doi: 10.1111/j.1467-9469.2006.00530.x

© Board of the Foundation of the Scandinavian Journal of Statistics 2006. Published by Blackwell Publishing Ltd, 9600 Garsington Road, Oxford OX4 2DQ, UK and 350 Main Street, Malden, MA 02148, USA Vol 34: 275–297, 2007

Some Results on the Control of the False Discovery Rate under Dependence

ALESSIO FARCOMENI

Dipartimento di Statistica, Probabilità e Statistiche Applicate, Università di Roma 'La Sapienza'

ABSTRACT. Controlling the false discovery rate (FDR) is a powerful approach to multiple testing, with procedures developed with applications in many areas. Dependence among the test statistics is a common problem, and many attempts have been made to extend the procedures. In this paper, we show that a certain degree of dependence is allowed among the test statistics, when the number of tests is large, with no need for any correction. We then suggest a way to conservatively estimate the proportion of false nulls, both under dependence and independence, and discuss the advantages of using such estimators when controlling the FDR.

Key words: dependence, false discovery rate, mixture model, multiple testing, time course DNA microarray

1. Introduction and motivation

Multiple testing methods that control the false discovery rate (FDR) have recently received much attention, following the seminal paper of Benjamini & Hochberg (1995). The main advantage of controlling the FDR in multiple testing is that this allows to achieve good power even when testing many (often thousands of) hypotheses. This happens in many modern applications, for instance, in bioinformatics.

While procedures controlling the FDR were designed initially for independent test statistics, it is actually more common to work with dependent test statistics. For instance, in DNA microarray analysis, the test statistics are a function of gene expression levels, and genes are dependent by their intrinsic nature.

Many efforts have been made to generalize the methods under dependence (Yekutieli & Benjamini, 1999; Benjamini & Yekutieli, 2001; Pollard & van der Laan, 2002; Sarkar, 2002; Storey, 2003, etc.); by verifying procedures developed under independence of test statistics to work under certain dependence conditions, estimation of the joint distribution, resampling methods.

In this paper, we argue that a certain degree of dependence is allowed among the test statistics, when the number of tests is large, with no need for any correction to the standard methods. We also provide asymptotic distributional results for the quantities of interest.

More in detail, we will consider the case of totally ordered test statistics, and show that the FDR is asymptotically (with the number of tests) controlled under conditions of weak dependence. Applications that possibly satisfy such assumptions include financial and biological time series, and problems in change-point modelling (e.g. in sequential analysis). Moreover, we show simulations of the case of weakly dependent test statistics that are not totally ordered, and see that still a correction is not needed. We will use our results to suggest a direct and new approach to microarray time-course data analysis, from which motivation for this work arise. Time course microarray data arise when repeated measures on gene expressions are taken over time; and the interest is in the time-dependent behaviour of the genes.

Secondly, we will discuss suitable estimators for the proportion of false nulls among the hypotheses, both under dependence and independence. Such estimators can be used to boost power of many multiple testing procedures. As we will point out, they also are of interest in conservatively estimating the weight of an unknown component in a mixture model; with applications in cosmology.

This paper is organized as follows: in section 1.1 we will review some background on FDR and multiple testing in general. In section 2, we will review the main assumptions made on the dependence among the test statistics. Section 3 provides asymptotic distributional results on the quantities of interest, and gives sufficient conditions on the dependence that lead to asymptotic control of the FDR if the test statistics are totally ordered. To actually do this, a robust estimator for the number of false null hypotheses is needed. We suggest two estimators that seem to be satisfactory in section 4.

1.1. Background

Consider a multiple testing situation in which m tests are being performed. Suppose M_0 of the m hypotheses are true, and M_1 are false. Table 1 shows the possible outcomes. R denotes the number of rejections. $N_{i|j}$, with $i,j \in \{0,1\}$, is the number of H_i accepted when H_j is true. Note that $N_{0|1}$ and $N_{1|0}$ give the actual number of errors, respectively, the number of Type I (false-positive) and II (false-negative) errors.

In the single test setting, one controls the probability of a false rejection. In the multiple tests setting, this is likely to lead to an explosion of the number of false-positive errors. Attention should be shifted from the single test to the performance on the whole set of tests, by bounding a function of $N_{1|0}$, the number of false-positive errors. Such functions are referred to as *Type I error rates*. A variety of possible Type I error rates have been proposed (see, e.g. Benjamini & Hochberg, 1995; Dudoit *et al.*, 2004; Lehmann & Romano, 2005; Sarkar, 2005). From a practical point of view, one still wants to get a sufficiently small $N_{1|0}$ while rejecting the maximum number of hypotheses, thus minimizing the number of false-negative errors.

A popular error measure is the FDR, proposed in Benjamini & Hochberg (1995). Define the false discovery proportion (FDP) to be the proportion of erroneously rejected hypotheses:

$$\mathrm{FDP} = \begin{cases} \frac{N_{1|0}}{R} & \text{if } R > 0 \\ 0 & \text{if } R = 0. \end{cases} \tag{1}$$

Benjamini & Hochberg (1995) define the FDR as:

$$\mathrm{FDR} = E(\mathrm{FDP}). \tag{2}$$

False discovery rate is motivated by modern applications, in which the number of tests is very large; and justified by the idea that any researcher is prepared to accept a high number of false rejections when a high number of rejections is made. This leads to a more liberal control on the number of false positives, and a much lower number of false negatives; with respect to procedures controlling more classical Type I error rates like the family-wise error rate (FWER) $(\Pr(N_{1|0} > 0)$. The FDR control provides also a weak control of the FWER, in the sense that if all the null hypotheses are true, FDR and FWER are equal.

Table 1. *Outcomes in testing m hypotheses*

	H_0 not rejected	H_0 rejected	Total
H_0 true	$N_{0\|0}$	$N_{1\|0}$	M_0
H_0 false	$N_{0\|1}$	$N_{1\|1}$	M_1
Total	$m-R$	R	m

© Board of the Foundation of the Scandinavian Journal of Statistics 2006.

It is common in the multiple testing literature to refer to testing of a simple hypothesis by means of significance levels, i.e. p-values. After the p-values are ordered, the problem of controlling any error measure reduces to fixing a cut-off T such that the error rate is at most equal to a pre-specified $\alpha \in [0, 1]$ when rejecting all the hypotheses corresponding to $p_j \leq T$. To simplify the exposure, we will say 'reject p_j such that ...' to mean 'reject the null hypotheses for which p_j is such that ...'.

Let the unknown proportion of false nulls M_1/m be denoted by a. We review now two procedures to control the FDR:

BH: reject $p_j < T_{\text{BH}}$, where T_{BH} is $\sup\{t : \hat{G}(t) = t/\alpha\}$, where α is the desired upper bound for the FDR and $\hat{G}(t) = 1/m \sum 1_{p_j < t}$ is the empirical distribution of the p-values.

Plug-in: Reject $p_j < T_{\text{PI}}$, where T_{PI} is $\sup\{t : \hat{G}(t) = (1 - \widehat{a})t/\alpha\}$, and $\hat{a}$ is any estimator of a.

The BH procedure was proposed in Benjamini & Hochberg (1995) The *plug-in* procedure was proposed in Genovese & Wasserman (2002) and independently in Benjamini & Hochberg (2000). Genovese & Wasserman (2002) show that it asymptotically controls the FDR and is more powerful than the BH procedure. An estimator for a is proposed in Storey (2002), and defined as:

$$\hat{a} = \frac{\hat{G}(t_0) - t_0}{1 - t_0} \tag{3}$$

for some $t_0 \in (0, 1)$. For a discussion of the role of $\hat{a}$ and other possible estimators, see section 4.

False discovery rate controlling procedures possess desirable asymptotic properties, in the sense that with mild conditions on the distribution of the p-values under the alternative hypothesis it can be proved that the number of rejections is asymptotically strictly bounded below by zero, if there exists at least one false null (Genovese & Wasserman, 2002).

Note that in multiple testing literature it is natural to think about asymptotics is in m, i.e. to have an increasing number of tests (and possibly fixed number of observations) (see, e.g. Finner & Roters, 2002, and references therein).

BH and plug-in, as many other procedures, are distribution free. They provide Type I error rate (namely, FDR) control without any assumption on the unknown distribution of p_j when H_0 is false. They only rely on the assumption of independence and the fact that $p_j \mid H_0$ is uniformly distributed on the interval $[0, 1]$.

Genovese & Wasserman (2002) and independently Sarkar (2004) also introduce the false-negatives rate (FNR), a Type II error rate defined as the dual of the FDR.

Benjamini & Hochberg (1997) and Genovese *et al.* (2007) describe algorithms that allow to control the FDR when a weight is assigned to each null hypothesis. This allows to control the propensity of rejection for certain hypotheses, and is useful if one has prior information on them. For instance, it is well known that in functional magnetic resonance imaging experiments false nulls are likely to be clustered. Genovese & Wasserman (2004b) put FDR control under the stochastic process framework we will exploit in this paper, introduce estimators for a, suggest ways to build confidence thresholds for the FDP and prove asymptotic results, providing some limiting distributions. In this paper, we show how such limiting distributions change under dependence. We also provide the asymptotic distribution of the FDP for the BH and plug-in procedure, which was *not* derived in Genovese & Wasserman (2004b), and to our knowledge is a novel result also under independence.

1.1.1. Dependent test statistics

Many authors deal with FDR control under dependence of the test statistics. Benjamini & Yekutieli (2001), prove that the BH procedure can never control the FDR at level higher than

$\alpha\sum_{i=1}^{m} 1/i$. Hence, taking into account a factor of $\sum_{i=1}^{m} 1/i$ will lead to control of the FDR under arbitrary dependence. Note that this is rather conservative. They also prove that, under conditions of positive regression dependency on the subset (PRDS) of true nulls, the BH procedure still controls the FDR at level α. The condition of PRDS introduced in Benjamini & Yekutieli (2001) is as follows: for any increasing set D and for each i in the set of true nulls, let $\Pr(X \in D \mid X_i = x)$ be non-decreasing in x, where X is the vector of m random variables associate with the test statistics. Recall that a set is said to be increasing if for any $x \in D$ and $y \geq x$, $y \in D$. Distributions satisfying this property include multivariate normal distributions with positive correlations, all unidimensional latent variable distributions, and few other cases. Sarkar (2002) also extends the results of Benjamini & Yekutieli (2001) by generalizing their results to a whole class of step-up/step-down procedures to control the FDR.

The conditions of weak dependence we will give are not completely overlapping with the PRDS condition.

Storey *et al.* (2004b) show several theorems that require almost sure pointwise convergence of the empirical distributions of the subsequence of p-values for which the null is true and the subsequence of p-values for which the alternative hypothesis is true. They argue that this may be true also under dependence; and in fact Bickel (2004) shows a process with long-range correlations that satisfies the conditions of Storey *et al.* (2004b).

Yekutieli & Benjamini (1999) and Pollard & van der Lann (2002) propose resampling-based procedures to control the FDR when the test statistics are correlated.

2. Assumptions on the dependence

In this section, we summarize the main assumptions we will make on the dependence throughout the paper. In the first part, we will use some of such concepts to prove weak convergence of opportune random sequences, which will imply asymptotic FDR control. For a detailed discussion on the convergence of dependent random sequences, for instance, see Wu (2004).

2.1. Association

Definition 1 (Association)
A vector of random variables $X_1, \ldots, X_n$ is associated if $\operatorname{cov}[g_1(X_1, \ldots, X_n), g_2(X_1, \ldots, X_n)] \geq 0$, *for all monotonically coordinate-wise non-decreasing functions g_1 and g_2.*

Positive association is introduced in Esary *et al.* (1967). Kumar & Proschan (1983) introduce the dual concept, negative association, simply putting the reverse inequality in the definition. A review of the main ideas and examples are given in Tong (1980).

Main properties of (positively) associated random variables are:

$$P\left(\bigcap_{i=1}^{n}\{X_i \leq z_i\}\right) \geq \Pi P(X_i \leq z_i)$$

and similarly

$$P\left(\bigcap_{i=1}^{n}\{X_i > z_i\}\right) \geq \Pi P(X_i > z_i), \qquad \text{for } z_i \in R, \quad i = 1, \ldots, n.$$

Moreover, it is straightforward to prove that $E[\prod X_i] \geq \prod E[X_i]$. If $n = 2$ the statement follows from the definition. If $n > 2$, suppose the statement is true for $n-1$ random variables. Then,

© Board of the Foundation of the Scandinavian Journal of Statistics 2006.

$$E\left[\prod X_i\right] = \operatorname{cov}\left(X_n, \prod_{i=1}^{n-1} X_i\right) + E[X_n]E\left[\prod_{i=1}^{n-1} X_i\right]$$
$$\geq E[X_n]E\left[\prod_{i=1}^{n-1} X_i\right] \geq \prod E[X_i],$$

where we used the inductive hypothesis at the last step. If the random variables are negatively associated, all the inequalities hold reversed.

As stated in Benjamini & Yekutieli (2001), PRDS condition is very similar but not completely overlapping with association. For a review of many such general conditions on dependence, see Lehmann (1966).

Multivariate normal random variables with all positive correlations are associated. Independent random variables are both positively and negatively associated.

Multivariate exponential random variables, as defined in Marshall & Olkin (1967), are always associated.

Multivariate normal random variables are negatively associated when the extra diagonal elements of the variance/covariance matrix are all non-positive. Multinomial, multivariate hypergeometric, Dirichlet random variables are always negatively associated (for other examples, refer to Kumar & Proschan, 1983).

Other applications include testing on the parameters of monotone latent variable models: suppose X is assumed to be the marginal distribution of some (X, U), with the elements of X conditionally independent on $U = u$ and stochastically monotone in u. Then positive association holds (Sarkar & Chang, 1997).

A similar setting is in multivariate modelling with random effects, in which there is one underlying common random effect inducing a positive association (see Alfó & Trovato, 2004, and references therein).

Cases of interest in which multiple testing procedures valid under negative association should be used also include multiple tests on ranks and any case in which the joint distribution of the test statistics is a permutation distribution (Kumar & Proschan, 1983).

2.2. Mixing

The concept of (unidimensional) mixing is used to describe how fast the dependence between random variables in a sequence decreases to zero as the *lag* between them increases.

Main references on mixing are Doukan (1994) and Billingsley (1999). Many different notions of mixing (α-, β-, ρ-mixing) are reviewed in Doukan (1994). We will use conditions only on the α-mixing coefficients:

Definition 2 (α-mixing)
The kth α-mixing coefficient is defined as:

$$\alpha(k) = \sup_j \{ |P(E_1)P(E_2) - P(E_1 \cap E_2)| : \\ E_1 \in \mathcal{M}_1^j, E_2 \in \mathcal{M}_{j+k}^{+\infty} \}; \tag{4}$$

where $\mathcal{M}_i^j$ is the $\sigma-$algebra generated by $\{p_i, \ldots, p_j\}$.

Conditions are in the form $\alpha(k) \to 0$, usually at a specific rate (see Arcones & Yu, 2000), to provide limit theorems.

© Board of the Foundation of the Scandinavian Journal of Statistics 2006.

Mixing is a general condition on dependence, often used in time-series analysis. Mixing conditions are usually implied by less general, more easily proved, conditions; like m-dependence. A sequence of random variables is m-dependent of order h when X_i is independent of X_{i+h} for any i. For other equivalent conditions, see Bradley (1993).

3. Asymptotic control of the FDR under dependence

Let $\{p_i\}_{i \in \mathcal{N}}$ be a random sequence of p-values from tests: in this section, we assume that the test statistics are totally ordered, and we denote with H_i the indicator of the ith hypothesis to be false. Let $\Pr(p_i < t \mid H_i = 1) \sim F(t)$, where $F(t) \neq t$.

We show that what is needed to obtain asymptotic control is just a little more than the convergence of the empirical process of the p-values, so we will make use of many conditions on the dependence among the random variables under which it is known that such process converges. We denote with U the uniform distribution, and assume $p_j \mid H_j = 0 \sim U$, while $p_j \mid H_j = 1 \sim F$, where F is arbitrary on [0, 1]; and $p_j \sim G$ marginally.

We show that under such conditions for the sequence of p-values, the plug-in procedure (with a good estimator of a) is able to control the FDR at the desired level α. This is a generalization of some of the results of Genovese & Wasserman (2004b). We will assume for the time being that M_1/m converges to a.

We will prove our main results under any of the following conditions:

1. If the p-values are independent (Genovese & Wasserman, 2004b).
2. If $\alpha(k)$ are the mixing coefficients of the p-values, there exists $\delta > 0$ such that $\alpha(k) \leq Ck^{-3-\delta}$ for some constant C, and the vector $(p_j)_{j \geq 1}$ is stationary.
3. If $(p_j)_{j \geq 1}$ is stationary, associated and $\sum_k k^{13/2+\delta} \mathrm{cov}(p_1, p_k) < +\infty$ for some $\delta > 0$.
4. If $(p_j)_{j \geq 1}$ is stationary and $\sum_k \alpha(k) < +\infty$.
5. If $(p_j)_{j \geq 1}$ is stationary, associated and
$$\sum_{k=2}^{+\infty} [P(p_1 \leq s, p_k \leq t) - G(s)G(t)] < +\infty.$$
If $M_0 = m$ (as in the notation of Table 1), Yu (1993) proves in a different setting that this is equivalent to $\sum \mathrm{cov}^{1/3}(p_1, p_k) < +\infty$.
6. If the test statistics $(T(X_j))_{j \geq 1}$ form a stationary sequence of normal random variables with short-range dependence, i.e. such that $\sum |\mathrm{cov}(T(X_1), T(X_n))| < \infty$ or equivalently if the corresponding p-values are such that
$$\sum_{k=2}^{+\infty} |P(p_1 \leq t, p_k \leq t) - G^2(t)| < +\infty \qquad \text{for any } t \in (0, 1).$$

Note that condition 4 is more general than condition 2. Note also that, in case the test statistics are actually normal, condition 5 is exactly equivalent to condition 6 under association. A further generalization would be given by considering long-range dependence. This would require a completely different approach, and is then grounds for possible further work.

First, note that the FDP can be seen as a stochastic process indexed by the threshold t:

$$\Gamma(t) = \frac{\sum (1 - H_i) 1_{\{p_i < t\}}}{\sum 1_{\{p_i < t\}} + \prod (1 - 1_{\{p_i < t\}})}, \tag{5}$$

note in fact that $(1 - H_i)1_{\{p_i < t\}}$ is one if and only if $H_i = 0$ and $p_i < t$, i.e. if the ith hypothesis has been rejected while being true. $\Gamma(t)$ is a stochastic process as, for each fixed $t \in [0, 1]$, $\Gamma(t)$ is a random variable. If t is a fixed cut-off (i.e. we actually reject $p_j < t$), then $\Gamma(t)$ is the realized FDP, the proportion of false rejections.

© Board of the Foundation of the Scandinavian Journal of Statistics 2006.

As also Genovese & Wasserman (2004b) note, the cut-offs T are usually random, which implies the non-trivial problem of evaluating a stochastic process at a random point:

> One of the essential difficulties in studying a procedure T is that $\Gamma(T)$ is the evaluation of the stochastic process $\Gamma(\cdot)$ at a random variable T. Both depend on the observed data, and in general they are correlated. In particular, if $\hat{Q}(t)$ estimates $\text{FDR}(t)$ for each fixed t, it does not follow that $\hat{Q}(T)$ estimates well $\text{FDR}(T)$ at a random T. The stochastic process point of view provides a suitable framework for addressing this problem.

Technical details related to this problem are described in the proofs in appendix, when needed. We will take the approach used by Genovese & Wasserman (2004b) at first, and then directly work with the stochastic process evaluated at the random point. This will provide an easier proof of the results of Genovese & Wasserman (2004b), together with the asymptotic distribution of $\Gamma(T)$ under independence, which has not been derived yet. We will use this approach also to provide the asymptotic distribution of $\Gamma(T)$ under assumptions (2–6).

We are now ready to state our main result: if the dependence decreases fast enough, then the BH and plug-in procedures remain valid without any modification. First, we prove FDR is controlled if the quantity a is known. Then we extend the results to the case in which a conservative estimator is used.

Theorem 1

Let $\{p_i\}_{i\in\mathcal{N}}$ be a random sequence of p-values from tests. Let H_i be the indicator of the ith hypothesis to be false. Let $\Pr(p_i < t \mid H_i = 1) \sim F(t)$, where $F(t) \neq t$. Assume any of the specified conditions (1–6) holds. Assume the quantity a is known.

Then, $E[\Gamma(T_{PI})] = \alpha + o(1)$, where $\Gamma(t)$ is the FDP for threshold t and T_{PI} is the plug-in threshold with $\hat{a} := a$.

The results are distribution free: they do not depend on the true $F(\cdot)$. Note that the assumption that the p-values under the alternative are all equally distributed is a condition that is unlikely to be ever satisfied. Nevertheless, as $F(\cdot)$ is almost completely unspecified one may think of $F(\cdot)$ as an alternative-specific mixture, like: $F(t) = \int \Pr(p_k < t \mid H_k = 1, \theta)\, \mathrm{d}M(\theta)$. In this sense each $p_k \mid H_k = 1$ has a specific distribution, sharing a functional form with the other alternative p-values, but with a subject-specific parameter vector (e.g. the mean and variance) sampled from a certain, unspecified, $M(\theta)$. This is likely to be satisfied in many real applications, and parallels the idea of hypothesis-specific mixture used to define $G(t)$.

Note that in almost all cases $F(t)$ will be stochastically smaller than $U[0, 1]$. If it was not so, it would be more likely to observe high p-values under the alternative than under the null. Nevertheless, the theorem is true for any $F \neq U$. In case U is stochastically smaller than F, there still is FDR control; even if the number of rejections will very likely be equal to zero.

Note moreover that our results are valid only as the number of tests grows; while the results of Benjamini & Yekutieli (2001) are valid for any number of tests, and that the conditions imposed are slightly different.

For a proof of the theorem refer to the appendix A. The proof shows also our distributional results. We summarize two of them in the following corollary.

Corollary 1

Let $Q(t) = (1 - a)t/G(t)$ and $\rho_{ij}(k) = \Pr(p_1 < s, p_k < t \mid H_1 = i, H_k = j)$. We have that: $\sqrt{m}(\Gamma(t) - Q(t))$ converges weakly to a centred Gaussian process with covariance kernel $K(s, t)$ given by:

© Board of the Foundation of the Scandinavian Journal of Statistics 2006.

$$
\begin{aligned}
&\frac{a(1-a)}{G^2(s)G^2(t)}[(1-a)stF(s\wedge t)+aF(s)F(t)(s\wedge t)] \\
&+\frac{2}{G^2(s)G^2(t)}[a^2F(s)F(t)\sum_{k=2}^{+\infty}[\rho_{00}(k)-(1-a)^2st] \\
&-a(1-a)sF(t)\sum_{k=2}^{+\infty}[\rho_{01}(k)-(1-a)saF(t)] \\
&-a(1-a)tF(s)\sum_{k=2}^{+\infty}[\rho_{10}(k)-(1-a)taF(s)] \\
&+(1-a)^2st\sum_{k=2}^{+\infty}[\rho_{11}(k)-a^2F(s)F(t)]. \qquad (6)
\end{aligned}
$$

This implies that $\Gamma(t)$ converges in probability to $Q(t)$. Moreover, $\sqrt{m}(T_{PI}-Q^{-1}(\alpha))$ converges in distribution to an $N(0,V)$, with

$$
\begin{aligned}
V=&\frac{\alpha^2(G(Q^{-1}(\alpha))-G^2(Q^{-1}(\alpha))}{G^2(Q^{-1}(\alpha))(Q'(Q^{-1}(\alpha)))^2} \\
&+\frac{2\alpha^2\sum_{k=2}^{+\infty}(P(p_1<Q^{-1}(\alpha),p_k<Q^{-1}(\alpha))-G^2(Q^{-1}(\alpha)))}{G^2(Q^{-1}(\alpha))(Q'(Q^{-1}(\alpha)))^2}.
\end{aligned}
$$

This implies that T_{PI}, the deciding point, converges in probability to $Q^{-1}(\alpha)$.

The convergence of T_{PI} to $Q^{-1}(\alpha)$ is particularly interesting, as the latter is the deciding point we would set if we exactly knew the marginal distribution $G(\cdot)$ and the quantity a.

We now prove that the plug-in method is asymptotically valid under conditions (1–6):

Corollary 2

Let $a_0\in[0,a]$ and let $\hat{a}$ be a consistent estimator of a_0. Let the conditions in Theorem 1 hold. Then the plug-in method will asymptotically control the FDR at α level. If $\hat{a}:=0$, so that $a_0:=0$, then the result holds for the BH method.

Proof. Proof will follow from theorem 1 and same reasoning as Theorem 5.2 in Genovese & Wasserman (2004b).

Note that if $a_0=0$, then the results hold for the BH method.

We are now ready to provide the asymptotic distribution of the stochastic process $\Gamma(t)$ evaluated at the random point T. That is, we provide the distribution of the FDP obtained for the BH and plug-in procedure, under independence (condition 1) and dependence (conditions 2–6). A proof is given in appendix C.

Theorem 2

Let D be a fixed point in $(0,1)$ and suppose T is any random point in $(0,1)$ such that $T\xrightarrow{P}D$. Suppose any of the assumptions (1–6) holds. Then, $\sqrt{m}(\Gamma(T)-Q(D))$ converges in distribution to an $N(0,K(D,D))$, where $K(s,t)$ is defined in (6) and $Q(t)=(1-a)t/G(t)$.

The previous theorem is of interest for proving asymptotic control of the FDR for any multiple testing procedure. What is needed is just the convergence of the deciding point T

© Board of the Foundation of the Scandinavian Journal of Statistics 2006.

and conditions (1–6). Moreover, the results can be used to provide asymptotic confidence intervals for the error measure under the assumptions (1–6), as long as a suitable estimate of $K(D, D)$ is provided. This can be carried out through an estimate of the distribution of p_j when H_0 is false, like the one proposed in Genovese & Wasserman (2004b).

Theorem 3

Suppose a is known, and T_{PI} is the plug-in threshold with $\hat{a} := a$. Suppose any of the assumptions (1–6) holds. Then, $\sqrt{m}(\Gamma(T) - \alpha)$ converges in distribution to an $N(0, K(Q^{-1}(\alpha), Q^{-1}(\alpha)))$, where $K(s, t)$ is defined in (6).

Proof. From theorem 1 we know that $T_{\mathrm{PI}} \xrightarrow{P} Q^{-1}(\alpha)$. We can now apply theorem 2 to obtain the thesis.

By looking at appendix A, it can be seen that $\Gamma(T)$ and $Q(T)$, rescaled, converge to normal random variables with the same mean but different variance. It is straightforward to provide asymptotic distributional results for the FDR of the plug-in procedure applied with any consistent estimator of $a_0 \in [0, a]$, and those are omitted for brevity. In all cases, as seen, asymptotic control of the FDR is achieved.

3.1. General weak dependence assumptions

Doukhan & Louhichi (1999) and then Nze *et al.* (2002) define a more general framework for weak dependence, which includes processes that satisfy mixing and association conditions, together with cases in which these two properties fail to hold, like Bernoulli shifts driven by discrete innovations. They define the set $\mathcal{L}_1 = \{h : h \text{ is Lipschitz}, |h||_\infty \leq 1\}$, and they define a weakly dependent sequence $\{X_n\}_{n \in \mathcal{N}}$ to satisfy

$$|\operatorname{cov}(h(X_{i_1}, \ldots, X_{i_u}), k(X_{j_1}, \ldots, X_{j_v})| \leq \theta_r \psi_i(h, k, u, v), \quad i = 1, 2,$$

where k and h are in $\mathcal{L}_1$, θ_r is a sequence of numbers decreasing to zero, $r = j_1 - i_u$, and $\psi_1(h, k, u, v) = \max(\mathrm{Lip}(h), \mathrm{Lip}(k))(u + v)$, $\psi_2(h, k, u, v) = \mathrm{Lip}(h)\mathrm{Lip}(k) \min(u, v)$; where $\mathrm{Lip}(h)$ is the Lipschitz dimension of h. It is apparent that this is a definition similar to the one of mixing processes.

Conditions of weak dependence can be given for the same results of theorem 1 to hold:

Theorem 4

If ψ_1 function is used and $\theta_r = O(r^{-5-v})$ or ψ_2 function is used and $\theta_r = O(r^{-15/2-v})$, then the empirical process weakly convergences in $D[0, 1]$ *to a centred Gaussian process indexed in* $[0, 1]$. *The covariance kernel of $\sqrt{n}(\widehat{G}(t) - G(t))$ is*

$$K_G(s, t) = 2 \sum_{k=0}^{+\infty} \operatorname{cov}(1_{p_1 \leq s}, 1_{p_k \leq t}).$$

Generalization of the covariance kernels of the quantities of interest is straightforward and omitted for brevity. A proof is given in the appendix B.

Note that the covariance kernel $K_{\mathrm{G}}(s, t)$ reduces to the corresponding covariance kernel defined in appendix A under conditions (1–6).

© Board of the Foundation of the Scandinavian Journal of Statistics 2006.

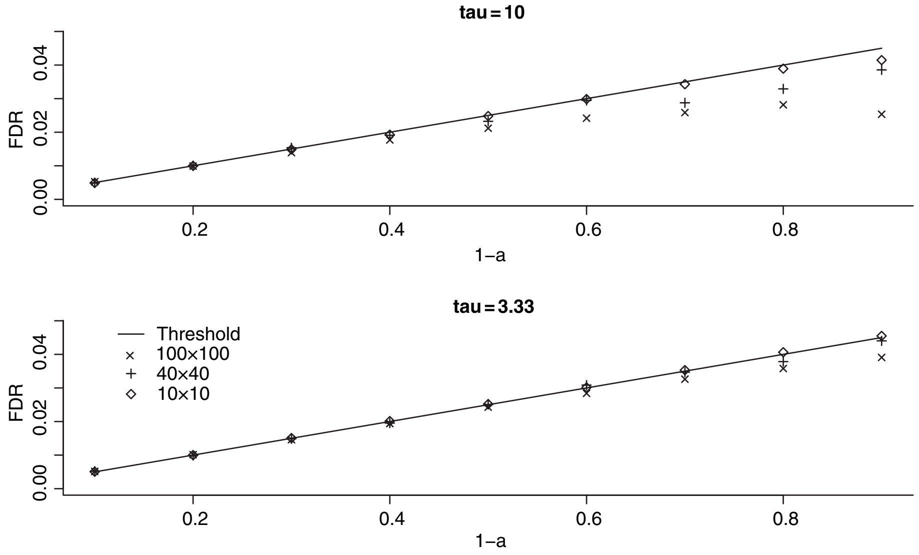

Fig. 1. False discovery rate for BH method, normal random variables with $\mathrm{cov}(X_{i,j}, X_{i',j'}) = \exp(-\frac{1}{\tau}\, d((i,j),(i',j')))$, 1000 iterations, a random proportion of $1-a$ random variables have a zero mean, the remaining have a non-zero mean sampled from uniform on $(0, 5)$.

3.2. Simulations

In this section, we will simulate a case of normal random variables which are *not* totally ordered, and see that the results of theorem 1 still hold.

We will now suppose to have normal random variables $X_{11}, \ldots, X_{nn}$, scattered on a quadratic grid. We will use a simplified version of a kernel commonly used in spatial statistics:

$$\mathrm{cov}(X_{ij}, X_{i'j'}) = \exp\left(-\frac{1}{\tau} d((i,j),(i',j'))\right), \tag{7}$$

where $d(\cdot,\cdot)$ is the opportune euclidean distance function and τ is just a tuning parameter. The higher τ, the more slowly decaying the correlation. We will then assign a proportion of a test statistics to a mean generated as a random uniform between zero and five, and the rest to a zero mean.

Figure 1 shows the average FDR obtained by applying the BH method. Figure 2 shows the average FDR obtained using the plug-in method. In both figures and all the following the ‘threshold’ is the nominal level at which the FDR is to be controlled, which is always $(1-a)\alpha$ if the BH method is used, and α if the plug-in method is used. In all cases α was taken equal to 0.05.

As long as the parameter τ is small, the methods control sharply the FDR (as in the independent case). When the parameter τ gets too big, dependence does not decrease fast enough and BH method becomes overly conservative, while the plug-in method violates the threshold. This behaviour is more and more evident as τ increases (other simulations not shown).

In a certain sense, it is happening what the theorem states: if the dependence is weak enough (compared with m), then the procedures will not be affected. When the dependence becomes stronger (compared with m: the violation of the threshold is more critical for smaller m), something can happen.

© Board of the Foundation of the Scandinavian Journal of Statistics 2006.

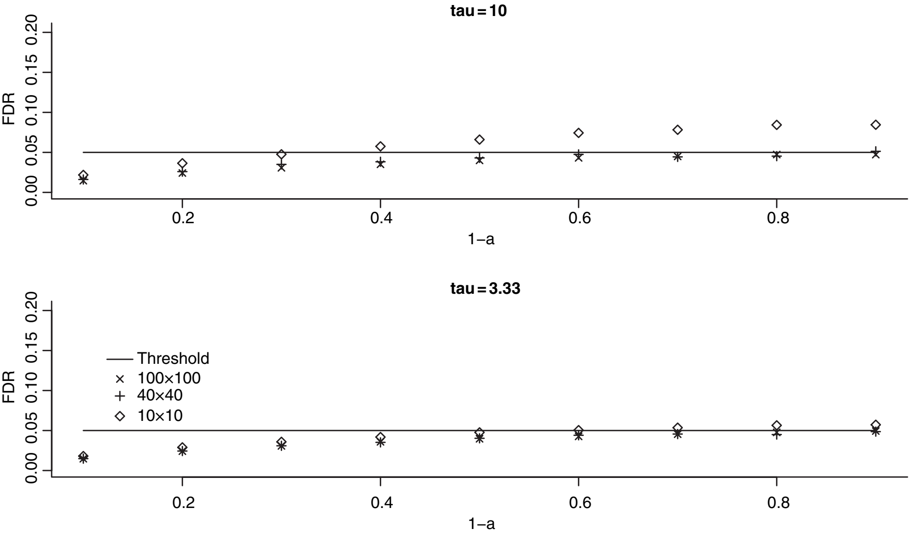

Fig. 2. False discovery rate for plug-in method, normal random variables with $\mathrm{cov}(X_{i,j}, X_{i',j'}) = \exp(-\frac{1}{\tau}\ d((i,j),(i',j')))$, 1000 iterations, a random proportion of $1-a$ random variables have a zero mean, the remaining have a non-zero mean sampled from uniform on $(0,5)$. Storey's estimator used for a.

Although, it is promising that the effect on BH is to overprotect against violation of the threshold. This means that the hypotheses of theorem 1 are only sufficient. Note that all the simulated random variables satisfy the PRDS condition of Benjamini & Yekutieli (2001).

On the other hand, the plug-in method violates the threshold and gets bigger than α when τ is high. It can be shown that this problem is determined by the estimator of a used. The other classical estimators, like the one in Swanepoel (1999) or the one in Woodroofe & Sun (1999), are also seen to break down under dependence. Simulations of the other estimators are not shown for reasons of space. We will propose in the next section estimators for a that do not seem to break down under dependence, and show they lead to control of the FDR.

Moreover, simulations suggest that the asymptotic results, for mild dependence, are achieved for small values of m ($m=100$ seems to be, in fact, enough).

Finally, we simulate the normal random variables $X_{11},\ldots,X_{nn}$, with a different covariance kernel, defined as:

$$\mathrm{cov}(X_{ij}, X_{i'j'}) = \exp\left(-\frac{1}{\tau}d((i,j),(i',j'))\right)\cos\left(\frac{1}{\tau}\ d((i,j),(i',j'))\right). \tag{8}$$

Note that such random variables are *not* PRDS. Figures 3 and 4 show the average FDR obtained applying the BH and plug-in methods. Even if the test statistics are not PRDS, the same considerations as before can be given.

In the previous section, we proved that under wide hypotheses on the dependence of the test statistics the BH method remains valid. This section gives an illustration of this behaviour through simulations. On the other hand, the same results were proved for the plug-in procedure with a suitable estimator for a. The simulations show that the common estimator used for a is not robust under dependence. An explicit proposal for a suitable estimator will be given in section 4.

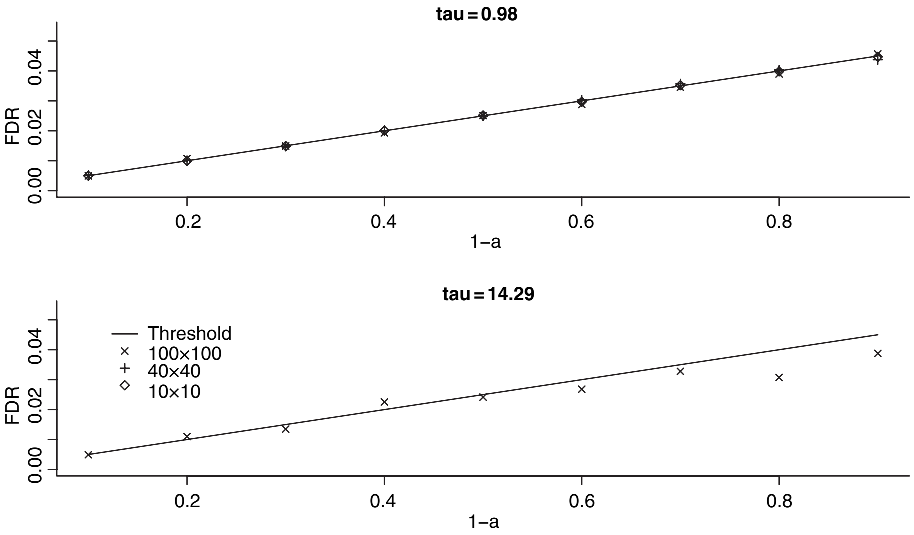

Fig. 3. False discovery rate for BH method, normal random variables with $\mathrm{cor}(X_{i,j}, X_{i',j'}) = \exp(-\frac{1}{\tau}\ d((i,j),(i',j'))\cos(\frac{1}{\tau}\ d((i,j),(i',j'))))$, 1000 iterations, a random proportion of $1-a$ random variables have a zero mean, the remaining have a non-zero mean sampled from uniform on $(0, 5)$.

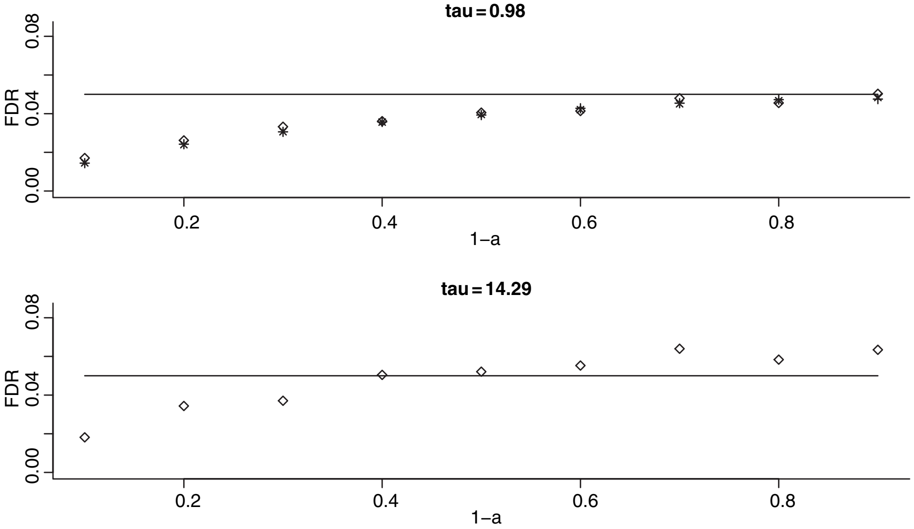

Fig. 4. False discovery rate for plug-in method, normal random variables with $\mathrm{cor}(X_{i,j}, X_{i',j'}) = \exp(-\frac{1}{\tau}\ d((i,j),(i',j'))\cos(\frac{1}{\tau}\ d((i,j),(i',j'))))$, 1000 iterations, a random proportion of $1-a$ random variables have a zero mean, the remaining have a non-zero mean sampled from uniform on $(0, 5)$. Storey's estimator used for a.

3.3. Applications

Totally ordered test statistics satisfying any of the assumptions (1–6) arise in general when applying multiple testing methods to stationary time series. We define a vector to be totally ordered whenever there is a natural total ordering of its elements, that is, whenever any

© Board of the Foundation of the Scandinavian Journal of Statistics 2006.

two elements can be put in a relationship of ordering. For instance, test statistics that arise through time are totally ordered (they arise one after another) while statistics arising through space are only partially ordered. Often there is no ordering of the vector of test statistics (e.g. in DNA microarrays, one test statistic arises for each gene and there is no natural ordering of the genes).

Assumptions (1–6) are implied by other (more strict) conditions, like m-dependence. Gaussian processes with covariance decreasing to 0 satisfy them, together with strictly stationary autoregressive integrated moving average models, block-dependent random variables with bounded block dimension.

Other cases in which any of the assumptions (1–6) may be valid are in cases in which there is an ‘increasing domain asymptotics’, a common concept in spatial statistics. In such cases, it may be safe to assume that the dependence among test statistics getting further and further fades fast enough.

Among the other possibilities, there are outlier detection and change-point tests in time series. We briefly describe now a direct application to time-course microarray data.

DNA microarrays are a recent technology that allow to measure the expression levels of thousands of genes at a time. The primary goal is usually to identify genes that are differentially (over or under) expressed among two or more biological conditions. The main output of a microarray experiment is a list of genes that are suspected to be related with the biological condition of interest. There are many detailed reviews of the methodology and problems: Amaratunga & Cabrera (2004), Parmigiani *et al.* (2003), Brown & Botstein (1999), Duggan *et al.* (1999). Dudoit *et al.* (2003) discuss the use of multiple testing methods in this context. DNA time-course microarrays are a specific kind of microarray experiments, in which the expression level of the genes is monitored over time. Thereby, the recorded expression levels of a single gene form a time series. Interest may be in finding periodicity in the behaviour of the genes (as in Whitfield *et al.*, 2002), or in finding which genes at which time points are differentially expressed. We will focus on this second task. For simplicity, here we assume without loss of generality that each of a set of G genes is analysed at T time points. DNA time-course data are usually treated with a modelling approach, as in Storey *et al.* (2004a) or Wichert *et al.* (2004). While a modelling approach is usually optimal, as it estimates the dependence in the time-series data and uses it to get to a better fit; it may be desirable for the researcher to directly test on genes, at given time points, since it is obviously easier and faster. Using the results of theorem 1 we can select lists of which genes were differentially regulated for each given time point. In fact, the expression levels for each gene can be assumed to be a weakly dependent time series. Given that, it is reasonable to assume that the expression level of a single gene is more dependent with the expression of the same gene at a given different time point than with the expression of any different gene at the different time point, so weak dependence characterizes the dependency structure. Hence, an FDR controlling procedure can be applied to the $G \times T$ test statistics arising from the $G \times T$ recorded gene expressions, and then the discoveries can be put into separate lists grouped by time point.

4. Estimating the proportion of false nulls

A good estimator for a can improve the BH method in terms of power, leading to the plug-in method. Many other procedures can benefit from a good estimator of a. For instance, it is well known that, if M_0 were known, rejection of $p_j < \alpha/M_0$ would yield control of FWER at level α. Of course, one may not have either strong or weak control in this way. See Hochberg & Tamhane (1987) for a discussion. Here, we propose the estimate $\hat{M}_0 = m(1-\hat{a})$. Note that

© Board of the Foundation of the Scandinavian Journal of Statistics 2006.

in many procedures m is used as a strongly conservative estimate of M_0, that is $\hat{a} := 0$ (as in the BH procedure or the standard Bonferroni correction).

Then, a good estimator $\hat{a}$ is one that underestimates the true proportion a, although getting as close as possible to the upper bound. It is straightforward to see, in fact, that $\hat{a} > a$ may lead to loss of control of the FDR.

For this reason, we will consider estimators that are conservative with high probability: $P(\hat{a} \leq a) \geq 1 - \alpha_1$, for α_1 small. A similar approach is taken in Meinshausen & Rice (2005), who derive conservative estimators for the quantity a under independence. As an aside, note that such estimators are of interest by themselves, providing a conservative statement about the weight of a component in a mixture (here, the component is $p_i \mid H_1$). This has application for instance in cosmology, where high energy photon arrival times are recorded. Each arrival time is either noise (hence, a uniform arrival time) or signal, a pulsed radiation. Arrival time of the pulsed radiation is distributed according to an unknown density. More formally, let $f(x) = (1 - a)g_1(x) + ag(x)$, where $g_1(\cdot)$ is a known density, and $g(\cdot)$ is an unknown density. The proportion of pulsed radiation (on the number of arrival times), a, describes the *strength* of the pulsed signal, and is of interest to cosmologists. The use of a conservative estimator for a allows not to over estimate the strength of the pulsed signal with high probability. See Swanepoel (1999) and references therein for a detailed discussion of the problem.

Estimators for a are derived in Swanepoel (1999), Woodroofe & Sun (1999), Storey (2002), Genovese & Wasserman (2004a) and Meinshausen & Rice (2005). As also Meinshausen & Rice (2005) note, the conservative property of many of them may be lost under independence.

We derive here estimators that are conservative with high probability, under independence. We then argue by simulations that one of them is conservative also under dependence.

In this and the following sections we do not assume that the test statistics are totally ordered, and no asymptotics is involved: the results are valid for any value of m.

4.1. Estimator based on a confidence interval

If the test statistics are independent, let $\varepsilon_m = \sqrt{\log(2/\alpha_1)/2m}$. Define

$$\bar{M}_0 = m * \min\left(\frac{\inf_s (1 - \hat{G}(s) + \varepsilon_m)}{(1 - s), 1}\right),$$

and let

$$\hat{a} = (1 - \hat{M}_0/m). \tag{9}$$

To prove this is a conservative estimator, we will just need to prove that $\bar{M}_0$ is an upper bound for M_0 with high probability for any m. We will use the Dvoretzky–Kiefer–Wolfowitz–Massart (DKWM) inequality:

Theorem 5 (Dvoretzky *et al.*, 1956; Massart, 1990)
Let $X_1, \ldots, X_n$ be a sequence of independent identically distributed random variables. Let $F(z)$ be the CDF of X_1, and $\hat{F}(z)$ the empirical distribution of the sequence $X_1, \ldots, X_n$. Then, $\Pr\{\sup_{z \in \mathcal{R}} |F(z) - \hat{F}(z)| > \varepsilon\} \leq 2\,\mathrm{e}^{-n\varepsilon^2}$.

We are now ready to prove that the proposed estimator is conservative:

© Board of the Foundation of the Scandinavian Journal of Statistics 2006.

Theorem 6

If the test statistics are independent the estimator $\hat{a}$ defined in (9) is such that $\Pr(\hat{a} \leq a) \geq 1-\alpha_1$.

Proof. This estimator is based on the fact that

$$a \geq \frac{G(s)-s}{1-s} \tag{10}$$

for any $s \in (0,1)$, as easily seen by looking at the definition of $G(\cdot)$. By theorem 5, $\Pr(1-a \leq \overline{M_1}/m) \geq 1-\alpha_1$.

This directly provides a $1-\alpha_1$ confidence interval for a: $[\hat{a}, 1]$.

Note that, while the estimator is conservative with high probability for any value of m, as m gets bigger and bigger the confidence bound provided by DKWM inequality gets smaller and smaller; leading to more powerful procedures.

4.2. Iterative estimators

The 'iterative plug-in method', which we will describe in this section, seems to be robust with respect to dependence. As we need to estimate M_1, the number of false nulls, we thought the most natural estimator was the number of rejected hypotheses. The proportion a is then estimated iteratively as the proportion of rejected nulls in the previous plug-in step; till a does not change in two subsequent iterations, with a BH method at the first iteration (i.e. the first estimator is always set to 0):

1. Let $R_0 := 0$.
2. At the ith iteration, apply plug-in method with $\hat{a}_i = R_{i-1}/m$, and reject R_i hypotheses. Let $\hat{a}_{i+1} = R_i/m$.
3. Iterate until R_{i-1} and R_i are equal. Reject R_i hypotheses.

It is straightforward to prove that the number of iterations is finite: there are only $m+1$ possible values for $\hat{a}$. Then, the random variable given by the difference between the current and the previous estimate is discrete and puts a non-null probability mass at 0, which is our stopping rule. Let us stress that such estimator does not possess any theoretical property, in particular it is not guaranteed to lead to FDR control in the plug-in method.

A similar estimator was independently derived in Benjamini *et al.* (2004). They prove, under independence, that if the process is stopped after just a single iteration, the estimator is in fact conservative. They also suggest an iterative estimator similar to our proposal, and note that, as in our case, this kind of estimators possess an interesting internal coherence property: the number of hypotheses one rejects is also the number of hypotheses believed to be false. Here we suggest by the simulation in the next section that such estimators are worth to be considered when dealing with dependent test statistics.

4.3. Simulation of the plug-in method

Table 2 compares the plug-in method with different (conservative) estimators for a. CI stands for our DKWM confidence interval estimator, with different choices of α_1. Note that if $\alpha_1 \to 0$ we get back the BH procedure. We compare the estimators with the two-step Benjamini *et al.* (2004) estimator, in the usual simulation setting under independence, for different values of m and a. It can be seen that, when m is small the CI estimators compare well with the two-step estimator; while they outperform it when m is big. This is due to the fact that the length

© Board of the Foundation of the Scandinavian Journal of Statistics 2006.

Table 2. *False-negative rate (false discovery rate) for the plug-in method with different estimators for a*

a	0.1	0.5	0.9
$m=100$			
BH(CI($\alpha_1 \to 0$))	0.0583 (0.0472)	0.304 (0.0243)	0.777 (0.0054)
Benjamini *et al.* (2004)	0.0580 (0.0493)	0.290 (0.0356)	0.736 (0.0116)
CI($\alpha_1 = 0.01$)	0.0583 (0.0475)	0.293 (0.0327)	0.742 (0.0109)
CI($\alpha_1 = 0.05$)	0.0582 (0.0477)	0.291 (0.0346)	0.736 (0.0118)
CI($\alpha_1 = 0.1$)	0.0582 (0.0477)	0.290 (0.0354)	0.734 (0.0123)
$m=1600$			
BH(CI($\alpha_1 \to 0$))	0.0593 (0.0447)	0.308 (0.0250)	0.780 (0.0049)
Benjamini *et al.* (2004)	0.0589 (0.0466)	0.293 (0.0350)	0.742 (0.0114)
CI($\alpha_1 = 0.01$)	0.0589 (0.0464)	0.289 (0.0388)	0.722 (0.0161)
CI($\alpha_1 = 0.05$)	0.0589 (0.0467)	0.288 (0.0396)	0.719 (0.0168)
CI($\alpha_1 = 0.1$)	0.0588 (0.0469)	0.287 (0.0400)	0.717 (0.0173)
$m=100{,}000$			
BH(CI($\alpha_1 \to 0$))	0.0594 (0.0450)	0.308 (0.0250)	0.781 (0.0050)
Benjamini *et al.* (2004)	0.0590 (0.0472)	0.293 (0.0351)	0.743 (0.0113)
CI($\alpha_1 = 0.01$)	0.0588 (0.0488)	0.283 (0.0433)	0.698 (0.0222)
CI($\alpha_1 = 0.05$)	0.0588 (0.0488)	0.283 (0.0435)	0.696 (0.0227)
CI($\alpha_1 = 0.1$)	0.0588 (0.0489)	0.283 (0.0437)	0.695 (0.0230)

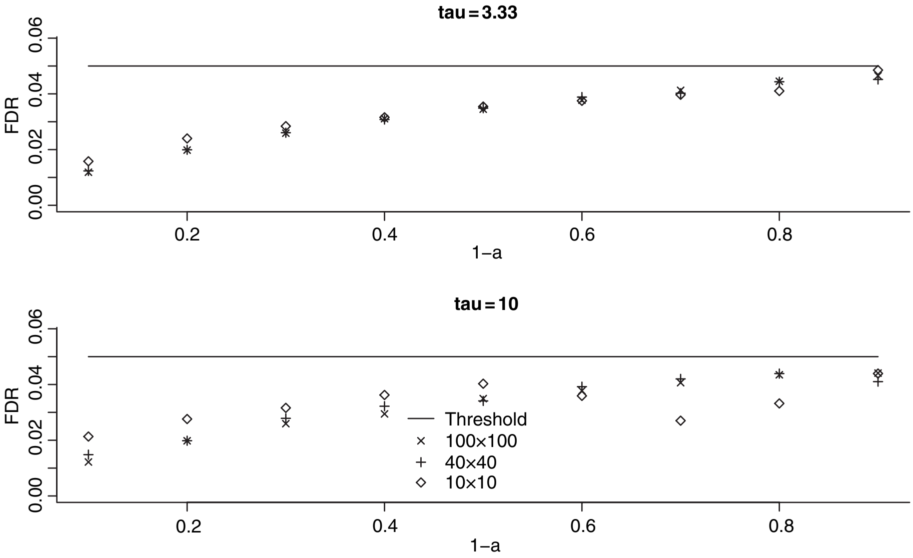

Fig. 5. False discovery rate for plug-in method, normal random variables with $\mathrm{cov}(X_{i,j}, X_{i',j'}) = \exp(-\frac{1}{\tau}\, d((i,j),(i',j')))$, 1000 iterations, a random proportion of $1-a$ random variables have a zero mean, the remaining have a non-zero mean sampled from uniform on $(0, 5)$. Iterative estimator used for a.

of the confidence interval is infinitesimal with m. The CI estimator is easily seen to converge with m to the Storey (2002) estimator, which is powerful even if not necessarily conservative. Finally, the choice of α_1 does not seem to be of much influence to the performance of the plug-in procedure.

We then turn to dependent test statistics. In Fig. 5 the results of the plug-in procedure with the iterative estimator are shown, with $a_0 = 0$, for multivariate normal random variables with variance kernel given by (7). The average number of steps was always between 2 and 7. Note that this procedure manages to control the FDR at level 5% when the correlation is

© Board of the Foundation of the Scandinavian Journal of Statistics 2006.

Table 3. *False-negative rate (false discovery rate) for different methods, multivariate normal random variables with covariance defined in (7)*

τ	50	20	10	2
BH	0.0575 (0.0204)	0.0563 (0.0239)	0.0583 (0.0253)	0.0570 (0.0378)
Iterative	0.0572 (0.0254)	0.0557 (0.0346)	0.0577 (0.0324)	0.0567 (0.0402)
BY	0.0683 (0.0016)	0.0667 (0.0062)	0.0690 (0.0043)	0.0680 (0.0078)
Stor.	0.0453 (0.2825)	0.0473 (0.2227)	0.0524 (0.1666)	0.0550 (0.0706)
CI ($\alpha_1 = 0.05$)	0.0546 (0.1123)	0.0532 (0.0991)	0.0564 (0.0735)	0.0569 (0.0558)

Table 4. *False-negative rate (false discovery rate) for different methods, multivariate normal random variables with covariance defined in (7)*

τ	1.67	1.25	1	0.83
BH	0.0590 (0.0432)	0.0582 (0.0435)	0.0584 (0.0432)	0.0577 (0.0438)
Iterative	0.0587 (0.0460)	0.0579 (0.0483)	0.0581 (0.0479)	0.0574 (0.0486)
BY	0.0695 (0.0074)	0.0692 (0.0104)	0.0691 (0.0094)	0.0688 (0.0101)
Stor.	0.0572 (0.0678)	0.0568 (0.0631)	0.0574 (0.0599)	0.0568 (0.0496)
CI ($\alpha_1 = 0.05$)	0.0587 (0.0470)	0.0580 (0.0497)	0.0581 (0.0448)	0.0587 (0.0447)

very strong (robustness), while it behaves just like the old one-step procedure when the correlation is weak (it just seems to be a little bit more conservative). If we compare Fig. 5 with Fig. 2, we can appreciate the robustness of iterative estimators. The same could be seen with many other possible choices of the parameter τ.

Finally, Tables 3 and 4 show the average FNR (FDR) obtained with different procedures, applied to a grid of $m = 100$ dependent normals with simplified exponential covariance structure defined in (7). Recall that the FNR is a Type II error rate, hence the lower the FNR the higher the power.

BH stands for the classical method of Benjamini & Hochberg (1995), Iterative stands for the plug-in method applied with the iterative estimator, BY for the classical method in which the level α is divided by $\sum_{i=1}^{m} 1/i$, as suggested in Benjamini & Yekutieli (2001). Stor. stands for plug-in method with Storey's estimator in (3).

It is seen that plug-in with Storey's estimator brings unacceptably high FDR, which are still over level α in many cases if one uses our CI estimator (designed for independent test statistics). On the other hand, FDR is controlled when using the iterative estimator proposed in previous section.

Note that early stopping (e.g. at the second step) of the iterations yields a more conservative procedure, so any early stopping would have given FDR control as well.

Acknowledgements

The author is grateful to the Editor, an Associate Editor and two referees for helpful remarks that lead to improvement of the paper, and for encouraging comments. Acknowledgements also go to Larry Wasserman and Chris Genovese, who supervised part of this research; and to Marco Perone Pacifico for careful reading of a first draft.

References

Alfó, M. & Trovato, G. (2004). Semiparametric mixture models for multivariate count data, with application. *Econ. J.* **7**, 426–454.

Amaratunga, D. & Cabrera, J. (2004). *Exploration and analysis of DNA microarray and protein array data*. Wiley, New York.

Arcones, M. A. & Yu, B. (2000). Limit theorems for empirical processes under dependence. In *Chaos expansions, multiple Wiener-Itô integrals and their applications* (ed. C. Houdre), 205–221. CRC Press Inc., Boca Raton, FL.

© Board of the Foundation of the Scandinavian Journal of Statistics 2006.

Benjamini, Y. & Hochberg, Y. (1995). Controlling the false discovery rate: A practical and powerful approach to multiple testing. *J. Roy. Statist. Soc. Ser. B* **57**, 289–300.

Benjamini, Y. & Hochberg, Y. (1997). Multiple hypothesis testing with weights. *Scand. J. Statist.* **24**, 407–418.

Benjamini, Y. & Hochberg, Y. (2000). The adaptive control of the false discovery rate in multiple hypothesis testing with independent test statistics. *J. Educ. Behav. Statist.* **25**, 60–83.

Benjamini, Y. & Yekutieli, D. (2001). The control of the false discovery rate in multiple testing under dependency. *Ann. Statist.* **29**, 1165–1188.

Benjamini, Y., Krieger, A. M. & Yekutieli, D. (2004). *Two staged linear step up FDR controlling procedures.* Technical report, Department of Statistics, Tel Aviv University, Tel Aviv.

Bickel, D. R. (2004). *On 'Strong control, conservative point estimation and simultaneous conservative consistency of false discovery rates': does a large number of tests obviate confidence intervals of the FDR?* Technical Report, Office of Biostatistics and Bioinformatics, Medical College of Georgia, Augusta, GA.

Billingsley, P. (1999). *Convergence of probability measures*. Wiley, New York.

Bradley, R. C. (1993). Equivalent mixing conditions for random fields. *Ann. Probab.* **21**, 1921–1926.

Brown, P. O. & Botstein, D. (1999). Exporing the new world of genome with DNA microarrays. *Nat. Genet.* **21**, 33–37.

Csörgo, S. & Mielniczuk, J. (1996). The empirical process of a short-range dependent stationary sequence under gaussian subordination. *Probab. Theory Related Fields* **104**, 15–25.

Doukan, P. (1994). *Mixing*, Lectures Notes in Statistics, 85. Springer-Verlag, New York.

Doukhan, P. & Louhichi, S. (1999). A new weak dependence condition and applications to moment inequalities. *Stochast. Process. Appl.* **84**, 313–342.

Dudoit, S., van der Laan, M. J. & Pollard, K. S. (2004). Multiple testing. Part I. Single-step procedures for control of general type I error rates. *Statist. Appl. Genet. Mol. Biol.* **3**, 1.

Dudoit, S., Shaffer, P. J. & Boldrick, J. C. (2003). Multiple hypothesis testing in microarray experiments. *Statist. Sci.* **18**, 71–103.

Duggan, D., Bittner, M., Chen, Y., Meltzer, P. & Trent, J. (1999). Expression profiling using cDNA microarrays. *Nat. Genet.* **21**, 10–14.

Dvoretzky, A., Kiefer, J. C. & Wolfowitz, J. (1956). Asymptotic minimax character of the sample distribution function and of the classical multinomial estimator. *Ann. Math. Statist.* **33**, 642–669.

Esary, J. D., Proschan, F. & Walkup, D. W. (1967). Association of random variables, with applications. *Ann. Math. Statist.* **38**, 1466–1474.

Finner, H. & Roters, M. (2002). Multiple hypotheses testing and expected number of type I errors. *Ann. Statist.* **30**, 220–238.

Genovese, C. R. & Wasserman, L. (2002). Operating characteristics and extensions of the FDR procedure. *J. Roy. Statist. Soc. Ser. B* **64**, 499–518.

Genovese, C. R. & Wasserman, L. (2004a). *Exceedance control of the false discovery proportion.* Technical Report. Department of Statistics, Carnegie Mellon University, Pittsburgh.

Genovese, C. R. & Wasserman, L. (2004b). A stochastic process approach to false discovery control. *Ann. Statist.* **32**, 1035–1061.

Genovese, C. R., Roeder, K. & Wasserman, L. (2007). False discovery control with p-value weighting. *Biometrika* (in press).

Hochberg, Y. & Tamhane, A. C. (1987). *Multiple comparisons procedures*. Wiley, New York.

Kumar, J. D. & Proschan, F. (1983). Negative association of random variables with applications. *Ann. Statist.* **11**, 286–295.

Lehmann, E. L. (1966). Some concepts of dependence. *Ann. Math. Statist.* **37**, 1137–1153.

Lehmann, E. L. & Romano, J. P. (2005). Generalizations of the familywise error rate. *Ann. Statist.* **33**, 1138–1154.

Marshall, A. W. & Olkin, I. (1967). A multivariate exponential distribution. *J. Amer. Statist. Assoc.* **62**, 30–44.

Massart, P. (1990). The tight constant in the Dvoretzky–Kiefer–Wolfowitz inequality. *Ann. Probab.* **18**, 1269–1283.

Meinshausen, N. & Rice, J. (2005). *Estimating the proportion of false null hypotheses among a large number of independently tested hypotheses.* Technical Report, Seminar für Statistik, Zurich.

Nze, P. A., Buhlmann, P. & Doukhan, P. (2002). Weak dependence beyond mixing and asymptotics for nonparametric regression. *Ann. Statist.* **30**, 397–430.

Oliveira, P. & Suquet, C. (1995). $L^2(0,1)$ weak convergence of the empirical process for dependent variables. Lecture notes in statistics 103. In *Wavelets and Statistics* (eds A. Antoniadis & G. Oppenheim), 331–344. Springer, New York.

Parmigiani, G., Garret, E. S., Irizarry, R. & Zeger, S. L. (2003). *The analysis of gene expression data: methods and software*. Springer, New York.

Pollard, K. S. & van der Laan, M. J. (2002). Resampling-based multiple testing: Asymptotic control of type I error and applications to gene expression data. *J. Statist. Plann. Inference* **125**, 85–100.

Sarkar, S. K. (2002). Some results on false discovery rate in stepwise multiple testing procedures. *Ann. Statist.* **30**, 239–257.

Sarkar, S. K. (2004). FDR-controlling stepwise procedures and their false negatives rate. *J. Stat. Plann. Inference* **125**, 119–137.

Sarkar, S. K. (2005). *Stepup procedures controlling generalized FWER and generalized FDR.* Technical Report, Department of Statistics, Temple University, Philadelphia.

Sarkar, S. K. & Chang, C. K. (1997). The Simes method for multiple hypotheses testing with positively dependent test statistics. *J. Amer. Statist. Assoc.* **26**, 1601–1608.

Silvestrov, D. S. (1971). Limit distributions for compositions of random functions. *Dokl. Akad. Nauk SSSR* **199**, 1251–1252. English translation in *Soviet Math. Dokl.* **12**, 1282–1285.

Silvestrov, D. S. (1972). Remarks on the limit of composite random functions. *Teor. Veroyatn. Primen.* **17**, 707–715. English translation in *Theory Probab. Appl.* **17**, 669–677.

Silvestrov, D. S. (2004). *Limit theorems for randomly stopped stochastic processes*. Springer-Verlag, New York.

Storey, J. D. (2002). A direct approach to false discovery rates. *J. Roy. Statist. Soc. Ser. B* **64**, 479–498.

Storey, J. D. (2003). The positive false discovery rate: a Bayesian interpretation and the q-value. *Ann. Statist.* **31**, 2013–2035.

Storey, J. D., Leek, J. T., Xiao, W., Dai, J. Y. & Davis, R. W. (2004a). *A significance method for time course microarray experiments applied to two human studies*. Technical Report, UW Biostatistics Working Paper Series, Working Paper 232.

Storey, J. D., Taylor, J. E. & Siegmund, D. (2004b). Strong control, conservative point estimation and simultaneous conservative consistency of false discovery rates: a unified approach. *J. Roy. Statist. Soc. Ser. B* **66**, 187–205.

Swanepoel, J. W. H. (1999). The limiting behavior of a modified maximal symmetric $2s-$spacing with applications. *Ann. Statist.* **27**, 24–35.

Tong, Y. L. (1980). *Probability inequalities in multivariate distributions*. Academic Press, London.

Van der Vaart, A. W. (1998). *Asymptotic statistics*. Cambridge University Press, Cambridge.

Whitfield, M. L., Sherlock, G., Saldanha, A. J., Murrary, J. I., Ball, C. A., Alexander, K. E., Matese, J. C., Perou, C. M., Hurt, M. M., Brown, P. O. & Botstein, D. (2002). Identification of genes periodically expressed in the human cell cycle and their expression in tumors. *Mol. Biol. Cell.* **13**, 1977–2000.

Wichert, S., Fokianos, K. & Strimmer, K. (2004). Identifying periodically expressed transcripts in microarray time series data. *Bioinformatics* **20**, 5–20.

Woodroofe, M. & Sun, J. (1999). Testing uniformity versus a monotone density. *Ann. Statist.* **27**, 338–360.

Wu, W. B. (2004). *Empirical processes of dependent random variables*. Technical Report, Department of Statistics, University of Chicago, Chicago.

Yekutieli, D. & Benjamini, Y. (1999). Resampling-based false discovery rate controlling multiple test procedures for correlated test statistics. *J. Statist. Plann. Inference* **82**, 171–196.

Yoshihara, K. (1975). Billingsley's theorems on empirical processes of strong mixing sequences. *Yokohama Math. J.* **23**, 77–83.

Yu, H. (1993). A Glivenko-Cantelli lemma and weak convergence for empirical processes of associated sequences. *Probab. Theory Related Fields* **95**, 357–370.

Received July 2005, in final form June 2006

Alessio Farcomeni, Dipartimento di Statistica, Probabilità e Statistiche Applicate, Università di Roma 'La Sapienza', Piazzale Aldo Moro, 5, 00185 Rome, Italy.
E-mail: alessio.farcomeni@uniroma1.it

Appendix A

Proof of theorem 1

We will make use of the following lemmas. Let $\Lambda_0(t) = \sum(1-H_i)1_{\{p_i<t\}}$ and $\Lambda_1(t) = \sum H_i 1_{\{p_i<t\}}$.

© Board of the Foundation of the Scandinavian Journal of Statistics 2006.

Lemma 1

Let $\hat{G}(t)=1/m\sum 1_{\{p_i<t\}}$ be asymptotically equicontinuous. Then $\Lambda_j(t), j=0,1$ will be too.

Proof. $\forall(s,t)\in\mathcal{R}^2$ we have that $|\Lambda_j(t)-\Lambda_j(s)|\leq|\hat{G}-\hat{G}(s)|$.

This is easily seen. For instance, let WLOG $t>s$ and $j=1$. We have that

$$\frac{1}{m}\sum H_i[1_{\{p_i<t\}}-1_{\{p_i<s\}}]\leq\frac{1}{m}\sum[1_{\{p_i<t\}}-1_{\{p_i<s\}}],$$

as H_i can either be 0 or 1.

Hence we can prove the asymptotic equicontinuity of $\Lambda_j(t)$ by applying the definition:

$$\limsup_n \Pr{}^*(\sup_i \sup_{s,t\in T_i}|\Lambda_j(t)-\Lambda_j(s)|\geq\varepsilon)\leq\limsup_n \Pr{}^*(\sup_i \sup_{s,t\in T_i}|\hat{G}-\hat{G}(s)|\geq\varepsilon)\leq\eta,$$

where T_i is an opportune partition of $[0,1]$.

Lemma 2

Assume any of the conditions holds. The vector $(W_0(t), W_1(t))$ will be convergent in distribution for any t; where $W_0(t)=\sqrt{m}(\Lambda_0(t)-(1-a)t)$ and $W_1(t)=\sqrt{m}(\Lambda_1(t)-aF(t))$, and where $F(t)$ is the CDF of p_j under the alternative hypothesis.

Proof. We will use the Cramer–Wold device (see Van der Vaart, 1988) to prove the convergence of the vector. We will assume any of the mixing conditions holds. If association holds, it is straightforward to prove the results are the same. Let $(c_0,c_1)\in\mathcal{R}^2$. It is easy to see that $c_0\Lambda_0(t)+c_1\Lambda_1(t)$ is equal to $1/m\sum(c_0+H_i(c_1-c_0))1_{\{p_i<t\}}$.

Define the sequence $\{\xi_i\}_{i\in\mathcal{N}}$ to be:

$$\xi_i=\begin{cases}H_i & \text{if } \{i \bmod 2\}=0\\ p_i & \text{if } \{i \bmod 2\}=1.\end{cases}$$

By definition, the sequence will be $\{p_1,H_1,p_2,H_2,p_3,\ldots\}$.

We have then that $(c_0+H_i(c_1-c_0))1_{\{p_i<t\}}=(c_0+\xi_{2i}(c_1-c_0))1_{\{\xi_{2i-1}<t\}}$.

It is easy to see that the ξ sequence is still mixing, and the same conditions will be true. In particular, called $\alpha'(n)$ the mixing coefficients for the ξ_is at lag n, under Assumption 2 we have that $\sum\sqrt{\alpha'(n)}\leq\sqrt{C'}\sum n^{\frac{-3-\delta'}{2}}<+\infty$. Hence, the series of the partial sums $\sum\xi_i$ will be weakly convergent (see, e.g. Billingsley, 1999). Same results are immediately proved under the other conditions.

As $(c_0+\xi_{2i}(c_1-c_0))1_{\{\xi_{2i-1}<t\}}$ is a measurable function which depends only on finitely many coordinates of the vector $\{\xi_i\}$, the same results on the mixing coefficients will hold for it (again, see Billingsley, 1999).

Hence $1/m\sum(c_0+H_i(c_1-c_0))1_{\{p_i<t\}}$, opportunely rescaled, will converge in distribution. By Cramer–Wold device, also the two-dimensional vector $(W_0(t), W_1(t))$ will converge in distribution.

By condition 1 (Van der Vaart, 1988), condition 2 (Yoshihara, 1975), condition 3 (Yu, 1993), condition 4 or 5 (Oliveira & Suquet, 1995), condition 6 (Csörgo & Mielniczuk, 1996), the empirical process $\sqrt{m}(\hat{G}(t)-G(t))$ will be convergent to a centred Gaussian random process. The convergence is to something very close to a Brownian bridge on the scale of G.

© Board of the Foundation of the Scandinavian Journal of Statistics 2006.

The only difference will be given by the covariance kernel, which will be the usual kernel to which is added a convergent series. The covariance kernel is given by:

$$K(s,t)=G(s\wedge t)-G(s)G(t)+2\sum_{k=2}^{+\infty}[P(p_1<s,p_k<t)-G(s)G(t)].$$

As the empirical process is convergent, $\hat{G}_{(t)}$ is asymptotically equicontinuous and $\forall t_1,\ldots,t_k$ the vector $(\hat{G}(t_1),\ldots,\hat{G}(t_k))$ will converge in distribution.

By lemma 1, then also $\Lambda_j(t), j=0,1$ will be asymptotically equicontinuous. As each component is asymptotically equicontinuous, then also the vector $(\Lambda_0(t),\Lambda_1(t))$ will be asymptotically equicontinuous.

Moreover, by lemma 3, $\forall t_1,\ldots,t_k$ the vector $[(\Lambda_0(t_1),\Lambda_1(t_1)),\ldots,(\Lambda_0(t_k),\Lambda_1(t_k))]$ will converge in distribution. It is straightforward, in fact, to extend the result of lemma 3 to the vector case.

Then, it happens that (W_0,W_1) will converge to a centred two-dimensional Gaussian process. Let $\rho_{ij}(k)=\Pr(p_1<s,p_k<t\,|\,H_1=i,H_k=j)$. The covariance kernel is obtained by direct calculation of $E[W_i(s)W_j(t)]$, $i=0,1$, $j=0,1$, and is given by

$$K_2(s,t)=\begin{bmatrix}K_{0,0}(s,t) & K_{0,1}(s,t)\\ K_{1,0}(s,t) & K_{1,1}(s,t)\end{bmatrix},$$

where

$$K_{0,0}(s,t)=(1-a)(s\wedge t)-(1-a)^2st+2\sum_{k=2}^{+\infty}[\rho_{00}(k)-(1-a)^2st],$$

$$K_{0,1}(s,t)=-(1-a)saF(t)+2\sum_{k=2}^{+\infty}[\rho_{01}(k)-(1-a)saF(t)],$$

$$K_{1,0}(s,t)=-(1-a)taF(s)+2\sum_{k=2}^{+\infty}[\rho_{10}(k)-(1-a)taF(s)],$$

and

$$K_{1,1}(s,t)=aF(s\wedge t)-a^2F(s)F(t)+2\sum_{k=2}^{+\infty}[\rho_{11}(k)-a^2F(s)F(t)].$$

The computations are tedious, not very complex and omitted for brevity.

Note now that

$$\Gamma(t)=\frac{\Lambda_0(t)}{\Lambda_0(t)+\Lambda_1(t)}=r(\Lambda_0(t),\Lambda_1(t));$$

where $r(\cdot,\cdot):l^\infty\ x\ l^\infty\to l^\infty$.

For technical reasons, we restrict the $\Gamma(t)$ process in $[\delta,1]$, for any $\delta>0$. The variance of $\Gamma(t)$ is in fact infinite at $t=0$. Let $Q(t)=(1-a)t/G(t)$.

The function $r(\cdot,\cdot)$ is such that $r((1-a)U,aF)=Q$ (where U is such that $U(t)=t$ for any $t\in[0,1]$), and it is also Fréchet differentiable at that point, with derivative: $r'_{(1-a)U,aF}(V_0,V_1)=aFV_0-(1-a)UV_1/G^2$.

Hence, by functional δ-method, $\sqrt{m}(\Gamma(t)-Q(t))$ will converge to the process defined by the evaluation of $r'_{(1-a)U,\,aF}(V_0,V_1)$ at the limit of (W_0,W_1). As this is nothing but a linear combination of the two elements of the vector, with coefficients aF/G^2 and $-(1-a)U/G^2$, it will be a Gaussian process on $(0,1]$, with mean 0 and covariance kernel $K_3(s,t)$ given by:

© Board of the Foundation of the Scandinavian Journal of Statistics 2006.

$$
\begin{aligned}
&\frac{a(1-a)}{G^2(s)G^2(t)}[(1-a)stF(s\wedge t)+aF(s)F(t)(s\wedge t)]\\
&+\frac{2}{G^2(s)G^2(t)}[a^2F(s)F(t)\sum_{k=2}^{+\infty}[\Pr(p_1<s,p_k<t\,|\,H_1=0,H_k=0)-(1-a)^2st]\\
&-a(1-a)sF(t)\sum_{k=2}^{+\infty}[\Pr(p_1<s,p_k<t\,|\,H_1=0,H_k=1)-(1-a)saF(t)]\\
&-a(1-a)tF(s)\sum_{k=2}^{+\infty}[\Pr(p_1<s,p_k<t\,|\,H_1=1,H_k=0)-(1-a)taF(s)]\\
&+(1-a)^2st\sum_{k=2}^{+\infty}[\Pr(p_1<s,p_k<t\,|\,H_1=1,H_k=1)-a^2F(s)F(t)].
\end{aligned}
$$

Again, the long computations are omitted. Hence, the FDP process centred at $Q(t)$ has a centred Gaussian limiting distribution with covariance kernel $K_3(s,t)$, as stated in Corollary 1.

Applying again the functional δ-method we see that $\sqrt{m}(\hat{Q}(t)-Q(t))$, where $\hat{Q}(t)=(1-a)t/\hat{G}_{(t)}$, will be convergent to a Gaussian process on $(0,1]$ with covariance kernel

$$
K_4(s,t)=\frac{Q(s)Q(t)}{G(s)G(t)}\left(G(s\wedge t)-G(s)G(t)+2\sum_{k=2}^{+\infty}(P(p_1<s,p_k<t)-G(s)G(t))\right),
$$

and then also

$$
\sqrt{m}(\hat{Q}^{-1}(u)-Q^{-1}(u)) \tag{11}
$$

is convergent to a Gaussian process with covariance kernel

$$
K_5(s,t)=\frac{K_4(Q^{-1}(s),Q^{-1}(t))}{Q'(Q^{-1}(s))Q'(Q^{-1}(t))}.
$$

As $T_{\mathrm{PI}}=\hat{Q}^{-1}(\alpha)$, by applying the δ-method to (11) it can be seen that $\sqrt{m}(T_{\mathrm{PI}}-Q^{-1}(\alpha))$ will converge in distribution to an $N(0,V)$, with $V=K_5(Q^{-1}(\alpha),Q^{-1}(\alpha))$; as stated in corollary 1. By one last application of the δ-method we see that $\sqrt{m}(Q(T_{\mathrm{PI}})-\alpha)$ will converge in distribution to an $N(0,(Q'(Q^{-1}(\alpha)))^2V)$.

We cannot conclude now that FDR → α, as FDR is the expected value of the stochastic process $\Gamma(t)$ evaluated at the *random* point T_{PI}. We need to make some further considerations: let now $0<\delta<Q^{-1}(\alpha)$, and note that

$$
\begin{aligned}
|\Gamma(T_{\mathrm{PI}})-\alpha| &\leq |\Gamma(T_{\mathrm{PI}})-Q(T_{\mathrm{PI}})|+|Q(T_{\mathrm{PI}})-\alpha|\\
&\leq \sup_t|\Gamma(t)-Q(t)|\,1_{\{T_{\mathrm{PI}}<\delta\}}+\sup_t|\Gamma(t)-Q(t)|\,1_{\{T_{\mathrm{PI}}\geq\delta\}}+|Q(T_{\mathrm{PI}})-\alpha|\\
&\leq 1_{\{T_{\mathrm{PI}}<\delta\}}+\frac{1}{\sqrt{m}}\sup_{t>\delta}|\sqrt{m}(\Gamma(t)-Q(t))|+|Q(T_{\mathrm{PI}})-\alpha|\\
&=O_P(m^{-1/2}),
\end{aligned}
$$

as T_{PI} converges in probability to $Q^{-1}(\alpha)$.

As $0<\Gamma(t)<1$ for any m, the sequence is uniformly integrable.

Hence, $E[\Gamma(T_{\mathrm{PI}})]=\alpha+o(1)$.

© Board of the Foundation of the Scandinavian Journal of Statistics 2006.

Appendix B

Proof of theorem 4

By Doukhan & Louhichi (1999) and Nze *et al.* (2002) the conditions imply that $\sqrt{n}(\widehat{G}(t) - G(t))$ converges in $D[0, 1]$ to a centred Gaussian process, with covariance kernel of is

$$K_{\mathrm{G}}(s,t) = 2\sum_{k=0}^{+\infty} \mathrm{cov}(1_{p_1 \leq s}, 1_{p_k \leq t}).$$

This implies also lemma 1. Lemma 2 follows from the same reasoning as the proof above, in particular it is straightforward to see that, in the notation used above, the series of the partial sums $\sum \xi$ is weakly convergent.

The rest of the theorem follows from iterated application of functional and classical δ-method, which can be done in light of the weak convergence of $\sqrt{n}(\hat{G} - G(t))$.

Appendix C

Proof of theorem 2

We will make use of the following theorem from Silvestrov (1971, 1972), reviewed in Silvestrov (2004).

Theorem 7

Suppose $X_n(t)$ is a stochastic process weakly converging to a process $X(t)$ for any $t \in [0, T]$. Suppose T_n is a random variable weakly converging to a random variable T. Suppose the following conditions hold:

1. *$(T_n, X_n(t)$ weakly converges to $(T, X(t))$ for any $t \in U$, where $U \subseteq [0, T]$, dense in the interval, and contains the point 0.*
2. *The process $X(t)$ is continuous at the point T with probability 1.*
3. *The process $X_n(t)$ is J-compact:*

$$\lim_{c\to 0} \limsup_{\varepsilon \to 0} \Pr(\sup_{0\vee(t-c)\leq t' \leq t \leq t'' \leq (t+c)\wedge 1} \min(|X_n(t') - X_n(t)|, |X_n(t) - X_n(t'')|) > \delta) = 0$$

for any $\delta > 0$.

Then, $X_n(T_n)$ weakly converges to $X(T)$.

It is actually easy to prove the assumptions of theorem 7 in our setting: let $X_m(t) = \sqrt{m}(\Gamma(t) - Q(t))$ and $T_m = T$. From theorem 1, we know that $X_m(t)$ converges weakly to $X(t) = W(t)$, where $W(t)$ is a Gaussian process on $(0, 1)$, centred at 0 and with covariance kernel $K(s, t)$ given in (6). We assumed that $T_m \xrightarrow{P} T$, which implies that the vector $(T_m, X_m(t)$ converges in distribution to $(T, X(t))$ for any $t \in (0, 1)$. As $X(t)$ is a Gaussian process, it is continuous with probability 1 at any point $t \in (0, 1)$. Finally, J-compactness is implied by asymptotic equicontinuity of $X_m(\cdot)$.

© Board of the Foundation of the Scandinavian Journal of Statistics 2006.