The Annals of Statistics
2011, Vol. 39, No. 3, 1551–1579
DOI: 10.1214/10-AOS869

ASYMPTOTIC BAYES-OPTIMALITY UNDER SPARSITY OF SOME MULTIPLE TESTING PROCEDURES

BY MAŁGORZATA BOGDAN[1], ARIJIT CHAKRABARTI, FLORIAN FROMMLET[2] AND JAYANTA K. GHOSH

Wrocław University of Technology and Purdue University, Indian Statistical Institute, University of Vienna, and Purdue University and Indian Statistical Institute

Within a Bayesian decision theoretic framework we investigate some asymptotic optimality properties of a large class of multiple testing rules. A parametric setup is considered, in which observations come from a normal scale mixture model and the total loss is assumed to be the sum of losses for individual tests. Our model can be used for testing point null hypotheses, as well as to distinguish large signals from a multitude of very small effects. A rule is defined to be asymptotically Bayes optimal under sparsity (ABOS), if within our chosen asymptotic framework the ratio of its Bayes risk and that of the Bayes oracle (a rule which minimizes the Bayes risk) converges to one. Our main interest is in the asymptotic scheme where the proportion p of "true" alternatives converges to zero.

We fully characterize the class of fixed threshold multiple testing rules which are ABOS, and hence derive conditions for the asymptotic optimality of rules controlling the Bayesian False Discovery Rate (BFDR). We finally provide conditions under which the popular Benjamini–Hochberg (BH) and Bonferroni procedures are ABOS and show that for a wide class of sparsity levels, the threshold of the former can be approximated by a nonrandom threshold.

It turns out that while the choice of asymptotically optimal FDR levels for BH depends on the relative cost of a type I error, it is almost independent of the level of sparsity. Specifically, we show that when the number of tests m increases to infinity, then BH with FDR level chosen in accordance with the assumed loss function is ABOS in the entire range of sparsity parameters $p \propto m^{-\beta}$, with $\beta \in (0, 1]$.

1. Introduction. Multiple testing has emerged as a very important problem in statistical inference because of its applicability in understanding large data sets involving many parameters. A prominent area of the application of multiple testing is microarray data analysis, where one wants to simultaneously test expression

Received February 2010; revised December 2010.

[1]Supported in part by Grants 1 P03A 01430 and NN 201414139 of the Polish Ministry of Science and Higher Education.

[2]Supported in part by the WWTF Grant MA09-007.

MSC2010 subject classifications. Primary 62C25, 62F05; secondary 62C10.

Key words and phrases. Multiple testing, FDR, Bayes oracle, asymptotic optimality.

levels of thousands of genes (see, e.g., [18, 19, 24, 31, 34, 35, 41] or [42]). Various ways of performing multiple tests have been proposed in the literature over the years, typically differing in their objective. Among the most popular classical multiple testing procedures, one finds the Bonferroni correction, aimed at controlling the family wise error rate (FWER) and the Benjamini–Hochberg procedure [2], which controls the false discovery rate (FDR). A wide range of empirical Bayes (e.g., see [6, 17–19] and [44]) and full Bayes tests (see, e.g., [6, 12, 31] and [35]) have also been proposed and are used extensively in such problems.

In the classical setting, a multiple testing procedure is considered to be *optimal* if it maximizes the number of true discoveries, while keeping one of the type I error measures (like FWER, FDR or the expected number of false positives) at a certain, fixed level. In this context, it is shown in [25] that the Benjamini–Hochberg procedure (henceforth called BH) is optimal within a large class of step-up multiple testing procedures controlling FDR. In recent years many new multiple testing procedures, which have some optimality properties in the classical sense, have been proposed (e.g., [11, 29, 32] or [33]). In [20] an asymptotic analysis is performed and new step-up and step-up-down procedures, which maximize the asymptotic power while controlling the asymptotic FDR, are introduced. Also, in [41] and [43] two classical oracle procedures for multiple testing are defined. The oracle procedure proposed in [41] maximizes the expected number of true positives where the expected number of false positives is kept fixed. This procedure requires the knowledge of the true distribution for all test statistics and is rather difficult to estimate without further assumptions on the process generating the data. The oracle proposed in [43] assumes that the data is generated according to a two-component mixture model. It aims at maximizing the marginal false nondiscovery rate (mFNR), while controlling the marginal false discovery rate (mFDR) at a given level. In [43] a data-driven adaptive procedure is developed, which asymptotically attains the performance of the oracle procedure for any fixed (though unknown) proportion p of alternative hypothesis.

In this paper we take a different point of view and analyze the properties of multiple testing rules from the perspective of Bayesian decision theory. We assume for each test fixed losses δ_0 and δ_A for type I and type II errors, respectively, and define the overall loss of a multiple testing rule as the sum of losses incurred in each individual test. We feel that such an approach is natural in the context of testing, where the main goal is to detect significant signals, rather than estimate their magnitude. In the specific case where $\delta_0 = \delta_A = 1$, the total loss is equal to the number of misclassified hypotheses. Also, we consider the asymptotic scheme, under which the proportion p of "true" alternatives among all tests converges to zero as the number of tests m goes to infinity, and restrict our attention to the signals on the *verge of detectability*, which can be asymptotically detected with the power in $(0, 1)$.

In recent years, substantial efforts have been made to understand the properties of multiple testing procedures under sparsity, that is, in the case where p is very small (e.g., [7, 13, 14, 26, 30]). A major theoretical breakthrough was made in [1], where it has been shown that the Benjamini–Hochberg procedure can be used for estimating a sparse vector of means, while the level of sparsity can vary considerably. In [1] independent normal observations $X_i, i = 1, \dots, m$, with unknown means μ_i and known variance are considered. Among the studied parameter spaces are the $l_0[p_m]$ balls, which consist of those real m-vectors for which the fraction of nonzero elements is at most p_m. A data-adaptive thresholding estimator for the unknown vector of means is proposed using the Benjamini–Hochberg rule at the FDR level $\alpha_m \geq \frac{\gamma}{\log m}$ for some $\gamma > 0$ and all $m > 1$. If the FDR control level α_m converges to $\alpha_0 \in [0, 1/2]$, this estimator is shown to be asymptotically minimax, simultaneously for a large class of loss functions (and in fact for many different types of sparsity classes including the l_0 balls), as long as p_m is in the range $[\frac{\log^5 m}{m}, m^{-\xi}]$, with $\xi \in (0, 1)$.

In this paper we provide new theoretical results, which illustrate the asymptotic optimality properties of BH under sparsity in the context of Bayesian decision theory. BH is a very interesting procedure to analyze from this point of view, since, despite its frequentist origin, it shares some of the major strengths of Bayesian methods. Specifically, as shown in [18] and [23], BH can be understood as an empirical Bayes approximation to the procedure controlling the "Bayesian" False Discovery Rate (BFDR). This approximation relies mainly on estimation of the distribution generating the data by the empirical distribution function. In this way, similarly to standard Bayes methods, it gains strength by combining information from all the tests. The major issue addressed in this paper is the relationship between BFDR control and optimization of the Bayes risk. Our research was motivated mainly by the good properties of BH with respect to the misclassification rate under sparsity, documented in [5, 6] and [23]. The present paper lends theoretical support to these experimental findings, by specifying a large range of loss functions for which BH is asymptotically optimal in a Bayesian decision theoretic context.

The outline of the paper is as follows. In Section 2 we define and discuss our model, and we introduce the decision theoretic and asymptotic framework of the paper. The Bayes oracle, which minimizes the Bayes risk, is presented, which applies a *fixed threshold* critical region for each individual test. Conditions are formulated under which the asymptotic power of this rule is larger than 0, but smaller than 1. Two different levels of sparsity, the extremely sparse case and a slightly denser case, are defined, which play a prominent role throughout the paper.

In Section 3 we compute the asymptotic risk of the Bayes oracle, and we formally define the concept of asymptotic Bayes optimality under sparsity (ABOS). We then study fixed threshold tests in great detail and fully characterize the class of fixed threshold testing rules being ABOS. In the subsequent Section 4 we study fixed threshold multiple testing rules which make use of the unknown model parameters to control the Bayesian False Discovery Rate (BFDR) exactly at a given level α. We provide conditions for such rules to be ABOS and also consider ABOS of the closely related fixed threshold tests using the asymptotic approximation of the BH threshold c_{GW}, introduced by Genovese and Wasserman [23]. Specifically, in Corollary 4.1 we show that if $p \propto m^{-\beta}$ for some $\beta > 0$, then the asymptotically optimal BFDR levels depend mainly on the ratio of loss functions for type I and type II errors and are independent of β.

The main results of the paper are included in Section 5, where we specify conditions under which the Bonferroni rule as well as the Benjamini–Hochberg procedure are ABOS. Specifically, Theorem 5.1 shows that when FDR levels $\alpha_m \to \alpha_\infty < 1$ satisfy the conditions of optimality of BFDR controlling rules, then the difference between the random threshold of BH and the Genovese–Wasserman threshold c_{GW} converges to 0 for any sequence of sparsity parameters $p_m \propto m^{-\beta}$, with $\beta \in (0, 1)$. Theorem 5.2 shows that for the same FDR levels BH is ABOS whenever $p_m \propto m^{-\beta}$, with $\beta \in (0, 1]$. Thus, our results show that BH adapts to the unknown level of sparsity. However, we also show that the optimal FDR controlling level depends on the relative cost of a type I error—it should be chosen to be small if the relative cost of the type I error is large. Specifically, within our asymptotic framework, the Benjamini–Hochberg rule controlling the FDR at a fixed level $\alpha \in (0, 1)$ is ABOS for a wide range of sparsity levels, provided that the ratio of losses for type I and type II errors converges to zero at a slow rate which can vary widely. When the loss ratio is constant, similar optimality results hold if the FDR controlling level slowly converges to zero.

Section 6 contains a discussion and directions for further research. The proof of the asymptotic optimality of BH can be found in Section 7, while the remaining lengthy proofs can be found in the supplemental report [3].

2. Statistical model and asymptotic framework.

Suppose we have m independent observations $X_1, \ldots, X_m$, and assume that each X_i has a normal $N(\mu_i, \sigma_\varepsilon^2)$ distribution. Here μ_i represents the effect under investigation, and σ_ε^2 is the variance of the random noise (e.g., the measurement error). We assume that each μ_i is an independent random variable, with distribution determined by the value of the unobservable random variable ν_i, which takes values 0 and 1 with probabilities $1 - p$ and p, respectively, for some $p \in (0, 1)$. We denote by H_{0i} the event that $\nu_i = 0$, while H_{Ai} denotes the event $\nu_i = 1$. We will refer to these events as the null and alternative hypotheses. Under H_{0i}, μ_i is assumed to have a $N(0, \sigma_0^2)$ distribution (where $\sigma_0^2 \geq 0$), while under H_{Ai} it is assumed to have a $N(0, \sigma_0^2 + \tau^2)$ distribution (where $\tau^2 > 0$). Hence, we are really modeling the μ_i's as i.i.d. r.v.'s from the following mixture distribution:

$$\mu_i \sim (1 - p)N(0, \sigma_0^2) + pN(0, \sigma_0^2 + \tau^2). \tag{2.1}$$

This implies that the marginal distribution of X_i is the scale mixture of normals, namely,

$$X_i \sim (1 - p)N(0, \sigma^2) + pN(0, \sigma^2 + \tau^2), \tag{2.2}$$

where $\sigma^2 = \sigma_\varepsilon^2 + \sigma_0^2$.

Note that in the case where $\sigma_0^2 = 0$, H_{0i} corresponds to the point null hypothesis that $\mu_i = 0$, and H_{Ai} says that $\mu_i \neq 0$ [since under H_{Ai} $P(\mu_i = 0) = 0$)]. Thus this model can be used for simultaneously testing if the means of the X_i's are zero or not. Allowing $\sigma_0^2 > 0$ greatly extends the scope of the applications of the proposed mixture model under sparsity. In many multiple testing problems it seems unrealistic to assume that the vast majority of effects are exactly equal to zero. For example, in the context of locating genes influencing quantitative traits, it is typically assumed that a trait is influenced by many genes with very small effects, so called polygenes. Such genes form a background, which can be modeled by the null component of the mixture. In this case the main purpose of statistical inference is the identification of a small number of significant "outliers," whose impact on the trait is substantially larger than that of the polygenes. These important "outlying" genes are modeled by the nonnull component of the mixture.

In the remaining part of the paper we will assume that the variance of X_i under the null hypothesis, σ^2, is known. This assumption is often used in the literature on the asymptotic properties of multiple testing procedures (see, e.g., [1] or [13]). Some discussion concerning the general issue of estimation of parameters in sparse mixtures is provided in Section 6.

REMARK 2.1. Note that given μ_i, the distribution of X_i is a location shift of the distribution under the null. This is the setting in which multiple testing is typically analyzed in the classical context. In our extended Bayesian model, the choice of a normal $N(0, \sigma_0^2 + \tau^2)$ prior for μ_i under the alternative results in a corresponding normal $N(0, \sigma^2 + \tau^2)$ marginal distribution for X_i, which differs from the null distribution only by a larger scale parameter. The proposed mixture model for X_i is a specific example of the two-groups model, which was discussed in a wider nonparametric context, for example, in [6, 17, 19] and [24]. Similar Gaussian mixture models for multiple testing were considered, for example, in [7] and [16]. Restricting attention to scale mixtures of normal distributions allows us to reduce the technical complexity of the proofs and to concentrate on the main aspects of the problem. Moreover, we believe that the proposed model is applicable in many practical situations, when there are no prior expectations concerning the sign of μ_i. Our asymptotic results may be extended to the situation when the distribution of μ_i under the alternative is not symmetric about 0. Namely, the techniques presented in the related report [22] can be used for a similar asymptotic analysis when the "alternative" normal distribution $N(0, \sigma_0^2 + \tau^2)$ of μ_i in the model (2.1) is replaced by a general scale distribution, with the scale parameter playing role of τ. A manuscript dealing with this case is in preparation.

We consider a Bayesian decision theoretic formulation of the multiple testing problem of testing H_{0i} versus H_{Ai}, for $i = 1, \dots, m$ simultaneously. For each i,

TABLE 1
Matrix of losses

	Choose H_{0i}	**Choose H_{Ai}**
H_{0i} true	0	δ_0
H_{Ai} true	δ_A	0

there are two possible "states of nature," namely H_{0i} with $X_i \sim N(0, \sigma^2)$ or H_{Ai} with $X_i \sim N(0, \sigma^2 + \tau^2)$, that occur with probabilities $(1 - p)$ and p, respectively. Table 1 defines the matrix of losses for making a decision in the ith test.

We assume that the overall loss in the multiple testing procedure is the sum of losses for individual tests. Thus our approach is based on the notion of an additive loss function, which goes back to [27] and [28], and seems to be implicit in most of the current formulations.

Under an additive loss function, the compound Bayes decision problem can be solved as follows. It is easy to see that the expected value of the total loss is minimized by a procedure which simply applies the Bayesian classifier to each individual test. For each i, this leads to choosing the alternative hypothesis H_{Ai} in cases such that

$$\frac{\phi_A(X_i)}{\phi_0(X_i)} \geq \frac{(1-p)\delta_0}{p\delta_A}, \tag{2.3}$$

where ϕ_A and ϕ_0 are the densities of X_i under the alternative and null hypotheses, respectively.

After substituting in the formulas for the appropriate normal densities, we obtain the optimal rule

$$\text{Reject } H_{0i} \qquad \text{if } \frac{X_i^2}{\sigma^2} \geq c^2, \tag{2.4}$$

where

$$c^2 = c^2_{\tau,\sigma,f,\delta} = \frac{\sigma^2 + \tau^2}{\tau^2}\left(\log\left(\left(\frac{\tau}{\sigma}\right)^2 + 1\right) + 2\log(f\delta)\right) \tag{2.5}$$

with $f = \frac{1-p}{p}$ and $\delta = \frac{\delta_0}{\delta_A}$. We call this rule a *Bayes oracle*, since it makes use of the unknown parameters of the mixture, τ and p, and therefore is not attainable in finite samples.

Using standard notation from the theory of testing, we define the probability of a type I error as

$$t_{1i} = P_{H_{0i}}(H_{0i} \text{ is rejected})$$

and the probability of a type II error as

$$t_{2i} = P_{H_{Ai}}(H_{0i} \text{ is accepted}).$$

Note that under our mixture model the marginal distributions of X_i under the null and alternative hypotheses do not depend on i, and the threshold of the Bayes oracle is also the same for each test. Hence, when calculating the probabilities of type I errors and type II errors for the Bayes oracle, we can, and will henceforth, suppress i from t_{1i} and t_{2i}. The same remark also applies to any fixed threshold procedure which, for each i, rejects H_{0i} if $X_i^2/\sigma^2 > K$ for some constant K.

2.1. *The asymptotic framework.* We now want to motivate the asymptotic framework which will be formally introduced below as Assumption (A). Let $\gamma = (p, \tau^2, \sigma^2, \delta_0, \delta_A)$ be the vector of parameters defining the Bayes oracle (2.5). In our asymptotic analysis, we will consider infinite sequences of such γ's. A natural example of such a situation arises when the number of tests m increases to infinity, and the vector γ varies with the number of tests m. But here we are actually trying to understand, in a unified manner, the general limiting problem when γ varies through a sequence.

The threshold (2.5) depends on τ and σ only through $u = (\frac{\tau}{\sigma})^2$. Note that u is a natural scale for measuring the strength of the signal in terms of the variance of X_i under the null. We also introduce another parameter $v = uf^2\delta^2$, which can be used to simplify the formula for the optimal threshold

$$c_{u,v}^2 = \left(1 + \frac{1}{u}\right)\bigl(\log v + \log(1 + 1/u)\bigr). \tag{2.6}$$

Observe that under the alternative $\frac{X_i}{\sigma}$ has a normal $N(0, 1 + u)$ distribution. Thus the probability of a type II error using the Bayes oracle is given by

$$t_2 = P\left(Z^2 < \frac{1}{u+1} c_{u,v}^2\right), \tag{2.7}$$

where Z is a standard normal variable.

From (2.7) it follows that given an arbitrary infinite sequence of γ's, the limiting power of the Bayes oracle is nonzero only if the corresponding sequence $\frac{c_{u,v}^2}{u+1}$ remains bounded. We will restrict ourselves to such sequences, since otherwise even the Bayes oracle cannot guarantee nontrivial inference in the limit and all rules will perform poorly.

The focus of this paper is the study of the inference problem when $p \to 0$, and the goal is to find procedures which will efficiently identify signals under such circumstances. To clarify these ideas, consider the situation where $p \to 0$ and $\log(\delta) = o(\log p)$. It is immediately evident from (2.5) that in this situation $c^2 = c_{u,v}^2$ diverges to infinity. Hence $\frac{c_{u,v}^2}{u+1}$ remains bounded only when the signal magnitude u diverges to infinity, in which case $\frac{c_{u,v}^2}{u+1} \propto \frac{\log v}{u}$. This explains two of the three asymptotic conditions we impose in Assumption (A). The third condition $v \to \infty$ pragmatically ensures that δ is not allowed to converge to zero too quickly.

ASSUMPTION (A). A sequence of vectors $\{\gamma_t = (p_t, \tau_t^2, \sigma_t^2, \delta_{0t}, \delta_{At}); t \in \{1, 2, \ldots\}\}$ satisfies this assumption if the corresponding sequence of parameter vectors, $\theta_t = (p_t, u_t, v_t)$, fulfills the following conditions: $p_t \to 0$, $u_t \to \infty$, $v_t \to \infty$ and $\frac{\log v_t}{u_t} \to C \in (0, \infty)$, as $t \to \infty$.

REMARK 2.2. We do not allow $C = \infty$ in Assumption (A) because then the limit of the probability of a type II error for Bayes oracle is equal to one, and signals cannot be identified. If $C = 0$, then the oracle has a limiting power equal to one. Such a situation can occur naturally if the number of replicates used to calculate X_i increases to infinity as $p \to 0$ (see, e.g., [22]). However, in this article we will restrict ourselves to $C \in (0, \infty)$, that is, the case where the asymptotic power is smaller than one. The corresponding parametric region might be thought of as being at "the verge of detectability." The extension of the asymptotic results presented in this paper to the case when $C = 0$ as well as to some cases when p does not converge to zero can be found in [4], which is an extended version of this manuscript. Specifically, Theorems 3.1, 3.2 and 4.1 below hold in exactly the same form even when the condition $p \to 0$ is eliminated from Assumption (A).

REMARK 2.3. We will frequently consider the generic situation

$$\log \delta = o(\log p). \tag{2.8}$$

In that case Assumption (A) reduces to $p \to 0$, $u \to \infty$, $v \to \infty$ and $-\frac{2 \log p}{u} \to C \in (0, \infty)$ and specifies the relationship between the magnitude u of asymptotically detectable signals and the sparsity parameter p. Interestingly, the relationship $u \propto -\log p$, can be related to asymptotically least-favorable configurations for $l_0[p]$ balls discussed in Section 3.1 of [1]. Ignoring constants, the typical magnitudes of observations corresponding to such signals will be similar to the threshold of the minimax hard thresholding estimator corresponding to the parameter space $l_0[p]$.

Notation: We will usually suppress the index t of the elements of the vector γ_t and θ_t. Unless otherwise stated, throughout the paper the notation o_t will denote an infinite sequence of terms indexed by t, which go to zero when $t \to \infty$. In many cases t is the same as the number of tests m, and in such cases the notation o_t will be replaced by o_m.

In case of $m \to \infty$ we will consider specifically two different levels of sparsity. The first, the extremely sparse case, is characterized by

$$mp_m \to s \in (0, \infty] \quad \text{and} \quad \frac{\log(mp_m)}{\log m} \to 0. \tag{2.9}$$

Condition (2.9) is satisfied, for example, when $p_m \propto \frac{1}{m}$. In this situation the expected number of "signals" does not increase with m, which makes it impossible to consistently estimate the mixture parameters. The second, "denser" case is characterized by

$$p_m \to 0 \quad \text{and} \quad \frac{\log(mp_m)}{\log m} \to C_p \in (0, 1], \tag{2.10}$$

which includes $p_m \propto m^{-\beta}$ for $0 < \beta < 1$.

3. Asymptotic Bayes-optimality under sparsity. We start by computing type I and type II error rates of the Bayes oracle. As usual Φ denotes the cumulative distribution function and ϕ the density of the standard normal distribution.

LEMMA 3.1. *Under Assumption (A) the probabilities of type* I *and type* II *error using the Bayes oracle are given by the following equations*:

$$t_1 = e^{-C/2}\sqrt{\frac{2}{\pi v \log v}}(1 + o_t), \tag{3.1}$$

$$t_2 = \bigl(2\Phi(\sqrt{C}) - 1\bigr)(1 + o_t). \tag{3.2}$$

PROOF. Note that $t_1 = P(|Z| > c_{u,v})$. Moreover,

$$c_{u,v}^2 = (1 + z_{u,v}) \log v, \tag{3.3}$$

where $\lim_{u\to\infty, v\to\infty} z_{u,v}u = 1$. Therefore, we obtain

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

which, together with Assumption (A), yields

$$\phi(c_{u,v}) = e^{-C/2}\sqrt{\frac{1}{2\pi v}}(1 + o_t). \tag{3.4}$$

Now the proof follows easily by invoking the well-known approximation to the tail probability of the standard normal distribution

$$P(|Z| > c) = \frac{2\phi(c)}{c}\bigl(1 - z_1(c)\bigr), \tag{3.5}$$

where $z_1(c)$ is a positive function such that $z_1(c)c^2 = O(1)$ as $c \to \infty$.

The formula for type II error immediately follows from (2.7) and Assumption (A). $\square$

3.1. *The Bayes risk.* Under an additive loss function, the Bayes risk for a multiple testing procedure is given by

$$R = \delta_0 E(V) + \delta_A E(T), \tag{3.6}$$

where $E(V)$ and $E(T)$ are the expected numbers of false positives and false negatives, respectively. In particular, under our mixture model, the Bayes risk for a fixed threshold multiple testing procedure is given by

$$R = m\big((1-p)t_1\delta_0 + pt_2\delta_A\big). \tag{3.7}$$

Equations (3.1) and (3.2) easily yield the following asymptotic approximation to the optimal Bayes risk.

THEOREM 3.1. *Under Assumption (A), using the Bayes oracle, the risk takes the form*

$$R_{\text{opt}} = mp\delta_A\big(2\Phi(\sqrt{C}) - 1\big)(1 + o_t). \tag{3.8}$$

REMARK 3.1. It is important to note that under Assumption (A), the asymptotic form of the risk of the Bayes oracle in (3.1) is determined by its type II error component. In fact the probability of type II error, t_2, is much less sensitive to changes in the threshold value than the probability of type I error, t_1. In particular, it is easy to see that the same asymptotic form of t_2 [as in (3.2)] is achieved by any multiple testing rule rejecting the null hypothesis H_{0i} when $X_i^2/\sigma^2 > c_t^2$, with $c_t^2 = \log v + z_t$ and $z_t = o(\log v)$. Probability of type I error is substantially more sensitive to changes in the critical value, even if $z_t = o(\log v)$. If z_t is always positive, then the *rate* of convergence of the probability of type I error to zero is faster than that of the optimal rule, and the total risk is still determined by the type II component. Therefore the rule remains optimal as long as $z_t = o(\log v)$. However, if $z_t = o(\log v)$ can take negative values, the situation is quite subtle. In this case the rate of convergence of the probability of type I error to zero may be equal or slower than that of the optimal rule, making the overall risk of the rule substantially larger than R_{opt}. These observations are formally summarized in Theorem 3.2, which gives a characterization of the set of the asymptotically optimal fixed threshold multiple testing rules.

DEFINITION. Consider a sequence of parameter vectors γ_t, satisfying Assumption (A). We call a multiple testing rule asymptotically Bayes optimal under sparsity (ABOS) for γ_t if its risk R satisfies

$$\frac{R}{R_{\text{opt}}} \to 1 \qquad \text{as } t \to \infty,$$

where R_{opt} is the optimal risk, given by Theorem 3.1.

REMARK 3.2. This definition relates optimality to a particular sequence of γ vectors satisfying Assumption (A). However, the asymptotically optimal rule for a specific sequence γ_t is also typically optimal for a large set of "similar" sequences. The asymptotic results presented in the following sections of this paper characterize these "domains" of optimality for some of the popularly used multiple testing rules. Since Assumption (A) is an inherent part of our definition of optimality, we will refrain from explicitly stating it when reporting our asymptotic optimality results.

The following theorem fully characterizes the set of asymptotically Bayes-optimal multiple testing rules with fixed thresholds.

THEOREM 3.2. *A multiple testing rule of the form* (2.4) *with threshold* $c^2 = c_t^2 = \log v + z_t$ *is ABOS if and only if*

$$z_t = o(\log v) \tag{3.9}$$

and

$$z_t + 2\log\log v \to \infty. \tag{3.10}$$

The proof of Theorem 3.2 is provided in Section 8 of [3].

REMARK 3.3. Conditions (3.9) and (3.10) guarantee the asymptotic Bayes optimality of the components of risk corresponding to type II and type I errors, respectively.

In the following corollary we present multiple testing rules which are ABOS in the generic situation of Remark 2.3, where $u \propto -\log p$.

COROLLARY 3.1. *Assume* (2.8) *holds*, δ *is bounded from above*, $m \to \infty$ *and* $p \propto m^{-\beta}$, *with* $\beta > 0$. *Then a fixed threshold multiple testing rule* (2.4) *based on the threshold*

$$c^2 = c_m^2 = 2\beta \log m + d, \tag{3.11}$$

where $d \in \mathbb{R}$, *is ABOS.*

The proof is straightforward and is thus skipped.

REMARK 3.4. The optimal threshold, provided in Corollary 3.1, depends on the unknown parameter β. It may be pointed out that it is proved in Section 5 that the Benjamini–Hochberg multiple testing procedure adapts to this unknown sparsity and, under very mild conditions on δ and the FDR level α, is ABOS whenever $0 < \beta \leq 1$. Corollary 3.1 shows also that the universal threshold $2\log m$ of [15] is ABOS when $\beta = 1$. Thus, within our asymptotic framework, the universal threshold is asymptotically optimal when the expected number of true signals does not increase with m.

4. Controlling the Bayesian False Discovery Rate. In a seminal paper [2], Benjamini and Hochberg introduced the False Discovery Rate (FDR) as a measure of the accuracy of a multiple testing procedure

$$\text{FDR} = E\left(\frac{V}{R}\right). \tag{4.1}$$

Here R is the total number of null hypotheses rejected, V is the number of "false" rejections and it is assumed that $\frac{V}{R} = 0$ when $R = 0$. For tests with a fixed threshold, Efron and Tibshirani [18] define another very similar measure, called the Bayesian False Discovery Rate, BFDR,

$$\text{BFDR} = P(H_{0i} \text{ is true} | H_{0i} \text{ was rejected}) = \frac{(1-p)t_1}{(1-p)t_1 + p(1-t_2)}, \tag{4.2}$$

where t_1 and t_2 are the probabilities of type I and type II errors.

According to Theorem 1 of [40], in the case when individual test statistics are generated by the two-component mixture model and the multiple testing procedure uses the same fixed threshold for each of the tests, BFDR coincides with the positive False Discovery Rate pFDR of [40], defined as

$$\text{pFDR} = E\left(\frac{V}{R}\Big| R > 0\right) = \frac{\text{FDR}}{P(R > 0)}.$$

Note here that in our context it is enough to consider threshold tests that reject for high values of $\frac{X_i^2}{\sigma^2}$. This is due to the fact that from the MLR property and the Neyman–Pearson lemma, it can be easily proved that any other kind of test with the same type 1 error will have a larger BFDR and Bayesian False Negative Rate (BFNR).

Extensive simulation studies and theoretical calculations in [6, 23] and [5] illustrate that multiple testing rules controlling the BFDR at a small level $\alpha \approx 0.05$ behave very well under sparsity in terms of minimizing the misclassification error (i.e., the Bayes risk for $\delta_0 = \delta_A$). We also recall in this context that a test has BFDR α if and only if

$$(1-\alpha)(1-p)t_1 + \alpha p t_2 = \alpha p, \tag{4.3}$$

the left-hand side of (4.3) being the Bayes risk for $\delta_0 = 1 - \alpha$ and $\delta_A = \alpha$. So the definition of the BFDR itself has a strong connection to the Bayes risk and a "proper" choice of α might actually yield an optimal rule (for similar conclusions, see, e.g., [43]). One can show quite easily that under the mixture model (2.2), the BFDR of a test based on the threshold c^2 continuously decreases from $(1 - p)$ for $c = 0$ to 0 for $c \to \infty$ (see Lemma 9.1 of [3]). In other words, there exists a 1–1 mapping between thresholds $c \in [0, \infty)$ and BFDR levels $\alpha \in (0, 1 - p]$. So, if the BFDR control level is chosen properly, the corresponding threshold can satisfy the conditions of Theorem 3.2.

REMARK 4.1. In [10] it is argued that when the data are generated according to the two component mixture model, BFDR of any fixed threshold rule as well as of the Benjamini–Hochberg procedure is bounded from below by a constant $\beta^\star \geq 0$, where $\beta^\star$ depends on the actual mixture density. Lemma 9.1 of [3] shows under our mixture model (2.2) $\beta^\star = 0$, that is, the criticality phenomenon of [10] does not occur. This is generally true in any case when the ratio of tail probabilities $P(|X_i| > c)$ under the null and alternative distributions converges to 0 as $c \to \infty$.

Now, we give a full characterization of asymptotically optimal BFDR levels, which will be later used to prove ABOS of BH.

4.1. *ABOS of BFDR rules.* The general Theorem 4.1, below, gives conditions on α, which guarantee optimality for any given sequence of parameters γ_t, satisfying Assumption (A). Corollary 4.1 presents a special simple choice which works in the general setting. In the subsequent Corollary 4.2 we study the generic situation (2.8) of Remark 2.3. Finally, Corollary 4.3 considers the case where $\alpha = \text{const} \in (0, 1)$ and gives simple conditions for δ that guarantee optimality.

Consider a fixed threshold rule (based on $\frac{X_i^2}{\sigma^2}$) with the BFDR equal to α. Under the mixture model (2.2), a corresponding threshold value c_B^2 can be obtained by solving the equation

$$\frac{(1-p)(1-\Phi(c_B))}{(1-p)(1-\Phi(c_B)) + p(1-\Phi(c_B/\sqrt{u+1}))} = \alpha, \tag{4.4}$$

or equivalently, by solving

$$\frac{1-\Phi(c_B)}{1-\Phi(c_B/\sqrt{u+1})} = \frac{\alpha}{f(1-\alpha)} = \frac{r_\alpha}{f}, \tag{4.5}$$

where

$$r_\alpha = \frac{\alpha}{1-\alpha}. \tag{4.6}$$

Note that r_α converges to 0 when $\alpha \to 0$ and to infinity when $\alpha \to 1$.

Using Theorem 3.2, one can show that this test is asymptotically optimal only if $\frac{c_B}{\sqrt{u+1}}$ converges to $\sqrt{C}$, where C is the constant in Assumption (A). From (4.5), this in turn implies that a BFDR rule for a chosen α sequence can only be optimal if $\frac{r_\alpha}{f}$ goes to zero while satisfying certain conditions. When $\frac{r_\alpha}{f} \to 0$, a convenient asymptotic expansion for c_B^2 can be obtained, and optimality holds if and only if this asymptotic form conforms to the conditions specified in Theorem 3.2. The following theorem gives the asymptotic expansion for c_B^2 and specifies the range of "optimal" choices of r_α.

THEOREM 4.1. *Consider a fixed threshold rule with* BFDR $= \alpha = \alpha_t$. *Define* s_t *by*

$$\frac{\log(f\delta\sqrt{u})}{\log(f/r_\alpha)} = 1 + s_t, \tag{4.7}$$

where $r_\alpha = \frac{\alpha}{1-\alpha}$. *Then the rule is ABOS if and only if*

$$s_t \to 0 \tag{4.8}$$

and

$$2s_t \log(f/r_\alpha) - \log\log(f/r_\alpha) \to -\infty. \tag{4.9}$$

The threshold for this rule is of the form

$$c_B^2 = 2\log\left(\frac{f}{r_\alpha}\right) - \log\left(2\log\left(\frac{f}{r_\alpha}\right)\right) + C_1 + o_t, \tag{4.10}$$

where $C_1 = \log(\frac{2}{\pi D^2})$, *and* $D = 2(1 - \Phi(\sqrt{C}))$ *is the asymptotic power. The corresponding probability of a type I error is equal to*

$$t_1 = D\frac{r_\alpha}{f}(1 + o_t).$$

The proof of Theorem 4.1 can be found in Section 10 of [3].

REMARK 4.2. In comparison to (4.8), condition (4.9) imposes an additional restriction on positive values of s_t (i.e., large values of α). It is clear from the proof of Theorem 4.1 that the necessity of this additional requirement results from the asymmetric roles of type I and type II errors in the Bayes risk, as discussed in Remark 3.1.

REMARK 4.3. Condition (4.8), given in Theorem 4.1, says (after some algebra) that a sequence of asymptotically optimal BFDR levels $\alpha = \alpha_t$ satisfies $\frac{\alpha}{1-\alpha} = (\delta\sqrt{u})^{b_t - 1} f^{b_t}$ for some b_t, where $b_t \to 0$ as $t \to \infty$. Broadly speaking, this means that for optimality the BFDR levels need to be chosen small when the loss ratio is large. The seemingly evident dependence of α on u is not stressed in this article, since on the verge of detectability $u = \frac{2}{C}\log(f\delta)(1 + o_t)$ and, as seen in the following corollaries, the range of asymptotically optimal levels of α does not depend on C. A thorough discussion of the dependence of α on u in case when $C = 0$ can be found in [4].

COROLLARY 4.1. *A rule with BFDR at the level* $\alpha = \alpha_t$, *such that* $r_\alpha \propto (\delta\sqrt{u})^{-1}$, *is ABOS. Specifically, if* $m \to \infty$, $p \propto m^{-\beta}$ $(\beta > 0)$ *and* $\frac{\log\delta}{\log m} \to C_\delta \in [0, \infty]$, *then a rule with BFDR at the level* α *such that* $r_\alpha \propto (\delta\sqrt{\log(m\delta)})^{-1}$ *is ABOS.*

REMARK 4.4. Corollary 4.1 shows that while the proposed optimal BFDR level clearly depends on the ratio of losses δ, it is independent of the sparsity parameter β.

The proof of Corollary 4.1 is immediate by verifying that (4.8) and (4.9) are satisfied by such sequences of α's. Also the proofs of the following Corollaries 4.2 and 4.3, follow quite immediately from Theorem 4.1 and are thus omitted.

COROLLARY 4.2. *Assume the generic situation* (2.8) *of Remark* 2.3. *Then a fixed threshold rule with BFDR equal to* α *is ABOS if and only if* r_α *satisfies*

$$\log r_\alpha = o(\log p) \quad \textit{and} \quad r_\alpha \delta \to 0.$$

If we assume further that $m \to \infty$ *and* $p \propto m^{-\beta}$ $(\beta > 0)$, *such a rule is ABOS if and only if*

$$\log r_\alpha = o(\log m) \quad \textit{and} \quad r_\alpha \delta \to 0.$$

In case when $\delta = \text{const}$ *and* $p \propto m^{-\beta}$, *the BFDR rule is ABOS if and only if* $\alpha \to 0$ *such that* $\log \alpha = o(\log m)$.

COROLLARY 4.3. *A fixed threshold rule with BFDR equal to* $\alpha \in (0, 1)$ *is ABOS if and only if* $\delta \to 0$ *at such a rate that* $\frac{\log \delta}{\log p} \to 0$. *If we assume that* $m \to \infty$ *and* $p \propto m^{-\beta}$ $(\beta > 0)$, *such a rule is ABOS if and only if* $\delta \to 0$ *such that* $\log \delta = o(\log m)$.

Corollary 4.3, given above, states that a rule with BFDR at a fixed level α is asymptotically optimal for a wide range of loss functions, such that $\delta \to 0$. Note that the assumption that $\delta \to 0$ as $p \to 0$ agrees with the intuition that the cost of missing a signal should be relatively large if the true number of signals is small. Corollary 4.2 shows that when the loss ratio is constant, a BFDR rule is asymptotically optimal for a wide range of α levels, such that $\alpha \to 0$.

4.2. *Optimality of the asymptotic approximation to the BH threshold.* In [23] it is proved that when the number of tests tends to infinity, and the fraction of true alternatives remains fixed, then the random threshold of the Benjamini–Hochberg procedure can be approximated by

$$c_{\mathrm{GW}} : \frac{(1 - \Phi(c_{\mathrm{GW}}))}{(1 - p)(1 - \Phi(c_{\mathrm{GW}})) + p(1 - \Phi(c_{\mathrm{GW}}/\sqrt{u + 1}))} = \alpha. \tag{4.11}$$

Compared to the equation defining the BFDR rule (4.4), the function on the left-hand side of (4.11) lacks $(1 - p)$ in the numerator. In the case where $p \to 0$ this term is negligible, and one expects that the rule based on c_{GW} asymptotically approximates the corresponding BFDR rule for the same α. The following result shows that this is indeed the case.

THEOREM 4.2. *Consider the rule rejecting the null hypothesis* H_{0i} *if* $\frac{X_i^2}{\sigma^2} \geq c_{\mathrm{GW}}^2$, *where* c_{GW} *is defined in* (4.11). *This rule is ABOS if and only if the corresponding BFDR rule defined in* (4.4) *is ABOS. In this case we have*

$$c_{\mathrm{GW}}^2 = c_B^2 + o_t,$$

where c_B^2 *is the threshold of an asymptotically optimal BFDR rule, defined in Theorem* 4.1.

PROOF. Note that (4.11) is equivalent to

$$\frac{1-\Phi(c_{\mathrm{GW}})}{1-\Phi(c_{\mathrm{GW}}/\sqrt{u+1})} = \frac{pr_\alpha}{1+pr_\alpha} = \frac{r_{\alpha'}}{f}, \tag{4.12}$$

where $\alpha' = \alpha(1-p)$. Thus c_{GW} is the same as the threshold of a rule with BFDR at the level α'.

Define $s_{t'}$ by $\frac{\log(f\delta\sqrt{u})}{\log(f/r_{\alpha'})} = 1 + s_{t'}$. It follows easily that $s_{t'}$ satisfies (4.8) and (4.9) of Theorem 4.1 (with α replaced by α'), if and only if s_t defined in (4.7) satisfies (4.8) and (4.9). Thus the first part of the theorem is proved.

To complete the proof of the theorem, we observe that the optimality of a BFDR rule implies that $\frac{r_\alpha}{f} \to 0$, and the optimality of the rule based on c_{GW} implies that $\frac{r_{\alpha'}}{f} \to 0$. In either case, $pr_\alpha \to 0$ and thus (4.12) reduces to

$$\frac{1-\Phi(c_{\mathrm{GW}})}{1-\Phi(c_{\mathrm{GW}}/\sqrt{u+1})} = pr_\alpha(1+o_t) = \frac{r_\alpha}{f}(1+o_t). \tag{4.13}$$

Now, the asymptotic approximation to c_{GW}^2 can be obtained analogously to the asymptotic form of the threshold for an optimal BFDR rule, provided in (4.10). □

5. ABOS of classical frequentist multiple testing procedures. Similarly to the Bayes oracle, the BFDR rules discussed in Section 4 are not attainable, since they require the knowledge of the parameters of the mixture distribution (2.2). However, the results included in Section 4 can be used to prove the asymptotic optimality of classical multiple testing procedures, such as the Bonferroni rule and the Benjamini–Hochberg procedure (BH). In this section we consider a sequence of problems in which the number of tests $m \to \infty$ and the γ sequence is indexed by $t = m$.

5.1. *ABOS of the Bonferroni correction.* The Bonferroni correction is one of the oldest and most popular multiple testing rules. It is aimed at controlling the Family Wise Error Rate, FWER $= P(V > 0)$, where V is the number of false discoveries. The Bonferroni correction at FWER level α rejects all null hypothesis for which $Z_i = \frac{|X_i|}{\sigma}$ exceeds the threshold

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

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

$$c_{\text{Bon}}^2 = 2\log\left(\frac{m}{\alpha}\right) - \log\left(2\log\left(\frac{m}{\alpha}\right)\right) + \log(2/\pi) + o_m. \tag{5.1}$$

Comparison of this threshold with the asymptotic approximation to an optimal BFDR rule (4.10) suggests that the Bonferroni correction will have similar asymptotic optimality properties in the extremely sparse case (2.9). Indeed, these expectations are confirmed by the following Lemma 5.1, which will be used in the next section for the proof of ABOS of the Benjamini–Hochberg procedure under very sparse signals.

LEMMA 5.1. *Assume that $m \to \infty$ and (2.9) holds. The Bonferroni procedure at FWER level $\alpha_m \to \alpha_\infty \in [0, 1)$ is ABOS if α_m satisfies the assumptions of Theorem 4.1.*

PROOF. Under the assumptions of Lemma 5.1 and Theorem 4.1

$$c_{\text{Bon}}^2 = c_B^2 + 2\log z_m - 2\log(1-\alpha_\infty) + 2\log D + o_m,$$

where $z_m = mp_m$, $D = 2(1-\Phi(\sqrt{C}))$, and c_B^2 is the threshold of the rule controlling the BFDR at level α_m. From (2.9) it follows easily that $c_{\text{Bon}}^2 = c_B^2(1+o_m)$. By assumption, the rule based on the threshold c_B^2 is optimal, and hence c_{Bon}^2 satisfies condition (3.9) of Theorem 3.2. Condition (3.10) is satisfied, since by assumption $\log z_m$ is bounded below for sufficiently large m and thus ABOS of the Bonferroni correction follows. □

5.2. *ABOS of BH.* Let $Z_i = |\frac{X_i}{\sigma}|$ and $p_i = 2(1-\Phi(Z_i))$ be the corresponding p-value. We sort p-values in ascending order $p_{(1)} \le p_{(2)} \le \cdots \le p_{(m)}$ and denote

$$k = \max\left\{i : p_{(i)} \le \frac{i\alpha}{m}\right\}. \tag{5.2}$$

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

REMARK 5.1. BH gained large popularity after the seminal paper [2], where it was proved that it controls FDR. It was originally proposed in [37], and later used in [39] as a test for the global null hypothesis.

Let us denote $1 - \hat{F}_m(y) = \#\{|Z_i| \ge y\}/m$. It is easy to check (see, e.g., (2.2) of [38] or the equivalence theorem of [18]) that the Benjamini–Hochberg procedure rejects the null hypothesis H_{0i} when $Z_i^2 \ge \tilde{c}_{\text{BH}}^2$, where

$$\tilde{c}_{\text{BH}} = \inf\left\{y : \frac{2(1-\Phi(y))}{1-\hat{F}_m(y)} \le \alpha\right\}. \tag{5.3}$$

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 as

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

Comparing (5.3) and (4.11), we observe that the difference between $\tilde{c}_{\mathrm{BH}}$ and c_{GW} is in replacing the cumulative distribution function of $|Z_i|$ [appearing in (4.11)] by the empirical distribution function (in 5.3). Therefore, as shown in [23], for any fixed mixture distribution (2.2) $\tilde{c}_{\mathrm{BH}}$ converges in probability to c_{GW} as $m \to \infty$. The following Theorem 5.1, shows that the approximation of $\tilde{c}_{\mathrm{BH}}$ by c_{GW} works also within our asymptotic framework, where $p_m \to 0$ and $c_{\mathrm{GW}} \to \infty$.

THEOREM 5.1. *Assume that $p_m \to 0$ such that for sufficiently large m*

$$p_m > \frac{log^{\beta_p} m}{m} \qquad \text{for some constant } \beta_p > 1. \tag{5.4}$$

Moreover, assume that the sequence of FDR levels α_m satisfies

$$\alpha_m \to \alpha_\infty < 1 \tag{5.5}$$

and

$$\alpha_m \text{ satisfies the assumption of Theorem 4.1.} \tag{5.6}$$

Then for every $\varepsilon > 0$, every constant $\beta_1 > 0$ and sufficiently large m (dependent on ε and β_1)

$$P(|c_{\mathrm{BH}} - c_{\mathrm{GW}}| > \varepsilon) \le m^{-\beta_1}.$$

The proof of Theorem 5.1 is provided in Section 11 of [3].

Theorem 5.1 suggests asymptotic optimality of BH under a relatively "dense" scenario, specified in assumption (5.4). Indeed, the following Theorem 5.2, shows asymptotic optimality of BH and extends the "optimality" range of the sparsity parameter to all sequences p_m such that $mp_m \to s \in (0, \infty]$. Concerning type I error component of the risk, this extension was possible due to the precise and powerful results of [21] on the expected number of false discoveries using BH under the total null hypothesis. The optimality of the type II error component under the extremely sparse scenario (2.9) results directly from a comparison with the Bonferroni correction and Lemma 5.1.

THEOREM 5.2. *Assume that*

$$m \to \infty, \qquad p_m \to 0, \qquad mp_m \to s \in (0, \infty]. \tag{5.7}$$

Then BH at the FDR level $\alpha = \alpha_m$ is ABOS if (5.5) and (5.6) hold.

The proof of Theorem 5.2 is provided in Section 7.

REMARK 5.2. Theorem 5.2 states that under the sparsity assumption (5.7), BH behaves similarly to a BFDR control rule. Specifically, if assumptions (5.5) and (5.7) are satisfied, then the BH rule is ABOS under FDR-levels $\alpha \propto (\delta\sqrt{u})^{-1}$, as in Corollary 4.1. Furthermore, if $p \propto m^{-\beta}$, with $0 < \beta \le 1$, $\frac{\log\delta}{\log m} \to C_\delta \in [0, \infty]$ and $\delta\sqrt{\log(m\delta)} \to \infty$, then a rule with FDR at the level α such that $\alpha \propto (\delta\sqrt{\log(m\delta)})^{-1}$ is ABOS. Also, in the case when $p \propto m^{-\beta}$ $(0 < \beta \le 1)$ and $\delta \propto \frac{1}{\sqrt{\log m}}$, then BH at a fixed FDR level $\alpha \in (0, 1)$ is ABOS. Thus, while the asymptotically optimal FDR levels clearly depend on the ratio of losses δ, they are independent of the sparsity parameter β; that is, ABOS property of BH is highly adaptive with respect to the level of sparsity.

The next Theorem 5.3, deals with optimality of BH under the generic assumption (2.8) which here has the form $\log\delta = o(\log m)$.

THEOREM 5.3. *Suppose* $m \to \infty$ *and* $p \propto m^{-\beta}$, *with* $0 < \beta \le 1$. *Moreover, assume that* $\log\delta = o(\log m)$ *and* $\alpha \to \alpha_\infty < 1$. *Then BH is ABOS if*

$$\log\alpha = o(\log m) \quad \textit{and} \quad \alpha\delta \to 0.$$

PROOF. Given the assumptions we are in the situation of Corollary 4.2, and it is easy to verify that therefore all assumptions of Theorem 5.2 are fulfilled. Thus ABOS holds. □

COROLLARY 5.1. *Suppose* $m \to \infty$ *and* $p \propto m^{-\beta}$, *with* $0 < \beta \le 1$. *Moreover, assume that* $\delta = \text{const}$. *Then BH is ABOS if* α *converges to* 0, *such that* $\log\alpha = o(\log m)$.

COROLLARY 5.2. *Suppose* $m \to \infty$ *and* $p \propto m^{-\beta}$, *with* $0 < \beta \le 1$. *Moreover, assume that* $\alpha = \text{const}$. *Then BH is ABOS if* δ *converges to* 0, *such that* $\log\delta = o(\log m)$.

Theorem 5.2, Remark 5.2, Theorem 5.3 and its corollaries give some general suggestions on the choice of the optimal FDR level for BH. Note, however, that according to Theorem 3.2, BH can be asymptotically optimal even when the difference between its asymptotic threshold c_{GW} and the threshold of the Bayes oracle slowly diverges to infinity. The following lemma provides a more specific condition on α and δ, which guarantees that the difference between c_{GW} and the threshold of the Bayes oracle converges to a constant.

LEMMA 5.2. *Let $p_m \propto m^{-\beta}$, for some $\beta > 0$. Moreover, assume that δ satisfies the generic assumption* (2.8) *and that α satisfies the assumptions of Theorem* 4.1. *Then the difference between the asymptotic approximation to the BH threshold c_{GW}* (4.11) *and the threshold of the Bayes oracle* (2.5) *converges to a constant if and only if the FDR level α_m and the ratio of loss functions δ_m satisfy the condition*

$$r_{\alpha_m}\delta_m = \frac{s_m}{\log m}, \tag{5.8}$$

where $r_{\alpha_m} = \frac{\alpha_m}{1-\alpha_m}$ and $s_m \to C_s \in (0, \infty)$.

PROOF. Straightforward algebra shows that the difference between the threshold of the Bayes oracle and c_{GW} is equal to

$$2\log\log m + 2\log(\delta_m r_{\alpha_m}) + \log(2\beta/C) + \log(2\beta) + C - C_1 + o_m,$$

where C_1 is the constant provided in (4.10). From this Lemma 5.2 follows easily. □

REMARK 5.3. Theorem 5.1 states that if $\beta \in (0, 1)$, then the random threshold of BH can be well approximated by c_{GW}. Therefore, in this case Lemma 5.2 provides also the "best" asymptotically optimal choices of FDR levels for BH. Since under the assumptions of Theorem 5.1 α_m converges to a constant smaller than one, condition (5.8) can be written as $\alpha_m\delta_m \propto (\log m)^{-1}$. Specifically, if $\delta_m = \text{const}$, then the sequence of best FDR levels should satisfy $\alpha_m \propto (\log m)^{-1}$. Thus the choice $\alpha_m \propto (\log m)^{-1}$ is recommended when one aims at minimizing the misclassification rate. On the other hand, BH with the fixed FDR level $\alpha \in (0, 1)$ works particularly well if $\delta_m \propto (\log m)^{-1}$.

6. Discussion. We have investigated the asymptotic optimality of multiple testing rules under sparsity, using the framework of Bayesian decision theory. We formulated conditions for the asymptotic optimality of the universal threshold of [15] and the Bonferroni correction. Moreover, similarly to [1], we have proved some asymptotic optimality properties of rules controlling the Bayesian False Discovery Rate and the Benjamini and Hochberg procedure. Comparing with [1], we replaced a loss function based on estimation error with a loss function dependent only on the type of testing error. This resulted in somewhat different optimality properties of BH. Specifically, we have shown that the optimal FDR level for BH depends on the ratio between the loss for type I and type II errors and is almost independent of the level of sparsity. Within our chosen asymptotic framework BH with the FDR levels chosen in accordance with the assumed loss function is asymptotically optimal in the entire range of sparsity parameters p, such that $p \to 0$ and $mp \to s \in (0, \infty]$. This range of values of p covers the situation when $p \propto 1/m$, and in this way it substantially extends the range of sparsity levels under which the asymptotic minimax properties of BH were proved in [1].

In this paper we proposed a new asymptotic framework to analyze properties of multiple testing procedures. According to our definition a multiple testing rule is ABOS if the ratio of its risk to the risk of the Bayes oracle converges to 1 as the number of tests increases to infinity. Our asymptotic results are to a large extend based on exact inequalities for finite values of m. The refined versions of these inequalities can be further used to characterize the rates of convergence of the ratio of risks to 1 and to compare "efficiency" of different ABOS methods. We consider this as an interesting area for further research.

The results reported in this paper provide sufficient conditions for the asymptotically optimal FDR levels for BH. They leave, however, a lot of freedom in the choice of proportionality constants, which obviously play a large role for a given finite value of m. Based on the properties of BFDR controlling rules we expect that for any given m there exists FDR level α such that the risk of BH is equal to the risk of the Bayes oracle. This finite sample optimal choice of α would depend on the actual values of the mixture parameters p and u. In recent years many Bayesian and empirical Bayes methods for multiple testing have been proposed, which provide a natural way of approximating the Bayes oracle in the case where the parameters of the mixture distribution are unknown. The advantages of these Bayesian methods, both in parametric and nonparametric settings, were illustrated in, for example, [5, 6, 17, 36, 41]. In [6] it is shown that when p is *moderately small* both fully Bayesian and empirical Bayes methods perform very well with respect to the Bayes risk. However, analysis of the asymptotic properties of fully Bayesian methods in the case where $p_m \to 0$ remains a challenging task. In the case of empirical Bayes methods, the asymptotic results given in [8] illustrate that consistent estimation of the mixture parameters is possible when $p_m \propto m^{-\beta}$, with $\beta \in (0, 1)$. New results on the convergence rates of these estimates, presented in [7], raise some hopes that proofs of the optimality properties of the corresponding empirical Bayes rules can be found. It is, however, rather unclear whether the full or empirical Bayes methods can be asymptotically optimal in the extremely sparse case of $p_m \propto m^{-1}$. Note that in this situation the expected number of signals does not increase when $m \to \infty$ and consistent estimation of the alternative distribution is not possible. These doubts, regarding the asymptotic optimality of Bayesian procedures in the extremely sparse case, are partially confirmed by the simulation study in [6], where for very small p Bayesian methods are outperformed by BH and the Bonferroni correction at the traditional FDR and FWER levels $\alpha = 0.05$.

The Benjamini–Hochberg procedure can only be directly applied when the distribution under the null hypothesis is completely specified, that is, when σ is known. In the case of testing a simple null hypothesis (i.e., when $\sigma_0 = 0$), σ can be estimated using replicates. The precision of this estimation depends on the number of replicates and can be arbitrarily good. In the case where $\sigma_0 > 0$ (i.e., when we want to distinguish large signals from background noise), the situation is quite different. In this case, σ can only be estimated by pooling the information from all the test statistics. The related modifications of the maximum likelihood method for estimating parameters in the sparse mixture (2.2) are discussed in [6]. More sophisticated methods for estimating parameters of the normal null distribution in case of no parametric assumptions on the form of the alternative are provided in [16] and [26]. In [7] it is proved that for $\beta < 1/2$ the proposed estimators based on the empirical characteristic function are minimax rate optimal. Simulation results reported in [6] show that in the parametric setting of (2.2) and for very small p, the plug-in versions of BH at FDR level $\alpha = 0.05$ outperform Bayesian approximations to the oracle. We believe that this is due to the fact that BH does not require the estimation of p, which is rather difficult when p is very small. Despite this relatively good behavior of BH, it is rather unlikely that the plug-in versions of BH are asymptotically optimal in the case where $p \propto m^{-1}$. A thorough theoretical comparison of empirical Bayes versions of BH with Bayesian approximations to the Bayes oracle and an analysis of their asymptotic optimality remains an interesting problem for future research.

Model (2.2) assumes that the statistics for different tests are independent. In principle, the model and the methods proposed in this paper can be extended to cover the situation of dependent test statistics. However, in that case the optimal Bayes solution for the compound decision problem will be more difficult to obtain. In particular the optimal Bayes classifier for the ith test may depend on the values of all other test statistics, leading to a rather complicated Bayes oracle. We believe that under specific dependency structures BH may still retain its asymptotic optimality properties. The detailed analysis of this problem requires a thorough new investigation and remains an open problem for future research.

In this paper we have modeled the test statistics using a scale mixture of normal distributions. As already mentioned, we believe that the main conclusions of the paper will hold for a substantially larger family of two component mixtures, which are currently often applied to multiple testing problems (see, e.g., [7, 16, 17]). In a recent article [9], a new "continuous" one-group model for multiple testing was proposed. As in our case, the test statistics are assumed to have a normal distribution with mean equal to zero, but the scale parameters are different for different tests and modeled as independent random variables from the one-sided Cauchy distribution. As discussed in [9], the resulting Bayesian estimate of the vector of means shrinks small effects strongly toward zero and leaves large effects almost intact. In this way, it enables very good separation of large signals from background noise. In [9] it is demonstrated that the results from the proposed procedure for multiple testing often agree with the results from Bayesian methods based on the two-group model. A thorough analysis of the asymptotic properties of the method proposed in [9] in the context of multiple testing remains a challenging task. However, we believe that the suggested one-group model has its own, very interesting virtues and Carvalho, Polson and Scott [9] clearly demonstrate that the search for modeling strategies for the problem of multiple testing, as well as for the most meaningful optimality criteria, is still an open and active area of research.

7. Proof of Theorem 5.2. The proof of Theorem 5.2 consists of two parts. The first part shows the optimality of the type I error component of the risk (see Theorem 7.1) while the second part shows that of the type II error component (see Theorem 7.2). Combining these two facts, the result follows immediately. The proofs of Theorems 7.1 and 7.2 are based on a series of intermediate results.

7.1. *Bound on the type* I *error component of the risk.* The first and most essential step of the proof of the optimality of the type I error component of the risk relies on showing that, under certain conditions, the expected number of false discoveries of BH, EV, is bounded by $c_v \alpha K$, where α is the FDR level, K is the true number of signals and c_v is a positive constant. This result is very intuitive in view of the definition of FDR [see (4.1)]. The proof is, however, nontrivial, due to the difference between $E(\frac{V}{R})$ and $\frac{EV}{ER}$.

LEMMA 7.1. *Consider the BH rule at a fixed FDR level* $\alpha \le \alpha_0 < 1$. *Let* K *be the number of true signals. The conditional expected number of false rejections given that* $K = k$, *with* $k < m(\frac{1}{\alpha_0} - 1)$, *is bounded by*

$$E(V|K = k) \le \alpha\left(\frac{k}{1-\alpha} + \frac{1}{(1-\alpha)^2}\right). \tag{7.1}$$

Specifically, for $1 \le k < m(\frac{1}{\alpha_0} - 1)$

$$E(V|K = k) \le c_v \alpha k \tag{7.2}$$

with

$$c_v = \frac{2-\alpha_0}{(1-\alpha_0)^2}. \tag{7.3}$$

PROOF. Given the condition $K = k$, there are $(m - k)$ true nulls. Let the corresponding ordered p-values be $\tilde{p}_{(1)} \le \cdots \le \tilde{p}_{(m-k)}$. Imagine that we apply to these p-values the following procedure $\tilde{B}H_k$ which rejects the hypotheses whose p-values are smaller than $\tilde{p}_{(\tilde{k})}$, where

$$\tilde{k} = \max\left\{i : \tilde{p}_{(i)} \le \frac{\alpha(i+k)}{m}\right\}. \tag{7.4}$$

Let $\tilde{V}$ be the corresponding number of rejections. Then $E(V|K = k) \le E(\tilde{V})$, since the number of false rejections for the original BH, V, is not larger than $\tilde{V}$. Now, consider m i.i.d. p-values $q_1, \dots, q_m$ from the total null (i.e., each of the m nulls is true), which are independent of the given original p-values. Let $\tilde{q}_{(1)} \le \cdots \le \tilde{q}_{(m-k)}$ be the ordered values from the subsequence $q_1, \dots, q_{m-k}$. Then $\tilde{q}_{(1)}, \dots, \tilde{q}_{(m-k)}$ and $\tilde{p}_{(1)}, \dots, \tilde{p}_{(m-k)}$ have exactly the same distribution. Let V_1 and V_2 be the number of rejections of null when the procedure (7.4) is applied to the first $(m-k)$ or m q's, respectively. Then $E(V|K=k) \le E(\tilde{V}) = E(V_1) \le E(V_2)$.

Now the bound on k (see the assumption of Lemma 7.1) guarantees that $\alpha(i+k)/m$ on the right-hand side of (7.4) is smaller than 1 for all possible i. We can thus apply Lemma 4.2 of [21] directly, which yields

$$E(V_2) = \alpha \sum_{i=0}^{m-1} (k+i+1) \binom{m-1}{i} i! \left(\frac{\alpha}{m}\right)^i.$$

Routine calculations now lead to Lemma 7.1

$$E(V_2) \le \alpha \sum_{i=0}^{\infty} (k+i+1)\alpha^i = \alpha\left(\frac{k}{1-\alpha} + \frac{1}{(1-\alpha)^2}\right). \qquad \square$$

REMARK 7.1. Note that in the case where $\alpha_0 < 0.5$, the inequality $k < m(\frac{1}{\alpha_0} - 1)$ is always fulfilled.

The following lemma is an extension of Lemma 7.1 to the mixture model (2.2).

LEMMA 7.2. *Under assumptions* (5.5) *and* (5.7), *the expected number of false rejections is bounded by*

$$E(V) < C_2 \alpha_m m p_m,$$

where C_2 *is any constant satisfying*

$$C_2 > \begin{cases} \dfrac{2-\alpha_\infty}{(1-\alpha_\infty)^2}, & \text{when } s = \infty, \\ \dfrac{e^{-s}}{s(1-\alpha_\infty)^2} + \dfrac{2-\alpha_\infty}{(1-\alpha_\infty)^2}, & \text{when } s \in (0,\infty). \end{cases}$$

PROOF. Define $C_6 := \frac{1}{\alpha_\infty} - 1$ and $m_0 := \min(m, C_6 m)$. The following holds:

$$E(V) \le \sum_{k=0}^{m_0} E(V|K=k) P(K=k) + m P(K > m_0). \tag{7.5}$$

The first term can be bounded for m large enough using Lemma 7.1,

$$\sum_{k=0}^{m_0} E(V|K=k)P(K=k) \le \frac{\alpha_m}{(1-\alpha_m)^2}(1-p_m)^m + \tilde{c}_v \alpha_m m p_m,$$

where $\tilde{c}_v$ is any constant larger than $\frac{2-\alpha_\infty}{(1-\alpha_\infty)^2}$. Now observe that $\frac{1}{(1-\alpha_m)^2}(1-p_m)^m$ converges to 0 if $s = \infty$ or to $\frac{e^{-s}}{(1-\alpha_\infty)^2}$ otherwise. Hence, it follows that

$$\sum_{k=0}^{m_0} E(V|K=k)P(K=k) < C_2 m \alpha_m p_m,$$

for any constant C_2 satisfying the assumption of Lemma 7.2.

Finally, note that the second term of (7.5) vanishes for $\alpha_\infty < 0.5$. On the other hand, for $\alpha_\infty \in [0.5, 1)$, Lemma 7.1 of [1] yields

$$mP(K > m_0) = mP(K > C_6 m) \le m \exp\left(-\tfrac{1}{4} m p_m h(C_6/p_m)\right),$$

where $h(x) = \min(|x-1|, |x-1|^2)$. If $p_m \to 0$, then for any constant $C_7 \in (0, C_6)$ and sufficiently large m, the right-hand side is bounded from above by $m\exp(-C_7 m) \to 0$. Now, from the assumptions $mp_m \to s > 0$ and $\alpha_m \to \alpha_\infty > 0.5$, it follows that for any constant $\beta_2 > 0$ and sufficiently large m, the second term of (7.5) is smaller than $\beta_2 \alpha_m m p_m$, and Lemma 7.2 follows. □

Lemma 7.2 easily leads to the following Theorem 7.1, on the optimality of the type I error component of the risk of BH.

THEOREM 7.1. *Under assumptions* (5.5)–(5.7), *the type* I *error component of the risk of BH*, $R_1 = \delta_0 E(V)$, *satisfies* $\frac{R_1}{R_{\mathrm{opt}}} \to 0$, *where* R_{opt} *is the optimal risk defined in Theorem* 3.1.

PROOF. From Lemma 7.2

$$\frac{R_1}{R_{\mathrm{opt}}} = \frac{\delta_0 E(V)}{R_{\mathrm{opt}}} \le C_3 \alpha_m \delta_m (1 + o_m), \tag{7.6}$$

where $C_3 = \frac{C_2}{2\Phi(\sqrt{C})-1}$. Now, observe that the left-hand side of (4.9) [included in assumption (5.6)] can be written as

$$2\log(\delta_m r_{\alpha_m}) + \log u - \log\log(f/r_{\alpha_m}),$$

and under (4.8) and Assumption (A) it can be further reduced to

$$2\log(\delta_m r_{\alpha_m}) - \log C + o_m.$$

Thus assumptions (4.9) and (5.5) together imply that $\delta_m \alpha_m \to 0$, and from (7.6) it immediately follows that $\frac{R_1}{R_{\mathrm{opt}}} \to 0$. □

7.2. *Bound on the type* II *component of the risk*. To prove the optimality of the type II error component of the risk of BH, we consider the extremely sparse case (2.9) and the denser case (2.10) separately. Note that in the extremely sparse case, the optimality of the type II component of the risk of BH follows directly from a comparison with the more conservative Bonferroni correction, which according to Lemma 5.1 is ABOS in this range of sparsity parameters.

The proof of optimality for the denser case is based on the approximation of the random threshold of BH by the asymptotically optimal threshold c_{GW} [see (4.11)], given in Theorem 5.1. The corresponding "denser" case assumption (5.4) is substantially less restrictive than (2.10) and partially covers the extremely sparse case (2.9).

THEOREM 7.2. *Under the assumptions of Theorem 5.1 the type* II *error component of the risk of BH satisfies*

$$R_2 \le R_{\mathrm{opt}}(1 + o_m). \tag{7.7}$$

PROOF. Denote the number of false negatives under the BH rule by T. Let us fix $\varepsilon > 0$ and let $\tilde{c}_1 = c_{\mathrm{GW}} + \varepsilon$. Clearly,

$$E(T) \le E(T|c_{\mathrm{BH}} \le \tilde{c}_1)P(c_{\mathrm{BH}} \le \tilde{c}_1) + mP(c_{\mathrm{BH}} > \tilde{c}_1),$$

and furthermore

$$E(T|c_{\mathrm{BH}} \le \tilde{c}_1)P(c_{\mathrm{BH}} \le \tilde{c}_1) \le ET_1,$$

where T_1 is the number of false negatives produced by the rule based on the threshold $\tilde{c}_1$. Note that the rule based on $\tilde{c}_1$ differs from the asymptotically optimal rule c_{GW} only by a constant, and therefore, from Theorem 3.2, it is asymptotically optimal. Hence, it follows that $\delta_A ET_1 = R_{\mathrm{opt}}(1 + o_m)$. On the other hand, from Theorem 5.1, for any $\beta_1 > 0$ and sufficiently large m (dependent on ε and β_1)

$$P(c_{\mathrm{BH}} > \tilde{c}_1) \le m^{-\beta_1}.$$

Therefore,

$$R_2 = \delta_A ET \le R_{\mathrm{opt}}(1 + o_m) + \delta_A m^{1-\beta_1}.$$

Now, observe that under assumption (5.4)

$$\frac{\delta_A m^{1-\beta_1}}{R_{\mathrm{opt}}} = C_4 \frac{m^{-\beta_1}}{p} < C_4 \frac{m^{1-\beta_1}}{\log^{\beta_p} m},$$

where $C_4 = \frac{1}{2\Phi(\sqrt{C})-1}$. Thus, choosing, for example, $\beta_1 = 1$, we conclude that $\delta_A m^{1-\beta_1} = o(R_{\mathrm{opt}})$, and the proof is thus complete. □

Acknowledgments. We want to thank two anonymous referees and the Associate Editor for many constructive suggestions to improve this manuscript. We would also like to express our gratitude to David Ramsey for careful reading of the manuscript and helpful comments and to Krzysztof Bogdan for several helpful suggestions.

SUPPLEMENTARY MATERIAL

Supplement to "Asymptotic Bayes-optimality under sparsity of some multiple testing procedures" (DOI: 10.1214/10-AOS869SUPP; .pdf). Analysis of behavior of BFDR for scale mixtures of normal distributions and proofs of Theorems 3.2, 4.1 and 5.1.

REFERENCES

[1] ABRAMOVICH, F., BENJAMINI, Y., DONOHO, D. L. and JOHNSTONE, I. M. (2006). Adapting to unknown sparsity by controlling the false discovery rate. *Ann. Statist.* **34** 584–653. MR2281879

[2] 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** 289–300. MR1325392

[3] BOGDAN, M., CHAKRABARTI, A., FROMMLET, F. and GHOSH, J. K. (2010). Supplement to "Asymptotic Bayes-optimality under sparsity of some multiple testing procedures." DOI:10.1214/10-AOS869SUPP.

[4] BOGDAN, M., CHAKRABARTI, A., FROMMLET, F. and GHOSH, J. K. (2010). The Bayes oracle and asymptotic optimality of multiple testing procedures under sparsity. Available at arXiv:1002.3501.

[5] 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.

[6] BOGDAN, M., GHOSH, J. K. and TOKDAR, S. T. (2008). A comparison of the Benjamini–Hochberg procedure with some Bayesian rules for multiple testing. In *Beyond Parametrics in Interdisciplinary Research: Festschrift in Honor of Professor Pranab K. Sen. Inst. Math. Stat. Collect.* **1** 211–230. IMS, Beachwood, OH. MR2462208

[7] CAI, T. T. and JIN, J. (2010). Optimal rates of convergence for estimating the null density and proportion of nonnull effects in large-scale multiple testing. *Ann. Statist.* **38** 100–145. MR2589318

[8] CAI, T. T., JIN, J. and LOW, M. G. (2007). Estimation and confidence sets for sparse normal mixtures. *Ann. Statist.* **35** 2421–2449. MR2382653

[9] CARVALHO, C. M., POLSON, N. and SCOTT, J. G. (2008). The horseshoe estimator for sparse signals. Technical Report 2008-31, Dept. Statistical Science, Duke Univ.

[10] CHI, Z. (2007). On the performance of FDR control: Constraints and a partial solution. *Ann. Statist.* **35** 1409–1431. MR2351091

[11] CHI, Z. (2008). False discovery rate control with multivariate p-values. *Electron. J. Stat.* **2** 368–411. MR2411440

[12] DO, K.-A., MÜLLER, P. and TANG, F. (2005). A Bayesian mixture model for differential gene expression. *J. Roy. Statist. Soc. Ser. C* **54** 627–644. MR2137258

[13] DONOHO, D. and JIN, J. (2004). Higher criticism for detecting sparse heterogeneous mixtures. *Ann. Statist.* **32** 962–994. MR2065195

[14] DONOHO, D. and JIN, J. (2006). Asymptotic minimaxity of false discovery rate thresholding for sparse exponential data. *Ann. Statist.* **34** 2980–3018. MR2329475

[15] DONOHO, D. L. and JOHNSTONE, I. M. (1994). Minimax risk over l_p-balls for l_q-error. *Probab. Theory Related Fields* **99** 277–303. MR1278886

[16] EFRON, B. (2004). Large-scale simultaneous hypothesis testing: The choice of a null hypothesis. *J. Amer. Statist. Assoc.* **99** 96–104. MR2054289

[17] EFRON, B. (2008). Microarrays, empirical Bayes and the two-groups model. *Statist. Sci.* **23** 1–22. MR2431866

[18] EFRON, B. and TIBSHIRANI, R. (2002). Empirical Bayes methods and false discovery rates for microarrays. *Genet. Epidemiol.* **23** 70–86.

[19] EFRON, B., TIBSHIRANI, R., STOREY, J. D. and TUSHER, V. (2001). Empirical Bayes analysis of a microarray experiment. *J. Amer. Statist. Assoc.* **96** 1151–1160. MR1946571

[20] FINNER, H., DICKHAUS, T. and ROTERS, M. (2009). On the false discovery rate and an asymptotically optimal rejection curve. *Ann. Statist.* **37** 596–618. MR2502644

[21] FINNER, H. and ROTERS, M. (2002). Multiple hypotheses testing and expected number of type I errors. *Ann. Statist.* **30** 220–238. MR1892662

[22] FROMMLET, F., BOGDAN, M. and CHAKRABARTI, A. (2010). Asymptotic Bayes optimality under sparsity for general priors under the alternative. Available at arXiv:1005.4753v1.

[23] 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** 499–517. MR1924303

[24] GENOVESE, C. and WASSERMAN, L. (2004). A stochastic process approach to false discovery control. *Ann. Statist.* **32** 1035–1061. MR2065197

[25] GUO, W. and RAO, M. B. (2008). On optimality of the Benjamini–Hochberg procedure for the false discovery rate. *Statist. Probab. Lett.* **78** 2024–2030. MR2528629

[26] JIN, J. and CAI, T. T. (2007). Estimating the null and the proportional of nonnull effects in large-scale multiple comparisons. *J. Amer. Statist. Assoc.* **102** 495–506. MR2325113

[27] LEHMANN, E. L. (1957). A theory of some multiple decision problems. I. *Ann. Math. Statist.* **28** 1–25. MR0084952

[28] LEHMANN, E. L. (1957). A theory of some multiple decision problems. II. *Ann. Math. Statist.* **28** 547–572. MR0096338

[29] LEHMANN, E. L., ROMANO, J. P. and SHAFFER, J. P. (2005). On optimality of stepdown and stepup multiple test procedures. *Ann. Statist.* **33** 1084–1108. MR2195629

[30] MEINSHAUSEN, N. and RICE, J. (2006). Estimating the proportion of false null hypotheses among a large number of independently tested hypotheses. *Ann. Statist.* **34** 373–393. MR2275246

[31] 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** 990–1001. MR2109489

[32] PEÑA, E. A., HABIGER, J. D. and WU, W. (2011). Power-enhanced multiple decision functions controlling family-wise error and false discovery rates. *Ann. Statist.* **39** 556–583.

[33] ROQUAIN, E. and VAN DE WIEL, M. A. (2009). Optimal weighting for false discovery rate control. *Electron. J. Stat.* **3** 678–711. MR2521216

[34] SARKAR, S. K. (2006). False discovery and false nondiscovery rates in single-step multiple testing procedures. *Ann. Statist.* **34** 394–415. MR2275247

[35] SCOTT, J. G. and BERGER, J. O. (2006). An exploration of aspects of Bayesian multiple testing. *J. Statist. Plann. Inference* **136** 2144–2162. MR2235051

[36] SCOTT, J. G. and BERGER, J. O. (2010). Bayes and empirical-Bayes multiplicity adjustment in the variable-selection problem. *Ann. Statist.* **38** 2587–2619.

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

[38] SEN, P. K. (1999). Some remarks on Simes-type multiple tests of significance. *J. Statist. Plann. Inference* **82** 139–145. MR1736438

[39] SIMES, R. J. (1986). An improved Bonferroni procedure for multiple tests of significance. *Biometrika* **73** 751–754. MR0897872

[40] STOREY, J. D. (2003). The positive false discovery rate: A Bayesian interpretation and the q-value. *Ann. Statist.* **31** 2013–2035. MR2036398

[41] STOREY, J. D. (2007). The optimal discovery procedure: A new approach to simultaneous significance testing. *J. R. Stat. Soc. Ser. B Stat. Methodol.* **69** 347–368. MR2323757

[42] 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** 187–205. MR2035766

[43] SUN, W. and CAI, T. T. (2007). Oracle and adaptive compound decision rules for false discovery rate control. *J. Amer. Statist. Assoc.* **102** 901–912. MR2411657

[44] TAI, Y. C. and SPEED, T. P. (2006). A multivariate empirical Bayes statistic for replicated microarray time course data. *Ann. Statist.* **34** 2387–2412. MR2291504

M. BOGDAN
INSTITUTE OF MATHEMATICS
AND COMPUTER SCIENCE
WROCŁAW UNIVERSITY OF TECHNOLOGY
UL. JANISZEWSKIEGO 14A
50-370 WROCLAW
POLAND
E-MAIL: Malgorzata.Bogdan@pwr.wroc.pl

F. FROMMLET
DEPARTMENT OF STATISTICS
AND DECISION SUPPORT SYSTEMS
UNIVERSITY OF VIENNA
BRÜNNERSTRASSE 72
1210 VIENNA
AUSTRIA
E-MAIL: Florian.Frommlet@univie.ac.at

A. CHAKRABARTI
J. K. GHOSH
INDIAN STATISTICAL INSTITUTE
203 B.T. ROAD
KOLKATA 700108, WEST BENGAL
INDIA
E-MAIL: arc@isical.ac.in
jayanta@isical.ac.in

J. K. GHOSH
M. BOGDAN
DEPARTMENT OF STATISTICS
PURDUE UNIVERSITY
150 NORTH UNIVERSITY STREET
WEST LAFAYETTE, INDIANA 47907
USA
E-MAIL: ghosh@purdue.edu
bogdanm@stat.purdue.edu