X})$, dont la loi inconnue P appartient à un ensemble $\mathcal{P}$ de lois de probabilités sur $(\mathbb{X}, \mathcal{X})$. On suppose dans ce qui suit que les lois de probabilité considérées sont continues.

Une hypothèse H est définie comme un sous-ensemble de $\mathcal{P}$: H est dite *vraie sous P* si P appartient à H , et *fausse sous P* sinon. Étant donnée une collection finie d’hypothèses

$\mathcal{H}$, le but est de tester simultanément pour toute hypothèse H de $\mathcal{H}$, H contre $\mathcal{P} \setminus H$, c'est-à-dire " H est vraie sous P " contre " H est fausse sous P ", ou encore " $P \in H$ " contre " $P \notin H$ ". On introduit l'ensemble $\mathcal{T}(P) = \{H \in \mathcal{H}, P \in H\}$ des hypothèses vraies sous P . Une procédure de test multiple est une statistique, dépendant de la variable aléatoire observée X , définie comme une collection d'hypothèses rejetées $\mathcal{R} \subset \mathcal{H}$, dont le but est d'inférer $\mathcal{H} \setminus \mathcal{T}(P)$.

Rappelons que le *weak Family-Wise Error Rate* ($w\text{FWER}$) et le *Family-Wise Error Rate* (FWER) d'une procédure de test multiple $\mathcal{R}$ sont définis comme $w\text{FWER}(\mathcal{R}) = \sup_{P, \mathcal{T}(P)=\mathcal{H}} P(\mathcal{R} \cap \mathcal{T}(P) \neq \emptyset)$ et $\text{FWER}(\mathcal{R}) = \sup_{P \in \mathcal{P}} P(\mathcal{R} \cap \mathcal{T}(P) \neq \emptyset)$.

Se donnant un niveau α dans $]0, 1[$, les procédures de Bonferroni, de Holm ou de type min- p par exemple, sont connues pour avoir un $w\text{FWER}$ et un FWER contrôlés par α . Ces procédures sont construites à partir d'un ensemble de p -valeurs $\{p_H, H \in \mathcal{H}\}$, telles que :

$$\forall P \in H, u \in]0, 1[, P(p_H \leq u) \leq u.$$

Pour tout sous-ensemble $\mathcal{G}$ de $\mathcal{H}$, soit $q_{mp, \mathcal{G}, \alpha}$ définissant une fonction décroissante de $\mathcal{G}$, et telle que :

$$\forall P \in \cap_{H \in \mathcal{G}} H, P(\min_{H \in \mathcal{G}} p_H < q_{mp, \mathcal{G}, \alpha}) \leq \alpha.$$

Une procédure min- p est définie de la façon séquentielle suivante.

$$\left\{ \begin{array}{l} 1. \text{ Initialiser } \mathcal{R}_0 = \emptyset. \\ 2. \text{ Pour tout } n \geq 0, \text{ construire } \mathcal{R}_{n+1} = \mathcal{R}_n \cup \{H \in \mathcal{H} \setminus \mathcal{R}_n, p_H < q_{mp, \mathcal{H} \setminus \mathcal{R}_n, \alpha}\}. \\ 3. \text{ Définir } \mathcal{R} = \lim_{n \rightarrow \infty} \mathcal{R}_n. \end{array} \right. \quad (1)$$

Le choix de $q_{mp, \mathcal{G}, \alpha} = \alpha / \#\mathcal{G}$ conduit à la procédure de Holm, que l'on notera $\mathcal{R}_{Holm}$, dont la première étape correspond à la procédure bien connue de Bonferroni notée $\mathcal{R}_{Bonf}$.

Si la loi de $\min_{H \in \mathcal{G}} p_H$ ne dépend pas de P dans $\cap_{H \in \mathcal{G}} H$, est connue, et si sa fonction de répartition est notée $F_{\mathcal{G}}$, le choix de $q_{mp, \mathcal{G}, \alpha} = F_{\mathcal{G}}^{-1}(\alpha)$ conduit à une procédure que l'on notera $\mathcal{R}_{mp}$, moins conservative que $\mathcal{R}_{Holm}$ dans le sens où elle la contient.

Nous montrons que ces procédures de tests multiples sont en fait liées à des tests d'une hypothèse nulle H_0 telle que $H_0 \subset \cap_{H \in \mathcal{H}} H$, construits selon un principe général d'agrégation décrit ci-dessous, et qui sont, dans certains cas bien précis, adaptatifs au sens du minimax. Pour chaque hypothèse H de $\mathcal{H}$, un test individuel ϕ_H de l'hypothèse nulle H contre l'alternative $\mathcal{P} \setminus H$ est défini comme une statistique, à valeurs dans $\{0, 1\}$, dont la valeur 1 équivaut au rejet de H . Une collection de tests $\Phi_{\mathcal{H}} = \{\phi_H, H \in \mathcal{H}\}$ est ainsi définie et le test agrégé correspondant $\bar{\Phi}_{\mathcal{H}}$ consiste à rejeter H_0 si au moins l'une des hypothèses H de $\mathcal{H}$ est rejetée avec ϕ_H , c'est-à-dire : $\bar{\Phi}_{\mathcal{H}} = \sup_{H \in \mathcal{H}} \phi_H$.

Le risque de première espèce d'un test agrégé $\bar{\Phi}_{\mathcal{H}}$ de l'hypothèse nulle H_0 est défini par :

$$\text{ER}(\bar{\Phi}_{\mathcal{H}}, H_0) = \sup_{P \in H_0} P(\bar{\Phi}_{\mathcal{H}} = 1) = \sup_{P \in H_0} P(\sup_{H \in \mathcal{H}} \phi_H = 1).$$

Étant donné un test multiple $\mathcal{R}$ de $\mathcal{H}$, on introduit $\bar{\Phi}(\mathcal{R}) = \mathbb{1}_{\{\mathcal{R} \neq \emptyset\}}$, qui peut être vu comme un test agrégé de l'hypothèse nulle H_0 . Il est alors évident que :

$$w\text{FWER}(\mathcal{R}) = \text{ER}(\bar{\Phi}(\mathcal{R}), \cap_{H \in \mathcal{H}} H) \quad . \quad (2)$$

L'introduction du test agrégé $\bar{\Phi}(\mathcal{R})$ peut ainsi être considérée comme un moyen d'étendre le critère du risque de première espèce pour un test d'une hypothèse nulle à celui de *weak Family-Wise Error Rate* pour un test multiple. Suivant cette idée, nous pouvons généraliser la notion de vitesse de séparation uniforme pour un test d'une hypothèse nulle à celle de *vitesse de séparation par famille* ou *Family-Wise Separation Rate* (FWSR) pour un test multiple.

3 Vitesses de séparation par famille

Rappelons les définitions des vitesses de séparation uniforme et minimax pour les tests d'une hypothèse nulle H_0 , pour une distance d sur $\mathcal{P}$, introduites par Baraud (2002).

Définition 1. Soit β un niveau fixé dans $]0, 1[$, une classe $\mathcal{Q} \subset \mathcal{P}$, et un test $\bar{\Phi}$ d'une hypothèse nulle $H_0 \subset \mathcal{P}$. La *vitesse de séparation uniforme* de $\bar{\Phi}$ sur $\mathcal{Q}$ pour un risque de deuxième espèce β est définie par :

$$\text{SR}_d^\beta(\bar{\Phi}, \mathcal{Q}, H_0) = \inf \left\{ r > 0, \sup_{P \in \mathcal{Q}, d(P, H_0) \geq r} P(\bar{\Phi} = 0) \leq \beta \right\} \quad .$$

La *vitesse de séparation minimax* sur $\mathcal{Q}$, pour des risques de première et deuxième espèces α et β , est définie par :

$$m\text{SR}_d^{\alpha, \beta}(\mathcal{Q}, H_0) = \inf_{\bar{\Phi}, \text{ER}(\bar{\Phi}, H_0) \leq \alpha} \text{SR}_d^\beta(\bar{\Phi}, \mathcal{Q}, H_0) \quad .$$

Un test $\bar{\Phi}$ de niveau α , c'est-à-dire tel que $\text{ER}(\bar{\Phi}, H_0) \leq \alpha$, est dit *minimax* sur $\mathcal{Q}$ si $\text{SR}_d^\beta(\bar{\Phi}, \mathcal{Q}, H_0)$ est du même ordre que $m\text{SR}_d^{\alpha, \beta}(\mathcal{Q}, H_0)$, à une éventuelle constante multiplicative près. Il est dit *adaptatif au sens du minimax* sur une collection $\mathcal{C}$ de classes $\mathcal{Q}$ s'il est minimax (ou presque) sur toutes les classes $\mathcal{Q}$ de $\mathcal{C}$ simultanément, sans a priori sur la classe à laquelle P peut appartenir.

Considérons maintenant un test multiple $\mathcal{R}$ et le test agrégé correspondant $\bar{\Phi}(\mathcal{R})$. Étant donné un niveau β dans $]0, 1[$ et une classe $\mathcal{Q} \subset \mathcal{P}$, suivant l'idée de (2), une notion de *vitesse de séparation par famille faible* pourrait être définie par $\text{SR}_d^\beta(\bar{\Phi}(\mathcal{R}), \mathcal{Q}, \cap_{H \in \mathcal{H}} H)$. Cette notion ne prendrait alors en compte que les alternatives P de $\mathcal{Q}$ à une certaine distance de l'intersection $\cap_{H \in \mathcal{H}} H$. Or dans un problème de test multiple, chaque hypothèse H de $\mathcal{H}$ a son importance propre et doit être considérée individuellement. Afin de répondre à cette exigence, nous définissons finalement la notion de *vitesse de séparation par famille faible* de la façon suivante.

Définition 2. Soit β un niveau fixé dans $]0, 1[$ et une classe $\mathcal{Q} \subset \mathcal{P}$. La *vitesse de séparation par famille faible* ou *weak Family-Wise Separation Rate* ($w\text{FWSR}$) d'un test multiple $\mathcal{R}$ sur $\mathcal{Q}$, pour un risque de deuxième espèce β , est définie par :

$$w\text{FWSR}_d^\beta(\mathcal{R}, \mathcal{Q}) = \inf \left\{ r > 0, \sup_{P \in \mathcal{Q}, \exists H \in \mathcal{H}, d(P, H) \geq r} P(\mathcal{R} = \emptyset) \leq \beta \right\} .$$

Définissons maintenant la notion de *vitesse de séparation par famille (forte)*, posant les bases d'une approche minimax pour les tests multiples.

Définition 3. Soit β un niveau fixé dans $]0, 1[$ et une classe $\mathcal{Q} \subset \mathcal{P}$. La *vitesse de séparation par famille* ou *Family-Wise Separation Rate* (FWSR) d'un test multiple $\mathcal{R}$ sur $\mathcal{Q}$, pour un risque de deuxième espèce β , est définie par :

$$\text{FWSR}_d^\beta(\mathcal{R}, \mathcal{Q}) = \inf \left\{ r > 0, \sup_{P \in \mathcal{Q}} P(\exists H \in \mathcal{H} \setminus \mathcal{R}, d(P, H) \geq r) \leq \beta \right\} .$$

La *vitesse de séparation par famille minimax* ou *minimax Family-Wise Separation Rate* ($m\text{FWSR}$) sur $\mathcal{Q}$, pour un FWER α et un risque de deuxième espèce β , est définie par :

$$m\text{FWSR}_d^{\alpha, \beta}(\mathcal{Q}) = \inf_{\mathcal{R}, \text{FWER}(\mathcal{R}) \leq \alpha} \text{FWSR}_d^\beta(\mathcal{R}, \mathcal{Q}) .$$

Un test multiple $\mathcal{R}$, de FWER contrôlé par α , est dit *minimax* sur $\mathcal{Q}$ si $\text{FWSR}_d^\beta(\mathcal{R}, \mathcal{Q})$ est du même ordre que $m\text{FWSR}_d^{\alpha, \beta}(\mathcal{Q})$, à une éventuelle constante multiplicative près. Il est dit *adaptatif au sens du minimax* sur une collection $\mathcal{C}$ de classes $\mathcal{Q}$ s'il est minimax (ou presque) sur toutes les classes $\mathcal{Q}$ de $\mathcal{C}$ simultanément, sans a priori sur la classe à laquelle P peut appartenir.

Remarquons que si $\mathcal{H}$ est réduite à une seule hypothèse nulle H_0 , $m\text{FWSR}_d^{\alpha, \beta}(\mathcal{Q}) = m\text{SR}_d^{\alpha, \beta}(\mathcal{Q}, H_0)$. L'approche minimax définie ici pour les tests multiples peut donc être vue comme une généralisation de l'approche minimax classique pour les tests d'une hypothèse nulle. Le résultat suivant montre d'autres liens entre les deux approches.

Théorème 1. Soit $\mathcal{Q}$ un sous-ensemble de $\mathcal{P}$. S'il existe une distance d' telle que

$$\forall P \in \mathcal{Q}, \forall r > 0, \quad \exists H \in \mathcal{H}, d(P, H) \geq r \Leftrightarrow d'(P, \cap_{H \in \mathcal{H}} H) \geq r ,$$

alors pour tout β dans $]0, 1[$, $m\text{FWSR}_d^{\alpha, \beta}(\mathcal{Q}) \geq m\text{SR}_{d'}^{\alpha, \beta}(\mathcal{Q}, \cap_{H \in \mathcal{H}} H)$.

4 Étude d'un problème de tests multiples en régression

Soit $n \geq 1$, et τ une partition de $\{1, \dots, n\}$. Soit $X = (X_1, \dots, X_n)'$ de loi $P = P_{f, \tau}$, définie par : $X_i = f_i + \sigma \varepsilon_t$, pour tout t de τ et tout i de t , $(\varepsilon_t)_{t \in \tau}$ étant une collection de variables aléatoires i.i.d. gaussiennes standard, $f = (f_1, \dots, f_n)' \in \mathbb{R}^n$ étant inconnu, $\sigma > 0$ connu. Nous considérons la distance $d_\infty(P_{f, \tau}, P_{g, \tau}) = \|f - g\|_\infty = \max_{i=1, \dots, n} |f_i - g_i|$,

ainsi que la collection d'hypothèses $\mathcal{H} = \{H_{S_i}, i = 1, \dots, n\}$, où $H_{S_i} = \{P_{f,\tau}, f_i = 0\}$. Soit pour T , $k \in \{1, \dots, n\}$, $\mathcal{P}_{k,T} = \{P_{f,\tau}, f \in \mathbb{R}^n, |f|_0 \leq k \text{ et } \#\tau = T\}$, où $|f|_0$ désigne le nombre de coefficients non nuls de f .

En utilisant les résultats de Baraud (2002) et le Théorème 1, nous montrons la borne inférieure suivante.

Théorème 2. *Pour $\alpha, \beta \in]0, 1[$ tels que $\alpha + \beta \leq 1/2$,*

$$m\text{FWSR}_{d_\infty}^{\alpha, \beta}(\mathcal{P}_{k,T}) \geq \sigma \sqrt{\ln T} .$$

Soit pour tout $i = 1 \dots n$, la p -valeur p_i correspondant au test $\mathbb{1}_{\{|X_i|\sigma^{-1} > F^{-1}(1-\alpha/2)\}}$, où F est la fonction de répartition de la loi gaussienne standard.

Théorème 3. *Soit $\alpha, \beta \in]0, 1[$. Soit $\mathcal{R}_{Bonf}$ et $\mathcal{R}_{mp}$ les procédures de Bonferroni et min- p basées sur les p -valeurs p_i définies ci-dessus, dont le FWER est contrôlé par α . Alors,*

$$\text{FWSR}_{d_\infty}^\beta(\mathcal{R}_{mp}, \mathcal{P}_{k,T}) \leq \sigma \left(\sqrt{2 \ln((k \wedge T)/\beta)} + \sqrt{2 \ln(T/\alpha)} \right),$$

et si $\alpha \leq n/5$,

$$\text{FWSR}_{d_\infty}^\beta(\mathcal{R}_{Bonf}, \mathcal{P}_{k,T}) \geq \sigma \left(\sqrt{2 \ln(2n/\alpha)} - \frac{\ln(4 \ln(2n/\alpha)) + 2}{2\sqrt{2 \ln(2n/\alpha)}} - \sqrt{(-2/k) \ln(1-\beta)} \right).$$

On peut déduire de ces résultats que $m\text{FWSR}_{d_\infty}^{\alpha, \beta}(\mathcal{P}_{k,T})$ est de l'ordre de $\sigma \sqrt{\ln T}$, et que la procédure min- p est adaptative au sens du minimax sur toutes les classes $\mathcal{P}_{k,T}$ considérées. En revanche, lorsque $\ln T \ll \ln n$, c'est-à-dire dans un cadre de dépendance des bruits extrême, la procédure de Bonferroni ne l'est pas.

Bibliographie

- [1] Baraud, Y. (2002), Non-asymptotic minimax rates of testing in signal detection, *Bernoulli*, 5, 8, 577–606.
- [2] Dudoit, S., Van Der Laan, M. J. (2007), *Multiple testing procedures with applications to genomics*.
- [3] Ingster, Yu. I. (1993), Asymptotically minimax testing for nonparametric alternatives I-II-III, *Mathematical Methods of Statistics*, 2, 85–114, 171–189, 249–268.
- [4] Lehmann, E. L., Romano, J. P., Shaffer, J. P. (2005), On Optimality of Stepdown and Stepup Multiple Test Procedures, *The Annals of Statistics*, 3, 33, 1084–1108.
- [5] Romano, J. P., Shaikh, A., Wolf, M. (2011), Consonance and the closure method in multiple testing, *The International Journal of Biostatistics*, 1, 7, 1–25.