

Powerful multiple testing procedures derived from hyperrectangular confidence regions having a minimal volume

Patrick J C Tardivel, Rémi Servien, Didier Concordet

▶ **To cite this version:**

Patrick J C Tardivel, Rémi Servien, Didier Concordet. Powerful multiple testing procedures derived from hyperrectangular confidence regions having a minimal volume. Journal de la Societe Française de Statistique, 2021, 162 (1), pp.2-21. hal-01904873v3

HAL Id: hal-01904873
https://hal.science/hal-01904873v3

Submitted on 6 Jul 2021

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.

Journal de la Société Française de Statistique
Vol. 162 No. 1 (2021)

Powerful multiple testing procedures derived from hyperrectangular confidence regions having a minimal volume

Titre: Procédures de tests multiples puissantes dérivées de régions de confiances hyperrectangulaires de volume minimal

Patrick J.C. Tardivel [1], Rémi Servien [2] and Didier Concordet [2]

Abstract: We study the control of the FamilyWise Error Rate (FWER) in the linear Gaussian model when the $n \times p$ design matrix is of rank p. Single step multiple testing procedures controlling the FWER are derived from hyperrectangular confidence regions. In this study, we aim to construct procedure derived from hyperrectangular confidence region having a minimal volume. We show that minimizing the volume seems a fair criterion to improve the power of the multiple testing procedure. Numerical experiments demonstrate the performance of our approach when compared with the state-of-the-art single step and sequential procedures. We also provide an application to the detection of metabolites in metabolomics.

Résumé : Nous étudions le contrôle de la probabilité d'obtenir un ou plusieurs faux-positifs dans le cadre du modèle linéaire gaussien lorsque la matrice de planification $n \times p$ est de rang p. Les procédures de tests multiples non-séquentielles contrôlant cette probabilité sont dérivées de régions de confiance hyperrectangulaires. Dans cet article, nous construisons une procédure basée sur une région de confiance hyperrectangulaires de volume minimal. Nous montrons que la minimisation du volume est un critère judicieux pour augmenter la puissance d'une procédure de tests multiples. Des expériences numériques montrent que notre démarche fournit une procédure plus performante que les procédures séquentielles et non-séquentielles de l'état de l'art. Enfin, nous appliquons cette procédure à la détection de métabolites en métabolomique.

Keywords: family wise error rate, multiple testing procedure, confidence region, linear model
Mots-clés : probabilité d'avoir un ou plusieurs faux-positifs, procédure de tests multiples, region de confiance, modèle linéaire
AMS 2000 subject classifications: 62H15

1. Introduction

Let us consider the linear Gaussian model

$$Y = X\beta + \varepsilon, \tag{1}$$

where X is an $n \times p$ design matrix of rank p with $p < n$, $\varepsilon \sim \mathcal{N}(\mathbf{0}, \sigma^2 Id_n)$, and $\beta \in \mathbb{R}^p$ is an unknown parameter. We aim to test the hypotheses $\mathscr{H}_i : \beta_i = 0$, with $1 \leq i \leq p$. Several type I errors can be controlled in such multiple testing procedures. In this study, we focus on the

[1] Institute of Mathematics, Wrocław University, Wrocław, Poland
E-mail: tardivel@math.uni.wroc.pl
[2] INTHERES, Université de Toulouse, INRAE, ENVT, Toulouse, France
E-mails: remi.servien@inrae.fr, d.concordet@envt.fr

© Société Française de Statistique et Société Mathématique de France (2021) ISSN: 2102-6238

Familywise Error Rate (FWER), defined as the probability of wrongly rejecting at least one hypothesis $\mathscr{H}_i$. Let $\hat{\beta}^{\text{mle}} := (X'X)^{-1}X'Y$ be the maximum likelihood estimator of the model (1). The usual multiple testing procedures are based on the maximum likelihood estimator and reject $\mathscr{H}_i : \beta_i = 0$ when $|\hat{\beta}_i^{\text{mle}}|/\text{se}(\hat{\beta}_i^{\text{mle}}) > s$, where $s \geq 0$ is the same threshold as that for the hypotheses $\mathscr{H}_1, \ldots, \mathscr{H}_p$. Let ζ be a random vector having the same distribution as

$$\left(\hat{\beta}_1^{\text{mle}}/\text{se}(\hat{\beta}_1^{\text{mle}}), \ldots, \hat{\beta}_p^{\text{mle}}/\text{se}(\hat{\beta}_p^{\text{mle}})\right)$$

when $\beta = \mathbf{0}$. We observe that ζ is a Gaussian vector or a multivariate student, depending on whether the standard errors $\text{se}(\hat{\beta}_1^{\text{mle}}), \ldots, \text{se}(\hat{\beta}_p^{\text{mle}})$ are known or estimated (thus, σ is known or estimated). There are several ways to choose such a threshold s, assuring control of the FWER at a significance level $\alpha \in (0,1)$. For example, s can be chosen according to correlation-free inequalities such as the Bonferroni inequality (Dunn, 1959) or the Gaussian correlation inequality [1] (Royen, 2014; Šidák, 1967)

$$\mathbb{P}\left(\bigcup_{1\leq i\leq p}\{|\zeta_i| \geq s\}\right) \leq \sum_{i=1}^{p}\mathbb{P}(|\zeta_i| \geq s) \text{ and } \mathbb{P}(|\zeta_1| \leq s, \ldots, |\zeta_p| \leq s) \geq \prod_{i=1}^{p}\mathbb{P}(|\zeta_i| \leq s).$$

The inequality given on the left is available for Gaussian vector and multivariate student; the one on the right is available for Gaussian vector and for some particular cases of multivariate student (Sidak et al., 1971). These inequalities are extremely convenient and provide a threshold for controlling the FWER at the significance level α. The first and second inequalities, respectively, provide the thresholds $s_{\text{bonf}} := q_{1-\alpha/2p}$ and $s_{\text{sidak}} := q_{(1+\sqrt[p]{1-\alpha})/2}$, where q_η denotes the η quantile of a ζ_1. By taking into account the correlation, a smaller threshold $s_{\max}$ (thus, a better power) is given by setting $s_{\max}$ as the $1-\alpha$ quantile of $\max\{|\zeta_1|, \ldots, |\zeta_p|\}$.

Confidence regions and testing procedures are closely related (see *e.g.* (Lehmann and Romano, 2005) page 72). Historically, the famous Bonferroni and Dunn-Šidák corrections for multiple testing procedures (Dunn, 1959; Šidák, 1967) originated from the construction of hyperrectangular confidence regions (also called simultaneous confidence intervals). Actually, taking $s \in \{s_{\text{bonf}}, s_{\text{sidak}}, s_{\max}\}$ gives the hyperrectangular confidence region $[\hat{\beta}_1^{\text{mle}} \pm s \times \text{se}(\hat{\beta}_1^{\text{mle}})] \times \cdots \times [\hat{\beta}_p^{\text{mle}} \pm s \times \text{se}(\hat{\beta}_p^{\text{mle}})]$ which contains β with a probability larger than $1-\alpha$. Conversely, given a hyperrectangular confidence region $[\hat{\beta}_1^{\text{mle}} \pm s_1 \times \text{se}(\hat{\beta}_1^{\text{mle}})] \times \cdots \times [\hat{\beta}_p^{\text{mle}} \pm s_p \times \text{se}(\hat{\beta}_p^{\text{mle}})]$ containing β with a probability larger than $1-\alpha$ one may derive a multiple testing procedure which controls the FWER at significance level α. Specifically, for any $i \in \{1, \ldots, p\}$, this procedure rejects $\mathscr{H}_i : \beta_i = 0$ if $0 \notin [\hat{\beta}_i^{\text{mle}} \pm s_i \times \text{se}(\hat{\beta}_i^{\text{mle}})]$. Of course, when $p = 1$, the classical confidence interval $[\hat{\beta}_1^{\text{mle}} \pm s_{\max} \times \text{se}(\hat{\beta}_1^{\text{mle}})] = [\hat{\beta}_1^{\text{mle}} \pm q_{1-\alpha/2} \times \text{se}(\hat{\beta}_1^{\text{mle}})]$ has a minimal length. Otherwise, when $p \geq 2$, we claim that none of the thresholds $s_{\text{bonf}}, s_{\text{sidak}}$ and $s_{\max}$ provides a hyperrectangular confidence region with the smallest volume. It is natural to attempt minimizing the volume for a confidence region. To our knowledge, such a minimization has never been studied in the multivariate case (in the univariate case, the minimization of the expected length of a confidence

[1] The inequality $\mathbb{P}(|\zeta_1| \leq s_1, \ldots, |\zeta_p| \leq s_p) \geq \prod_{i=1}^{p}\mathbb{P}(|\zeta_i| \leq s_i)$ already proved in Šidák (1967) is a particular case of the Gaussian correlation inequality $\mathbb{P}(|\zeta_1| \leq s_1, \ldots, |\zeta_p| \leq s_p) \geq \mathbb{P}(|\zeta_i| \leq s_1, \ldots, |\zeta_i| \leq s_k)\mathbb{P}(|\zeta_i| \leq s_{k+1}, \ldots, |\zeta_i| \leq s_p)$. Recently, the Gaussian correlation inequality was proved by Royen (2014).

© Société Française de Statistique et Société Mathématique de France (2021) ISSN: 2102-6238

interval has been studied by Pratt (1961)). Another theoretical justification for the volume minimization is the following:

Let $A \subset \mathbb{R}^p$ be an acceptance region defined as $\mathbb{P}_{\beta=0}(\hat{\beta}^{\text{mle}} \in A) = 1-\alpha$. We consider the procedure rejecting the null hypothesis $\beta = \mathbf{0}$ against the alternative $\beta \neq \mathbf{0}$ when $\hat{\beta}^{\text{mle}} \notin A$. The power (and type II error) of this procedure depends on β in the alternative. We would like to control some kind of "average" power that would not depend on a specific value of β. To this end, we define an "average" type II error by integrating the type II error $\mathbb{P}_\beta(\hat{\beta}^{\text{mle}} \in A)$ over all the possible values of $\beta \neq \mathbf{0}$. Let f be the density of $\hat{\beta}^{\text{mle}}$ when $\beta = \mathbf{0}$, by proceeding so and by using Fubini-Tonelli theorem, we obtained the following identity

$$\begin{aligned}\int_{\mathbb{R}^p\setminus\{\mathbf{0}\}} \mathbb{P}_\beta(\hat{\beta}^{\text{mle}} \in A)d\beta &= \int_{\mathbb{R}^p\setminus\{\mathbf{0}\}} \left(\int_{\mathbb{R}^p} f(x-\beta)\mathbf{1}(x\in A)dx\right) d\beta \\ &= \int_{\mathbb{R}^p} \left(\int_{\mathbb{R}^p\setminus\{\mathbf{0}\}} f(x-\beta)d\beta\right) \mathbf{1}(x\in A)dx \\ &= \int_{\mathbb{R}^p} \mathbf{1}(x\in A)dx = \text{vol}(A).\end{aligned}$$

Therefore, a procedure having an acceptance region with a small volume has globally a small type II error. Consequently, comparatively to classical single step multiple testing procedures (which are also derived from hyperrectangular confidence regions), our procedure derived from a hyperrectangular confidence region having a minimal volume is globally better to detect that the parameter β is not null when actually $\beta \neq \mathbf{0}$.

Theoretical results showing that a multiple testing procedure has an optimal power are quite rare (Fromont et al., 2016; Lehmann et al., 2012; Romano et al., 2011). For example, in the particular case where the covariance of $\hat{\beta}^{\text{mle}}$ is a scalar matrix then, the procedure described in section 4 of Romano et al. (2011) has a maximal power for a specific class of alternatives. However, in the general setting, we are not aware of the existence of an optimal testing procedure.

We illustrate that deriving a multiple testing procedure from a hyperrectangular confidence region having a minimal volume is an intuitive way to improve power. In addition, we present a new operational procedure through a numerical method for volume minimization. This article is organized as follows.

Section 2 contains some basic properties about the optimal hyperrectangular confidence region. We exhibit some cases in which it is convenient to perform the computation of the optimal hyperrectangular confidence region.

Section 3 presents a method to numerically minimize the volume of the hyperrectangular confidence region.

Section 4 is devoted to simulation experiments: we compare our multiple testing procedure with the state-of-the-art single step and sequential procedures.

Section 5 details the analysis of metabolomic data, which motivated this study.

We use the following notations:

— The zero in bold $\mathbf{0}$ represents the null vector of $\mathbb{R}^n$ for some $n \geq 2$, the transpose matrix of A is denoted by A' and Id_p represents the $p \times p$ identity matrix.

— The sets $\mathscr{A}_0$ and $\mathscr{A}_1$ are, respectively, $\mathscr{A}_0 := \{i \in \{1,\dots,p\} \mid \beta_i = 0\}$ and $\mathscr{A}_1 := \{i \in \{1,\dots,p\} \mid \beta_i \neq 0\}$.

© Société Française de Statistique et Société Mathématique de France (2021) ISSN: 2102-6238

— The matrix $M_p(a,b)$ is a $p \times p$ matrix for which the diagonal elements are a and the non-diagonal elements are b.
— Given a random vector $V := (V_1, \ldots, V_p)$, $\mathrm{var}(V)$ denotes the covariance matrix of V and $\mathrm{var}(V_i)$ denotes the marginal variance of V_i.
— The matrix Σ is a $p \times p$ positive definite matrix and C is a $p \times p$ 'correlation' matrix, namely, C is a positive definite matrix, such that $C_{11} = \cdots = C_{pp} = 1$.
— The covariance matrix of $\hat{\beta}^{\mathrm{mle}}$ is $\sigma^2(X'X)^{-1}$ thus, for $i \in \{1, \ldots, p\}$, the standard error of $\hat{\beta}_i^{\mathrm{mle}}$ is equal to $\sigma\sqrt{[(X'X)^{-1}]_{ii}}$. When σ is not known an estimator of the standard error is $\hat{\sigma}\sqrt{[(X'X)^{-1}]_{ii}}$ where $\hat{\sigma} := \left(\|Y - X\hat{\beta}^{\mathrm{mle}}\|^2/(n-p)\right)^{1/2}$. For $i \in \{1, \ldots, p\}$, let us define $\mathrm{se}(\hat{\beta}_i^{\mathrm{mle}})$ as follows

$$\mathrm{se}(\hat{\beta}_i^{\mathrm{mle}}) := \begin{cases} \sigma\sqrt{[(X'X)^{-1}]_{ii}} \text{ when } \sigma \text{ is known} \\ \hat{\sigma}\sqrt{[(X'X)^{-1}]_{ii}} \text{ when } \sigma \text{ is estimated} \end{cases}.$$

— The random vector ζ has the same distribution as $\left(\hat{\beta}_1^{\mathrm{mle}}/\mathrm{se}(\hat{\beta}_1^{\mathrm{mle}}), \ldots, \hat{\beta}_p^{\mathrm{mle}}/\mathrm{se}(\hat{\beta}_p^{\mathrm{mle}})\right)$ when $\beta = \mathbf{0}$. Consequently, depending on whether σ is known or estimated, $\zeta \sim \mathcal{N}(\mathbf{0}, C)$ or $\zeta \sim t_{n-p}(\mathbf{0}, C)$ where C is a 'correlation' matrix.

2. Minimization of volume

We aim to construct a multiple testing procedure derived from a hyperrectangular confidence region for β having the following expression: $[\hat{\beta}_1^{\mathrm{mle}} \pm s_1 \times \mathrm{se}(\hat{\beta}_1^{\mathrm{mle}})] \times \cdots \times [\hat{\beta}_p^{\mathrm{mle}} \pm s_p \times \mathrm{se}(\hat{\beta}_p^{\mathrm{mle}})]$. To guarantee a significance level of $1-\alpha$ (with $\alpha \in (0,1)$) the thresholds $s_1, \ldots, s_p$ must satisfy the following equality:

$$\mathbb{P}(\beta \in [\hat{\beta}_1^{\mathrm{mle}} \pm s_1 \times \mathrm{se}(\hat{\beta}_1^{\mathrm{mle}})] \times \cdots \times [\hat{\beta}_p^{\mathrm{mle}} \pm s_p \times \mathrm{se}(\hat{\beta}_p^{\mathrm{mle}})]) = 1-\alpha$$
$$\Leftrightarrow \mathbb{P}(|\zeta_1| \leq s_1, \ldots, |\zeta_p| \leq s_p) = 1-\alpha.$$

Among these hyperrectangular confidence containing β with a probability $1-\alpha$, we aim to pick one for which the expected value of the volume [2] $2^p s_1 \ldots s_p \mathbb{E}(\mathrm{se}(\hat{\beta}_1^{\mathrm{mle}}) \ldots \mathrm{se}(\hat{\beta}_p^{\mathrm{mle}}))$ is minimal, which leads to the following optimisation problem:

$$\text{minimize} \prod_{i=1}^{p} s_i \text{ subject to } \mathbb{P}(|\zeta_1| \leq s_1, \ldots, |\zeta_p| \leq s_p) = 1-\alpha. \tag{2}$$

Note that, when $p = 1$, the optimal threshold is $s_1^* = q_{1-\alpha/2}$, the $1-\alpha/2$ quantile of ζ_1, which yields to the confidence interval $[\hat{\beta}_1^{\mathrm{mle}} \pm q_{1-\alpha/2} \times \mathrm{se}(\hat{\beta}_1^{\mathrm{mle}})]$. This optimization problem has at least one minimizer as proved in Proposition 3 given in appendix. We do not need the uniqueness of the minimizer of (2), but only pick a particular optimal threshold s^*. In the Gaussian framework, given a minimizer s^*, the following proposition holds:

[2] Let us notice that the volume is not random when σ is known.

© Société Française de Statistique et Société Mathématique de France (2021) ISSN: 2102-6238

Proposition 1. *Let us consider the case where ζ has a Gaussian distribution. Let C be the invertible correlation matrix of ζ, $s^* = (s_1^*, \ldots, s_p^*)$ be a solution of the optimisation problem (2) and T^{s^*} denote the truncated Gaussian vector on $S^* = [-s_1^*, s_1^*] \times \cdots \times [-s_p^*, s_p^*]$ having the following density:*

$$f_{T^{s^*}}(u) = \frac{1}{(1-\alpha)\sqrt{(2\pi)^p \det(C)}} \exp\left(-\frac{1}{2} u' C^{-1} u\right) \mathbf{1}(u \in S^*)$$

then all the diagonal coefficients of $C^{-1}\mathrm{var}(T^{s^*})$ are equal.

Assuming that the covariance matrix of T^{s^*} (here denoted by $\mathrm{var}(T^{s^*})$) was equal to C, all the diagonal coefficients of $C^{-1}\mathrm{var}(T^{s^*})$ would be equal, indicating that s^* is a solution of (2). Because the diagonal terms of $\mathrm{var}(T^{s^*})$ are always smaller than the diagonal terms of C, $\mathrm{var}(T^{s^*})$ cannot be equal to C. However, the condition given by Proposition 1 can be intuitively interpreted. The optimal (with respect to the volume) hyperrectangular should be such that the covariance of the truncated Gaussian vector ζ restrained to $[-s_1^*, s_1^*] \times \cdots \times [-s_p^*, s_p^*]$ is as close as possible to the non-constraint covariance of the random vector ζ. The Gaussian framework has some simple yet interesting cases where the computation of the optimal thresholds $s_1^*, \ldots, s_p^*$ can be performed by hand. Note that in the special case $p = 2$ (i.e. where $\hat{\beta}^{\mathrm{mle}}$ has only two components), basic algebra shows that $s_1^* = s_2^*$. This property does not hold true when $p > 2$.

According to Proposition 1, diagonal coefficients of $C^{-1}\mathrm{var}(T^{s^*})$ are equal is a necessary condition for s^* to be a minimizer of (2). We aim to illustrate this condition on the following three examples. For convenience, we denote $M_p(a,b)$, a $p \times p$ matrix whose diagonal coefficients are equal to a and whose non-diagonal coefficients are equal to b.

1) **In the independent case:** Let us set $C = Id_p$ and $s \in \mathbb{R}^p$ where $s_1 = \cdots = s_p > 0$ then diagonal coefficients of $C^{-1}\mathrm{var}(T^s)$ are equal. This equality suggests (but does not prove) that components of s^*, minimizer of (2), are all equal.

2) **In the equicorrelated case:** Let us set $C = M_p(1,\rho)$ and $s \in \mathbb{R}^p$ where $s_1 = \cdots = s_p > 0$. It follows that $C^{-1} = M_p(a,b)$ for some a and b and $\mathrm{var}(T^s) = M(c,d)$ for some c and d. Consequently all the diagonal coefficients of $C^{-1}\mathrm{var}(T^s) = M(a,b)M(c,d)$ are equal. Again, this equality suggests that components of s^* are all equal.

3) **In the block diagonal equicorrelated case:** Let us set $C = \mathrm{diag}(M_k(1,\rho), M_{p-k}(1,\rho'))$. It follows that C^{-1} is block diagonal with $C^{-1} = \mathrm{diag}(M_k(a,b), M_{p-k}(a',b'))$ for some a, b, a', b'. Let $s \in \mathbb{R}^p$ where $s_1 = \cdots = s_k = c_1$ and $s_{k+1} = \cdots = s_p = c_2$, one deduces that $\mathrm{var}(T^s)$ is block diagonal with $\mathrm{var}(T^s) = \mathrm{diag}(M_k(c,d), M_{p-k}(c',d'))$ for some c, d, c', d'. Consequently, whatever c_1 and c_2, the k first diagonal coefficients of $C^{-1}\mathrm{var}(T^s)$ are equal and the $p-k$ last diagonal coefficients of $C^{-1}\mathrm{var}(T^s)$ are equal. We only need to tune c_1, c_2 such that all diagonal coefficients of $C^{-1}\mathrm{var}(T^s)$ become equal. This equality suggests that $s_1^* = \cdots = s_k^*$ and $s_{k+1}^* = \cdots = s_p^*$.

According to Proposition 2, when correlation coefficients are non-negative, it is actually true that in settings 1) and 2) components of s^* are all equals and in setting 3) that $s_1^* = \cdots = s_k^*$ and $s_{k+1}^* = \cdots = s_p^*$.

Proposition 2. *Let us consider the case where ζ has a Gaussian distribution. Let C, the correlation matrix of ζ, be the block diagonal matrix $C = \mathrm{diag}(M_{k_1}(1,\rho_1), M_{k_2}(1,\rho_2), \ldots, M_{k_l}(1,\rho_l))$*

© Société Française de Statistique et Société Mathématique de France (2021) ISSN: 2102-6238

where $k_1+\cdots+k_l=p$ *and* $\rho_1,\dots,\rho_l\in[0,1)$ *and let* $\alpha\in[0,1)$. *The optimisation problem (2) has a minimizer* s^* *which satisfies*

$$s^*=(\underbrace{c_1,\dots,c_1}_{k_1\text{ components}},\dots,\underbrace{c_l,\dots,c_l}_{k_l\text{ components}}).$$

Except for the particular cases mentioned above, we do not have a closed form for the optimal thresholds. Therefore, we develop a numerical method to compute these optimal thresholds efficiently.

3. Numerical solver for the optimal thresholds

The optimal thresholds are provided by the solution of the following problem (equivalent to (2))

$$\min g(s)=\sum_{i=1}^{p}\ln(s_i)\text{ subject to }F(s)=\mathbb{P}(|\zeta_1|\le s_1,\dots,|\zeta_p|\le s_p)=1-\alpha.$$

Let $u\in(0,+\infty)^p$, the notation $1/u$ denotes $(1/u_1,\dots,1/u_p)$. The Lagrange multiplier theorem assures that at s^*, the minimizer of (2), the vector $\nabla g(s^*)$ is collinear to $\nabla F(s^*)$, where ∇ denotes the gradient. Consequently, the following equivalences hold:

$$\frac{1}{s^*}\propto\nabla F(s^*)\Leftrightarrow s^*\propto\frac{1}{\nabla F(s^*)}\Leftrightarrow s^*\propto s^*+\frac{1}{\nabla F(s^*)}\text{ where }u\propto v\text{ means that }u\text{ is collinear to }v.$$

Let us notice that whatever $s\in[0,+\infty)^p$ the components of $\nabla F(s)$ are strictly positive. This collinearity motivates us to consider the following iterative sequence:
Let us set $s^{(0)}=(c_{1-\alpha},\dots,c_{1-\alpha})$, where $c_{1-\alpha}$ is the $1-\alpha$ quantile of $\max\{|\zeta_1|,\dots,|\zeta_p|\}$ and let us define the iterative sequence $(s^{(i)})_{i\in\mathbb{N}}$, where $s^{(i+1)}$ is given by

$$\begin{cases}u^{(i)}=\left(s^{(i)}+\frac{1}{\nabla F(s^{(i)})}\right),\\ s^{(i+1)}=\lambda_{1-\alpha}u^{(i)}\text{ where }\lambda_{1-\alpha}\text{ is such that }F(\lambda_{1-\alpha}u_1^{(i)},\dots,\lambda_{1-\alpha}u_p^{(i)})=1-\alpha.\end{cases}$$

In the previous expression, because $\nabla F(s^{(i)})>0$, then $u^{(i)}>s^{(i)}$ (these two inequalities are given as per component). The parameter $\lambda_{1-\alpha}$, the $1-\alpha$ quantile of $\max\left\{|\zeta_1|/u_1^{(i)},\dots,|\zeta_p|/u_p^{(i)}\right\}$, shrinks $u^{(i)}$ in order to recover an element $s^{(i+1)}$ so that $\mathbb{P}(|\zeta_1|\le s_1^{(i+1)},\dots,|\zeta_p|\le s_p^{(i+1)})=1-\alpha$. Thus far, this numerical method is available whatever ζ having a continuous distribution on $\mathbb{R}^p$ and a covariance matrix Σ. However, the naive computation of the gradient $\nabla F(s^{(i)})$ through simulation is time expensive. Roughly, the components of $s^{(i)}$ are large; thus, $\mathbb{P}(s_j^{(i)}\le|\zeta_j|\le s_j^{(i)}+h)$ is very small, and consequently, a good estimation of $\big(F(s_1,\dots,s_{j-1},s_j+h,s_{j+1},\dots,s_p)-F(s_1,\dots,s_p)\big)/h$ through simulations is very time consuming. Fortunately, there is a trick to compute ∇F in the Gaussian and student frameworks. For example, let us explain how this trick provides the first component of ∇F. When ζ is Gaussian, the conditional distribution $\mathscr{L}(\zeta_1|\zeta_2=x_2,\dots,\zeta_p=x_p)$ is a Gaussian distribution with density $f_{m(x),\sigma^2}$ (the mean $m(x)$ depends on $x:=(x_2,\dots,x_p)$ while the variance σ^2 does not depend on x). Precisely, let $A=(C_{1j})_{2\le j\le p}$ and let

© Société Française de Statistique et Société Mathématique de France (2021) ISSN: 2102-6238

$B = (C_{ij})_{2\leq i,j\leq p}$, then $m(x) = AB^{-1}x$ and $\sigma^2 = C_{11} - AB^{-1}A'$. Let $f_{\zeta_1}(\,.\,|\,|\zeta_2| \leq s_2, \ldots, |\zeta_p| \leq s_p)$ be the density of the conditional distribution $\mathscr{L}(\zeta_1||\zeta_2| \leq s_2, \ldots, |\zeta_p| \leq s_p)$ and let $f_{\zeta_2,\ldots,\zeta_p}$ be the density of $(\zeta_2, \ldots, \zeta_p)$, the first component of ∇F is given hereafter

$$\frac{\partial F}{\partial s_1}(s) = 2 f_{\zeta_1}(s_1 \,|\, |\zeta_2| \leq s_2, \ldots, |\zeta_p| \leq s_p) = 2 \int_{-s_2}^{s_2} \ldots \int_{-s_p}^{s_p} f_{\zeta_2,\ldots,\zeta_p}(x) f_{m(x),\sigma^2}(s_1) dx_2 \ldots dx_p.$$

Finally, the last expression is easy to compute by using Monte-Carlo simulations of the Gaussian vector $(\zeta_2, \ldots, \zeta_p)$. When $\zeta \sim t_{n-p}(\mathbf{0}, C)$ then ζ has the same distribution as $Z/\sqrt{V/(n-p)}$, where $Z \sim \mathscr{N}(\mathbf{0}, C)$ is independent of $V \sim \chi^2_d$ with $d = n - p$. In the student framework, the first component of ∇F is given hereafter

$$\begin{aligned} \frac{\partial F}{\partial s_1}(s) &= 2 \quad f_{\zeta_1}(s_1 \,|\, |\zeta_2| \leq s_2, \ldots, |\zeta_p| \leq s_p) \\ &= \int_0^{+\infty} 2 f_{Z_1}\left(s_1\sqrt{\frac{v}{d}} \,|\, |Z_2| \leq s_2\sqrt{\frac{v}{d}}, \ldots, |Z_p| \leq s_p\sqrt{\frac{v}{d}} \right) f_V(v) dv. \end{aligned}$$

This integral is still easy to infer through simulation by adding, with respect to the Gaussian framework, Monte-Carlo simulations of V. Now, in the Gaussian framework, we illustrate the performance of our solver using two examples.

Setting 1: We set $C := \mathrm{diag}(M_{500}(1, 0.9), Id_{500})$. Because C is block diagonal equicorrelated, the optimal thresholds satisfy $s^*_1, \ldots, s^*_{500} = c_1$ and $s^*_{501}, \ldots, s^*_{1000} = c_2$ where c_1 and c_2 are handily computable. Thus, in this setting, it is easy to compare the theoretical optimal thresholds with the thresholds given by the solver of the problem (2).

Setting 2: We set $C = (C_{ij})_{1\leq i,j\leq 1000}$ with $C_{ij} = \sqrt{\min\{i,j\}/\max\{i,j\}}$. The matrix C is the correlation matrix of a Brownian motion discretized on the set $\{1, \ldots, 1000\}$.

The left panel of the figure 1 shows that in setting 1 after $i = 5$ iterations, the threshold $s^{(i)}$ almost recovers the optimal thresholds. The right panel shows that in setting 2, $s^{(5)}_1 \geq \cdots \geq s^{(5)}_{999}$ (there is a singularity for $s^{(5)}_{1000}$ that is not a numerical problem).

As described in the figure 1, in both settings, the gain between the volume of the hypercube associated with the initial threshold $s^{(0)}$ (for which $s^{(0)}_1 = \cdots = s^{(0)}_{1000}$) and that of the hyperrectangular region associated with the threshold $s^{(5)}$ is very large.

Given optimal thresholds $s^*_1, \ldots, s^*_p$, the solution of the problem (2), one derives a multiple testing procedure for the null hypotheses $\mathscr{H}_i : \beta_i = 0$, $i \in \{1, \ldots, p\}$. The hypothesis $\mathscr{H}_i$ is rejected when $0 \notin [\hat{\beta}^{\mathrm{mle}}_i \pm s^*_i \times \mathrm{se}(\hat{\beta}^{\mathrm{mle}}_i)]$ or equivalently, when $|\hat{\beta}^{\mathrm{mle}}_i|/\mathrm{se}(\hat{\beta}^{\mathrm{mle}}_i) > s^*_i$. Because the confidence region $[\hat{\beta}^{\mathrm{mle}}_1 \pm s^*_1 \times \mathrm{se}(\hat{\beta}^{\mathrm{mle}}_1)] \times \cdots \times [\hat{\beta}^{\mathrm{mle}}_p \pm s^*_p \times \mathrm{se}(\hat{\beta}^{\mathrm{mle}}_p)]$ contains β with a probability $1 - \alpha$, the previous procedure controls the Family-Wise Error Rate (FWER) at significance level α. Let us remind that $\mathscr{A}_0 := \{i \in \{1, \ldots, p\} \mid \beta_i = 0\}$, the FWER is the probability $\mathbb{P}\left(\bigcup_{i\in\mathscr{A}_0} |\hat{\beta}^{\mathrm{mle}}_i|/\mathrm{se}(\hat{\beta}^{\mathrm{mle}}_i) > s^*_i\right)$. The inequality given hereafter assures the control of the FWER at significance level α.

$$\begin{aligned} \mathbb{P}\left(\bigcup_{i\in\mathscr{A}_0} \frac{|\hat{\beta}^{\mathrm{mle}}_i|}{\mathrm{se}(\hat{\beta}^{\mathrm{mle}}_i)} > s^*_i \right) &= \mathbb{P}\left(\bigcup_{i\in\mathscr{A}_0} \beta_i \notin [\hat{\beta}^{\mathrm{mle}}_i \pm s^*_i \times \mathrm{se}(\hat{\beta}^{\mathrm{mle}}_i)] \right) \\ &\leq \mathbb{P}\left(\bigcup_{i=1}^{p} \beta_i \notin [\hat{\beta}^{\mathrm{mle}}_i \pm s^*_i \times \mathrm{se}(\hat{\beta}^{\mathrm{mle}}_i)] \right) = \alpha. \end{aligned}$$

© Société Française de Statistique et Société Mathématique de France (2021) ISSN: 2102-6238

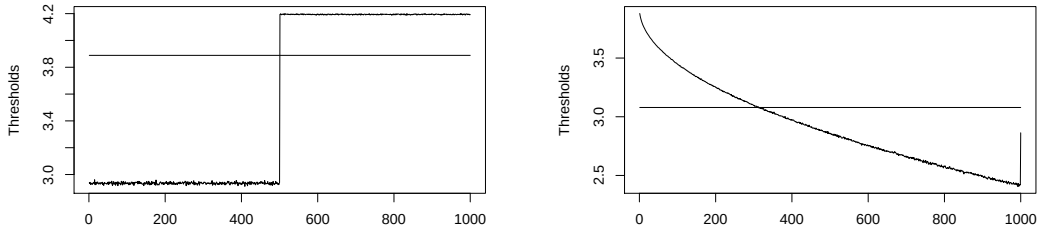


FIGURE 1. *The figure on the left provides the optimal thresholds associated with setting 1 described above. A handy computation of the optimal thresholds gives* $s_1^* = \cdots = s_{500}^* = 2.93$ *and* $s_{501}^* = \cdots = s_{1000}^* = 4.19$. *The y-axis of the figure provides the thresholds* $s_1^{(5)}, \ldots, s_{1000}^{(5)}$ *given by the iterative sequence* $s^{(i)}$ *after* $i = 5$ *iterations and* $s_1^{(0)} = \cdots = s_{1000}^{(0)} = 3.88$. *Observe that our solver almost recovers the optimal thresholds. With respect to the initial threshold* $s^{(0)}$, *the gain in volume is very large, as* $\sum_{j=1}^{1000} \log(s_j^{(0)}) = 1358.07$ *while* $\sum_{j=1}^{1000} \log(s_j^{(5)}) = 1255.20$. *The figure on the right provides the optimal thresholds associated with setting 2 described above. The y-axis of the figure provides the thresholds* $s_1^{(5)}, \ldots, s_{1000}^{(5)}$ *given by the iterative sequence* $s^{(i)}$ *after* $i = 5$ *iterations and* $s_1^{(0)} = \cdots = s_{1000}^{(0)} = 3.08$. *Again, the gain in volume is also large in setting 2, as* $\sum_{j=1}^{1000} \log(s_j^{(0)}) = 1124.60$ *while* $\sum_{j=1}^{1000} \log(s_j^{(5)}) = 1063.71$.

Intuitively, because the volume of the hyperrectangular confidence region is minimal, one should expect to recover a multiple testing more powerful than the classical single step procedures. The numerical experiments given in the following section confirm this intuition.

4. Comparison of multiple testing procedures

One of the most famous single-step multiple testing procedure controlling the FWER is the procedure described in Lehmann and Romano (2005) page 352 (hereafter, procedure 1).

Procedure 1: Whatever $i \in \{1, \ldots, p\}$, the null hypothesis $\mathscr{H}_i : \beta_i = 0$ is rejected in favour of the alternative $\beta_i \neq 0$ when $|\hat{\beta}_i^{\text{mle}}|/\text{se}(\hat{\beta}_i^{\text{mle}}) > s_{\max}$. The threshold $s_{\max}$ is the $1-\alpha$ quantile of $\max\{|\zeta_1|, \ldots, |\zeta_p|\}$.

Sequential procedures have better power than single step procedures, especially when β has many large components. Hereafter, we describe the StepDown (SD) counterpart of procedure 1.

The generic stepdown procedure defined by Romano and Wolf (2005), Lehmann and Romano (2005) p. 352, Dudoit and Van Der Laan (2007) p. 126 is a generalization of Holm's sequential procedure (Holm, 1979). To describe the generic stepdown procedure, let us denote $T_i = \hat{\beta}_i^{\text{mle}}/\text{se}(\hat{\beta}_i^{\text{mle}})$. The statistical tests are sorted from the most significant to the least significant, namely, $|T_{r(1)}| \geq \cdots \geq |T_{r(p)}|$. The rejection of the hypotheses $\mathscr{H}_{r(1)}, \ldots, \mathscr{H}_{r(p)}$ is done sequentially, as explained hereafter.

Procedure 1 SD: The hypothesis $\mathscr{H}_{r(1)}$ is rejected if $|T_{r(1)}| \geq t_{r(1)}$. The hypothesis $\mathscr{H}_{r(2)}$ is rejected if $|T_{r(1)}| > t_{r(1)}$ and $|T_{r(2)}| > t_{r(2)}$ and so on. In the previous expressions, the threshold $t_{r(s)}$ is the $1-\alpha$ quantile of $\max\{|\zeta_{r(s)}|, \ldots, |\zeta_{r(p)}|\}$ (see *e.g.* (Lehmann and Romano, 2005) pages 351 to 353).

© Société Française de Statistique et Société Mathématique de France (2021) ISSN: 2102-6238

We will compare procedures 1 and 1 SD with procedure 2, described hereafter, which is derived from the computation of the optimal thresholds. Note that procedure 1 and 1 SD are respectively more powerful that Bonferroni's precedure (Dunn, 1961) and Holm's procedure (Holm, 1979) . Therefore performances of Bonferroni and Holm procedures are not reported.

Procedure 2: Whatever $i \in \{1,\dots,p\}$, the null hypothesis $\mathscr{H}_i : \beta_i = 0$ is rejected in favour of the alternative $\beta_i \neq 0$ when $|\hat{\beta}_i^{\text{mle}}|/\text{se}(\hat{\beta}_i^{\text{mle}}) > s_i^*$. The thresholds $s_1^*,\dots,s_p^*$ are the optimal ones given in (2).

By construction of the thresholds $s_{\max}$ and $s_1^*,\dots,s_p^*$, the single step procedures 1 and 2 control the FWER at a significance level $\alpha \in (0,1)$. In addition, the procedure 1 SD also controls the FWER at a significance level $\alpha \in (0,1)$ (see *e.g.* (Lehmann and Romano, 2005) pages 351 to 353 or (Romano and Wolf, 2005)). A comparison of these three procedures based on the average power is carried out on the following setting:

We set $\text{var}(\hat{\beta}_1^{\text{mle}}) = \cdots = \text{var}(\hat{\beta}_p^{\text{mle}}) = 1$ and the Gaussian vector ζ has a $\mathscr{N}(\mathbf{0}, C)$ distribution, where $C := \text{diag}(M_{900}(1,\rho), Id_{100})$. We set $\beta \in \mathbb{R}^{1000}$ with $\text{card}(\mathscr{A}_1) = k \in \{50, 500\}$ and for all i $\in \mathscr{A}_1, \beta_i = t$, where $t > 0$. For the different values of $\rho \in \{0, 0.5, 0.9, 0.999\}$, the optimal thresholds $s_1^*,\dots,s_{1000}^*$ given by Proposition 2 are as follows.

— When $\rho = 0$ then $s_1^* = \cdots = s_{900}^* = c_1 \approx 4.0553$ and $s_{901}^* = \cdots = s_{1000}^* = c_2 \approx 4.0553$.
— When $\rho = 0.5$ then $s_1^* = \cdots = s_{900}^* = c_1 \approx 3.7628$ and $s_{901}^* = \cdots = s_{1000}^* = c_2 \approx 4.0961$.
— When $\rho = 0.9$ then $s_1^* = \cdots = s_{900}^* = c_1 \approx 2.9284$ and $s_{901}^* = \cdots = s_{1000}^* = c_2 \approx 4.3327$.
— When $\rho = 0.999$ then $s_1^* = \cdots = s_{900}^* = c_1 \approx 2.0601$ and $s_{901}^* = \cdots = s_{1000}^* = c_2 \approx 4.4542$.

Hereafter, the average power of a multiple testing procedure represents the proportion of hypotheses associated to non-null components of β which are correctly rejected. Let $I_1 = \{1,\dots,900\}$, let $I_2 = \{901,\dots,1000\}$ and let $V \sim \mathscr{N}(t,1)$. In this framework, the average power of procedure 2 is

$$\frac{\text{card}(\mathscr{A}_1 \cap I_1)}{\text{card}(\mathscr{A}_1)} \mathbb{P}_t(|V| > c_1) + \frac{\text{card}(\mathscr{A}_1 \cap I_2)}{\text{card}(\mathscr{A}_1)} \mathbb{P}_t(|V| > c_2).$$

We observe that the average power of procedure 2 depends on the location of the non-null components of β. Intuitively, when the non-null components of β are located on I_1 (*i.e.* $\mathscr{A}_1 \subset I_1$) then, because the thresholds $s_1^*,\dots,s_{900}^*$ are small, procedure 2 should be powerful. On the other hand, when the non-null components of β are located on I_2 (*i.e.* $\mathscr{A}_1 \subset I_2$) then, because the thresholds $s_{901}^*,\dots,s_{1000}^*$ are large, procedure 2 should not be powerful. Thus, to perform a fair comparison, instead of computing the average power for a particular $\mathscr{A}_1$, we will examine the expected value of the average power when '$\mathscr{A}_1$' is a random set uniformly distributed on the set of combination of k elements among 1000. Let U be a random set with a uniform distribution on the set $\{I \subset \{1,\dots,1000\} \mid \text{card}(I) = k\}$, the expected value of the average power of procedure 2 is given hereafter

$$\mathbb{E}_U\left(\frac{\text{card}(U \cap I_1)}{k} \mathbb{P}_t(|V| > c_1) + \frac{\text{card}(U \cap I_2)}{k} \mathbb{P}_t(|V| > c_2)\right) = \frac{9}{10} \mathbb{P}_t(|V| > c_1) + \frac{1}{10} \mathbb{P}_t(|V| > c_2). \quad (3)$$

In the figure 2, we compare the average power of procedures 1, 2 and 1 SD.

© Société Française de Statistique et Société Mathématique de France (2021) ISSN: 2102-6238

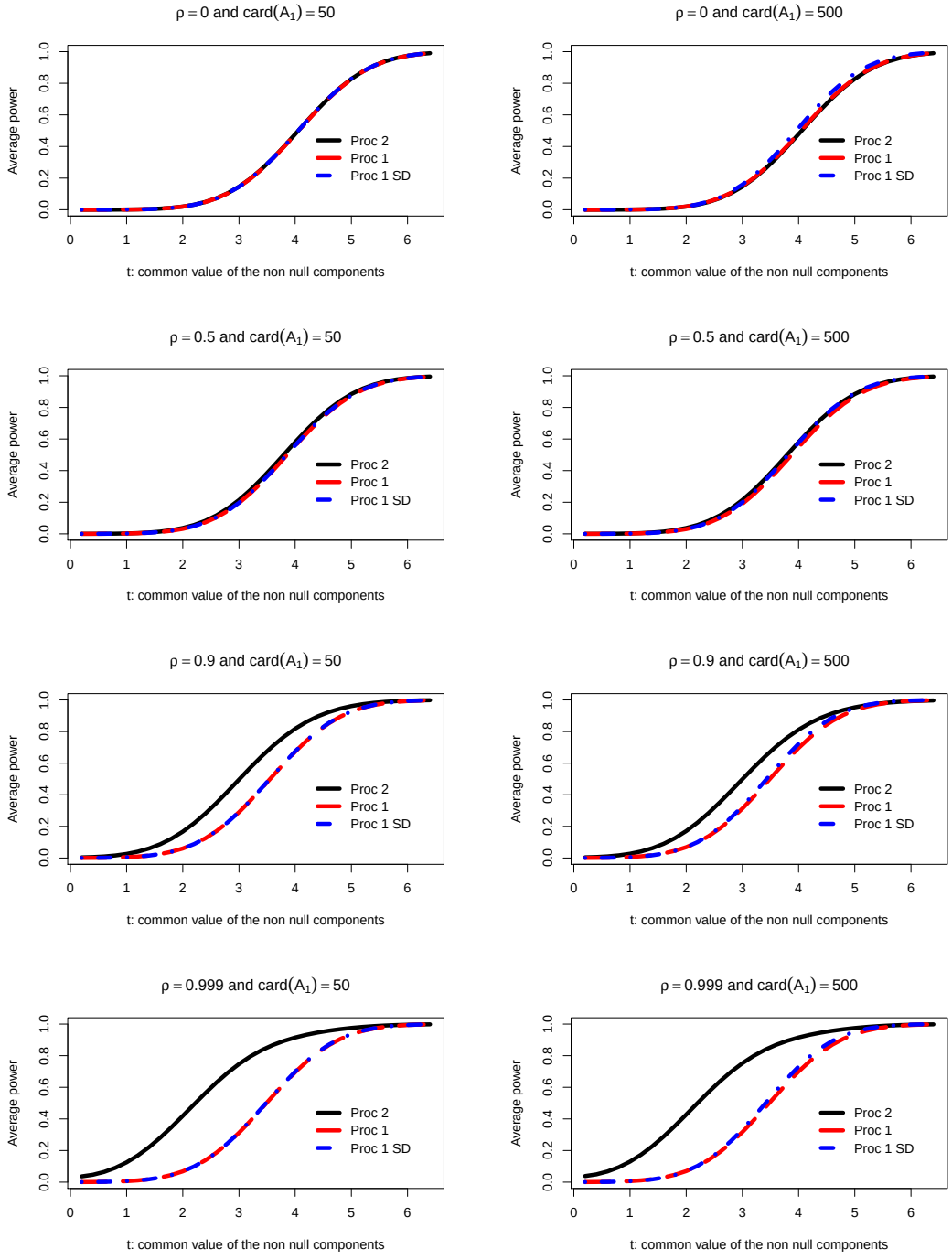


FIGURE 2. *This figure provides the average power of procedures 1, 1 SD and 2 (the average power of procedure 2 is reported in (3)). Average power of procedures 1 and 2 are explicit and thus can be computed without using simulations whereas the average power of procedure 1 SD is computed based on 10000 simulations. When* $\rho = 0$*, these three procedures have approximately the same power. When* ρ *increases, the difference between the average power of procedure 2 increases in comparison with the average power of the other procedures.*

© Société Française de Statistique et Société Mathématique de France (2021) ISSN: 2102-6238

A stepdown procedure is merely a sequential application of a single step procedure. Precisely, in the first stage of stepdown procedure 1 SD, the rejections are the ones given by the single step procedure 1. The second stage of stepdown procedure 1 SD is an application of single step procedure 1 on the hypotheses not rejected in the first stage, and so on. Intuitively, when β has many very large components, a large number of hypotheses are rejected in the first stage, implying that the number of hypotheses tested in the second stage becomes small, allowing the stepdown procedure to become powerful. The most favourable setting for the stepdown procedures is when $\text{card}(\mathscr{A}_1) = 500$ and t is large, as in this case, β has lot of large components. Note that our procedure is at least as competitive as the other ones in this situation, and depending on ρ, our procedure can be much more powerful than the state-of-the art procedures.

In summary, when some components of $\hat{\beta}^{\text{mle}}$ are strongly correlated, our method outperforms the other ones.

It could appear as appealing to construct a stepdown procedure based on the procedure 2. Unfortunately, as illustrated in the appendix the application of the generic stepdown method on the procedure 2 (as described in Lehmann and Romano (2005) page 353) does not control the FWER since the monotonicity assumption does not hold. The construction of a stepdown procedure based on the procedure 2 and controlling the FWER is, for the authors, an open question.

5. Application in metabolomics: detection of metabolites

Metabolomics is the science of detection of metabolites (small molecules) in biological mixtures (e.g. blood and urine). The most common technique for performing such characterization is proton nuclear magnetic resonance (NMR). Each metabolite generates a characteristic resonance signature in the NMR spectra with an intensity proportional to its concentration in the mixture. The number of peaks generated by a metabolite and their locations and ratio of heights are reproducible and uniquely determined: each metabolite has its own signature in the spectra. Each signature spectrum of each metabolite can be stored in a library that could contain hundreds of spectra. A major challenge in NMR analysis of metabolic profiles is automatic metabolite assignment from spectra. To identify metabolites, experts use spectra of pure metabolites and manually compare these spectra to the spectrum of the biological mixture under analysis. Such a method is time-consuming and requires domain-specific knowledge. Furthermore, complex biological mixtures can contain hundreds or thousands of metabolites, which can result in highly overlapping peaks. Figure 3 gives an example of an annotated spectrum of a mixture.

Recently, automatic methods have been proposed, for example, Metabohunter (Tulpan et al., 2011), BATMAN (Astle et al., 2012; Hao et al., 2012), Bayesil (Ravanbakhsh et al., 2015) or the software Chenomx (Weljie et al., 2006). Most of these methods are based on modelling using a Lorentzian shape and a Bayesian strategy. Nevertheless, most are time-consuming, and thus, cannot be applied to a large library of metabolites, and/or their statistical properties are not proven. Thus, the establishment of a gold-standard methodology with proven statistical properties for identification of metabolites would be very helpful for the metabolomic community.

Because the number of tests is not very large (one can expect to analyse a mixture with about 200 metabolites), and as NMR experts want to recover all metabolites present in the mixture, but to prevent a false discovery, we developed a multiple testing procedure controlling the FWER.

© Société Française de Statistique et Société Mathématique de France (2021) ISSN: 2102-6238

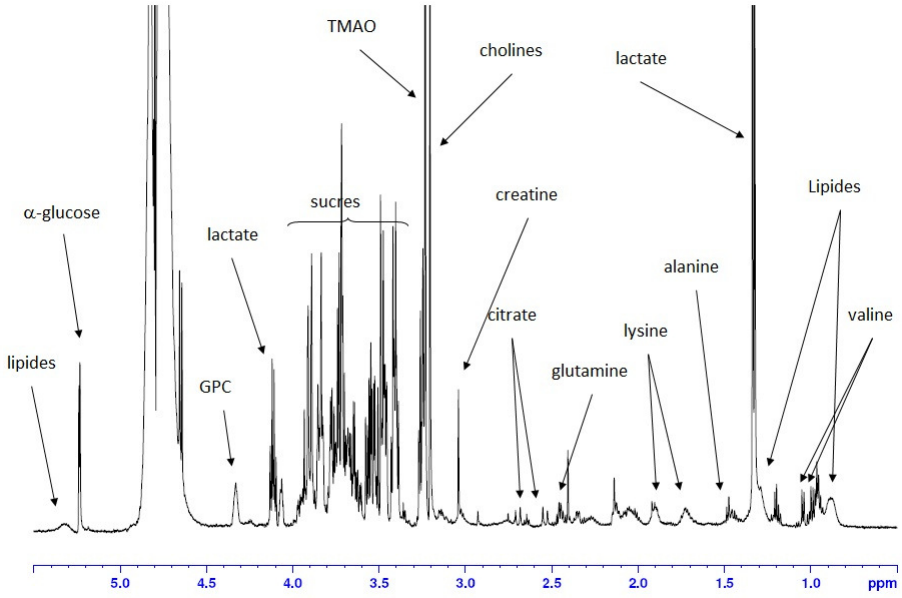


FIGURE 3. *An annotated mixture spectrum with overlaps between peaks of lipides and valine and between peaks of glutamine and lysine.*

5.1. Modelling

The spectrum of a metabolite (or a mixture) is a nonnegative function defined on a compact interval T. We assume that we have a library of known spectra containing all $p = 36$ metabolites $\{f_i\}_{1\leqslant i\leqslant p}$ (with $\int_T f_i(t)dt = 1$) that can be found in a mixture. This family of p spectra is assumed to be linearly independent. In the first approximation, the observed spectrum of the mixture Y can be modelled as a discretized noisy convex combination of the pure spectra:

$$Y_j = \left(\sum_{i=1}^{p} \beta_i f_i(t_j)\right) + \varepsilon_j \text{ with } 1 \leqslant j \leqslant n, t_1 < \cdots < t_n \text{ a subdivision of } T \text{ and } n = 6001.$$

The random vector $(\varepsilon_1, \ldots, \varepsilon_n)$ is a Gaussian vector $\mathcal{N}(\mathbf{0}, \Gamma)$, where Γ is a known and invertible covariance matrix. The covariance structure $(\varepsilon_1, \ldots, \varepsilon_n)$ is described in Tardivel et al. (2017).

5.2. Real dataset

The method for the detection of metabolites was tested on a known mixture. The NMR experts supplied us with a library of 36 spectra of pure metabolites and a mixture composed of these metabolites. We first analysed this mixture without any knowledge about the number of used metabolites and their proportions. The results are presented in Table 1.

After analysing this mixture we have compared our results with the real composition of the mixture supplied by NMR experts. The six metabolites present in the complex mixture was

© Société Française de Statistique et Société Mathématique de France (2021) ISSN: 2102-6238

Metabolites	Actual proportions	Significantly not null
Choline chloride	0.545	Yes
Creatinine	0.209	Yes
Benzoic acid	0.086	Yes
L-Proline	0.069	Yes
D-Glucose	0.060	Yes
L-Phenylalanine	0.029	Yes
30 other metabolites	0	No

TABLE 1. *This table presents the results for the 36 metabolites of the library. The actual proportions of each metabolite are presented in the first column. For each metabolite, evidence against the nullity of the proportion is given in the second column.*

detected, including those with small proportions. Among the 30 other metabolites which are not present in the mixture, no one was wrongly detected. Because the whole procedure is fast, lasting only a few seconds, it could be easily applied to a library containing several hundred metabolites. For more detailed results on this application to metabolomics, we refer the interested reader to Tardivel et al. (2017) where our procedure is compared to existing procedures on more complex datasets and to Lefort et al. (2019) where the package ASICS, derived from this procedure and available on the Bioconductor platform, is presented.

6. Conclusions

This study takes a new look at an old problem: the construction of multiple testing procedures derived from hyperrectangular confidence regions. Our purpose is to derive such a procedure based on hyperrectangular confidence regions having a minimal volume. These regions depend on an optimal threshold s^*, which is a solution of the constraint problem (2); we provide a solver giving a numerical solution to this problem. When $p \leq 1000$, the optimal threshold s^* (thus, the optimal hyperrectangular confidence region) is easily tractable. With respect to standard hypercube confidence regions, the gain in volume obtained with our method is huge. Based on simulations, we show that deriving a multiple testing procedure from a hyperrectangular region having a minimal volume is an intuitive way to increase the average power. Indeed, simulations show that our procedure is at least as powerful as the other procedures, and depending on the correlation matrix, our procedure can be much more powerful than the state-of-the-art procedures. However, it is still a challenge to provide a stepdown counterpart to our procedure.

7. Appendix

7.1. Proofs:

Proposition 3. *Let C be a $p \times p$ correlation matrix and let ζ be a Gaussian vector $\mathcal{N}(\mathbf{0}, C)$ or a multivariate student $t_{n-p}(\mathbf{0}, C)$. Then, there exists at least one element $s^* \in [0, +\infty)^p$ solution of the problem (2).*

Proof: We see that the volume cannot be minimal when $\|s\|_\infty$ is too large. Let $q_i > 0$ be the $1-\alpha$ quantile of $|\zeta_i|$, let $q = \min\{q_1, \ldots, q_p\} > 0$ and let us set $S := \{s \in \mathbb{R}^p \mid \mathbb{P}(|\zeta_1| \leq$

© Société Française de Statistique et Société Mathématique de France (2021) ISSN: 2102-6238

$s_1,\dots,|\zeta_1|\le s_p)=1-\alpha\}$. Whatever $s\in S$ the following inequality holds $\prod_{i=1}^p s_i\ge\|s\|_\infty q^{p-1}$, consequently, the function $s\in S\mapsto\prod_{i=1}^p s_i$ is coercive and continuous. Finally, since the function $s\in[0,+\infty)^p\mapsto\mathbb{P}(|\zeta_1|\le s_1,\dots,|\zeta_1|\le s_p)$ is continuous then the set S is closed and consequently the minimum of the problem (2) is reached. □

Proof of Proposition 1:
To simplify the computation of the gradients, we consider the following problem, which has the same solution as (2)

$$\min g(s)=\sum_{i=1}^p \ln(s_i) \text{ subject to } F(s)=\mathbb{P}(|\zeta_1|/s_1\le 1,\dots,|\zeta_p|/s_p\le 1)=1-\alpha.$$

As this problem reaches its minimum at s^*, $\nabla g(s^*)$ is collinear to $\nabla F(s^*)$. Let us set D the matrix $D=\mathrm{diag}(\mathrm{s}_1,\dots,\mathrm{s}_\mathrm{p})$, we have the following expression for $F(s_1,\dots,s_p)$, namely,

$$\begin{aligned}F(s_1,\dots,s_p) &= \int_{[-1,1]^p} R\exp\left(-\frac{1}{2}x'DC^{-1}Dx\right)\det(D)dx\\ &= \int_{[-1,1]^p} R\exp\left(-\frac{1}{2}x'DC^{-1}Dx+\ln(\det(D))\right)dx,\end{aligned}$$

with $R=1/((2\pi)^{p/2}\det(C)^{1/2})$. Next, the expression of the partial derivative

$$\frac{\partial}{\partial s_i}\left(-\frac{1}{2}x'DC^{-1}Dx+\ln(\det(D))\right)=\frac{1}{s_i}-\sum_{j=1}^p C^{-1}_{i,j}x_ix_js_j,$$

implies that the gradient of F is equal to

$$\begin{aligned}\frac{\partial F}{\partial s_i}(s_1,\dots,s_p) &= \frac{1}{s_i}F(s_1,\dots,s_p)-R\sum_{j=1}^p\int_{[-1,1]^p}(C^{-1}_{i,j}x_ix_js_j)\exp\left(-\frac{1}{2}x'DC^{-1}Dx\right)\det(D)dx,\\ &= \frac{1-\alpha}{s_i}-R\sum_{j=1}^p\int_{[-1,1]^p}(C^{-1}_{i,j}x_ix_js_j)\exp\left(-\frac{1}{2}x'DC^{-1}Dx\right)\det(D)dx.\end{aligned}$$

Thus, $\nabla F(s)=(1-\alpha)\nabla g(s)+v(s)$, where $v(s)\in\mathbb{R}^p$ is the following vector

$$v(s):=\left(\sum_{j=1}^p C^{-1}_{i,j}\int_{[-1,1]^p}x_ix_js^*_jR\exp\left(-\frac{1}{2}x'DC^{-1}Dx\right)\det(D)dx\right)_{1\le i\le p}.$$

Consequently, $\nabla g(s^*)$ and $\nabla F(s^*)$ are collinear if and only if $\nabla g(s^*)$ and $v(s^*)$ are collinear.

$$\begin{aligned}&\exists k\in\mathbb{R} \text{ such that } v(s^*)=k\nabla g(s^*),\\ \Leftrightarrow\quad &\forall i\in\{1,\dots,p\},\sum_{j=1}^p C^{-1}_{i,j}\int_{[-1,1]^p}x_is^*_ix_js^*_jR\exp\left(-\frac{1}{2}x'DC^{-1}Dx\right)\det(D)dx=k,\\ \Leftrightarrow\quad &\forall i\in\{1,\dots,p\},\sum_{j=1}^p C^{-1}_{i,j}\int_{u\in\mathbb{R}^p}u_iu_j\frac{R}{1-\alpha}\exp\left(-\frac{1}{2}u'C^{-1}u\right)\mathbf{1}(u\in S^*)du=\frac{k}{1-\alpha}.\end{aligned}\quad(4)$$

© Société Française de Statistique et Société Mathématique de France (2021) ISSN: 2102-6238

The expression (4) is obtained *via* the change of variables $\forall i \in \{1,\dots,p\}, u_i = x_i s_i^*$. To conclude, one recognizes that

$$\int_{u\in\mathbb{R}^p} u_i u_j \frac{R}{1-\alpha} \exp\left(-\frac{1}{2}u'C^{-1}u\right)\mathbf{1}(u\in S^*)du = \mathbb{E}\left(T_i^{s^*}T_j^{s^*}\right) = \operatorname{cov}\left(T_i^{s^*},T_j^{s^*}\right).$$

Thus, the diagonal coefficients of $C^{-1}\operatorname{var}(T_{s^*})$ are equal to $k/(1-\alpha)$. □

7.2. *Proof of Proposition 2*

Let us remind that $M_p(a,b)$ represents a $p\times p$ matrix whose diagonal coefficients are a and non-diagonal coefficients are b. Our main purpose is to prove Lemma 1.

Lemma 1. *Let ζ be a Gaussian vector of $\mathbb{R}^p$ having a $\mathcal{N}(\mathbf{0},M_p(1,\rho))$ distribution where $\rho\in[0,1)$ and let $\alpha\in[0,1)$. The following problem*

$$\text{minimize}\prod_{i=1}^p s_i \text{ subject to } \mathbb{P}(|\zeta_1|\le s_1,\dots,|\zeta_p|\le s_p) = 1-\alpha. \tag{5}$$

has a unique minimizer $s^* = (c_{1-\alpha},\dots,c_{1-\alpha})$ where $c_{1-\alpha}$ is the $1-\alpha$ quantile of the random variable $\max\{|\zeta_1|,\dots,|\zeta_p|\}$.

Proposition 2 is a straightforward consequence of lemma 1

Sketch of the proof

First le us notice the Gaussian vector ζ given in Lemma 1 has the same distribution as $(\sqrt{\rho}Z_0+\sqrt{1-\rho}Z_i)_{1\le i\le p}$ where $Z_0,\dots,Z_p$ are i.i.d $\mathcal{N}(0,1)$ random variables. Consequently, conditionally to $\{Z_0=z_0\}$, the Gaussian vector $(\sqrt{\rho}Z_0+\sqrt{1-\rho}Z_i)_{1\le i\le p}$ has i.i.d components having $\mathcal{N}(\sqrt{\rho}z_0,1-\rho)$ distribution. Lemmas 2 and 3 provide some geometrical results associated to this conditional distribution $\mathcal{N}(\theta,a^2Id_p)$ for some $\theta\in\mathbb{R}^p$ and some $a\ge 0$. More precisely, when $\theta=0$, Lemma 2 shows that among hyperrectangle with volume 2^pV, the hypercube $[-V^{1/p},V^{1/p}]^p$ has a maximal probability with respect to the Gaussian measure $\mathcal{N}(\mathbf{0},a^2Id_p)$. Lemma 2 is a consequence of Proposition 1 but a specific and easier proof for this lemma is given here. Lemma 3 extends the result given by Lemma 2 to the case in which $\theta\neq 0$. Technical details to prove this second lemma are different than the ones used for the first lemma. Actually, the result given by Lemma 2 and an application of the Gronwall's inequality (Gronwall, 1919) are key steps to prove the Lemma 3. The proof of Lemma 1 is a quite straightforward consequence of lemmas 2 and 3.

Proof of Lemma 1

Lemma 2. *Let W be a Gaussian vector having a $\mathcal{N}(\mathbf{0},a^2Id_p)$ distribution where $a>0$ and let $V>0$. The following optimization problem*

$$\text{maximize } \mathbb{P}(|W_1|\le s_1,\dots,|W_p|\le s_p) \text{ subject to } \prod_{i=1}^p s_i = V \tag{6}$$

has a unique maximizer $\bar{s}\in(0,+\infty)^p$ which is $\bar{s}=(V^{1/p},\dots,V^{1/p})$.

© Société Française de Statistique et Société Mathématique de France (2021) ISSN: 2102-6238

Proof: Similar arguments than the ones given in proof of Proposition 1 shows that the optimization problem (6) has at least one maximizer in $(0,+\infty)^p$. Let us denote ϕ and Φ be respectively the density and cumulative distribution function of a $\mathcal{N}(0,1)$ distribution.

Let us introduce the following optimization problem equivalent to (6) which allows to reduce technical computation difficulties.

$$\text{maximize } L(s)=\sum_{i=1}^{p}\ln(2\Phi(s_i/a)-1) \text{ subject to } g(s)=\sum_{i=1}^{p}\ln(s_i)=\ln(V).$$

At $\bar{s}$, maximizer of (6), according to Lagrange multipliers theorem, gradient vectors $\nabla L(\bar{s})$ and $\nabla g(\bar{s})$ are collinear. Consequently the following collinearity holds

$$\left(\frac{2\phi(\bar{s}_i/a)}{a(2\Phi(\bar{s}_i/a)-1)}\right)_{1\le i\le p}\propto\left(\frac{1}{\bar{s}_i}\right)_{1\le i\le p} \text{ where } \propto \text{ means collinear to.}$$

Thus there exists $\lambda\in\mathbb{R}$ such that the following equalities occurs

$$\forall i\in\{1,\dots,p\},\frac{2\bar{s}_i\phi(\bar{s}_i/a)}{a(2\Phi(\bar{s}_i/a)-1)}=\lambda. \tag{7}$$

Whatever $a>0$, the function $h:t>0\mapsto(2t\phi(t/a))/(a(2\Phi(t/a)-1))$ is strictly decreasing on $(0,+\infty)$. Consequently, if $s\in(0,+\infty)^p$ is a vector for which equalities $s_1=\cdots=s_p$ do not occur then the condition (7) is not met implying thus that s is not a maximizer. One may deduce that $\bar{s}$, maximizer of (6), satisfies the equalities $\bar{s}_1=\cdots=\bar{s}_p$ which implies that $\bar{s}=(V^{1/p},\dots,V^{1/p})$. □

Lemma 3. *Let $W_1,\dots,W_p$ be i.i.d random variable having $\mathcal{N}(\theta,a^2)$ distribution, where $\theta\in\mathbb{R}$ and $a>0$ and let $V>0$. The following optimization problem*

$$\text{maximize } \mathbb{P}_\theta\left(|W_1|\le s_1,\dots,|W_p|\le s_p\right) \text{ subject to } \prod_{i=1}^{p}s_i=V$$

has a unique minimizer $\bar{s}\in[0,+\infty)^p$ which is $\bar{s}=(V^{1/p},\dots,V^{1/p})$.

Proof: Let $\overline{S}$ be the hypercube $\overline{S}:=[-V^{1/p},V^{1/p}]^p$, let $s_1>0,\dots,s_p>0$ not simultaneously equal and such that $\prod_{i=1}^{p}s_i=V$, let S be the hyperrectangle $S:=[-s_1,s_1]\times\cdots\times[-s_p,s_p]$ and let W be the Gaussian vector $W:=(W_1,\dots,W_p)$. Since S is not an hypercube, according to Lemma 2, when $\theta=0$ we already know that $\mathbb{P}_0(W\in\overline{S})>\mathbb{P}_0(W\in S)$. We are going to show that this inequality remains true when $\theta\neq0$. Let us set G be the following function

$$\begin{aligned}\forall\theta\in\mathbb{R},G(\theta):=(2\pi)^{p/2}&\left(\mathbb{P}_\theta(W\in\overline{S})-\mathbb{P}_\theta(W\in S)\right)\\&=\int_{x\in\mathbb{R}^p}\exp\left(-\frac{1}{2}\sum_{i=1}^{p}(\theta-x_i)^2\right)\left(\mathbf{1}(x\in\overline{S})-\mathbf{1}(x\in S)\right)dx.\end{aligned}$$

We aim to prove that G is strictly positive. The derivative of G is the following function

$$\begin{aligned}G'(\theta)&=\int_{x\in\mathbb{R}^p}\sum_{i=1}^{p}(x_i-\theta)\exp\left(-\frac{1}{2}\sum_{i=1}^{p}(\theta-x_i)^2\right)\left(\mathbf{1}(x\in\overline{S})-\mathbf{1}(x\in S)\right)dx,\\&=-p\theta G(\theta)+\int_{x\in\mathbb{R}^p}\sum_{i=1}^{p}x_i\exp\left(-\frac{1}{2}\sum_{i=1}^{p}(\theta-x_i)^2\right)\left(\mathbf{1}(x\in\overline{S})-\mathbf{1}(x\in S)\right)dx.\end{aligned}$$

© Société Française de Statistique et Société Mathématique de France (2021) ISSN: 2102-6238

Since $\sum_{i=1}^{p} x_i$ is bounded over the set $\overline{S} \cup S$, there exists a constant $K \geq 0$ such that whatever $x \in \overline{S} \cup S$, $\left|\sum_{i=1}^{p} x_i\right| \leq K$. From this inequality one deduces the differential inequality $G'(\theta) \geq -p\theta G(\theta) - K|G(\theta)|$. Now, let us prove, by contradiction, that G is positive on $[0, +\infty)$. Let us assume that G is not positive on $[0, +\infty)$ then, because G is continuous and $G(0) > 0$, there exists $\theta_0 > 0$ such that $G(\theta_0) = 0$ and G is non-negative on $[0, \theta_0]$. Therefore on this interval $G'(\theta) \geq G(\theta)(-p\theta - K)$ and thus, according to the Gronwall's inequality (Gronwall, 1919), this differential inequality implies the following result

$$\forall \theta \in [0, \theta_0], G(\theta) \geq G(0) \exp\left(\int_0^{\theta} (-pt - K)dt\right).$$

One may deduce the contradiction $G(\theta_0) > 0$ and thus, because G is even, G is positive on $\mathbb{R}$. Consequently, whatever $\theta \in \mathbb{R}$, the inequality $\mathbb{P}_\theta(W \in \overline{S}) > \mathbb{P}_\theta(W \in S)$ occurs implying thus that $\bar{s} = (V^{1/p}, \ldots, V^{1/p})$ is the unique maximizer of the problem (3). □

Proof of Lemma 1:
First, let us notice that when $Z_0, \ldots, Z_p$ are i.i.d $\mathcal{N}(0,1)$ random variables, the Gaussian vector $(\sqrt{\rho}Z_0 + \sqrt{1-\rho}Z_i)_{1 \leq i \leq p}$ has the same distribution as ζ. Let us assume that a minimizer s^*, solution of the problem (5), does not satisfy inequalities $s_1^* = \cdots = s_p^*$ and let us set $v = (s_1^* \times \cdots \times s_p^*)^{1/p}$ (thus $s^* \neq (v, \ldots, v)$). According to Lemma 3, conditionally to the event $\{Z_0 = z_0\}$, the following inequality occurs

$$\begin{aligned}
&\mathbb{P}\left(\forall i \in \{1, \ldots, p\}, \left|\sqrt{\rho}Z_0 + \sqrt{1-\rho}Z_i\right| \leq s_i^* | Z_0 = z_0\right) \\
&\qquad = \mathbb{P}\left(\forall i \in \{1, \ldots, p\}, \left|\sqrt{\rho}z_0 + \sqrt{1-\rho}Z_i\right| \leq s_i^*\right) \\
&\qquad < \mathbb{P}\left(\forall i \in \{1, \ldots, p\}, \left|\sqrt{\rho}z_0 + \sqrt{1-\rho}Z_i\right| \leq v\right).
\end{aligned}$$

Let ϕ be the density of the $\mathcal{N}(0,1)$ distribution, since the previous inequality holds whatever $z_0 \in \mathbb{R}$, one may deduce that

$$\begin{aligned}
&\mathbb{P}(\forall i \in \{1, \ldots, p\}, |\zeta_i| \leq v) \\
&\qquad = \int_{z_0 \in \mathbb{R}} \mathbb{P}\left(\forall i \in \{1, \ldots, p\}, \left|\sqrt{\rho}Z_0 + \sqrt{1-\rho}Z_i\right| \leq v | Z_0 = z_0\right) \phi(z_0) dz_0 \\
&\qquad > \int_{z_0 \in \mathbb{R}} \mathbb{P}\left(\forall i \in \{1, \ldots, p\}, \left|\sqrt{\rho}Z_0 + \sqrt{1-\rho}Z_i\right| \leq s_i^* | Z_0 = z_0\right) \phi(z_0) dz_0 \\
&\qquad > \mathbb{P}(\forall i \in \{1, \ldots, p\}, |\zeta_i| \leq s_i^*) = 1 - \alpha.
\end{aligned}$$

Because $\mathbb{P}(|\zeta_1| \leq v, \ldots, |\zeta_p| \leq v) > 1 - \alpha$, there exists $t \in [0, 1)$ such that $\mathbb{P}(|\zeta_1| \leq tv, \ldots, |\zeta_p| \leq tv) = 1 - \alpha$. Finally, since $t^p v^p < s_1^* \times \cdots \times s_p^*$, one deduces that $s^* = (s_1^*, \ldots, s_p^*)$ is not a minimizers of (5). Consequently, every components of a minimizer s^* of (5) are equal which implies that $s_1^* = \cdots = s_p^* = c_{1-\alpha}$. □

7.3. Dunnett's procedure is optimal

As an illustration of Proposition 1, the thresholds prescribed by Dunnett's procedure are optimal.

© Société Française de Statistique et Société Mathématique de France (2021) ISSN: 2102-6238

Dunnett's procedure compares the mean of treatment groups with the mean of the control group (Dunnett, 1955). When each group (control and treatment) has the same number of observations the thresholds given by Dunnett's procedure are the optimal ones.

Let us denote p the number of treatment groups and n the number of observations, under the assumptions of Dunnett (Gaussianity, homoscedasticity), the empirical means $\hat{M}_0, \hat{M}_1, \ldots, \hat{M}_p$ of each group are independent and distributed according to $\mathcal{N}(\mu_0, \sigma^2/n)$,$\mathcal{N}(\mu_1, \sigma^2/n)$, ..., $\mathcal{N}(\mu_p, \sigma^2/n)$ distributions (with σ known to simplify). In Dunnett's procedure, whatever $i \in \{1, \ldots, p\}$, the null hypothesis $\mathcal{H}_i^0 : \mu_i = \mu_0$ is rejected for the alternative $\mathcal{H}_i^1 : \mu_i \neq \mu_0$ as soon as $|\hat{M}_i - \hat{M}_0| > c_{1-\alpha}\sigma\sqrt{2/n}$, where $c_{1-\alpha}$ is the $1-\alpha$ quantile of $\max\{|\zeta_1|, \ldots, |\zeta_p|\}$. In the later expression, $\zeta := (\zeta_1, \ldots, \zeta_p)$ is distributed according to $\mathcal{N}(\mathbf{0}, C)$ distribution, where C is a $p \times p$ equicorrelated correlation matrix defined by $C_{ij} = 1$ if $i = j$ and $C_{ij} = 1/2$ otherwise. Because C is an equicorrelated correlation matrix, the threshold prescribed in Dunnett's procedure (the same for each hypothesis) is optimal with respect to the volume.

7.4. Application of the generic stepdown method to procedure 2 does not control the FWER

The following example illustrates that an application of the generic stepdown method to procedure 2 does not control the FWER. Let us assume that the maximum likelihood estimator $\hat{\beta}^{\text{mle}}$ is distributed according to a $\mathcal{N}(\beta, C)$ distribution where C and β, as given hereafter,

$$C = (C_1|C_2|C_3) := \begin{pmatrix} 1 & 0.999 & 0 \\ 0.999 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix} \text{ and } \beta = (+\infty, 0, 0).$$

Let us apply the generic stepdown method to our procedure based on the optimal thresholds. The solution of the problem (2) with the matrix C and $\alpha = 0.05$ is $s_1^{(0)} = s_2^{(0)} = 2.10$ and $s_3^{(0)} = 2.43$. In the first step, the hypothesis $\mathcal{H}_i$ is rejected if $|\hat{\beta}_i^{\text{mle}}| > s_i^{(0)}$ with $i \in \{1, 2, 3\}$. Obviously, because $\beta_1 = +\infty$, the hypothesis $\mathcal{H}_1$ is rejected; let us assume that $\mathcal{H}_2$ and $\mathcal{H}_3$ are not rejected. Consequently, there is no false discovery in the first step; thus, $|\hat{\beta}_2^{\text{mle}}| \leq s_2^{(0)}$ and $|\hat{\beta}_2^{\text{mle}}| \leq s_3^{(0)}$. Let us set $\tilde{C} = (C_2|C_3) = Id_2$, the solution of the problem (2) with $\tilde{C}$ is $s_2^{(1)} = s_3^{(1)} = 2.23$. In the second step, the hypothesis $\mathcal{H}_i$ is rejected if $|\hat{\beta}_i^{\text{mle}}| > s_i^{(1)}$ with $i \in \{2, 3\}$. Let us assume that $\mathcal{H}_2$ and $\mathcal{H}_3$ are not rejected. Consequently, there is no false discovery in the second step; thus, $\hat{\beta}_2^{\text{mle}} \leq s_2^{(1)}$ and $\hat{\beta}_2^{\text{mle}} \leq s_3^{(1)}$. Finally, the probability of no false discovery is

$$\mathbb{P}(|\hat{\beta}_2^{\text{mle}}| \leq 2.10 \cap |\hat{\beta}_3^{\text{mle}}| \leq 2.23) < \mathbb{P}(|\hat{\beta}_2^{\text{mle}}| \leq 2.23 \cap |\hat{\beta}_2^{\text{mle}}| \leq 2.23) = 0.95.$$

Consequently, the FWER is not controlled at the significance level $\alpha = 0.05$. Note that, in this example, optimal thresholds are not decreasing in the sense that $(s_2^{(1)}, s_3^{(1)})$ is not per component smaller than $(s_2^{(0)}, s_3^{(0)})$. This fact is the reason why the application of the generic stepdown method to procedure 2 fails to control the FWER.

Acknowledgements

The authors are grateful for the real data provided by the following metabolomicians from Toxalim: Cécile Canlet, Laurent Debrauwer, and Marie Tremblay-Franco and are grateful to Holger

© Société Française de Statistique et Société Mathématique de France (2021) ISSN: 2102-6238

Rauhut for its careful reading. This work is part of the project GMO90+ supported by the grant CHORUS 2101240982 from the Ministry of Ecology, Sustainable Development and Energy in the national research program RiskOGM. Patrick Tardivel is partly supported by a PhD fellowship from GMO90+. We also received a grant for the project from the IDEX of Toulouse 'Transversalité 2014'.

References

Astle, W., De Iorio, M., Richardson, S., Stephens, D., and Ebbels, T. (2012). A bayesian model of NMR spectra for the deconvolution and quantification of metabolites in complex biological mixtures. *Journal of the American Statistical Association*, 107(500):1259–1271.

Dudoit, S. and Van Der Laan, M. J. (2007). *Multiple Testing Procedures with Applications to Genomics*. Springer.

Dunn, O. J. (1959). Confidence intervals for the means of dependent, normally distributed variables. *Journal of the American Statistical Association*, 54(287):613–621.

Dunn, O. J. (1961). Multiple comparisons among means. *Journal of the American Statistical Association*, 56(293):52–64.

Dunnett, C. W. (1955). A multiple comparison procedure for comparing several treatments with a control. *Journal of the American Statistical Association*, 50(272):1096–1121.

Fromont, M., Lerasle, M., and Reynaud-Bouret, P. (2016). Family-wise separation rates for multiple testing. *The Annals of Statistics*, 44(6):2533–2563.

Gronwall, T. H. (1919). Note on the derivatives with respect to a parameter of the solutions of a system of differential equations. *Annals of Mathematics*, pages 292–296.

Hao, J., Astle, W., De Iorio, M., and Ebbels, T. M. (2012). BATMAN - an R package for the automated quantification of metabolites from nuclear magnetic resonance spectra using a bayesian model. *Bioinformatics*, 28(15):2088–2090.

Holm, S. (1979). A simple sequentially rejective multiple test procedure. *Scandinavian Journal of Statistics*, 6(2):65–70.

Lefort, G., Liaubet, L., Canlet, C., Tardivel, P., Père, M.-C., Quesnel, H., Paris, A., Iannuccelli, N., Vialaneix, N., and Servien, R. (2019). Asics: an r package for a whole analysis workflow of 1d 1h nmr spectra. *Bioinformatics*, 35(21):4356–4363.

Lehmann, E., Romano, J. P., and Shaffer, J. P. (2012). On optimality of stepdown and stepup multiple test procedures. In *Selected Works of EL Lehmann*, pages 693–717. Springer.

Lehmann, E. L. and Romano, J. P. (2005). *Testing Statistical Hypotheses*. Springer Texts in Statistics. Springer, New York, third edition.

Pratt, J. W. (1961). Length of confidence intervals. *Journal of the American Statistical Association*, 56(295):549–567.

Ravanbakhsh, S., Liu, P., Bjordahl, T. C., Mandal, R., Grant, J. R., Wilson, M., Eisner, R., Sinelnikov, I., Hu, X., Luchinat, C., Greiner, R., and Wishart, D. S. (2015). Accurate, fully-automated NMR spectral profiling for metabolomics. *PLoS ONE*, 10(5):e0124219.

Romano, J. P., Shaikh, A., and Wolf, M. (2011). Consonance and the closure method in multiple testing. *The International Journal of Biostatistics*, 7(1).

Romano, J. P. and Wolf, M. (2005). Exact and approximate stepdown methods for multiple hypothesis testing. *Journal of the American Statistical Association*, 100(469):94–108.

Royen, T. (2014). A simple proof of the gaussian correlation conjecture extended to some multivariate gamma distributions. *Far East Journal of Theoretical Statistics*, 48(2):139–145.

Šidák, Z. (1967). Rectangular confidence regions for the means of multivariate normal distributions. *Journal of the American Statistical Association*, 62(318):626–633.

Sidak, Z. et al. (1971). On probabilities of rectangles in multivariate student distributions: their dependence on correlations. *The Annals of Mathematical Statistics*, 42(1):169–175.

Tardivel, P. J. C., Canlet, C., Lefort, G., Tremblay-Franco, M., Debrauwer, L., Concordet, D., and Servien, R. (2017). ASICS: an automatic method for identification and quantification of metabolites in complex 1D ^{1}H NMR spectra. *Metabolomics*, 13(10):109.

Tulpan, D., Léger, S., Belliveau, L., Culf, A., and Čuperlović-Culf, M. (2011). Metabohunter: an automatic approach for identification of metabolites from ^{1}H-NMR spectra of complex mixtures. *BMC Bioinformatics*, 12(1):400.

© Société Française de Statistique et Société Mathématique de France (2021) ISSN: 2102-6238

Weljie, A. M., Newton, J., Mercier, P., Carlson, E., and Slupsky, C. M. (2006). Targeted profiling: Quantitative analysis of ^{1}H-NMR metabolomics data. *Analytical Chemistry*, 78(13):4430–4442.

© Société Française de Statistique et Société Mathématique de France (2021) ISSN: 2102-6238