Bayes Oracle and the Asymptotic Optimality of the Multiple Testing Procedures Under Sparsity

by

M. Bogdan
Wroclaw University of Technology

A. Chakrabarti
Indian Statistical Institute

J. K. Ghosh
Purdue University

Technical Report #09-02

Department of Statistics
Purdue University

February 2009

Bayes oracle and the asymptotic optimality of the multiple testing procedures under sparsity.

[1]Małgorzata Bogdan, [2] Arijit Chakrabarti, [2,3] Jayanta K. Ghosh,

[1] *Institute of Mathematics and Computer Science, Wrocław University of Technology, Poland*
[2] *Indian Statistical Institute, Kolkata, India*
[3] *Department of Statistics, Purdue University, USA*

Abstract

We investigate the asymptotic optimality of the multiple testing rules, using the framework of the Bayesian decision theory. We concentrate on the multiple testing within the normal scale mixture model proposed in [31] and consider the asymptotic scheme under which the proportion of the alternative hypothesis p converges to zero. We characterize the set of the fixed threshold multiple testing rules which are asymptotically optimal and prove the asymptotic optimality of the rules, which control the Bayesian False Discovery Rate. We also provide the conditions under which the popular Benjamini and Hochberg and Bonferroni procedures are asymptotically optimal. Our results show that the optimal BFDR level should depend on the expected signal magnitude and the ratio of losses for type I and type II errors. Specifically, under our asymptotic scheme, the rule controlling BFDR at a fixed level α can be optimal only if the relative loss for missing the true signals increases when $p \longrightarrow 0$.

1. Introduction

Multiple testing is a very important problem in statistical inference because of its applicability in understanding large data sets involving many parameters. A prominent area of application of multiple testing is in the context of microarray data analysis, where one wants to simultaneously test expression levels of thousands of genes (e.g. see [12], [11], [38], [17], [23], [30], [31] or [37]). Various ways of performing multiple tests have been proposed in the literature over the years, typically differing in the target they want to achieve. Among the most popular multiple testing procedures one could mention the Bonferroni correction, aimed at controlling the Family Wise Error Rate, or the Benjamini and Hochberg procedure (BH, [3]), which controls the false discovery rate (FDR).

In recent years a substantial effort has been made to understand the properties of multiple testing procedures under sparsity, i.e. in the case when the proportion p of "true" alternatives among all tests is very small. A mile step in this direction was taken in [1], where the properties of FDR controlling procedures are analyzed from the point of view of the estimation of the sparse vector of means. Specifically, in [1] it is shown that BH adapts very well to the unknown sparsity parameter p

and is asymptotically minimax over a wide range of sparse parameter spaces and loss functions.

In the present paper we focus on the analysis of the properties of the multiple testing rules from the Bayes Decision Theoretic perspective, assuming fixed losses: δ_0 and δ_A for each of type I and type II errors, respectively. We believe that such an approach is natural in the context of testing, where the main goal is detecting the significant signals, rather than estimating their magnitude. In a specific example when $\delta_0 = \delta_A = 1$, the total loss is equal to the number of misclassified hypothesis. Very good properties of BH with respect to the misclassification rate under sparsity were signalized in [16] and illustrated by extensive simulation studies in [7] and [8]. [8] also contains some remarks on the relationship between the control of the Bayesian FDR (BFDR), defined in [11], and the control of the Bayes risk. Moreover, the results reported in [7] and [8] illustrate that for very small values of p the Bonferroni correction has very good properties with respect to the misclassification error.

In the present paper we support these experimental findings with the formal results on the asymptotic optimality of the multiple testing rules. We consider the multiple testing within the normal scale mixture model proposed in [31], [7] and [8]. The model differs from the model used by [1] in imposing a prior distribution on the unknown vector of means. As discussed in Section 2, this model can be used for testing the point hypothesis about the unknown means as well as for identifying large means embedded in the multitude of very small signals. We consider the asymptotic scheme when p goes to zero and the signals magnitude increases at such a rate that the asymptotic power of the optimal Bayes classifier is larger than zero. Our results confirm asymptotic optimality properties of BFDR controlling rules and the BH procedure but they are qualitatively different from the results reported in [1]. Specifically, we show that the rules controlling BFDR at a fixed level α can be asymptotically optimal only if the ratio of losses $\frac{\delta_0}{\delta_A}$ converges to zero. However, this restrictive assumption does not undermine good properties of rules controlling BFDR at a fixed level. It is quite natural to impose a large loss function for missed signals in case when p is very small or the signal magnitude is large. We also show that the optimal level of α strongly depends on the expected signal magnitude: in case when expected signals are large one should use relatively small α.

The paper contains the results on the asymptotic optimality of BH obtained under the assumption that the number of tests $m \to \infty$ and $p \to 0$ in such a rate that $p \geq \frac{log^\gamma}{m}$, for some constant $\gamma > 1$. For the case when $pm \to s \in (0, \infty)$ we have proved the optimality of the Bonferroni correction. In Section 7 we present some arguments and illustrative theoretical results suggesting that BH procedure is optimal within an entire range of $p \to 0$, such that $mp \to s \in (0, \infty]$. However, this conjecture still remains to be formally validated.

The outline of the paper is as follows. In Section 2 we define and discuss our model. In Section 3 we compute the type I and type II errors of the Bayes classifier and formulate the conditions under which the asymptotic power of this classifier is larger than 0. In Section 4 we give the definition of the asymptotic optimality and characterize fixed thresholds multiple testing rules, which are asymptotically optimal. In Section 5 we give the conditions under which a fixed threshold BFDR controlling is asymptotically optimal. In Section 6 we formulate similar conditions for the Bonferroni correction. Section 7 contains the results on the asymptotic optimality of the BH procedure, while Section 8 contains the discussion and directions for further research.

2. Statistical model

Suppose one has observations X_i, $i = 1, 2, \ldots, m$, where the X_i's are independent for each i having distribution $N(\mu_i, \sigma^2)$. We assume further that for $i = 1, \ldots, m$, μ_i's are iid from a mixture distribution $(1-p)\delta_{\{0\}} + pN(0, \tau^2)$, where τ^2 is constant. Our goal is to discover those i's for which $\mu_i \neq 0$, i.e for every $i \in \{1, \ldots, m\}$ we test the null hypothesis $H_{0i} : \mu_i = 0$ vs the alternative $H_{Ai} : \mu_i \neq 0$. In the next two paragraphs, we say a few words about the choice of the model, its motivation and relevance.

This kind of model is a popular one and is usually referred to as the two-groups model, since the marginal distribution of the X_i's are given by $X_i \sim (1-p)N(0, \sigma^2) + pN(0, \sigma^2 + \tau^2)$. The models of this kind have been used in, e.g., Scott and Berger (2006). People have employed this model to make inferences using parametric and nonparametric Empirical Bayes methods for estimating the hyperparameters from the data (see, e.g., Bogdan et al.(2007a, 2007b). In some cases, as in Efron (2008), one uses $X_i \sim (1-p)N(0, 1) + pF$ where F is any arbitrary distribution and nonparametric Empirical Bayes methods are used more directly in the analysis. A model like the last one is appropriate when, for example, one is testing against a nonparametric alternative.

The data of the kind we consider in this paper are often generated from DNA microarray experiments, where the gene expression levels (over or underexpression) of thousands of genes (m) are collected for a small or moderate number (n) of individuals and the X_i's correspond to the expression level for the i-th gene. In such a situation, the question of interest is which genes are "differentially expressed" i.e for which i's, $\mu_i \neq 0$. A common phenomenon in such experiments is that the fraction of non-zero μ_i's is relatively small, thus giving rise to high-dimensional data with "sparse" signals (see also in this context, say, Donoho and Jin (2004)). In this paper one of our main interest is the situation when the X_i's are generated by such sparse signals and the goal is to discover the signals. To reflect sparsity in the choice of the prior for the μ_i's, one takes p to be small. For p small, this model says that most (precisely $100(1-p)\%$) of the μ's are actually zero and the non-zero μ's have some similarity. The assumption of similarity of the non-zero μ_i's is also common in the context of microarray experiments (as remarked, for example, in Scott and Berger (2006)). To separate the signals from the noise, we take τ^2 to be relatively large, see e.g., Assumption (A) and Remark 2.1.

In this problem, we consider a decision theoretic formulation with additive losses, which goes back to Lehamann (1958) and seems to be implicit in most (but not all, e.g Scott and Berger (2008)) current formulations. Within this general framework we introduce a (Bayesian) Oracle that minimizes the expected loss when σ^2, p and τ^2 are known. The Oracle rejects H_{0i} if

$$\text{Reject } H_{0i} \text{ if } \frac{X_i^2}{\sigma^2} > (1+\frac{\tau^2}{\sigma^2})\left(\log(\frac{\tau^2}{\sigma^2}+1)+2\log(f\delta)\right) \ , \tag{2.1}$$

where $f = \frac{1-p}{p}$ and $\delta = \frac{\delta_0}{\delta_A}$.

3. Bayes oracle and asymptotic optimality

Let us introduce the following notation: $u = \left(\frac{\tau}{\sigma}\right)^2$ and $v = u(f\delta)^2$. Below we consider asymptotic behavior of the Bayes rule and some other multiple testing procedures under the assumption

(A): $u \to \infty$, $v \to \infty$, $\frac{\log v}{u} \to C$, where $0 \leq C < \infty$.

Comment. Observe that the variable u is a natural scale for strength of the signal in terms of the variance of $\bar{X}$ under the null, and f measures sparsity while v is natural reparametrization to simplify the threshold in (2.1). The significance of the relationship between u and v prescribed in assumption (A) is explained in detail in Remark 1. The reason for excluding the case $C = \infty$ in (A) is also given in remark 1 below and amounts to exclude the "nondetectable" signals.

We use the generic notation $o_{u,v}$ for a function of u and v which converges to zero under the assumption (A).

Theorem 1 *Under the assumption (A) the risk obtained by the Bayes oracle (2.1) takes the form*

$$R = mp\delta_A\sqrt{\frac{2\log v}{\pi u}}(1+o_{u,v}) \ , \tag{3.2}$$

when $C = 0$

or

$$R = mp\delta_A(2\Phi(\sqrt{C})-1)(1+o_{u,v}) \ , \tag{3.3}$$

when $0 < C < \infty$. Here $\Phi(\cdot)$ denotes the cumulative distribution function of the standard normal distribution.

Proof.
Let us denote by $c^2_{u,f,\delta}$ the threshold of the Bayes oracle (2.1), i.e. let

$$c^2_{u,f,\delta} = \frac{u+1}{u}\log\left((u+1)f^2\delta^2\right) \ .$$

Note that

$$\frac{c^2_{u,f,\delta}}{\log v} = \left(1+\frac{1}{u}\right)\left(1+\frac{\log(1+\frac{1}{u})}{\log v}\right) \ .$$

Since $\lim_{x\to 0}\frac{\log(1+x)}{x} = 1$ it follows that

$$c^2_{u,f,\delta} = \log v\left(1+\frac{1}{u}\right)\left(1+\frac{1}{u\log v}(1+o_{u,v})\right) = \log v(1+z_{u,v}) \ , \tag{3.4}$$

where $\lim_{u\to\infty,v\to\infty} z_{u,v}u = 1$. Note that $c^2_{u,f,\delta}\to\infty$ under condition (A).

Now, let us denote by t_1 the probability of the type I error and by t_2 the probability of the type II error. Note that the risk of the Bayes oracle

$$R = E(L) = m((1-p)t_1\delta_0 + pt_2\delta_A) \ . \tag{3.5}$$

At the first step we will approximate the probability of the type I error. Note that

$$t_1 = P(|Z| > c_{u,f,\delta}) \ , \tag{3.6}$$

where Z is the standard normal variable.

Let us observe that for any positive c

$$P(|Z|>c) = \frac{2\phi(c)}{c}(1-z_1(c)) \ , \tag{3.7}$$

where $\phi(\cdot)$ is the density of the standard normal distribution and $z_1(c)$ is a positive function such that $z_1(c)c^2 = O(1)$ as $c\to\infty$. This follows easily from well known tail estimates of standard normal distribution.

Moreover by (3.4) it holds that

$$\phi(c_{u,f,\delta})\sqrt{2\pi v} = \exp\left(\frac{-z_{u,v}\log v}{2}\right) \ .$$

This, together with assumption (A) yields

$$\phi(c_{u,f,\delta}) = e^{-C/2}\sqrt{\frac{1}{2\pi v}}(1+o_{u,v}) \ . \tag{3.8}$$

By (3.6), (3.7) and (3.8) we obtain

$$t_1 = e^{-C/2}\sqrt{\frac{2}{\pi v\log v}}(1+o_{u,v}) \ . \tag{3.9}$$

Consider now the probability of the type II error. Note that

$$t_2 = P\left(Z^2 < \frac{1}{u+1}c^2_{u,f,\delta}\right) \ . \tag{3.10}$$

Observe also that for any positive c

$$P(|Z|<c) = \sqrt{\frac{2}{\pi}}c(1-z_2(c)) \ , \tag{3.11}$$

where $z_2(c) = O(c^2)($ as $c \to 0)$ is a strictly positive quantity.

Using (3.10) and (3.4), a simple algebra shows that for the case $0 < C < \infty$

$$t_2 = (2\Phi(\sqrt{C}) - 1)(1 + o_{u,v}). \tag{3.12}$$

In the case where $C = 0$, (3.11) yields more specifically,

$$t_2 = \sqrt{\frac{2\log v}{\pi u}}(1 + o_{u,v}) \ . \tag{3.13}$$

Now (3.2) and (3.3) follow by combining (3.5), (3.9) and (3.13) or (3.12).

□

Definition. We call the multiple testing rule asymptotically optimal if under the condition (A) its risk $V = R(1 + o_{u,v})$, where R is the risk of the Bayes oracle.

Remark 1. When $C = 0$ both type I and type II error of the Bayes oracle converge to zero. When $C > 0$ type I error converges to 0 but type II error converges to a constant. From (3.10) and (3.4) it follows that when $u \to \infty$ and $v \to \infty$ in such a way that $\frac{\log v}{u} \to \infty$, the power of the Bayes oracle converges to zero. We exclude this range of "nondetectable" signals from our consideration. Now note that $\frac{\log v}{u}$ converges to a limit if and only if $\frac{2\log f\delta}{u}$ converges to the same limit. This observation seems to give us a bit more insight on this condition. Let us first try to understand the case when δ is a constant. Note that u and $\log f$ (or equivalently $\log(1/p)$) are the basic parameters in this case when we interpret what the limit means. The condition suggests that the magnitude of the signal and the amount of sparsity should be suitably related for non trivial inference to be possible. In particular, if the signals in the sparse sequence are significantly large (i.e $C < \infty$), they will be detected with high probability, i.e there should be some kind of parity between the amount of sparsity and the magnitudes of signals. When δ is not a constant, then the relationship between the sparsity of signals detectable by the Bayes Oracle and their magnitude depends on the relative costs of both types of errors.

Verge of detectability - least favorable balls in Abramovich et al.

Remark 2. The asymptotic form (under assumption (A)) of the risk of the Bayes oracle is determined by t_2. In particular, it is easy to see that the same asymptotic form of the type 2 error (as in (3.12) or (3.13)) is achieved by any multiple testing rule with the threshold value of the form $c^2 = \log v(1 + o_{u,v})$. Probability of type I error is somewhat more sensitive to the change in critical value. If the $o_{u,v}$ term is always positive then the type I error converges to zero not slower than for the optimal rule and the total risk is still determined by type II component. However, if the $o_{u,v}$ term can take negative values, the rate of convergence of type I error to zero may be substantially slower than the rate of convergence of type II component of the risk and the rule may not longer be optimal. These observations are formally summarized in the following Theorem.

Theorem 2 *Assume condition (A) holds. A multiple testing rule with the threshold value* $c^2 = \log v + z_{u,v}$ *is asymptotically optimal if and only if*

$$z_{u,v} = o(\log v) \tag{3.14}$$

and

$$z_{u,v} + 2\log\log v \to \infty. \tag{3.15}$$

Proof.

Let us first prove sufficiency of (3.14) and (3.15) for optimality of the multiple testing rule. The condition $z_{u,v} = o(\log v)$ implies that $t_2 = A\sqrt{\frac{\log v}{u}}(1 + o_{u,v})$ and the constant A is equal to $\sqrt{\frac{2}{\pi}}$ or $\frac{2\Phi(\sqrt{C})-1}{\sqrt{C}}$ according as C is zero or strictly positive. Note that (3.7) and the fact that $z_{u,v} = o(\log v)$ together imply that type I error is given by

$$t_1 = P(|Z| > c) = \sqrt{\frac{2}{\pi}}\frac{\exp(-z_{u,v}/2)}{\sqrt{v \log v}}(1 + o_{u,v}). \tag{3.16}$$

Now assume that the constant C specified in assumption (A) is equal to 0. Now excluding the multiplier m, the type I error component of the risk (see (3.5)) is equal to $R_1 = (1-p)t_1\delta_0$ while that for type II error is $R_2 = pt_2\delta_A$. The ratio of R_1 and R_2 becomes

$$\frac{R_1}{R_2} = \frac{\delta f \sqrt{u}\exp(-z_{u,v}/2)}{\sqrt{v}\log v}(1 + o_{u,v})$$

By the definition of v this is equal to

$$\frac{R_1}{R_2} = \exp(-z_{u,v}/2 - \log\log v)(1 + o_{u,v}) \tag{3.17}$$

and it converges to zero if $z_{u,v} + 2\log\log v \to \infty$. This shows that the overall risk is given by $R = mR_2(1 + o_{u,v})$, which in turn, equals the expressions in (3.2).

For the case when $C > 0$, exactly similar steps will give the required result, the only difference being that the ratio $\frac{R_1}{R_2}$ will be a different multiple of the expression in (3.17). This completes the proof of the sufficiency part.

We will now prove the necessity of both the conditions (3.14) and (3.15) for optimality to hold. At first we prove the necessity of condition (3.14)

Assume that (3.14) does not hold. Noting that $z_{u,v} \geq -\log v$, clearly this can happen if either (i) $\frac{z_{u,v}}{\log v}$ converges to a point in $[-1, \infty] - \{0\}$ or (ii) $\frac{z_{u,v}}{\log v}$ does not converge anywhere under assumption (A).

At first we consider the case :

- $\frac{z_{u,v}}{\log v} \to -1$ and there exist a constant C_1, $0 \leq C_1 < \infty$, and a subsequence of critical values $c_{u,v}$ such that $c_{u,v} \to C_1$. Observe that in this situation type I error

$$t_1 = P(|Z| > c_{u,v}) \to 2(1 - \Phi(C_1)) = C_2 > 0$$

and the type II error

$$t_2 = P\left(|Z| < \frac{c_{u,v}}{\sqrt{u+1}}\right) = \sqrt{\frac{2}{\pi}}\frac{C_1}{\sqrt{u}}(1+o_{u,v}) \ .$$

Thus the total risk

$$\begin{aligned} R_t &= m\left(\delta_0(1-p)C_2 + \delta_A p\frac{\sqrt{2}C_1}{\sqrt{\pi u}}\right)(1+o_{u,v}) \\ &= \frac{m\delta_A p}{\sqrt{u}}(\sqrt{v}C_2 + \sqrt{\frac{2}{\pi}}C_1) = \frac{C_2 m\delta_A p\sqrt{v}}{\sqrt{u}}(1+o_{u,v}) \ . \end{aligned}$$

In the result, in case when $C = 0$, the ratio of the total risk to the optimal risk provided in (3.2)

$$\frac{R_t}{R_{opt}} = C_2\sqrt{\frac{\pi v}{2\log v}}(1+o_{u,v}) \to \infty$$

and the corresponding rule is clearly not optimal.
Similarly, for $C \in (0,\infty)$,

$$\frac{R_t}{R_{opt}} = \frac{C_2}{2\Phi(\sqrt{C})-1}\sqrt{\frac{Cv}{\log v}}(1+o_{u,v}) \to \infty \ .$$

- Now, consider the case: $z_{u,v} = k\log v(1+o_{u,v})$, $k \in [-1,0)$, and $c_{u,v} \to \infty$. Observe that in this situation $z_{u,v} + 2\log\log v \to -\infty$ and the calculations leading to (3.17) easily show that

$$\frac{R_t}{R_{opt}} \geq \frac{R_1}{R_{opt}} \to \infty \ .$$

- Next, we consider the case when $k \in (0,\infty]$. Then there exists a constant $\epsilon > 0$ such that for sufficiently large v, $c_{u,v}^2 > \log v(1+\epsilon)$. Thus the type II error

$$t_2 = P\left(|Z| < \frac{c_{u,v}}{\sqrt{u+1}}\right) \geq P\left(|Z| < \sqrt{\frac{\log v(1+\epsilon)}{u+1}}\right) \ .$$

In the result

$$t_2 \geq \sqrt{\frac{2\log v(1+\epsilon)}{\pi u}}(1+o_{u,v}) \quad \text{when} \quad C = 0$$

or

$$t_2 \geq (2\Phi(\sqrt{C(1+\epsilon)}-1)(1+o_{u,v}) \quad \text{when} \quad C > 0 \ .$$

This implies that in both cases the asymptotic ratio of R_2 to R_{opt} is larger than 1 and the rule with the threshold $c_{u,v}$ is not optimal.

- Finally, we consider (ii). In this situation there will be at least two distinct points in $[-1, \infty]$ (and hence at least one point different from zero) to each of which some subsequence pair $(\frac{z_1'}{\log v}, \frac{z_2'}{\log v})$ converges. By the previous argument, at least along one such pair optimal risk properties will not hold and hence also not for the whole sequence.

To conclude the proof observe that the necessity of condition (3.15) easily follows from the necessity of condition (3.14) and the calculations in the proof of sufficiency. □

Corollary 1 *The mBIC type threshold*

$$c_{mBIC}^2 = \log \frac{n\tau^2}{\sigma^2} - 2\log p + d \ ,$$

where $d \in \mathbf{R}$, is asymptotically optimal when $p \to 0$, $\log(\delta) = o(\log v)$ and $\log(\delta) - \log\log v \to -\infty$. (In particular δ may be constant).

Corollary 2 *If $\frac{\tau}{\sigma} = const$, $\delta = const$ and $pm \to s$, where $0 < s < \infty$, then the standard version of mBIC, proposed in Bogdan et al. (2004),*

$$c_{mBIC}^2 = \log n + 2\log m + d$$

is asymptotically optimal.

4. Controlling Bayesian False Discovery Rate

Bayesian False Discovery Rate, BFDR, is defined as

$$BFDR = P(H_{0i} \text{ is true} | H_{0i} \text{ was rejected}) = \frac{(1-p)t_1}{(1-p)t_1 + p(1-t_2)} \ . \tag{4.18}$$

Let us define $t_{u,v,\delta} = u\delta^2 \log v$.

Lemma 1 *Assume that the Assumption (A) holds. If $t_{u,v,\delta} \to \infty$ then BFDR of the Bayes oracle is of the form*

$$BFDR_{BO} = \frac{e^{-C/2}}{D}\sqrt{\frac{2}{\pi t_{u,v,\delta}}}(1 + o_{u,v}) \ , \tag{4.19}$$

where $D = 2(1 - \Phi(\sqrt{C}))$.

If $t_{u,v,\delta} \to C_1^2$, where $0 < C_1 < \infty$ then

$$BFDR_{BO} = \frac{1}{1 + \sqrt{\frac{\pi}{2}}e^{C/2}DC_1}(1 + o_{u,v}) \tag{4.20}$$

Proof.

$$BFDR_{BO} = \frac{1}{1+\frac{1-t_2}{ft_1}} , \qquad (4.21)$$

where $f = \frac{1-p}{p}$. The thesis of the lemma follows by observing that (3.9) yields

$$ft_1 = e^{-C/2}\sqrt{\frac{2}{\pi u \delta^2 \log v}}(1 + o_{u,v}) \ . \qquad (4.22)$$

and (3.12) results in

$$1 - t_2 = 2(1 - \Phi(\sqrt{C})) + o_{u,v} \ . \qquad (4.23)$$

□

Remark 4. We have shown that for $\delta = const$ BFDR of the Bayes oracle converges to zero at the rate $1/\sqrt{u \log v}$. Specifically, in case when $p \to 0$ and $u = -c \log p$ (i.e. we are at the verge of detectability) BFDR of the Bayes oracle converges to zero at the rate $(-\log p)^{-1}$. Note, also that the Bayes oracle has a fixed limiting BFDR if the ratio of losses converges to 0 in such a rate that $\delta^2 u \log v \to C_1$, where $C_1 < \infty$. This condition requires that the relative loss for type II increases slightly quicker than $u = \sqrt{n}\frac{\tau}{\sigma}$ (i.e. missed signals should be penalized more when we expect that they are relatively large). For the signals of the magnitude $u = -c \log p$ this condition requires that δ is of the of rate $(-\log p)^{-1}$. Note that in this situation signals of the form $u = -c \log p$ are also at the verge of detectability.

Now, consider the rule controlling BFDR at the level $\alpha_{u,v}$. A threshold value for this rule c_B^2 can be obtained by solving the equation

$$\frac{(1-p)(1-\Phi(c_B))}{(1-p)(1-\Phi(c_B)) + p\left(1 - \Phi\left(\frac{c_B}{\sqrt{u+1}}\right)\right)} = \alpha_{u,v} \ . \qquad (4.24)$$

Theorem 3 *Assume (A) holds. Moreover, assume that* $\alpha_{u,v} \to \alpha_0 < 1$,

$$\frac{f}{\alpha_{u,v}} \to \infty \qquad and \qquad \frac{\log\left(\frac{f}{\alpha_{u,v}}\right)}{u} \to C_0 < \infty \ . \qquad (4.25)$$

Then the threshold value of the rule controlling BFDR at the level $\alpha_{u,v}$ *is given by the equation*

$$c_B^2 = 2\log\left(\frac{f}{\alpha_{u,v}}\right) - \log\left(2\log\left(\frac{f}{\alpha_{u,v}}\right)\right) + C_1 + o_{u,v} \ , \qquad (4.26)$$

where $C_1 = \ln\left(\frac{2(1-\alpha_0)^2}{\pi D^2}\right)$ *and* $D = 2(1 - \Phi(\sqrt{2C_0}))$ *is the asymptotic power. The corresponding type I error is of the form*

$$t_1 = \frac{D}{(1-\alpha_0)}\frac{\alpha_{u,v}}{f}(1 + o_\theta) \ .$$

Proof.

Equation (4.24) is equivalent to

$$\frac{1-\Phi(c_B)}{1-\Phi\left(\frac{c_B}{\sqrt{u+1}}\right)}=\frac{\alpha_{u,v}}{f(1-\alpha_{u,v})} \quad . \tag{4.27}$$

Now, observe that (4.25) and (4.27) can both hold only if

$$\limsup \frac{c_B}{\sqrt{u+1}}<\infty \quad . \tag{4.28}$$

Indeed, assume that $c_B = z_{u,v}\sqrt{u+1}$, where

$$\limsup z_{u,v}=\infty \quad . \tag{4.29}$$

The tail approximation (3.7) yields that

$$\frac{1-\Phi(c_B)}{1-\Phi\left(\frac{c_B}{\sqrt{u+1}}\right)}=\frac{\exp(-z_{u,v}^2u/2)}{\sqrt{u+1}}(1+o_{u,v}) \quad .$$

This together with (4.27) yield

$$z_{u,v}^2=\frac{2\log\left(\frac{f(1-\alpha_{u,v})}{\sqrt{u+1}\alpha_{u,v}}\right)}{u}+o_{u,v} \quad ,$$

and (4.29) contradicts(4.25).

Now, consider a subsequence of c_B such that $\frac{c_B}{\sqrt{u+1}} \to C_3 < \infty$ and the corresponding asymptotic power $D_1 = 2(1-\Phi(C_3)) > 0$. In this case (4.27) reduces to

$$1-\Phi(c_B)=\frac{D_1\alpha_{u,v}(1+o_{u,v})}{2f(1-\alpha_{u,v})} \quad .$$

Now (4.26) follows by using the tail approximation (3.7)to the left-hand side and some simple algebra. □

Comparing the asymptotic threshold of the BFDR controlling rule specified in (4.26) to the threshold of the asymptotically optimal rule provided in Theorem 2 we may expect that the BFDR controlling rule is asymptotically optimal if $\alpha \approx \frac{1}{\delta\sqrt{u}}$. The consecutive Lemma 2 states how close these two quantities need to be for the BFDR rule to be asymptotically optimal.

Lemma 2 *Under Assumptions of Theorem 3 BFDR controlling rule at the level $\alpha_{u,v}$ is asymptotically optimal if and only if*

$$\frac{\log(f\delta\sqrt{u})}{\log(f/\alpha_{u,v})}=1+s_{u,v} \quad , \tag{4.30}$$

where

$$s_{u,v}\to 0 \tag{4.31}$$

and

$$2s_{u,v}\log(f/\alpha_{u,v})-\log\log(f/\alpha_{u,v})\to-\infty \quad . \tag{4.32}$$

Proof. The threshold of BFDR controlling rule provided in (4.26) can be written as

$$c_B^2 = \log v - \log(u\delta^2\alpha^2 \log(f/\alpha)) + C + o_{u,v} \ .$$

Thus, from Theorem 2 we obtain that the rule is asymptotically optimal if and only if

$$\log(u\delta^2\alpha^2 \log(f/\alpha)) = o(\log v) \tag{4.33}$$

and

$$\frac{u\delta^2\alpha^2 \log(f/\alpha)}{(\log v)^2} \to 0 \ . \tag{4.34}$$

Note that

$$\log(u\delta^2\alpha^2 \log(f/\alpha)) = \log v - 2\log(f/\alpha) + \log\log(f/\alpha) \ .$$

Thus the condition (4.33) is fulfilled if and only if

$$\frac{2\log(f/\alpha)}{\log v} = \frac{2\log(f/\alpha)}{2\log(f\delta\sqrt{u})} \to 1$$

i.e. if and only if the condition (5.42) holds.

Now, observe that the condition (4.34) is equivalent to the condition

$$\frac{v\log(f/\alpha)}{(f/\alpha)^2(\log v)^2} \to 0 \ . \tag{4.35}$$

Moreover, using notation from (5.41) we have that

$$\log v = 2(1+s_{u,v})\log(f/\alpha)$$

and

$$v = (f/\alpha)^{2(1+s_{u,v})} \ .$$

Thus, assuming that $s_{u,v} \to 0$, (4.35) is equivalent to

$$\frac{(f/\alpha)^{2s_{u,v}}}{\log(f/\alpha)} \to 0$$

which is fulfilled if and only if (5.43) holds. □

Remark 5. The condition (4.34) provides much more room for small than for large α. This phenomenon reflects different sensitivities of type I and type II errors to the change of the threshold, already discussed in Remark 2.

Lemma 3 *Under Assumptions of Theorem 3 the rule controlling BFDR at the level* $\alpha_{u,v} = \frac{s}{\delta\sqrt{u}}$, *with* $s \in (0,\infty)$, *is asymptotically optimal. The corresponding critical value is of the form*

$$c_{BFDR}^2 = \log v - \log\log v + C_4 + o_{u,v} \ , \tag{4.36}$$

where $C_4 = \ln\left(\frac{2(1-\alpha_0)^2}{\pi D^2 s^2}\right)$. *The corresponding type I error is of the form*

$$t_{1B} = \frac{Ds}{(1-\alpha_0)\sqrt{v}}(1+o_{u,v}) \ . \tag{4.37}$$

Note that the threshold of the proposed optimal BFDR controlling rule is smaller by $O(\log\log v)$ term from the threshold of the Bayes oracle, provided in Theorem 1. In the result, the rate of the corresponding type I error is larger than for the Bayes oracle (it lacks $\sqrt{\log v}$ term in the denominator). However, the total risk is still determined by essentially larger type II error component and is of the same rate as for the Bayes oracle.

Lemma 2 states that BFDR controlling rule is asymptotically optimal if the corresponding BFDR level α decreases when the signal magnitude u and the relative cost of type I error δ increase. The following lemma throws some light on the behavior of the BFDR controlling rule with a fixed BFDR level α.

Lemma 4 *Under Assumptions of Theorem 3 BFDR controlling rule at a fixed level α is optimal if and only if the ratio of loss functions converges to 0 in such a rate that*

$$\frac{\log(\delta^2 u)}{\log p} \to 0 \ . \tag{4.38}$$

and

$$\frac{\delta^2 u}{\log p} \to 0 \tag{4.39}$$

Proof. Observe that

$$\frac{\log(f\delta\sqrt{u})}{\log(f/\alpha)} - 1 = \frac{\log(\delta\alpha\sqrt{u})}{\log(f/\alpha)} \ .$$

Note also that that when α is fixed assumptions of Theorem 3 imply that $p \to 0$. Thus the condition (5.42) of Lemma 2 reduces to (??). To complete the proof observe that the condition (5.43) of Lemma 2 implies that

$$2\log(\delta\sqrt{u}) - \log\log f \to \infty \ ,$$

which is true if and only if (5.45) holds. □

Lemma 3 states that under Assumption (A) the rule controlling BFDR at a fixed level α can be optimal only if the relative cost of type II error increases when $p \to 0$ and $u \to \infty$. Specifically, it implies that such a rule can not be optimal with respect to minimizing the misclassification rate. This result illustrates different aspects of the behavior of BFDR controlling rule than those discussed in Abramovich et al (2006). Note that Abramovich et al (2006) proved that in the context of the estimation of μ, the rule controlling FDR at the level $\alpha \to \alpha_0 \leq 0.5$ is asymptotically minimax. To show some similarities between our results and the results of Abramovich et al (2006) we now present two lemmas describing the behavior of BFDR controlling rules for the signals on the verge of detectability. First of these lemmas states that for the signals at the verge of detectability, BFDR controlling rule with a fixed BFDR level $\alpha \in (0,1)$ is asymptotically optimal if and only if the ratio of loss functions δ decreases to 0 at a very slow rate. The second lemma, dual to the first one, states that for the signals at the verge of detectability and when $\delta = const$, the BFDR controlling rule is asymptotically optimal if and only if BFDR level α decreases to zero at the same, very slow, rate. This last result can be specifically used to judge the optimality of the BFDR controlling rule with respect to minimizing the misclassification error.

Lemma 5 *Assume that* $p \to 0$ *and* $-\frac{2\log p}{u} \to C$, *with* $0 < C < \infty$. *Then the rule controlling BFDR at level* α, $0 < \alpha < 1$ *is asymptotically optimal if and only if* $\delta \to 0$ *in such a rate that* $\frac{\log \delta}{\log p} = 0$.

Proof. Direct consequence of Lemma 3. □

Note that under Assumptions of Lemma 4 $\frac{2\log(f\delta)}{u} \to C$, i.e. u is at the verge of detectability.

Lemma 6 *Assume that* $p \to 0$, $\delta = const$, *and* $-\frac{2\log p}{u} \to C$, *with* $0 < C < \infty$. *Then the rule controlling BFDR at level* α, *is asymptotically optimal if and only if* $\alpha \to 0$ *in such a rate that* $\frac{\log \alpha}{\log p} = 0$.

Proof. The proof is analogous to the proof of Lemma 3 and therefore omitted. □

Note that, as expected, the level of BFDR for the optimal Bayes rule, specified in Remark 4, satisfies the Assumption of Lemma 5. Also, the ratio of loss functions yielding a Bayes oracle with a fixed BFDR, specified in Remark 4, satisfies the Assumption of Lemma 4. However, the conditions of Lemma 4 and 5 illustrate much wider range of flexibility in the choice of these parameters, so the BFDR controlling is still optimal.

Lemma 7 *Let* c_{GW} *denote the approximation to the BH threshold provided in [16]:*

$$c_{GW} : \frac{(1 - \Phi(c_{GW}))}{(1-p)(1-\Phi(c_{GW})) + p\left(1 - \Phi\left(\frac{c_{GW}}{\sqrt{u+1}}\right)\right)} = \alpha_{u,v} \ . \qquad (4.40)$$

When $p \to 0$ *and the assumptions of Theorem 3 are fulfilled then*

$$c_{GW}^2 = c_B^2 + o_{u,v} \ ,$$

where c_B *is the BFDR controlling rule at the level* α.

Proof.

The proof follows the line of the proof of Theorem 3 and therefore is omitted. □

Corollary 3 *The multiple testing rule based on the threshold* c_{GW} *is asymptotically optimal when* $p \to 0$ *and the assumptions of Lemma 2 are fulfilled.*

5. Optimality of the Bonferroni correction

The Bonferroni correction at the level α rejects all the null hypothesis for which $Z_i = \frac{|X_i|}{\sigma}$ exceeds the threshold

$$c_{Bon} : 1 - \Phi(c_{Bon}) = \frac{\alpha}{2m} \ .$$

Under the assumption that $m \to \infty$, the threshold of the Bonferroni correction can be written as

$$c_{Bon}^2 = 2\log\left(\frac{m}{\alpha}\right) - \log\left(2\log\left(\frac{m}{\alpha}\right)\right) + \log(2/\pi) + o_{u,v} \ .$$

By the comparison with the asymptotic expansion of the BFDR controlling rule (4.26) and the line of proofs of Lemmas 2-6 one can easily derive the optimality properties of the Bonferroni correction for the case when $pm \to c$, with $c \in (0, \infty)$.

Lemma 8 *Assume that $m \to \infty$ and $pm \to c$, with $c \in (0, \infty)$. Moreover assume that the Assumption (A) holds. Bonferroni correction at the level $\alpha_{u,v}$ is asymptotically optimal if and only if*

$$\frac{\log(m\delta\sqrt{u})}{\log(m/\alpha_{u,v})} = 1 + s_{u,v} \ , \tag{5.41}$$

where

$$s_{u,v} \to 0 \tag{5.42}$$

and

$$2s_{u,v}\log(m/\alpha_{u,v}) - \log\log(m/\alpha_{u,v}) \to -\infty \ . \tag{5.43}$$

Lemma 9 *Under the assumptions of Lemma (8) Bonferroni correction at the level $\alpha_{u,v} = \frac{s}{\delta\sqrt{u}}$, with $s \in (0, \infty)$, is asymptotically optimal.*

Lemma 10 *Under the assumptions of Lemma (8) Bonferroni correction at a fixed FWER level α is optimal if and only if the ratio of loss functions converges to 0 in such a rate that*

$$\frac{\log(\delta^2 u)}{\log m} \to 0 \ . \tag{5.44}$$

and

$$\frac{\delta^2 u}{\log m} \to 0 \tag{5.45}$$

Lemma 11 *Assume that $m \to \infty$ and $pm \to c$, with $c \in (0, \infty)$. Moreover assume that $\frac{2\log m}{u} \to C$, with $0 < C < \infty$. Then the Bonferroni at the level α, $0 < \alpha < 1$ is asymptotically optimal if and only if $\delta \to 0$ in such a rate that $\frac{\log\delta}{\log m} = 0$.*

Lemma 12 *Assume that $m \to \infty$ and $pm \to c$, with $c \in (0, \infty)$. Moreover assume that $\frac{2\log m}{u} \to C$, with $0 < C < \infty$ and $\delta = const$ Then the Bonferroni correction at the level α, is asymptotically optimal if and only if $\alpha \to 0$ in such a rate that $\frac{\log\alpha}{\log m} = 0$.*

6. Optimality of the Benjamini and Hochberg procedure

In this section we will prove the optimality of the Benjamini Hochberg procedure for p converging to zero in such a rate that for sufficiently large m, $p_m > \frac{log^\gamma m}{m}$, for some constant $\gamma > 1$. We also provide an auxiliary result, which suggests that BH procedure with α converging to zero is asymptotically optimal also when $mp \to C$, with $C \in (0, \infty)$.

Let $Z_i^2 = \frac{X_i^2}{\sigma^2}$ and $p_i = 2(1 - \Phi(Z_i))$ be the corresponding p-value. Let us sort p-values in the ascending order $p_{(1)} \leq p_{(2)} \leq \ldots \leq p_{(m)}$ and let

$$k_F = argmax_i \left\{ p_{(i)} \leq \frac{i\alpha}{m} \right\} . \tag{6.46}$$

BH at the FDR level α rejects all hypotheses for which the corresponding p-values are smaller or equal than $p_{(k)}$.

Let us denote $1 - \hat{F}_m(y) = \#\{Y_i \geq y\}/m$. It is easy to check (eg. see [11]) that the Benjamini-Hochberg procedure rejects the null hypothesis H_{0i} when $Z_i^2 \leq \tilde{c}_{BH}^2$ where

$$\tilde{c}_{BH} = \inf\{y : \frac{2(1 - \Phi(y))}{1 - \hat{F}_m(y)} \leq \alpha\} .$$

Remark: BH as the empirical Bayes version of BFDR.

Note also that BH rejects the null hypothesis H_{0i} whenever Z_i^2 exceeds the threshold of the Bonferroni correction. Therefore we define the random threshold for BH procedure as

$$c_{BH} = \min\{c_{Bon}, \tilde{c}_{BH}\} .$$

The proof of the optimality of BH will consist of two steps. In the first step we will show that the corresponding type II component of the risk $R_A = \delta_A EL_A$, where L_A is the number of false negatives, is of the optimal rate.

To prove the theorem on the optimality of the type II component of the risk we will at first prove two auxiliary lemmas.

Lemma 13 *Consider the multiple testing rule based on the GW threshold, defined in (4.40). Assume that $p \to 0$ and that the level α is chosen in such a way that the assumptions of Theorem 3 are satisfied. Then for any constant $\xi \in (0, 1)$ it holds*

$$P\left(\frac{1 - \hat{F}_m(c_{GW})}{1 - F(c_{GW})} > (1 + \xi)\right) \leq \exp\left\{-\frac{1}{4} mpC_1\xi^2(1 + o_{u,v})\right\} \tag{6.47}$$

and

$$P\left(\frac{1 - \hat{F}_m(c_{GW})}{1 - F(c_{GW})} < (1 - \xi)\right) \leq \exp\left\{-\frac{1}{4} mpC_1\xi^2(1 + o_{u,v})\right\} , \tag{6.48}$$

where

$$C_1 = \frac{D}{1 - \alpha_0} > 0 . \tag{6.49}$$

Proof. Let $1-F(c_{GW}) = (1-p)t_1(c_{GW})+p(1-t_2(c_{GW}))$, where $t_1(c_{GW})$ and $t_2(c_{GW})$ denote the type I and type II errors of the procedure based on GW threshold. Lemma 7 implies that $t_1(c_{GW})$ and $t_2(c_{GW})$ are of the same rate that the corresponding type I and type II errors of the BFDR controlling rule. Therefore, using Theorem 3 we obtain

$$1 - F(c_{GW}) = pC_1(1 + o_{u,v}) \ . \tag{6.50}$$

Now, observe that $Y = m(1 - \hat{F}_m(c_{GW}))$ is a Binomial $B(m, 1 - F(c_{GW})$ random variable. Therefore, by the Bennett's inequality (e.g. see [33], page 440) it follows that

$$P(Y > m(1 - F(c_{GW}))(1 + \xi)) \leq \exp\left\{-\frac{1}{4}mpC_1\xi^2(1 + o_{u,v})\right\}$$

and

$$P(Y < m(1 - F(c_{GW}))(1 - \xi)) \leq \exp\left\{-\frac{1}{4}mpC_1\xi^2(1 + o_{u,v})\right\}$$

and the proof of Lemma 13 is completed. □

Lemma 14 *Assume that α and p satisfy the assumptions of Lemma 13. Let c_{BH} be the BH threshold at the level α and let $\tilde{c}_1$ be the GW threshold at the level $\alpha_1 = \alpha(1 - \xi)$, where ξ is a constant included in $(0, 1)$. It holds that*

$$P(c_{BH} \geq \tilde{c}_1) \leq \exp\left\{-\frac{1}{4}mpC_1\xi^2(1 + o_{u,v})\right\} \ .$$

Proof. By the defintion of c_{BH} it holds that

$$\begin{aligned} P(c_{BH} < \tilde{c}_1) &\geq P\left(\frac{2(1 - \Phi(\tilde{c}_1))}{(1 - \hat{F}_m(\tilde{c}_1))} \leq \alpha\right) \\ &= P\left(\frac{2(1 - \Phi(\tilde{c}_1))}{1 - F(\tilde{c}_1)} \frac{1 - F(\tilde{c}_1)}{1 - \hat{F}_m(\tilde{c}_1)} \leq \alpha\right) = P\left(\frac{1 - \hat{F}_m(\tilde{c}_1)}{1 - F(\tilde{c}_1)} \geq 1 - \xi\right) \ . \end{aligned}$$

Now the thesis of Lemma 14 follows by invoking Lemma 13. □

Theorem 4 *Consider BH rule at the level α, which satisfies the assumption of Lemma 2. Assume that $m \to \infty$, $p_m \to 0$ in such a rate that for some constant $\gamma > 1$, $p_m \geq \frac{(\log m)^\gamma}{m}$. Moreover, assume that the signal magnitude u_m satisfies $\frac{2(\log \delta_m - \log p_m)}{u_m} \to C \in [0, \infty)$ (i.e. Assumption (A) holds) and that*

$$u_m \leq \exp\{2C_2mp\} \ , \tag{6.51}$$

for some constant $C_2 \in \left(0, \frac{1}{4}C_1\right)$, where C_1 is defined in Lemmma 13. Then the type II error component of the risk of BH satisfies

$$R_A \leq R(1 + o_{u,v}) \ ,$$

where R is the optimal risk provided in Theorem 1.

Remark. The condition (6.57) can be relaxed to $C_2 m p_m + \frac{1}{2}\log\left(\frac{\log v}{u}\right) \to \infty$.

Proof. Note that that the expected number of false negatives satisfies

$$E(L_A) \leq E(L_A | c_{BH} \leq \tilde{c}_1) P(c_{BH} \leq \tilde{c}_1) + mP(c_{BH} \geq \tilde{c}_1) \ .$$

Now observe that

$$E(L_A | c_{BH} \leq \tilde{c}_1) P(c_{BH} \leq \tilde{c}_1) \leq E(L_1 | c_{BH} \leq \tilde{c}_1) P(c_{BH} \leq \tilde{c}_1) \leq EL_1 \ ,$$

where L_1 is the expected number of false negatives produced by the rule based on the threshold $\tilde{c}_1$. Note also that α_1 satisfies the assumptions of Lemma 2 and therefore the rule based on $\tilde{c}_1$ is asymptotically optimal, i.e. $\delta_A EL_1 = R(1 + o_m)$. Thus

$$R_A = \delta_A EL_A \leq R(1 + o_{u,v}) + m\delta_A P(c_{BH} \geq \tilde{c}_1)$$

and by Lemma 14

$$R_A = R(1 + o_m) + m\delta_A \exp\left\{-\frac{1}{4} mpC_1 \xi^2 (1 + o_m)\right\} \ . \qquad (6.52)$$

The thesis of Theorem 4 follows by observing that under (6.57) the second component on the righthand-side of (6.58) is $o(R)$. □

To prove the optimality of the type I error component of the risk of BH we will at first show that with a very large probability c_{BH} can be bounded from below with an asymptotically optimal BFDR controlling rule.

Lemma 15 *Assume that $p_m \to 0$ in such a rate that $p_m > \frac{(\log m)^\gamma}{m}$, for some constant $\gamma > 1$. Moreover assume that the FDR level α_m satisfies the assumptions of Lemma 13. Let c_{BH} be the BH threshold at the level α_m and let c_2 be the GW threshold at the level $\alpha_{2m} = \alpha_m(1 + \xi)$, where ξ is a constant included in $\left(0, \min\left(1, \frac{1}{\alpha_0} - 1\right)\right)$. It holds that for every constant $\eta > 0$ and sufficiently large m*

$$P(c_{BH} < \tilde{c}_2) \leq m^{-\eta} \ .$$

Proof. Let c_0 be the solution to the equation

$$1 - F(c_0) = \frac{\log m}{\sqrt{m}} \ .$$

At first we consider the case when $c_2 < c_0$ (or equivalently when $p_m > \frac{1}{C_1}\frac{\log m}{\sqrt{m}}(1 + o_m)$, where C_1 is defined in (6.49)). Note that

$$P(c_{BH} < c_2) = P\left(\frac{2(1 - \Phi(c_{BH}))}{1 - F(c_{BH})} \geq \alpha(1 + \xi)\right) \ . \qquad (6.53)$$

Now, observe that by the definition of c_{BH}

$$\frac{2(1 - \Phi(c_{BH}))}{1 - \hat{F}_m(c_{BH})} \leq \alpha \ .$$

Thus

$$\begin{aligned} P(c_{BH} < c_2) &\leq P\left(\sup_{c\in[0,c_2]} \frac{1-\hat{F}_m(c)}{1-F(c)} \geq (1+\xi)\right) \\ &= P(\sup_{c\in[0,c_2]} F(c) - \hat{F}_m(c) \geq (1-F(c))\xi) \\ &\leq P(\sup_{c\in[0,\infty]} |F(c) - \hat{F}(m)| \geq \xi\frac{\log m}{\sqrt{m}}) \ . \end{aligned} \quad (6.54)$$

Now, by invoking Rubin-Sethuraman [27] innequality for moderate deviations we obtain

$$P(c_{BH} < c_2) \leq \exp(-2\xi^2 \log^2 m)$$

and the thesis of the Lemma 15) easily follows.

Now, consider the case when $c_2 > c_0$. In this case

$$P(c_{BH} < c_2) = P(c_{BH} < c_0) + P(c_{BH} \in [c_0, c_2]) \ .$$

Repeating the arguments from (6.53) and (6.54) we can easily show that

$$P(c_{BH} < c_0) \leq \exp(-2\xi^2 \log^2 m) \ .$$

Now, observe that, similarly as in (6.54),

$$P(c_{BH} \in [c_0, c_2]) \leq P\left(\sup_{c\in[c_0,c_2]} \frac{1-\hat{F}_m(c)}{1-F(c)} \geq (1+\xi)\right) \ .$$

Using the standard uniform transformation $U_i = F\left(\frac{|X_i|}{\sigma}\right)$, we obtain

$$P(c_{BH} \in [c_0, c_2]) \leq P\left(\sup_{t\in[z_{1m},z_{2m}]} \frac{1-\hat{G}_m(t)}{1-t} \geq (1+\xi)\right) \ ,$$

where $z_{1m} = F(c_0) = 1 - \frac{\log m}{\sqrt{m}}$, $z_{2m} = F(c_2) = 1 - C_1 p_m(1 + o_m)$, and $\hat{G}_m(t)$ is the empirical cdf of the uniform distribution on $[0, 1]$.

Let $u_i = \frac{i}{m}$, $k_{1m} = \lfloor m - \sqrt{m}\log m \rfloor$, and $k_{2m} = \lceil m(1 - C_1 p_m) \rceil$. Observe that

$$k_{2m} - k_{1m} \leq \sqrt{m}\log m \ .$$

By the monotonicity of $G_m(t)$ and t it holds

$$P\left(\sup_{t\in[z_{1m},z_{2m}]} (1-\hat{G}_m(t)) \geq (1-t)(1+\xi)\right) \leq$$
$$\sum_{i=k_{1m}}^{k_{2m}} P\left(1-\hat{G}_m(u_i) \geq (1-u_i-\frac{1}{m})(1+\xi)\right) \ . \quad (6.55)$$

Observe that

$$1 - u_i \geq C_1\frac{\log(m)^\gamma}{m} - \frac{1}{m}$$

and therefore

$$(1 - u_i - \frac{1}{m}) = (1 - u_i)(1 + o_m) \ .$$

Now, from Benett's inequality, we obtain that for every $i \in (k_{1m}, k_{2m})$

$$P(1 - \hat{G}_m(u_i) \geq (1 - u_i)(1 + \xi)(1 + o_m)) \ \leq \ \exp\left(-\frac{1}{4}m(1 - u_i)\xi^2\right)$$
$$\exp\left(-\frac{1}{4}(\log m)^\gamma \xi^2 (1 + o_m)\right) \qquad , \tag{6.56}$$

and the thesis of Lemma (15) easily follows. □

Theorem 5 *Assume that the assumptions of Theorem 4 and Lemma 15 hold. Moreover, assume that for some positive constant* γ_2

$$\delta\sqrt{u} \leq m^{\gamma_2} \ . \tag{6.57}$$

Then the Benjamini-Hochberg procedure is asymptotically optimal.

Proof. Let L_0 be the number of false rejections produced by BH. Note that

$$E(L_0) \leq E(L_0 | c_{BH} \geq c_2) P(c_{BH} \geq c_2) + m P(c_{BH} \leq c_2) \ .$$

Now observe that

$$E(L_0 | c_{BH} \geq c_2) P(c_{BH} \leq c_2) \leq E(L_2 | c_{BH} \geq c_2) P(c_{BH} \geq c_2) \leq E L_2 \ ,$$

where L_2 is the expected number of false rejections produced by the rule based on the threshold $\tilde{c}_2$. Note also that α_2 satisfies the assumptions of Lemma 2 and therefore the rule based on c_2 is asymptotically optimal and $\delta_0 E L_2 = o(R)$. Thus, the null component of the risk

$$R_0 = \delta_0 E L_0 \leq o(R) + m\delta_0 P(c_{BH} \leq c_2)$$

and by Lemma 15, for sufficiently large m,

$$R_0 \leq o(R) + \delta_0 m^{-\gamma_2 + 1} \ , \tag{6.58}$$

The thesis of Theorem 5 follows by observing that under (6.57) the second component on the righthand-side of (6.58) is $o(R)$. □

Now we will consider the performance of BH under the assumption that $m \to \infty$, $p \to 0$ and $mp \to C_5$, where $0 < C_5 < \infty$. According to our knowledge the results on the optimality of BH in this limiting case have not been proven yet. Note that under this assumption the Bonferroni controlling FWER at the level $\alpha = \frac{s}{\delta\sqrt{u}}$ is asymptotically optimal and differs from the asymptotically optimal BFDR controlling rule only by a constant. It is also easy to check that in this case a fixed threshold FDR controlling rule is asymptotically optimal. Moreover, based on the comparison with Bonferroni correction it holds that the type II error of the component of the risk of BH is of the optimal rate. All these facts together suggest that the BH procedure is asymptotically optimal also in this limiting case. Below we present an illustrative theorem, which states that if $\alpha \to 0$ then for any $\epsilon > 0$ the BH threshold can be bounded by two asymptotically optimal rules with a probability larger than $1 - \epsilon$.

Let us denote by c_{BH} the threshold of the Benjamini and Hochberg procedure at the level $\alpha = \frac{s}{\delta\sqrt{u}}$.

Theorem 6 *Assume that* $m \to \infty$, $p \to 0$ *and*

$$mp \to C_5 \ , \tag{6.59}$$

where $0 < C_5 < \infty$. *Moreover, assume that* $\delta\sqrt{u} \to \infty$ *and that the Assumption (A) holds. Then for every* $\epsilon > 0$ *there exists a constant* $D_2 > 0$ *such that for sufficiently large* m

$$P(|c^2_{BH} - c^2_{BFDR}| > D_2) < \epsilon \ . \tag{6.60}$$

Proof.

Observing that under the assumption (6.59) c_{Bon} is different from c_{BFDR} only by a constant, it is enough to prove that for any $\epsilon > 0$ there exists a constant D_2 such that for sufficiently large m

$$P(c^2_{BH} < c^2_{BFDR} - D_2) < \epsilon \ . \tag{6.61}$$

Let us denote by R_0 and R_A the number of null and alternative hypothesis rejected by BH. Note that BH controls FDR at the level $\alpha = \frac{s}{\delta\sqrt{u}}$. Therefore

$$P\left(\frac{R_0}{R_0 + R_A} \geq 0.5\right) = P(R_0 \geq R_A) \leq 2\alpha \ . \tag{6.62}$$

Note also that $R_A \leq m_A$, where m_A is the total number of alternative hypothesis. Now, observe that m_A is a Binomial random variable $B(m, p)$. Thus, by Benett's inequality and assumption (6.59), we get that for any natural number $D_2 > C_5$

$$P(m_A \geq D_2) \leq \exp\left(-\frac{1}{4}(D_2 - C_5)(1 + o_m)\right) \ . \tag{6.63}$$

(6.62) and (6.63) yield

$$P(R_0 + R_A \geq 2D_2) \leq 2\alpha + \exp\left(-\frac{1}{4}(D_2 - C_5)(1 + o_m)\right) \ . \tag{6.64}$$

Let $D_3 = 2D_2$ and let z_{D_3} be the BH threshold for D_3-th largest test statistic. Note that z_{D_3} is given by the equation

$$2(1 - \Phi(z_{D3})) = \frac{\alpha D_3}{m} = \frac{sD_3}{C_5\sqrt{v}} \ .$$

Thus

$$z_{D3}^2 = \log v - \log\log v - 2\log\left(\frac{sD_3}{C_5}\right) + o_m \ .$$

Now observe that the event $\{c_{BH}^2 < z_{D_3}^2\}$ implies that $R_0 + R_A \geq D_3$. Therefore, by (6.64), we obtain that

$$P(c_{BH}^2 < z_{D3}) \leq 2\alpha + \exp\left(-\frac{1}{4}(D_3/2 - C_5)(1 + o_m)\right)$$

and the proof is completed by observing that $\alpha \to 0$ as $m \to \infty$. □

7. Discussion

We have investigated the asymptotic optimality of the multiple testing rules under sparsity, using the framework of the Bayesian decision theory. Similarly as in [1] we have proved some asymptotic optimality properties of the rules controlling the false discovery rate. However, in our setup the proofs of the asymptotic optimality of FDR controlling rules are substantially simpler than the proofs in [1]. Moreover, our results provide some hints on how the "optimal" FDR level should be chosen depending on the expected magnitude of true signals and the ratio of loss functions for type I and type II errors.

One could ask why we investigate the properties of the Benjamini-Hochberg procedure or the Bonferroni correction within the framework of the Bayes decision theory. It seems that a more natural way to mimic the Bayes oracle is to use a plug-in method based on the estimated mixture parameters or to use a full Bayes approach and integrate the unknown parameters with respect to some prior distribution. The advantages of these approaches, both in parametric and nonparametric settings, were illustrated e.g. in [37], [10], [7] and [8]. However, as noted in [7], the applicability of these methods is limited by serious problems with the estimation of the mixture parameters when the mixing parameter p is very small. These difficulties occur even in the simple parametric context and are related to the problems with the nonidentifiability of the mixture parameters. The strength of BH lies in the fact that it requires only the estimate of the cdf of the mixture, which for large m can be quite accurate even when p is approaching $\frac{1}{m}$. Therefore, as shown by the simulation study reported in [7], BH can outperform the rules directly mimicking the Bayes oracle when p is very small. The results reported in this article give a thorough theoretical explanation of this phenomenon.

Our results are derived under the assumption that the data are generated by the scale mixture of normal distributions. This allowed us to reduce the technical complexity of the proofs and concentrate on the main aspects of the problem. Further extensions to the situation where the prior on μ_i comes from a general scale family are possible and are the topic of our ongoing research.

Acknowledgment. MB would like to thank Florian Frommlet for many helpful suggestions.

References

[1] Abramovich F., Benjamini Y., Donoho D. L., Johnstone I. M. Adapting to unknown sparsity by controlling the false discovery rate *Ann. Statist.* 2006 **34**, 584-653.

[2] Donoho D., Jin J. Higher criticism for detecting sparse heterogeneous mixtures. *Ann. Stat.* 2004, **32** 962-994.

[3] Benjamini, Y. and Hochberg, Y. (1995). Controlling the false discovery rate: a practical and powerful approach to multiple testing. *J. Roy. Statist. Soc. Ser. B*, **57**(1), 289–300.

[4] Benjamini, Y. and Hochberg, Y. (2000). On the adaptive control of the false discovery fate in multiple testing with independent statistics. *J. Educ. Behav. Stat.*, **25**, 60–83.

[5] Benjamini, Y. and Yekutieli, D. (2001). The control of the false discovery rate in multiple testing under dependency. *Ann. Statist.*, **29**(4), 1165–1188.

[6] Bogdan, M., Ghosh, J. K., and Doerge, R. W. (2004). Modifying the schwarz bayesian information criterion to locate multiple interacting quantitive trait loci. *Genetics*, **167**, 989–999.

[7] Bogdan, M., Ghosh, J. K., Ochman, A. and Tokdar S.T. (2007) On the Empirical Bayes approach to the problem of multiple testing. *Quality and Reliability Engineering International*, **23**, 727–739.

[8] Bogdan, M., Ghosh, J. K. and Tokdar S. T. (2008) A comparison of the Simes-Benjamini-Hochberg procedure with some Bayesian rules for multiple testing. IMS Collections, **Vol.1**, Beyond Parametrics in Interdisciplinary Research: Fetschrift in Honor of Professor Pranab K. Sen, edited by N. Balakrishnan, Edsel Peña and Mervyn J. Silvapulle, pp. 211–230, 2008, Beachwood Ohio.

[9] Donoho, D. and Jin, J. (2004). Higher criticism for detecting sparse heterogeneous mixtures. *Ann. Statist.*, **32**(3), 962–994.

[10] Efron, B. (2008). Microarrays, Empirical Bayes and the Two-Groups Model *Statist. Sci.*, **23**, 1-22.

[11] Efron, B. and Tibshirani, R. (2002). Empirical bayes methods and false discovery rates for microarrays. *Genetic Epidemiology*, **23**, 70–86.

[12] Efron, B., Tibshirani, R., Storey, J. D., and Tusher, V. (2001). Empirical Bayes analysis of a microarray experiment. *J. Amer. Statist. Assoc.*, **96**(456), 1151–1160.

[13] Elmore, R., Hall, P., and Neeman, A. (2005). An application of classical invariant theory to identifiability in nonparametric mixtures. *Ann. Inst. Fourier (Grenoble)*, **55**(1), 1–28.

[14] Escobar, M. D. and West, M. (1995). Bayesian density estimation and inference using mixtures. *J. Amer. Statist. Assoc.*, **90**(430), 577–588.

[15] Ferguson, T. S. (1973). A Bayesian analysis of some nonparametric problems. *Ann. Statist.*, **1**, 209–230.

[16] Genovese, C. and Wasserman, L. (2002). Operating characteristics and extensions of the false discovery rate procedure. *J. R. Stat. Soc. Ser. B Stat. Methodol.*, **64**(3), 499–517.

[17] Genovese, C. and Wasserman, L. (2004). A stochastic process approach to false discovery control. *Ann. Statist.*, **32**(3), 1035–1061.

[18] Ghosh, J. K. and Ramamoorthi, R. V. (2003). *Bayesian nonparametrics.* Springer Series in Statistics. Springer-Verlag, New York.

[19] Ghosh, J. K. and Sen, P. K. (1985). On the asymptotic performance of the log likelihood ratio statistic for the mixture model and related results. In *Proceedings of the Berkeley conference in honor of Jerzy Neyman and Jack Kiefer, Vol. II (Berkeley, Calif., 1983)*, Wadsworth Statist./Probab. Ser., pages 789–806, Belmont, CA. Wadsworth.

[20] Holm, S. (1979). A simple sequentially rejective multiple test procedure. *Scand. J. Statist.*, **6**(2), 65–70.

[21] Lehmann, E. L. and Romano, J. P. (2005). Generalizations of the familywise error rate. *Ann. Statist.*, **33**(3), 1138–1154.

[22] Meinshausen, N. and Rice, J. (2006). Estimating the proportion of false null hypotheses among a large number of independently tested hypotheses. *Ann. Statist.*, **34**(1), 373–393.

[23] Müller, P., Parmigiani, G., Robert, C., and Rousseau, J. (2004). Optimal sample size for multiple testing: the case of gene expression microarrays. *J. Amer. Statist. Assoc.*, **99**(468), 990–1001.

[24] Müller, P., Parmigiani, G., and Rice, K. (2006). Fdr and bayesian multiple comparison. In ??, editor, *Proc. Valencia/ISBA 8-th World Meeting on Bayesian Statistics*, page ??, ?? ??, ??

[25] Newton, M. A. (2002). On a nonparametric recursive estimator of the mixing distribution. *Sankhyā Ser. A*, **64**(2), 306–322. Selected articles from San Antonio Conference in honour of C. R. Rao (San Antonio, TX, 2000).

[26] Otto, S. P. and Jones, C. D. (2000). Detecting the undetected: Estimating the total number of loci underlying a quantitative trait. *Genetics*, **156**, 2093–2107.

[27] Rubin, H. amd Sethuraman, J. (1965). Probabilities of moderate deviations. *Sankhya*, **27**, 325–346.

[28] Sarkar, S. and Chen, J. (??). A bayesian stepwise multiple testing procedure. Technical report, available at http://astro.temple.edu/ sanat/reports.html.

[29] Sarkar, S. K. (2002). Some results on false discovery rate in stepwise multiple testing procedures. *Ann. Statist.*, **30**(1), 239–257.

[30] Sarkar, S. K. (2006). False discovery and false nondiscovery rates in single-step multiple testing procedures. *Ann. Statist.*, **34**(1), 394–415.

[31] Scott, J. G. and Berger, J. O. (2006). An exploration of aspects of Bayesian multiple testing. *J. Statist. Plann. Inference*, **136**(7), 2144–2162.

[32] Seeger, P. (1968). A note on a method for the analysis of significance en masse. *Technometrics*, **10**, 586–593.

[33] Shorack, G. R. and Wellner, J. A. (1986) *Empirical processes with applications to Statistics*, Wiley Series in Probability and Mathematical Statistics.

[34] Simes, R. J. (1986). An improved Bonferroni procedure for multiple tests of significance. *Biometrika*, **73**(3), 751–754.

[35] Storey, J. D. (2002). A direct approach to false discovery rates. *J. R. Stat. Soc. Ser. B Stat. Methodol.*, **64**(3), 479–498.

[36] Storey, J. D. (2003). The positive false discovery rate: a Bayesian interpretation and the q-value. *Ann. Statist.*, **31**(6), 2013–2035.

[37] Storey, J. D. (2007). The optimal discovery procedure: a new approach to simultaneous significance testing. *J. R. Statist. Soc. B*, **69**, 347-368.

[38] Storey, J. D., Taylor, J. E., and Siegmund, D. (2004). Strong control, conservative point estimation and simultaneous conservative consistency of false discovery rates: a unified approach. *J. R. Stat. Soc. Ser. B Stat. Methodol.*, **66**(1), 187–205.