Submitted to the Annals of Statistics

THE BAYES ORACLE AND ASYMPTOTIC OPTIMALITY OF MULTIPLE TESTING PROCEDURES UNDER SPARSITY

By Małgorzata Bogdan[*,†], Arijit Chakrabarti[‡], Florian Frommlet[§] and Jayanta K. Ghosh[†,‡]

Wrocław University of Technology[*], *Purdue University*[†], *Indian Statistical Institute*[‡] *and University of Vienna*[§]

We investigate the asymptotic optimality of a large class of multiple testing rules using the framework of Bayesian Decision Theory. We consider a parametric setup, in which observations come from a normal scale mixture model and assume that the total loss is the sum of losses for individual tests. Our model can be used for testing point null hypotheses of no signals (zero effects), as well as to distinguish large signals from a multitude of very small effects. The optimality of a rule is proved by showing that, 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 under which the proportion p of "true" alternatives converges to zero. We fully characterize the class of fixed threshold multiple testing rules which are asymptotically optimal and hence derive conditions for the asymptotic optimality of rules controlling the Bayesian False Discovery Rate (BFDR). We also provide conditions under which the popular Benjamini-Hochberg and Bonferroni procedures are asymptotically optimal and show that for a wide class of sparsity levels, the threshold of the former can be approximated very well by a non-random threshold.

Our results show that for optimal performance the BFDR (or FDR) controlling level should be chosen to be small if the expected signal magnitude or the relative cost of a type I error is large. We also show that for a wide range of sparsity levels (i.e. rates of convergence of p to zero) and expected signal magnitudes, the Benjamini-Hochberg rule controlling the FDR at a fixed level $\alpha \in (0,1)$ is asymptotically optimal; provided only that the ratio of losses for type I and type II errors converges to zero at a slow rate which can vary quite widely. When the loss ratio is constant, similar optimality results hold if the FDR controlling level slowly converges to zero as $p \to 0$. As far as we know, this is the first proof of the decision theoretic asymptotic optimality of the Benjamini-Hochberg rule in the context of hypothesis testing.

AMS 2000 subject classifications: Primary 62C25,62F05; secondary 62C10
Keywords and phrases: Multiple testing, FDR, Bayes oracle,asymptotic optimality

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 levels of thousands of genes (e.g. see [15], [14], [35], [18], [25], [26], [27] or [34]). 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 could mention 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 [13], [14], [15], [37] and [4]) and full Bayes tests (see e.g. [25], [8], [27] and [4]) have also been proposed and are used extensively in such problems.

In recent years, substantial efforts have 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. We cite a few among many important papers, ([9], [10], [24], [20], [5]). A major theoretical breakthrough was made in [1], where, in a problem of estimating a sparse vector of means, a data-dependent thresholding estimator for the unknown means is proposed, the threshold being determined by applying the Benjamini-Hochberg procedure (henceforth denoted as BH). Specifically, in [1] it is shown that this estimator 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 this paper we analyze the properties of multiple testing rules from the perspective of Bayesian Decision Theory. We assume fixed losses δ_0 and δ_A for type I and type II errors, respectively, for each test and define the overall loss of a multiple testing rule as the sum of the 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. The main result of this paper is the proof of the asymptotic optimality properties of BH within this Bayesian perspective. 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 [14] and [17], 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 [17], [3] and [4]. 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.

We consider multiple testing where our observations are assumed to come from a normal scale mixture model (see (2.6) below). This model has been used earlier in the context of multiple testing (see, e.g., [27], [3] and [4]) and differs from the model used in [1] by imposing a normal prior distribution on the unknown vector of means. As discussed in Section 2, depending on the form of the assumed mixture distribution, this model can be used for testing point null hypotheses to decide if the unknown means are zero, as well as for identifying large signals embedded in a multitude of very small effects. In this situation, each individual test tries to decide which component of the mixture generated the corresponding data. For the rest of the paper we will use the generic term "signal" to refer to the unknown mean value under the alternative. Under an additive loss function, we first find the Bayes rule which minimizes the overall risk (the Bayes risk), and this rule is henceforth referred to as the Bayes oracle. The Bayes oracle turns out to be a rule which applies a *fixed threshold* critical region (of the form $Y_i > K$) for each individual test. This threshold and the properties of the Bayes oracle depend on three parameters: the sparsity level p, ratio of losses $\delta = \frac{\delta_0}{\delta_A}$ and average squared signal magnitude u, defined as the ratio between the variances of the non-null and null components of the mixture. In our asymptotic considerations, we mainly consider the scenario where p goes to zero and $\log \delta = o(\log p)$. We observe that in this situation the Bayes oracle has an asymptotic power larger than zero if and only if u increases to infinity, such that $-\frac{u}{\log p} \to C_u \in (0, \infty]$. We concentrate our attention on such detectable signals and classify a multiple testing rule as asymptotically optimal if, in this setting, the ratio of its Bayes risk to that of the Bayes oracle converges to one. We place special emphasis on the case where $C_u < \infty$. In this situation, the asymptotic power of the Bayes oracle is smaller than one and we classify the signals $u \propto -\log p$ as signals on "the verge of detectability". Specifically, if $p \propto m^{-\beta}$, where m is the number of tests and

$\beta \in (0, \infty)$, then signals on the verge of detectability satisfy

$$\frac{u}{\log m} \to C_u \in (0, \infty) \quad . \tag{1.1}$$

This result is quite natural, since, under sparsity, the magnitude of the largest test statistic corresponding to the null hypothesis is of the order $2 \log m$. Thus, signals increasing to infinity at a rate slower than $\log m$ cannot be distinguished from the largest components of the noise. In a frequentist setting, a similar scaling for asymptotically detectable signals was proposed in [9] (with u replaced by the square of a true mean).

In the first part of this paper we study fixed threshold tests in great detail and fully characterize the class of asymptotically optimal fixed threshold testing rules. Using this, we specify conditions for the asymptotic optimality of the "universal threshold" $2 \log m$ of [11] and the closely related Bonferroni correction. We also provide conditions for the asymptotic optimality of fixed threshold multiple testing rules which control the Bayesian False Discovery Rate (BFDR) at a given level α. It turns out that the optimal choice of α depends on the expected signal magnitude, u, and the ratio between the losses, δ. Broadly speaking, α should decrease when u or δ increases. Our results also show that for a wide range of choices of the sparsity level p and expected signal magnitude u, a rule controlling the BFDR at a fixed level α is asymptotically optimal if the ratio between losses, δ, converges to zero at a suitably slow rate. In the case where the sequence of sparsity levels satisfies $p_m \propto m^{-\beta}$ and signals are on the verge of detectability, our results take an especially simple form. Specifically, we prove that under this scenario a rule controlling the BFDR at a fixed FDR level $\alpha \in (0, 1)$ is asymptotically optimal for a wide range of loss ratios δ_m satisfying

$$\delta_m \to 0 \qquad \text{and} \qquad \frac{\log \delta_m}{\log m} \to 0 \tag{1.2}$$

(see Corollary 5.5). The assumption that $\delta_m \to 0$ as $p_m \to 0$ agrees with the intuition that the cost of missing a signal should be relatively large if the true number of signals is small.

The final results of the paper are included in Section 6, where we prove some optimality properties of the Benjamini-Hochberg procedure. Here we assume that $m \to \infty$ and $p_m \to 0$ in such a way that $mp_m \to C_p \in (0, \infty]$. We distinguish between two cases. In the case where

$$p_m \geq \frac{log^{\beta} m}{m}, \quad \text{for some} \quad \beta > 1 \quad , \tag{1.3}$$

the proof can be based on a comparison with a fixed threshold BFDR control rule. Specifically, the approximation of the random threshold used in the BH procedure by the threshold of the BFDR control rule works at this level of sparsity. The assumption (1.3) is similar to the one used in [1] for proving the optimality of BH, though our results are proved in a substantially different asymptotic context. Furthermore, BH is also shown to be asymptotically optimal for the extremely sparse case, where $p_m = z_m/m$, such that z_m converges to a finite positive constant or diverges to infinity in such a way that $\log z_m = o(\log m)$. In this situation the type I error component of the risk is bounded by invoking the results of [16] on the expected number of type I errors under BH, while the bound on the type II error component of the risk follows from a comparison of BH with the Bonferroni rule, which is asymptotically optimal for this range of p_m.

Our results show that for a wide range of choices of the mixture parameters, the BH rule shares the asymptotic optimality properties of the BFDR control rules discussed above, and adapts very well to the unknown sparsity. Specifically, in Corollary 6.2 we show that for any sequence of sparsity levels p_m, satisfying

for some constants $a_1 \in (0,\infty), a_2 \in (0,\infty)$ and $a_3 \in (0,1)$

$$a_1 m^{-1} \leq p_m \leq a_2 m^{-a_3} \quad , \tag{1.4}$$

and any sequence of loss ratios satisfying (1.2), a BH rule with a fixed FDR level $\alpha \in (0,1)$ is asymptotically optimal for all signals on the verge of detectability (i.e. signals satisfying (1.1)). In comparison to [1], our general results give some hints on how the optimal FDR level should be chosen, depending on the expected magnitude of the signal and the ratio between the losses. As far as we know, this is the first thorough discussion of the decision theoretic optimality of the Benjamini-Hochberg procedure in the context of hypothesis testing.

As already mentioned, we place great emphasis on discussing asymptotic optimality for signals "on the verge of detectability". We believe that rules which perform well in this region could be used as a kind of "gold standard" when not much information about the magnitude of the signal is available. But our optimality results are of a more general nature and specify optimality conditions for the whole range of detectable signals.

The outline of the paper is as follows. In Section 2 we define and discuss our model. In Section 3 we introduce the decision theoretic and asymptotic framework of the paper. We present the Bayes oracle, which minimizes the Bayes risk, and formulate the conditions under which the asymptotic power of this rule is larger than 0. We also provide a formula for the optimal Bayes risk. In Section 4 we give a definition of asymptotic optimality in terms of the Bayes risk and characterize the fixed threshold multiple testing rules which are asymptotically optimal. Section 4 also contains two examples of asymptotically optimal rules, which are related to the "universal threshold" $2\log m$ of [11]. In Section 5 we discuss the Bayesian False Discovery Rate and give conditions under which controlling the BFDR is asymptotically optimal. We also provide conditions for the asymptotic optimality of the Bonferroni correction and relate the asymptotic approximation of the Benjamini-Hochberg random threshold to the threshold of the BFDR control rule. Section 6 contains results on the asymptotic optimality of the BH procedure, while Section 7 contains a discussion and directions for further research. The majority of the proofs can be found in the Appendix.

2. Statistical model. In this section we introduce the normal scale mixture model, in the context of which we study multiple testing. We will explain below that this model, previously applied in [27] and [4], can be used both for testing point null hypotheses, as well as for distinguishing a small number of relatively large signals from a multitude of very small effects. We believe that the latter case is much more realistic in large scale multiple testing applications, e.g. microarray studies.

Suppose we have m independent observations $X_1, \ldots, X_m$ and assume that each X_i has a normal $N(\mu_i, \sigma_\epsilon^2)$ 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 modelling the μ_i's as iid rv's from the following mixture distribution:

$$\mu_i \sim (1-p)N(0, \sigma_0^2) + pN(0, \sigma_0^2 + \tau^2) \ . \tag{2.5}$$

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

$$X_i \sim (1-p)N(0, \sigma^2) + pN(0, \sigma^2 + \tau^2) \ , \tag{2.6}$$

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

We will use the term "sparse mixture" to refer to the situation when $p \approx 0$.

Note that in the case where $\sigma_0^2 = 0$, H_{0i} corresponds to the point null hypothesis that $\mu_i = 0$. 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. E.g., 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 non-null 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., [9] or [1]). However, in practical applications σ is often unknown and needs to be estimated. In the case of a simple null hypothesis (i.e. when $\sigma_0^2 = 0$), σ^2 can be precisely estimated by using replicates of X_i. In the case where $\sigma_0^2 > 0$ the situation is more difficult, but σ^2 can still be estimated by pooling the information from all the test statistics and applying Empirical Bayes methods (see e.g., [4]). Some discussion on the issue of estimating the parameters in sparse mixtures is provided in Section 7.

REMARK 2.1. The proposed mixture model for X_i is a specific example of the two-groups model, which was discussed in a wider nonparametric context e.g in [15], [13], [18] and [4]. Restricting attention to normal mixtures allows us to reduce the technical complexity of the proofs and to concentrate on the main aspects of the problem. We believe that similar results also hold in a substantially more general setting, e.g. when the normal distribution is replaced by another suitable scale distribution, with a large scale parameter under the alternative.

3. The Bayes oracle. We consider a Bayesian decision theoretic formulation of the multiple testing problem of testing H_{0i} versus H_{Ai}, for each $i = 1, \ldots, m$ simultaneously. For each i, there are two possible "states of nature", namely H_{0i} or H_{Ai}, that occur with probabilities $(1-p)$ and p, respectively. As indicated in Section 2, under H_{0i}, $X_i \sim N(0, \sigma^2)$, while under H_{Ai}, $X_i \sim N(0, \sigma^2 + \tau^2)$. Table 1 defines the matrix of losses for making a decision in the i^{th} test.

TABLE 1
Matrix of losses

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

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 [21] and [22], 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

$$\frac{\phi_A(X_i)}{\phi_0(X_i)} \geq \frac{(1-p)\delta_0}{p\delta_A}\,, \tag{3.7}$$

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:

$$\text{Reject} \quad H_{0i} \quad \text{if} \quad \frac{X_i^2}{\sigma^2} \geq c^2\,, \tag{3.8}$$

where

$$c^2 = c^2_{\tau,\sigma,f,\delta} = \frac{\sigma^2+\tau^2}{\tau^2}\left(\log\left(\left(\frac{\tau}{\sigma}\right)^2+1\right)+2\log(f\delta)\right) \tag{3.9}$$

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 and therefore is not attainable in finite samples.

REMARK 3.1. The Bayes oracle for multiple tests as defined above was introduced independently in [36] and [4]. Two other oracles for multiple tests have recently been proposed in [34] and [36]. They are both based on the principles of classical statistics and aim to maximize the number of true discoveries, while keeping the expected number of false positives or false discovery rate at a given level. Interestingly, Sun and Cai [36] point out a relationship between their oracle and the Bayes oracle (3.7). In practical applications both of the methods proposed in [34] and [36] require estimation of the parameters of the mixture distribution. The asymptotic results given in [36] illustrate the optimality of the corresponding multiple testing procedure in the proposed classical context and for any fixed (though unknown) $p \in (0,1)$.

The oracle (3.7), considered in this manuscript, is motivated by traditional Bayesian decision theory and minimizes a weighted average of the misclassification errors of both types. We are interested in identifying multiple testing procedures that are asymptotically as good as this oracle. We consider the case where p tends to zero, which requires rather subtle methods. Specifically, under this scenario only relatively large signals have a chance of being detected by a Bayes oracle. The corresponding assumption (A), specifying the range of detectable signals, is proposed in Section 3.1. In a frequentist setting, a similar scaling of the asymptotically detectable signals was introduced in [9]. Some extensions of the latter are obtained in [19].

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.

In the remainder of this section we provide formulas for the probabilities of type I and type II errors using the Bayes oracle and calculate the corresponding Bayes risk. We also introduce the asymptotic framework used in this article.

3.1. *Type II errors and 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 (3.9). 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 (3.9) depends on τ and σ only through $u = \left(\frac{\tau}{\sigma}\right)^2$. 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

$$c_{u,v}^2 = \left(1 + \frac{1}{u}\right)(\log v + \log(1 + 1/u)) \; . \tag{3.10}$$

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

$$t_2 = P\left(Z^2 < \frac{1}{u+1} c_{u,v}^2\right) \; , \tag{3.11}$$

where Z is a standard normal variable.

From (3.11) it follows that given an arbitrary infinite sequence of γ's, the limiting power of the Bayes oracle is non-zero 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 non-trivial 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 (3.9) 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 $\frac{c_{u,v}^2}{u+1} \approx \frac{\log v}{u}$. This explains two of the three asymptotic conditions we impose below 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, $\theta_t = (u_t, v_t)$, fulfills the following conditions: $u_t \to \infty$, $v_t \to \infty$ and $\frac{\log v_t}{u_t} \to C \in [0, \infty)$, as $t \to \infty$.

REMARK 3.2. While this article is mainly focused on the case $p \to 0$, the asymptotic results which follow can also be applied in other situations where Assumption (A) is satisfied. We do not allow $C = \infty$ in Assumption (A), because then using the Bayes oracle the limit of the probability of a type II error is equal to one and signals cannot be identified. For other values of C the limiting power for the detection of signals is in $(0, 1]$. We call the corresponding parametric region detectable. If $C = 0$, then the oracle has a limiting power equal to one (see equation (3.13) below). As discussed in Section 4.1, such a situation can occur naturally if the number of replicates used to calculate X_i increases to infinity as $p \to 0$. In the case where $C \in (0, \infty)$, the asymptotic power is smaller than one and we refer to the corresponding parametric region as "the verge of detectability".

When $p \to 0$ and $\log(\delta) = o(\log(p))$, Assumption (A) reduces to $-\frac{u}{\log p} \to C_u \in (0, \infty]$ and specifies the relationship between the magnitude u of asymptotically detectable signals and the sparsity parameter p. Interestingly, in this case, signals on the verge of detectability, $u \propto -\log p$, can be related to asymptotically least-favorable configurations for $l_0[p]$ balls (defined in Section 5 below) 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 $l_0[p]$.

REMARK 3.3. A similar relationship between the sparsity parameter p and the signal magnitude can also be shown to be necessary for ensuring non-trivial inference for mixtures of other types of scale families, for example the gamma family or a generalization of the double exponential, namely $f(x) = exp(-|x|^\alpha)$, $\alpha > 0$.

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.

LEMMA 3.1. *Under Assumption (A) the probability of a type II error using the Bayes oracle is given by the following equations:*

$$t_2 = (2\Phi(\sqrt{C}) - 1)(1 + o_t) , \tag{3.12}$$

when $C \in (0, \infty)$, *and*

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

when $C = 0$.

PROOF. Lemma 3.1 easily follows from (3.11) and Assumption (A). □

3.2. *Type I errors.*

LEMMA 3.2. *Under Assumption (A), the probability of a type I error using the Bayes oracle is given by*

$$t_1 = e^{-C/2} \sqrt{\frac{2}{\pi v \log v}} (1 + o_t) . \tag{3.14}$$

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

$$c_{u,v}^2 = \log v (1 + z_{u,v}) , \tag{3.15}$$

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

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

where $\phi(\cdot)$ is the density of the standard normal distribution. This, together with Assumption (A), yields

$$\phi(c_{u,v}) = e^{-C/2} \sqrt{\frac{1}{2 \pi v}} (1 + o_t) . \tag{3.16}$$

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

$$P(|Z| > c) = \frac{2 \phi(c)}{c} (1 - z_1(c)) , \tag{3.17}$$

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

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

$$R = \sum_{i=1}^{m} \{(1-p)t_{1i}\delta_0 + pt_{2i}\delta_A\}. \tag{3.18}$$

In particular, the Bayes risk for a fixed threshold multiple testing procedure is given by

$$R = m((1-p)t_1\delta_0 + pt_2\delta_A)\ . \tag{3.19}$$

Equations (3.12), (3.13) and (3.14) 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*

$$R_{opt} = mp\delta_A \sqrt{\frac{2\log v}{\pi u}}(1+o_t)\ , \tag{3.20}$$

when $C = 0$

or

$$R_{opt} = mp\delta_A(2\Phi(\sqrt{C}) - 1)(1+o_t)\ , \tag{3.21}$$

when $0 < C < \infty$.

REMARK 3.4. It is important to note that under Assumption (A), the asymptotic form of the risk under the Bayes oracle R_{opt} is determined by the component of risk corresponding to type II errors, $mpt_2\delta_A$. This is due to the fact that the probability of a type I error is much more sensitive to a change in the threshold than the probability of a type II error. Specifically, it is easy to check that a "slight" decrease in the threshold leads to an increase in the rate of convergence of the component of risk corresponding to type I errors such that it equals the rate of convergence of R_{opt}, without affecting the rate and the constant corresponding to the type II error component. Thus the risk of the resulting "balanced" rule would have the same rate of convergence as the Bayes oracle, but with a larger constant of proportionality.

4. Asymptotically optimal rules.
In this section we formally define the asymptotic optimality of multiple testing rules and then characterize the class of asymptotically optimal rules with fixed thresholds.

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

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

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

REMARK 4.1. 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 optimal multiple testing rules with fixed thresholds.

THEOREM 4.1. *A multiple testing rule of the form (3.8) with threshold* $c^2 = c_t^2 = \log v + z_t$ *is asymptotically optimal if and only if*

$$z_t = o(\log v) \tag{4.22}$$

and

$$z_t + 2\log\log v \to \infty \ . \tag{4.23}$$

The proof of Theorem 4.1 is given in Appendix 8.1.

REMARK 4.2. Conditions (4.22) and (4.23) guarantee the asymptotic optimality of the components of risk corresponding to type II and type I errors, respectively.

REMARK 4.3. We have observed that the Bayes oracle is a fixed threshold test. So it is natural that an optimal multiple testing procedure will be of this kind or will behave (at least asymptotically) like a fixed threshold test, with the threshold depending on the unknown parameters. So a study of the optimality of fixed threshold tests may give important clues about the optimality of more general tests. In Section 6 this is shown to be true in the context of proving the optimality of the popular Benjamini-Hochberg [2] procedure.

4.1. *Examples.* Here we present two multiple testing rules, which are asymptotically optimal when $p_m \propto \frac{1}{m}$. Both rules are closely related to the universal threshold $2\log m$ of [11], which, according to [9] and [12], has some optimality properties under sparsity. [9] and [1] consider a range of sparsity given by $p_m \propto m^{-\beta}$, with $\beta < 1$. Here we consider more extreme sparsity, $\beta = 1$, and prove the asymptotic optimality of a universal threshold for signals at the verge of detectability. The second of the rules considered is a modification of the universal threshold, which is asymptotically optimal when each of the tests is based on n replicates.

LEMMA 4.1. *Assume that $\delta =$ constant, $m \to \infty$ and $pm \to s$, where $0 < s < \infty$. Then the multiple testing rule (3.8) based on the threshold*

$$c^2 = c_m^2 = 2\log m + d \ , \tag{4.24}$$

where $d \in \mathbf{R}$, is asymptotically optimal for signals on the verge of detectability, $u = \beta_1 \log m(1 + o_m)$, with $\beta_1 \in (0, \infty)$.

LEMMA 4.2. *Let $\sigma_0^2 = 0$. Assume that each test statistic X_i is based on $n = n_m$ replicates and $\sigma^2 = \frac{\sigma_s^2}{n}(1 + o_m)$, where σ_s^2 represents the variance of X_i for one replicate. Moreover, assume that $m \to \infty$, $\tau =$ constant, $\delta =$ constant, $pm \to s \in (0, \infty)$ and $\frac{2\log m}{n} \to s_1 \in [0, \infty)$. The multiple testing rule (3.8) based on the threshold*

$$c^2 = c_{m,n}^2 = \log n + 2\log m + d \ , \tag{4.25}$$

with $d \in \mathbf{R}$, is asymptotically optimal.

The main difference between the rules defined in (4.24) and (4.25) is the different ranges of scaled signal magnitudes u for which they are optimal. The rules proposed in Lemma 4.1 are asymptotically optimal for the smallest detectable signals, which are of the order of $\log m$. On the other hand, rules of the form (4.25) are asymptotically optimal when $u = \left(\frac{\tau}{\sigma}\right)^2$ is proportional to n, which can be of a substantially larger order than $\log m$ (since s_1 can be equal to 0). Note however that such a situation can only occur if $\sigma_0^2 = 0$ (i.e.

when we test $H_{0i}: \ \mu_i = 0$). If $\sigma_0^2 > 0$ then the variance of the test statistic under the null hypothesis, σ^2, does not converge to 0 when $n \to \infty$ and u is bounded from above by $\left(\frac{\tau}{\sigma_0}\right)^2$. Thus the rule (4.25) is not recommended for the detection of outlying signals from the background noise.

5. Controlling the Bayesian False Discovery Rate. In the previous section we described two rules, which are asymptotically optimal when the expected number of signals converges to a finite constant and hence remains bounded as the number of tests increases. However, this assumption is often unrealistic. In many applications the main reason for performing a large number of statistical tests is the belief that increasing the number of tests will enable the detection of a larger number of true signals.

In this context, we refer to a recent paper [1], where it has been shown that the well known Benjamini-Hochberg procedure (BH, [2]), originally proposed in [29] and later in [31], can be used to estimate a sparse vector of means, where the level of sparsity can vary considerably. In [1], independent normal observations $X_i, i = 1, \ldots, m$ with unknown means μ_i and known variance are considered. Among the studied parameter spaces are $l_0[p_m]$ balls, which consist of those real m-vectors for which the fraction of non-zero elements is at most p_m. A data-adaptive thresholding estimator for the unknown vector of means is proposed using the Benjamini-Hochberg rule controlling the FDR at $\alpha_m \geq \frac{\beta_1}{\log m}$ for each $m \geq 1$ and some constant $\beta_1 > 1$. If the FDR control level α_m converges to $\alpha_0 \in [0, 1/2]$, this estimator is shown to be asymptotically minimax for a large class of loss functions (and in fact for many different types of sparsity classes including l_0 balls), as long as p_m is in the range $\left[\frac{\log^5 m}{m}, m^{-\beta_2}\right]$, with $\beta_2 \in (0, 1)$.

Here we want to use the framework presented in Sections 3 and 4 to investigate the asymptotic optimality of BH for a broad range of sparsity levels by studying its Bayes risk with respect to an additive loss function. We reemphasize that minimizing Bayes risk with respect to such a loss function seems to be a natural optimality criterion in the context of testing, where the main goal is to correctly detect signals, rather than estimating their magnitude. However, it is not easy to show the optimality of BH directly, because it is a random thresholding rule (see Section 6). On the other hand, it was proved by Genovese and Wassermann (GW) in [17] that when p remains fixed, as the number of tests increases, this random threshold can be approximated by a non-random one (defined in equation (5.54) below).

When $p \approx 0$, the approximate threshold of [17] is basically the same as that of a fixed threshold rule controlling the Bayesian False Discovery Rate (BFDR, [14], defined below) at the same level. If a BFDR control rule is optimal under sparsity, the same can be expected of the corresponding rule using the threshold of [17]. The optimality of BH in turn may be proved by showing that, even under sparsity, GW thresholds of the form (5.54) are tight estimates of the BH random threshold. In Section 6 we will actually see that this can be done for a broad class of sparsity levels.

In the present section we first recall the definition of the Bayesian False Discovery Rate (BFDR) and then briefly motivate why one might expect that controlling the BFDR leads to an optimal rule. We provide general necessary and sufficient conditions under which fixed threshold rules controlling the BFDR at level α will be asymptotically optimal in terms of the Bayes risk. We then show that under sparsity the same conditions ensure the optimality of a threshold rule using the GW threshold. As a simple consequence of our general results, we finally show in Section 5.5 that, in addition, rules based on the Bonferroni correction are asymptotically optimal in the extremely sparse case.

5.1. *The False Discovery Rate and 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:

$$FDR = E\left(\frac{V}{R}\right) \quad . \tag{5.26}$$

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 [14] define another very similar measure, called the Bayesian False Discovery Rate, BFDR:

(5.27)

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

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

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 [17], [4] and [3] 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

$$(1-\alpha)(1-p)t_1 + \alpha p t_2 = \alpha p \ , \tag{5.28}$$

the l.h.s. of (5.28) 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., [36]). To support this statement, Lemma 8.1 in Appendix 8.2 shows that under the mixture model (2.6), the BFDR of a test based on the threshold c^2 continuously decreases from $(1-p)$ for $c=0$ to 0 for $c \to \infty$. 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 4.1. Naturally, such "optimal" BFDR control levels must be sufficiently similar to the BFDR of the Bayes oracle.

Keeping the above in mind, in the next two sub-sections we explore the relationship between BFDR control rules and the Bayes oracle. Specifically, we calculate the BFDR of the Bayes oracle and specify the conditions under which BFDR control rules are asymptotically optimal.

5.2. *BFDR of the Bayes oracle.* Before going to the mathematical derivations, we first observe some simple facts that indicate the form of dependency of the BFDR of the Bayes oracle on the ratio of losses $\delta = \frac{\delta_0}{\delta_A}$. Suppose the BFDR of the Bayes oracle is α. Using definition (5.27), we can easily see that the corresponding ratio between the type I and type II components of the risk satisfies the following relationship:

$$\frac{\delta_0(1-p)t_1}{\delta_A p t_2} = \delta\left(\frac{\alpha}{1-\alpha}\right)\left(\frac{1-t_2}{t_2}\right). \tag{5.29}$$

We recall that under Assumption (A), the lhs of (5.29) converges to zero. Since t_2 is strictly smaller than 1, this can only happen when $\delta\left(\frac{\alpha}{1-\alpha}\right)$ (and hence $\delta\alpha$) converges to zero. This implies that if δ remains fixed, then α tends to zero. Also, the Bayes oracle can have a constant or a non-zero limiting BFDR, only if δ converges to 0.

Let us define $t_{u,v,\delta} = \delta\sqrt{u \log v}$ and denote the BFDR of the Bayes oracle by $BFDR_{BO}$. Lemma 5.1 and the remark below show the exact dependence of $BFDR_{BO}$ on δ and u.

LEMMA 5.1. *Suppose Assumption (A) holds. If* $t_{u,v,\delta} \to \infty$*, then* $BFDR_{BO}$ *converges to zero at a rate specified by the formula:*

$$BFDR_{BO} = \sqrt{\frac{2}{\pi}} \frac{e^{-C/2}}{Dt_{u,v,\delta}}(1 + o_t) \ , \tag{5.30}$$

where $D = 2(1 - \Phi(\sqrt{C}))$ *is the asymptotic power.*

If $t_{u,v,\delta} \to C_1$*, where* $0 \leq C_1 < \infty$*, then* $BFDR_{BO}$ *converges to a constant and is given by,*

$$BFDR_{BO} = \frac{1}{1 + \sqrt{\frac{\pi}{2}} e^{C/2} DC_1}(1 + o_t) \tag{5.31}$$

PROOF. Note that

$$BFDR = \frac{1}{1 + \frac{1-t_2}{ft_1}} \ , \tag{5.32}$$

where $f = \frac{1-p}{p}$. The lemma follows easily by observing that (3.14) yields

$$ft_1 = \sqrt{\frac{2}{\pi}} \frac{e^{-C/2}}{t_{u,v,\delta}}(1 + o_t), \tag{5.33}$$

while (3.13) and (3.12) give

$$1 - t_2 = 2(1 - \Phi(\sqrt{C})) + o_t \ . \tag{5.34}$$

□

REMARK 5.1. From equations (5.32), (5.33) and (5.34), it is clear that the BFDR of the Bayes oracle is essentially a decreasing function of $t_{u,v,\delta} = (\delta\sqrt{u})^{1+b(v)} f^{b(v)}$, where $b(v) = \frac{\log(\log v)}{\log v}$. This effectively says that the BFDR of the Bayes oracle decreases when both δ and u increase (since $b(v) \approx 0$ for large v, even large variation in the level of sparsity (f) will not typically alter this fact). In particular, Lemma 5.1 shows that under Assumption (A) and for $\delta =$ *constant*, the BFDR of the Bayes oracle converges to zero at the rate $1/\sqrt{u \log v}$. Specifically, in the case when $p \to 0$ and $u = -c \log p$ ($c > 0$) (i.e. on the verge of detectability) the BFDR of the Bayes oracle converges to zero at the rate $(-\log p)^{-1}$. The Bayes oracle has a non-zero limiting BFDR only if the ratio of losses converges to 0 at 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 errors increases to infinity slightly quicker than u (i.e. there is a higher penalty for missed signals when they are sparse and relatively large). Specifically, for signals at the verge of detectability, $u = -c \log p$, the Bayes oracle has a fixed limiting BFDR if δ is of the order $(-\log p)^{-1}$.

5.3. *Asymptotic optimality of BFDR control rules.* In section 5.2 we computed the BFDR of the Bayes oracle. These results give some indication of how the BFDR level α should be chosen to obtain control rules that are asymptotically optimal in terms of the Bayes risk. In this section we give a full characterization of asymptotically optimal BFDR levels. There is some flexibility in the choice of α, although the general behavior, as expected, is closely related to the behavior of the BFDR of the Bayes oracle under similar circumstances. The general Theorem 5.1, below, gives conditions on α, which guarantee optimality for any given sequence of parameters γ_t, satisfying Assumption (A). In Section 5.4 we present some clearer results, which characterize BFDR control rules which are asymptotically optimal on the verge of detectability and in its close neighborhood. Specifically, Corollaries 5.5 and 5.6 give simple conditions on α and δ, which make these parameters only dependent on the number of tests m.

Consider a fixed threshold rule (based on $\frac{X_i^2}{\sigma^2}$) controlling the BFDR at the level α. Under the mixture model (2.6), a corresponding threshold value c_B^2 can be obtained by solving the equation

$$\frac{(1-p)(1-\Phi(c_B))}{(1-p)(1-\Phi(c_B))+p\left(1-\Phi\left(\frac{c_B}{\sqrt{u+1}}\right)\right)}=\alpha \quad , \tag{5.35}$$

or equivalently, by solving

$$\frac{1-\Phi(c_B)}{1-\Phi\left(\frac{c_B}{\sqrt{u+1}}\right)}=\frac{\alpha}{f(1-\alpha)}=\frac{r_\alpha}{f} \quad , \tag{5.36}$$

where

$$r_\alpha=\frac{\alpha}{1-\alpha} \quad . \tag{5.37}$$

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

Using Theorem 4.1, 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 (5.36), this in turn implies that a BFDR control 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 4.1. The following theorem give the asymptotic expansion for c_B^2 and specifies the range of "optimal" choices of r_α.

THEOREM 5.1. *Consider a rule controlling the BFDR at level $\alpha = \alpha_t$. Define s_t by*

$$\frac{\log(f\delta\sqrt{u})}{\log(f/r_\alpha)} = 1 + s_t \ . \tag{5.38}$$

Then the rule is asymptotically optimal if and only if

$$s_t \to 0 \tag{5.39}$$

and

$$2s_t \log(f/r_\alpha) - \log\log(f/r_\alpha) \to -\infty \ . \tag{5.40}$$

The threshold for this rule is of the form

$$c_B^2 = 2\log\left(\frac{f}{r_\alpha}\right) - \log\left(2\log\left(\frac{f}{r_\alpha}\right)\right) + C_1 + o_t \ , \tag{5.41}$$

where $C_1 = \log\left(\frac{2}{\pi D^2}\right)$ *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 5.1 can be found in Appendix 8.3.

REMARK 5.2. In comparison to (5.39), condition (5.40) imposes an additional restriction on positive values of s_t (i.e large values of α). It is clear from the proof of Theorem 5.1 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 Section 4.

REMARK 5.3. Condition (5.39), given above, says (after some algebra) that a sequence of "optimal" BFDR levels $\alpha = \alpha_t$ satisfies $\alpha = (1 + (\delta\sqrt{u})^{1+b_t} f^{b_t})^{-1}$ (i.e $r_\alpha = (\delta\sqrt{u})^{-1-b_t} f^{-b_t}$) for some b_t, where $b_t \to 0$ as $t \to \infty$. Thus asymptotically, the optimal BFDR levels will generally be smaller as δ and u get larger (or one increases while the other is fixed). Since b_t is small, again variation in f will typically have a minimal effect. Thus the general behavior of optimal BFDR levels is similar to what we observed for the BFDR of the Bayes oracle. Below we present several corollaries of Theorem 5.1, which provide additional, more explicit conditions for the optimality of BFDR control rules, each of which corroborates this broad finding.

COROLLARY 5.1. *A rule controlling the BFDR at the level* $\alpha = \alpha_t$, *such that* $r_\alpha = \frac{s}{\delta\sqrt{u}}(1 + o_t)$, *with* $s \in (0, \infty)$, *is asymptotically optimal.*

The proof of Corollary 5.1 is immediate by verifying that (5.39) and (5.40) are satisfied by such a sequence of α's.

As a special case we discuss the situation described in Lemma 4.2, where each of the test statistics X_i is based on $n = n_t$ replicates and the focus is on testing the simple null hypothesis $\mu_i = 0$. In this case any BFDR level $\alpha = \alpha_t \propto \frac{1}{\sqrt{n}}$ is asymptotically optimal:

COROLLARY 5.2. *Assume that* $\sigma_\epsilon^2 = \frac{c_\sigma}{n}$, *with* $c_\sigma \in (0, \infty)$. *Moreover, assume that* $p \to 0$, $n \to \infty$, $-\frac{\log(p)}{n} \to s \in [0, \infty)$, $\sigma_0^2 = 0$, $\delta = const$ *and* $\tau = const$. *Then a rule controlling the BFDR at the level* $\alpha_n = \frac{s_1}{\sqrt{n}}$, $s_1 \in (0, \infty)$, *is asymptotically optimal.*

PROOF. This is a direct consequence of Corollary 5.1. □

The following two corollaries shed some more light on asymptotic optimality of rules controlling the BFDR at a fixed level $\alpha \in (0, 1)$ or when δ remains fixed.

COROLLARY 5.3. *A rule controlling the BFDR at a* **fixed** *level* $\alpha \in (0, 1)$ *is asymptotically optimal if and only if the ratio of loss functions converges to 0 at such a rate that*

$$\frac{\log(\delta\sqrt{u})}{\log p} \to 0 \ . \tag{5.42}$$

and

$$\frac{\delta^2 u}{\log p} \to 0 \tag{5.43}$$

PROOF. Note that the term s_t defined in Theorem 5.1 is given by

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

We first show necessity. By Theorem 5.1, optimality holds only if (5.39) is fulfilled, i.e. $\frac{\log(\delta r_\alpha\sqrt{u})}{\log(f/r_\alpha)} \to 0$. Under Assumption (A), this can happen only if $p \to 0$, since α is a constant. When $p \to 0$, condition (5.39) reduces to

(5.42). To complete the proof of necessity, observe that condition (5.40) from Theorem 5.1 implies that for fixed α

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

which yields (5.43) when $p \to 0$.
To prove sufficiency, first observe that Assumption (A) and (5.43) together imply that $p \to 0$, since $v \to \infty$ and (5.43) is equivalent to saying that $\frac{v}{f^2 \log p}$ goes to zero. Hence for fixed $\alpha \in (0,1)$, (5.42) and (5.43) imply that s_t satisfies properties (5.39) and (5.40), respectively. □

REMARK 5.4. Under Assumption (A), conditions (5.42) and (5.43) can occur together only if $\delta \to 0$. To this end, observe that Assumption (A) and (5.43) together imply that $p \to 0$. When $p \to 0$, (5.42) is equivalent to saying that $\frac{\log v}{-2\log p} \to 1$. Under Assumption (A), this implies that $\frac{u}{-2\log p}$ converges to the required $C \in (0,\infty]$. This and (5.43) together imply that $\delta \to 0$.

COROLLARY 5.4. *Suppose δ is fixed. Then a rule controlling the BFDR at level $\alpha = \alpha_t$ is asymptotically optimal if and only if α converges to 0 at such a rate that*

$$\frac{\log(\alpha\sqrt{u})}{\log(f/\alpha)} \to 0 \ . \tag{5.44}$$

and

$$\frac{\alpha^2 u}{\log(f/\alpha)} \to 0 \tag{5.45}$$

PROOF. The proof of this result is very similar to the proof of Corollary 5.3 and is therefore omitted. □

Corollary 5.3, given above, states that under Assumption (A), a rule controlling the BFDR at a fixed level α can be optimal only if $p \to 0$ and, due to Remark 5.4, the relative cost of type II errors increases. In particular, this implies that such a rule will not be asymptotically optimal in the problem of minimizing the overall misclassification rate (since in this case $\delta_0 = \delta_A$). This result provides important insight and brings new aspects of BFDR control procedures under sparsity into light in the context of multiple testing.

5.4. *Optimal BFDR control on the verge of detectability.* In this section we present several results describing the behavior of BFDR control rules on the verge of detectability and in its neighborhood. Optimality on the verge of detectability is particularly important, since it guarantees asymptotically optimal performance in a very difficult scenario where signals are so small that they are barely detectable. Hence, procedures which are optimal on the verge of detectability are expected to be robust and give good overall performance when no prior information about the magnitudes of signals is available.

The first of our results below states that for signals on the verge of detectability, a rule controlling the BFDR at level $\alpha \in (0,1)$ is asymptotically optimal if and only if the ratio between loss functions δ decreases to 0 at a relatively slow rate. The second lemma, dual to the first one, states that if $\delta = constant$, then a BFDR control rule is asymptotically optimal for signals on the verge of detectability if and only if the BFDR level α decreases to zero at the same, very slow rate. This last result explains the good performance of BFDR control rules observed with respect to controlling the misclassification error for small α's, reported in [3] and [4].

LEMMA 5.2. *Suppose Assumption (A) holds with $C \in (0,\infty)$ and $p \to 0$. A rule controlling the BFDR at a fixed level $\alpha \in (0,1)$ is asymptotically optimal if and only if $\delta \to 0$ at such a rate that $\frac{\log \delta}{\log p} \to 0$.*

PROOF. This is a direct consequence of Corollary 5.3. □

LEMMA 5.3. *Assume that $p \to 0$, $\delta = constant$, and $-\frac{2\log p}{u} \to C$, with $0 < C < \infty$. A rule controlling the BFDR at level $\alpha = \alpha_t \in (0,1)$ is asymptotically optimal if and only if $\alpha \to 0$ at such a rate that $\frac{\log \alpha}{\log p} \to 0$.*

PROOF. This is a direct consequence of Corollary 5.4. □

The conditions specified in Lemmas 5.2 and 5.3 make δ and α dependent on the unknown sparsity parameter p. To make these results more applicable, we now consider the situation in which the number of tests m goes to infinity and p_m is such that

$$-\frac{\log p_m}{\log m} \to K, \quad \text{for some constant} \quad K \in (0,\infty)\ . \tag{5.46}$$

Note that this large set includes all decreasing sequences p_m such that $m^{-c_1} \le p_m \le m^{-c_2}$, where c_1 and c_2 are any constants satisfying $0 < c_2 < c_1 < \infty$.

Corollary 5.5. *Consider the whole class of sparsity sequences p_m satisfying (5.46). A rule controlling the BFDR at a fixed level $\alpha \in (0,1)$ is asymptotically optimal for signals on the verge of detectability if and only if*

$$\delta_m \to 0 \qquad \textit{and} \qquad \frac{\log \delta_m}{\log m} \to 0 \ . \tag{5.47}$$

Remark 5.5. It is easy to check that under (5.46) and (5.47) signals on the verge of detectability are of the form

$$u_m = \beta \log m(1 + o_m) \ , \quad \text{with} \quad \beta \in (0, \infty) \ . \tag{5.48}$$

Corollary 5.6. *Consider the class of sparsity sequences p_m satisfying (5.46). Assume that the ratio between losses δ is fixed. Then a BFDR control rule is asymptotically optimal for signals on the verge of detectability (5.48) if and only if $\alpha = \alpha_m$ satisfies*

$$\alpha_m \to 0 \qquad \textit{and} \qquad \frac{\log \alpha_m}{\log m} \to 0 \ . \tag{5.49}$$

Conditions (5.47) and (5.49) allow some freedom in the choice of δ and α. Interestingly, some of these choices guarantee asymptotic optimality for signals which are substantially larger than those on the verge of detectability. Corollaries 5.7 and 5.8, given below, specify the range of magnitudes of signals for which such rules are asymptotically optimal.

Corollary 5.7. *Suppose the number of tests $m \to \infty$. Consider the class of sparsity sequences p_m satisfying (5.46). Suppose $\delta_m \to 0$, such that $\log \delta_m = o(\log m)$. A rule controlling the BFDR at a fixed level $\alpha \in (0,1)$ is asymptotically optimal if and only if the sequence of the magnitudes of signals u_m satisfies*

$$\frac{u_m}{\log m} \to C_u \in (0, \infty) \quad \textit{(verge of detectability)} \tag{5.50}$$

or

$$\frac{u_m}{\log m} \to \infty \ \ \textit{and} \ \ u_m = o\left(\frac{\log m}{\delta_m^2}\right) \ . \tag{5.51}$$

Proof. Corollary 5.5 follows directly from Corollary 5.3. □

Corollary 5.8. *Suppose the number of tests $m \to \infty$. Consider the class of sparsity sequences p_m satisfying (5.46). Suppose that the ratio between loss functions δ is fixed. A rule controlling the BFDR at the level*

$\alpha_m \to 0$, such that $\log \alpha_m = o(\log m)$, is asymptotically optimal if and only if the sequence of magnitudes of signals u_m satisfies

$$\frac{u_m}{\log m} \to C_u \in (0, \infty) \quad \textit{(verge of detectability)} \tag{5.52}$$

or

$$\frac{u_m}{\log m} \to \infty \quad \textit{and} \quad u_m = o\left(\frac{\log m}{\alpha_m^2}\right) \quad . \tag{5.53}$$

PROOF. Corollary 5.6 is a direct consequence of Corollary 5.4. □

5.5. *Optimality of the asymptotic approximation to the BH threshold.* In [17] 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

$$c_{GW} : \frac{(1-\Phi(c_{GW}))}{(1-p)(1-\Phi(c_{GW})) + p\left(1-\Phi\left(\frac{c_{GW}}{\sqrt{u+1}}\right)\right)} = \alpha \quad . \tag{5.54}$$

Compared to the equation defining the BFDR control rule (5.35), the function on the left-hand-side of (5.54) 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 control rule for the same α. The following theoretical result shows that this is indeed the case.

THEOREM 5.2. *Suppose $p \to 0$. Consider the rule rejecting the null hypothesis H_{0i} if $\frac{X_i^2}{\sigma^2} \geq c_{GW}^2$, where c_{GW} is defined in (5.54). This rule is asymptotically optimal if and only if the corresponding BFDR control rule defined in (5.35) is asymptotically optimal. In this case we have*

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

where c_B^2 is the threshold of an asymptotically optimal BFDR control rule, defined in Theorem 5.1.

PROOF. Note that (5.54) is equivalent to

$$\frac{1-\Phi(c_{GW})}{1-\Phi\left(\frac{c_{GW}}{\sqrt{u+1}}\right)} = \frac{p r_\alpha}{1+p r_\alpha} = \frac{r_{\alpha'}}{f}, \tag{5.55}$$

where $\alpha' = \alpha(1-p)$. Thus c_{GW} is the same as the threshold of a rule controlling the 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 (5.39) and (5.40) of Theorem 5.1 (with α replaced by α'), if and only if s_t defined in (5.38) satisfies (5.39) and (5.40). Thus the first part of the theorem is proved.

To complete the proof of the theorem, we observe that the optimality of a BFDR control 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 (5.55) reduces to

$$\frac{1-\Phi(c_{GW})}{1-\Phi\left(\frac{c_{GW}}{\sqrt{u+1}}\right)} = pr_\alpha(1+o_t) = \frac{r_\alpha}{f}(1+o_t) \quad . \tag{5.56}$$

Now, the asymptotic approximation to c^2_{GW} can be obtained analogously to the asymptotic form of the threshold for an optimal BFDR control rule, provided in (5.41). □

5.6. *Optimality 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

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

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

$$c^2_{Bon} = 2\log\left(\frac{m}{\alpha}\right) - \log\left(2\log\left(\frac{m}{\alpha}\right)\right) + \log(2/\pi) + o_m \quad . \tag{5.57}$$

Comparison of this threshold with the asymptotic approximation to an optimal BFDR control rule (5.41) suggests that the Bonferroni correction will have similar asymptotic optimality properties in the "extremely" sparse case $p_m \propto \frac{1}{m}$ (see also the comparison with the "universal threshold" discussed in Section 4.1). Indeed, these expectations are confirmed by the following lemma, which actually specifies a slightly larger set of sequences of sparsity parameters under which the Bonferroni correction is asymptotically optimal. Lemma 5.4 will be used in the next section for the proof of the optimality of the Benjamini-Hochberg procedure under very sparse signals.

LEMMA 5.4. *Assume that* $m \to \infty$ *and* $p_m = \frac{z_m}{m}$, *where* z_m *converges to a finite positive constant or diverges to infinity at such a rate that*

$$\frac{\log z_m}{\log m} \to 0 \ . \tag{5.58}$$

The Bonferroni procedure at FWER level $\alpha_m \to \alpha_\infty \in [0, 1)$ *is asymptotically optimal if* α_m *satisfies the assumptions of Theorem 5.1.*

PROOF. Observe that under the assumptions of Lemma 5.4 and Theorem 5.1

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

where $D = 2(1 - \Phi(\sqrt{C}))$ and c_B^2 is the threshold of the rule controlling the BFDR at level α_m. From (5.58) it follows easily that $c_{Bon}^2 = c_B^2(1 + o_m)$. By assumption, the rule based on the threshold c_B^2 is optimal, and hence c_{Bon}^2 satisfies condition (4.22) of Theorem (4.1). Condition (4.23) is satisfied, since by assumption $\log z_m$ is bounded below for sufficiently large m and thus the optimality of the Bonferroni correction follows. □

6. Optimality of the Benjamini-Hochberg procedure. In this section we report results on the asymptotic optimality of the Benjamini-Hochberg procedure (BH). We consider a sequence of problems in which the number of tests $m \to \infty$ and the γ sequence is indexed by $t = m$. In Section 6.1 we present BH and the formula for its random threshold c_{BH}. In Section 6.2, we prove the asymptotic optimality properties of BH under a wide range of sparsity parameters $p_m \to 0$, such that $mp_m \to s \in (0, \infty]$. The proof of the optimality of the type I error component of the risk is based on the precise results of [16] on the expected number of type I errors under the total null hypothesis and holds over the whole range of the sparsity parameters considered. The proof of the optimality of the type II error component is broken into two parts. In the extremely sparse case, described by Lemma 5.4, the optimality of BH follows from a comparison with the asymptotically optimal Bonferroni correction. For the remaining range of sparsity parameters, the proof follows from the approximation of the random threshold of BH by the asymptotically optimal threshold c_{GW} (see (5.54)), given by [17]. Theorem 6.3, given below, extends the results of [17] to our sparse asymptotic scenario and illustrates the accuracy of this approximation.

Our results establish that, under the considered range of sparsity parameters, BH is asymptotically optimal if the chosen FDR control level α_m depends on u_m and δ_m in the same way as specified in Theorem 5.1. Specifically, we show that BH at any fixed FDR level will be optimal for a wide class of sparsity levels and magnitudes of signals, as long as the loss ratio goes to zero slowly. A similar result is proved for the case of a fixed loss ratio, as long as the FDR control level goes to zero slowly. In Section 6.2.3 we give very transparent results describing the optimality properties of BH for signals on the verge of detectability. Sections 6.2.4 and 6.2.5 contain a comparison between the thresholds of the asymptotically optimal BH rule and the Bayes oracle and a summary of results on the expected numbers of true and false rejections by rules which are optimal on the verge of detectability.

6.1. *Random threshold of the Benjamini-Hochberg procedure.* Let $Z_i^2 = \frac{X_i^2}{\sigma^2}$ and $p_i = 2(1 - \Phi(Z_i))$ be the corresponding p-value. We sort p-values in ascending order $p_{(1)} \leq p_{(2)} \leq \ldots \leq p_{(m)}$ and denote

$$k = argmax_i \left\{ p_{(i)} \leq \frac{i\alpha}{m} \right\} . \tag{6.59}$$

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

Let us denote $1 - \hat{F}_m(y) = \#\{|Z_i| \geq y\}/m$. It is easy to check (eg. see [14]) that the Benjamini-Hochberg procedure rejects the null hypothesis H_{0i} when $Z_i^2 \geq \tilde{c}_{BH}^2$, where

$$\tilde{c}_{BH} = \inf \left\{ y : \frac{2(1 - \Phi(y))}{1 - \hat{F}_m(y)} \leq \alpha \right\} . \tag{6.60}$$

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

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

Comparing (6.60) and (5.54), we observe that the difference between $\tilde{c}_{BH}$ and c_{GW} is in replacing the cumulative distribution function of $|Z_i|$ (appearing in (5.54)) by the empirical distribution function (in 6.60). The proof of the optimality of BH, presented in the next section, is partially based on the investigation of the accuracy of this approximation.

6.2. *Optimality of BH.* To prove the optimality of the BH rule, we distinguish two cases. The first, the extremely sparse case, is characterized by

$$mp_m \to s \in (0, \infty] \quad \text{and} \quad \frac{\log(mp_m)}{\log m} \to 0 . \tag{6.61}$$

Specifically, condition (6.61) is satisfied by very sparse signals with $p_m \propto \frac{1}{m}$. The second, "denser" case is characterized by

$$p_m \to 0 \quad \text{and} \quad \frac{\log(mp_m)}{\log m} \to C_p \in (0,1] \ . \tag{6.62}$$

The theorem on the optimality of BH presented in this section requires the following assumptions.

1. **Number of tests m and the sparsity $\mathbf{p_m}$:**

$$m \to \infty, \quad p_m \to 0, \quad mp_m \to s \in (0, \infty] \tag{6.63}$$

2. **FDR level $\alpha_\mathbf{m}$:**

$$\alpha_m \to \alpha_\infty < 1, \text{ and} \tag{6.64}$$

$$\alpha_m \text{ satisfies the conditions of Theorem (5.1),} \tag{6.65}$$

   i.e a rule controlling the BFDR at level α_m is asymptotically optimal

3. **Additional assumption for the denser case:**
   If (6.62) holds, then assume

$$u_m \le m^\beta, \quad \text{for some } \beta > 0 \ . \tag{6.66}$$

THEOREM 6.1. *Under Assumptions (6.63) - (6.65), and the additional assumption (6.66) in the case where (6.62) holds, BH is asymptotically optimal.*

The proof of Theorem 6.1 consists of two parts. The first part, concerned with the optimality of the type I error component of the risk, is based on the precise and powerful results of [16] on the expected number of false discoveries using BH under the total null hypothesis. This part of the proof does not require distinguishing between the extremely sparse and the denser case.

6.2.1. *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 (5.26)). The proof is however nontrivial, due to the difference between $E\left(\frac{V}{R}\right)$ and $\frac{EV}{ER}$.

LEMMA 6.1. *Consider the BH rule at a fixed FDR level* $\alpha \leq \alpha_0 < 1$. *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*

$$E(V|K = k) \leq \alpha \left(\frac{k}{1-\alpha} + \frac{1}{(1-\alpha)^2} \right) . \tag{6.67}$$

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

$$E(V|K = k) \leq c_v \alpha k \tag{6.68}$$

with

$$c_v = \frac{2 - \alpha_0}{(1-\alpha_0)^2} . \tag{6.69}$$

The proof of Lemma 6.1 is given in Appendix 8.4.

REMARK 6.1. 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 6.1 to the mixture model (2.6).

LEMMA 6.2. *Under assumptions (6.63)-(6.65), the expected number of false rejections is bounded by*

$$E(V) < C_1 \alpha_m m p_m ,$$

where C_1 *is any constant satisfying*

$$C_1 > \begin{cases} \frac{2-\alpha_\infty}{(1-\alpha_\infty)^2}, & \text{when } s = \infty \\ \frac{e^{-s}}{s(1-\alpha_\infty)} + \frac{2-\alpha_\infty}{(1-\alpha_\infty)^2} & \text{when } s \in (0, \infty) \end{cases} .$$

The proof of Lemma 6.2 is provided in Appendix 8.4.

Lemma 6.2 easily leads to the following Theorem 6.2 on the optimality of the type I error component of the risk of BH.

THEOREM 6.2. *Under assumptions (6.63)-(6.65), the type I error component of the risk of BH,* R_0, *satisfies* $\frac{R_0}{R_{opt}} \to 0$, *where* R_{opt} *is the optimal risk defined in Theorem 3.1.*

PROOF. From Lemma 6.2

$$\frac{R_0}{R_{opt}} = \frac{\delta_0 E(V)}{R_{opt}} \leq C_1 \frac{\alpha_m \delta_m \sqrt{u_m}}{\sqrt{\log v_m}} .$$

Theorem 6.2 now easily follows by invoking (5.39) and (5.40) (included in assumption (6.65)). □

6.2.2. *Bound on the type II component of the risk.* To prove the optimality of the type II component of the risk of BH, we consider the extremely sparse case (6.61) and the denser case 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 5.4 is asymptotically optimal 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 (5.54)), given in Theorem 6.3 below. The corresponding "denser" case assumption (6.71) is substantially less restrictive than (6.62) and partially covers the extremely sparse case (6.61). Theorem 6.3 extends the results of [17] to the case where $p_m \to 0$ and illustrates the precision of this approximation.

THEOREM 6.3. *Assume that*

$$p_m \to 0 \ , \tag{6.70}$$

such that for sufficiently large m

$$p_m > \frac{log^{\beta_p} m}{m}, \quad \textit{for some constant} \quad \beta_p > 1 \ . \tag{6.71}$$

Moreover, assume that the sequence of FDR levels α_m satisfies the assumptions of Theorem (5.1). Then for every $\epsilon > 0$, every constant $\beta_u > 0$ and sufficiently large m (dependent on ϵ and β_u)

$$P(|c_{BH} - c_{GW}| > \epsilon) \leq m^{-\beta_u} \ ,$$

where c_{GW} is the asymptotically optimal threshold defined in (5.54).

The proof of Theorem 6.3 is given in the Appendix.

Using Theorem 6.3, we can easily show the asymptotic optimality of the type II component of the risk of BH.

LEMMA 6.3. *Suppose (6.70) and (6.71) hold and assumptions (6.64)-(6.66) are also true. The type II error component of the risk of BH satisfies*

$$R_A \leq R_{opt}(1 + o_m) \ . \tag{6.72}$$

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

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

and furthermore

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

where L_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 (4.1), it is asymptotically optimal. Hence, it follows that $\delta_A EL_1 = R_{opt}(1+o_m)$. On the other hand, from Theorem 6.3, for any $\beta_u > 0$ and sufficiently large m (dependent on ϵ and β_u)

$$P(c_{BH} > \tilde{c}_1) \leq m^{-\beta_u} \ .$$

Therefore,

$$R_A = \delta_A EL_A \leq R_{opt}(1+o_m) + \delta_A m^{1-\beta_u} \ .$$

Now, using assumptions (6.71) and (6.66), and choosing e. g. $\beta_u = \beta + 1$, we conclude that $\delta_A m^{1-\beta_u} = o(R_{opt})$ and the proof is thus complete. □

REMARK 6.2. According to Theorem 6.1, the BH rule is asymptotically optimal under the scenarios described in Corollaries 5.1-5.4 if the assumptions (6.63), (6.64) and (6.66) are satisfied.

6.2.3. *Optimality on the verge of detectability.* Theorem 6.1 states that under the sparsity assumption (6.63), BH behaves similarly to a BFDR control rule. Specifically, this result shows that the optimal FDR level α should depend on the expected magnitude of a signal u and the ratio between losses δ according to the formula $\alpha \approx \frac{1}{\delta\sqrt{u}}$. In this section we present some more specific results, which describe the behavior of BH in the important case of signals on the verge of detectability.

COROLLARY 6.1. *Suppose that Assumption (6.63) holds. Moreover, assume that*

$$-\frac{2\log p_m}{u_m} \to C \ , \tag{6.73}$$

where $C \in (0,\infty)$. The BH rule controlling at a fixed level $\alpha \in (0,1)$ is asymptotically optimal if δ_m converges to zero such that $\log\delta_m = o(\log p_m)$. This last condition also guarantees that signals satisfying (6.73) are on the verge of detectability. For a fixed δ, the BH rule controlling the FDR at level α_m is asymptotically optimal if α_m converges to zero such that $\log\alpha_m = o(\log p_m)$.

PROOF. The proof follows easily using Lemmas 5.2 (for the fixed α case) and 5.3 (for the fixed δ case) and checking that, under the given conditions, assumptions (6.64)-(6.66) are also satisfied. □

The following corollaries are analogous to Corollaries 5.5-5.6 for BFDR control rules and hold when

$$a_1 m^{-1} \le p_m \le a_2 m^{-a_3} \quad , \tag{6.74}$$

for some constants $a_1 \in (0,\infty), a_2 \in (0,\infty)$ and $a_3 \in (0,1)$.

COROLLARY 6.2. *Consider the class of sparsity sequences p_m satisfying (6.74). The BH rule controlling the FDR at a fixed level $\alpha \in (0,1)$ is asymptotically optimal for signals on the verge of detectability if and only if*

$$\delta_m \to 0 \qquad \textit{and} \qquad \frac{\log \delta_m}{\log m} \to 0 \ . \tag{6.75}$$

REMARK 6.3. It is easy to check that under (6.74) and (6.75), signals on the verge of detectability are of the form

$$u_m = \beta \log m(1+o_m) \ , \quad \text{with} \quad \beta \in (0,\infty) \ . \tag{6.76}$$

COROLLARY 6.3. *Consider the class of sparsity sequences p_m satisfying (6.74). Assume that the ratio between losses δ is fixed. The BH rule is asymptotically optimal for signals on the verge of detectability $u_m = \beta \log m(1+o_m)$ if and only if $\alpha = \alpha_m$ satisfies*

$$\alpha_m \to 0 \qquad \textit{and} \qquad \frac{\log \alpha_m}{\log m} \to 0 \ . \tag{6.77}$$

6.2.4. *A comparison of the threshold for the optimal BH rule and the Bayes Oracle.* Here we briefly discuss the relationship between the BH threshold and the Oracle threshold for signals on the verge of detectability.

Suppose $p \to 0$ and $p > \frac{\log^\beta m}{m}$, where $\beta > 1$. Moreover, assume that the ratio between losses $\delta = Constant$ and that the signals are on the verge of detectability $u_m \propto -\log p$. The BH rule is asymptotically optimal, provided the FDR control level α converges to zero such that $\frac{\log \alpha}{\log p} \to 0$. Moreover, under these conditions the BH threshold can be sandwiched between two optimal GW thresholds at neighboring FDR levels. It follows from the proof of Theorem 6.3 that for such an optimal α sequence, for any $\beta_u > 0$, with probability greater than $(1 - m^{-\beta_u})$ (for all sufficiently large m), the BH threshold is given by

$$c_{BH}^2 = 2\log(1/p) - 2\log\alpha - \log(2\log(1/p)) + \text{Constant} + o_m. \tag{6.78}$$

On the other hand, the Bayes Oracle threshold is given by

$$c_{BO}^2 = 2\log(1/p) + \log(2\log(1/p)) + \text{Constant} + o_m. \tag{6.79}$$

So $c_{BO}^2 - c_{BH}^2 = 2\log(\log(1/p)) + 2\log\alpha + \text{Constant} + o_m$ and it is clear that whether the threshold of the Bayes Oracle is larger or smaller than that of BH depends on the rate of convergence of α to zero.

Now observe that under the above assumptions $-\frac{\log p}{\log m} \to (0,1]$ and the above difference becomes $2\log(\log m) + 2\log\alpha + \text{Constant} + o_m$. In this case, for example, $c_{BO}^2 - c_{BH}^2$ goes to ∞ if $\alpha \to 0$ such that $\alpha \geq \frac{\log(\log m)}{\log m}$, while it goes to $-\infty$ if, for example, $\alpha \to 0$ in such a way that $\alpha \leq \frac{1}{(\log m)\log(\log m)}$ (as long as $\log\alpha = o(\log m)$).

6.2.5. *Expected numbers of rejections for rules which are optimal on the verge of detectability.* Assume that p_m satisfies assumption (6.63) and $\delta = Const$. Now consider a multiple testing rule which is asymptotically optimal on the verge of detectability $u_m \propto -\log p_m$. Note that the threshold of such a rule is of the order of $-2\log p_m$ and is proportional to the magnitude of signals on the verge of detectability. Thus, according to Lemma 3.1, the expected number of true rejections is proportional to the expected number of true signals $mp_m = z_m$ (with proportionality coefficient $D = 2(1-\Phi(\sqrt{C})) \in (0,1)$). On the other hand, $-2\log p_m$ is of the order of the expected value of the z_m^{th} largest statistic under the total null hypothesis. Thus, one might expect that the expected number of falsely rejected null hypotheses is also approximately of the order of z_m. However, it turns out that the second term in the asymptotic expansions for the asymptotically optimal rules (6.78) and (6.79) has a substantial influence on the probability of a type I error and, according to formula (3.14), the expected number of false rejections is of a slightly smaller order, of order $\frac{z_m}{\log 1/p_m}$. Thus, in the case where $z_m \to \infty$, the expected number of false rejections may converge to infinity, but the corresponding false discovery rate still converges to zero.

Similarly, it is easy to check that for a BH rule with a fixed FDR $\alpha \in (0,1)$ and signals on the verge of detectability, the expected numbers of true and false discoveries are both proportional to z_m. Recall that such a rule is asymptotically optimal if $\delta_m \to 0$ and $\log(\delta_m) = o(\log p_m)$.

7. 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 [11] and the Bonferroni correction. Moreover, as in [1], we have proved some asymptotic optimality properties of rules controlling the false discovery rate. Comparing with [1], we replaced a loss function based on the error in estimation with a loss function dependent only on the type of testing error. This resulted in somewhat different optimality properties of BH. Specifically, we have proved that BH controlling the FDR at a fixed level α can be asymptotically optimal only if the relative cost of type II errors increases when $p \to 0$. However, this assumption does not undermine the desirable properties of BH controlling at a fixed FDR level, since it is quite natural to impose a large loss for missed signals in the case where p (equivalently, the total number of signals) is very small. Our results also provide some hints on how the "optimal" FDR level should be chosen, depending on the expected magnitude of true signals and the ratio between the loss for type I and type II errors.

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 e.g. [34], [13], [3], [4]. A further discussion on the multiplicity adjustment, inherent to the appropriately designed Bayesian methods of model selection, can be found in [28]. In [4] it is shown that both fully Bayesian and empirical Bayes methods substantially outperform the Benjamini-Hochberg procedure for *moderately small* values of p. 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 [6] illustrate that consistent estimation of the mixture parameters is possible when $p_m \propto m^{-\beta}$, with $\beta < 1$. New results on the convergence rates of these estimates, presented in [20], raise some hopes that proofs of the optimality properties of the corresponding empirical Bayes rules can be found. It is, however, not clear whether 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 [4], where Bayesian methods are outperformed by BH and the Bonferroni correction for very small p.

The Benjamini-Hochberg procedure can only be directly applied when the distribution under the null hypothesis is completely specified, i.e. 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 using, for example, empirical Bayes methods (see e.g. [4]). While the simulation results reported in [4] show that for very small p BH can outperform Bayesian approximations to the oracle even in this context, it is rather unlikely that such a plug-in version of BH is 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 topic 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. [13]). Such two-group models assume a sharp distinction between the mechanisms generating the null and alternative hypotheses. In a recent article [7], 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 [7], the resulting Bayesian estimate of the vector of means shrinks small effects strongly towards zero and leaves large effects almost intact. In this way, it enables very good separation of large signals from background noise. In [7] 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 [7] 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 [7] clearly demonstrates 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.

Acknowledgement. We would like to express our gratitude to David Ramsey for a careful reading of the manuscript and helpful suggestions.

Much of the work on the manuscript was done during the visits of the first author to the Bayesian and Interdisciplinary Unit of the Indian Statistical Institute in Kolkata and to the Statistical Bioinformatics Center in the Department of Statistics at Purdue University. The hospitality of both departments is gratefully appreciated. This work is partially funded by grant 1 P03A 01430 of the Polish Ministry of Science and Higher Education to M. Bogdan.

8. Appendix.

8.1. *Proof of Theorem 4.1.* We first prove the sufficiency of (4.22) and (4.23) for the optimality of the multiple testing rule.

The condition $z_t = o(\log v)$ implies that $t_2 = A\sqrt{\frac{\log v}{u}}(1+o_t)$ and the constant A is equal to $\sqrt{\frac{2}{\pi}}$ or $\frac{2\Phi(\sqrt{C})-1}{\sqrt{C}}$ according to whether C is zero or strictly positive. Note that (3.17) and the fact that $z_t = o(\log v)$ together imply that the probability of a type I error is given by

$$t_1 = P(|Z| > c_t) = \sqrt{\frac{2}{\pi}}\frac{\exp(-z_t/2)}{\sqrt{v}\log v}(1+o_t). \tag{8.80}$$

Now, assume that the constant C specified in Assumption (A) is equal to 0. Excluding the multiplier m, the type I error component of the risk (see (3.19)) is equal to $R_1 = (1-p)t_1\delta_0$, while the component corresponding to type II errors is $R_2 = pt_2\delta_A$. The ratio between R_1 and R_2 becomes

$$\frac{R_1}{R_2} = \frac{\delta f\sqrt{u}\exp(-z_t/2)}{\sqrt{v}\log v}(1+o_t).$$

By the definition of v this is equal to

$$\frac{R_1}{R_2} = \exp(-z_t/2 - \log\log v)(1+o_t), \tag{8.81}$$

which converges to zero if $z_t + 2\log\log v \to \infty$. This shows that the overall risk is given by $R = mR_2(1+o_t)$, which is equivalent to the expression in (3.20).

In the case where $C > 0$, analogous steps give the required result. The only difference is that the ratio $\frac{R_1}{R_2}$ is a different multiple of the expression in (8.81). This completes the proof of the sufficiency part.

We will now prove that under Assumption (A), both conditions (4.22) and (4.23) are necessary for optimality to hold. First we prove the necessity of condition (4.22)

Assume that (4.22) does not hold. Noting that $z_t \geq -\log v$ (since $c_t^2 \geq 0$), clearly this can happen if either (i) $\frac{z_t}{\log v}$ converges to a point in $[-1, \infty] - \{0\}$ or (ii) $\frac{z_t}{\log v}$ does not converge anywhere.

First we consider case (i). This case leads to three distinct possibilities, dealt with separately in the following:

- $\frac{z_t}{\log v} \to -1$ and c_t does not diverge to infinity. In this case, there exists a constant C_1, $0 \leq C_1 < \infty$, and a subsequence of critical values $c_{\tilde{t}}$, such that $c_{\tilde{t}} \to C_1$. Observe that in this situation the probability of a type I error is given by

  $$t_1 = P(|Z| > c_{\tilde{t}}) \to 2(1 - \Phi(C_1)) = C_2 > 0 \ .$$

  Thus, the type I error component of the risk satisfies

  $$R_1 = m\delta_0(1-p)C_2(1 + o_{\tilde{t}}) = \frac{C_2 m \delta_A p \sqrt{v}}{\sqrt{u}}(1 + o_{\tilde{t}}) \ . \quad (8.82)$$

  As a result, in the case where $C = 0$, the ratio of the type I error component of the risk to the optimal risk given in (3.20) is given by

  $$\frac{R_1}{R_{opt}} = C_2 \sqrt{\frac{\pi v}{2 \log v}}(1 + o_{\tilde{t}}) \to \infty$$

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

  $$\frac{R_1}{R_{opt}} = \frac{C_2}{2\Phi(\sqrt{C}) - 1}\sqrt{\frac{Cv}{\log v}}(1 + o_{\tilde{t}}) \to \infty \ .$$

- Now consider the case: $z_t = s \log v(1 + o_t)$, $s \in [-1, 0)$ and $c_t \to \infty$. It is easy to see that in this case

  $$t_1 = \sqrt{\frac{2}{\pi}} \frac{\exp(-z_t/2)}{\sqrt{v}\sqrt{\log v + z_t}}(1 + o_t).$$

  Note that

  $$\frac{R}{R_{opt}} \geq \frac{R_1}{R_{opt}} = A \frac{\exp(-z_t/2 - \log(\log v))}{\sqrt{1 + \frac{z_t}{\log v}}}(1 + o_t),$$

where the constant A depends on the value of C. Observing that in this case $z_t + 2\log\log v \to -\infty$, it is clear that optimality does not hold.

- Next, we consider the case where $s \in (0, \infty]$. There exists a constant $\epsilon > 0$, such that for sufficiently large t, $c_t^2 > \log v(1+\epsilon)$. Thus, for all sufficiently large t, the probability of a type II error is given by

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

Now observe that

$$P\left(|Z| < \sqrt{\frac{\log v(1+\epsilon)}{u+1}}\right) = \sqrt{\frac{2\log v(1+\epsilon)}{\pi u}}(1+o_t) \text{ when } C = 0$$

and

$$P\left(|Z| < \sqrt{\frac{\log v(1+\epsilon)}{u+1}}\right) = (2\Phi(\sqrt{C(1+\epsilon)}-1)(1+o_t) \text{ when } C > 0 .$$

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

Now we consider case (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 of $\frac{z_t}{\log v}$ converges. By an analogous argument to case (i), optimal risk properties will not hold for a subsequence which converges to a point in $[-1, \infty] - \{0\}$, and hence neither for the whole sequence.

To conclude the proof, we prove the necessity of condition (4.23). Suppose (4.23) does not hold. If (4.22) does not hold either, optimality cannot hold, since (4.22) is necessary for optimality, as proved above. If on the other hand, (4.22) does hold, then the calculations leading to formula (8.81) remain valid. This implies that $\frac{R_1}{R_2}$ does not converge to zero (since (4.23) does not hold) and hence optimality does not hold.

8.2. *The BFDR is a decreasing function of the threshold c.* Notation:

$\Phi_\sigma(x)$ and $\phi_\sigma(x)$ - cdf and pdf of $N(0, \sigma^2)$

$\Phi(x)$ and $\phi(x)$ - cdf and pdf of $N(0, 1)$

Proposition 8.1. *If $\sigma > 1$, then the function*

$$H(x) = \frac{1 - \Phi_\sigma(x)}{1 - \Phi(x)}$$

is increasing on $[0, \infty)$ and

$$\lim_{x \to \infty} H(x) = \infty \quad . \tag{8.83}$$

Proof. First we prove (8.83). Observe that

$$1 - \Phi_\sigma(x) = \int_x^\infty \phi_\sigma(x) dx$$

and that

$$\frac{\phi_\sigma(x)}{\phi(x)} = \frac{1}{\sigma} e^{\frac{x^2(\sigma^2 - 1)}{2\sigma^2}} \quad .$$

Thus, for $\sigma > 1$, $\phi_\sigma(x) = \phi(x) g(x)$, where $g(x)$ is increasing on $[0, \infty)$ and $\lim_{x \to \infty} g(x) = \infty$.

Thus $\forall D > 0,\ \exists x_0 > 0$ such that $\forall x > x_0$

$$H(x) = \frac{\int_x^\infty \phi(x) g(x) dx}{\int_x^\infty \phi(x) dx} > D \quad ,$$

which completes the proof of (8.83).

To show that $H(x)$ is increasing on $[0, \infty)$, consider an arbitrary pair of numbers c_1 and c_2, such that $0 \leq c_1 < c_2$. Note that

$$H(c_1) = \frac{\int_{c_1}^\infty \phi(x) g(x) dx}{\int_{c_1}^\infty \phi(x) dx} = \frac{\int_{c_1}^{c_2} \phi(x) g(x) dx + \int_{c_2}^\infty \phi(x) g(x) dx}{\int_{c_1}^{c_2} \phi(x) dx + \int_{c_2}^\infty \phi(x) dx} \quad .$$

Observe that

$$\frac{\int_{c_1}^{c_2} \phi(x) g(x) dx}{\int_{c_1}^{c_2} \phi(x) dx} < g(c_2),$$

while

$$\frac{\int_{c_2}^\infty \phi(x) g(x) dx}{\int_{c_2}^\infty \phi(x) dx} > g(c_2) \quad .$$

This implies that $H(c_1) < H(c_2)$, since for positive numbers a, b, c, d,

$$\frac{a}{b} < \frac{c}{d} \quad \text{implies that} \quad \frac{a + c}{b + d} < \frac{c}{d} \quad .$$

∎

LEMMA 8.1. *For any fixed $p \in (0,1)$ and $u > 0$, the function*

$$BFDR(c) = \frac{(1-p)(1-\Phi(c))}{(1-p)(1-\Phi(c)) + p(1-\Phi(c/\sqrt{u+1}))}$$

is monotonically decreasing from $(1-p)$ for $c = 0$ to 0 for $c \to \infty$.

PROOF. Noting that

$$BFDR(c) = \left(1 + \frac{1-\Phi_{\sqrt{u+1}}(c)}{f(1-\Phi(c))}\right)^{-1},$$

the conclusion of Lemma 8.1 easily follows from Proposition 8.1. □

8.3. *Proof of Theorem 5.1.* The proof of Theorem 5.1 is based on the following two lemmas.

LEMMA 8.2. *Under Assumption (A), a BFDR control rule (5.35) is asymptotically optimal only if*

$$\frac{f}{r_\alpha} \to \infty \tag{8.84}$$

and

$$\frac{2\log\left(\frac{f}{r_\alpha}\right)}{u} \to C \quad . \tag{8.85}$$

Its threshold value is given by (5.41), i.e.

$$c_B^2 = 2\log\left(\frac{f}{r_\alpha}\right) - \log\left(2\log\left(\frac{f}{r_\alpha}\right)\right) + C_1 + o_t \quad , \tag{8.86}$$

where $C_1 = \ln\left(\frac{2}{\pi D^2}\right)$ and $D = 2(1-\Phi(\sqrt{C})$.

The corresponding probability of a type I error is given by

$$t_1 = D\frac{r_\alpha}{f}(1 + o_t) \quad . \tag{8.87}$$

PROOF. From Theorem 4.1, the multiple testing rule based on c_B^2 can be asymptotically optimal only if

$$t_1 = 2(1-\Phi(c_B)) \to 0 \tag{8.88}$$

and

$$\frac{c_B^2}{u+1} \to C \quad . \tag{8.89}$$

From (5.36), it follows that conditions (8.88) and (8.89) can both hold only if (8.84) is satisfied.

Moreover, (5.36) and (8.89) imply that

$$2(1-\Phi(c_B)) = D\frac{r_\alpha}{f}(1+o_t) \quad , \tag{8.90}$$

which proves (8.87).

Applying the normal tail approximation (3.17) to the left hand side of (8.90), we obtain

$$\sqrt{\frac{2}{\pi}}\frac{\exp(-c_B^2/2)}{c_B} = D\frac{r_\alpha}{f}(1+o_t) \quad .$$

Thus,

$$c_B^2 = 2\log(f/r_\alpha) + C_1 - 2\log c_B + o_t$$

and we subsequently obtain

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

This implies that condition (8.89) can hold only if

$$\frac{2\log\left(\frac{f}{r_\alpha}\right)}{u} \to C \quad . \tag{8.91}$$

□

Lemma 8.2 shows that conditions (8.84) and (8.85) are necessary for the optimality of a BFDR control rule. In Lemma 8.3 we prove that they are also sufficient for (8.86) to hold.

LEMMA 8.3. *Suppose Assumption (A) holds and α satisfies conditions (8.84) and (8.85). Then the threshold c_B^2 (defined in (5.35)) of a rule controlling the BFDR at level α is given by (8.86).*

PROOF. First we will show that condition (8.85) implies that $\frac{c_B}{\sqrt{u+1}} = z_t$ is bounded from above as $t \to \infty$. If this is false, one can find a subsequence $\tilde{t}$ such that $z_{\tilde{t}} \to \infty$. Using the tail approximation (3.17), along this subsequence the following holds

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

This together with (5.36) yields

$$z_{\tilde{t}}^2 = \frac{2\log\left(\frac{f}{r_\alpha\sqrt{u+1}}\right)}{u} + o_{\tilde{t}}.$$

This, together with the fact that $z_{\tilde{t}} \to \infty$, contradicts (8.85).

As $\frac{c_B}{\sqrt{u+1}}$ is bounded from above, it always remains within a compact interval, as it is non-negative. So given any subsequence, there will be a further subsequence that converges to a finite constant. Now consider an arbitrary subsequence $\tilde{t}$ such that $\frac{c_B}{\sqrt{u+1}} \to C_3 < \infty$ along this subsequence and let the corresponding asymptotic power be $D_1 = 2(1-\Phi(C_3)) > 0$. Along this subsequence, (5.36) reduces to

$$1-\Phi(c_B) = \frac{D_1 r_\alpha (1+o_{\tilde{t}})}{2f} \ .$$

Since we assume (8.84), using the normal tail approximation (3.17), for this subsequence the corresponding thresholds are of the form

$$c_B^2 = 2\log\left(\frac{f}{r_\alpha}\right) - \log\left(2\log\left(\frac{f}{r_\alpha}\right)\right) + C_1 + o_{\tilde{t}} \ , \tag{8.92}$$

where $C_1 = \log\left(\frac{2}{\pi {D_1}^2}\right)$. Now observe that (8.92) and (8.85) imply that $C_3 = \sqrt{C}$. So we have proved that every convergent subsequence of $\frac{c_B}{\sqrt{u+1}}$ converges to $\sqrt{C}$. From the compactness of the sequence, this implies that the sequence itself will in fact converge to $\sqrt{C}$. The proof of the lemma is now complete, since this implies that the asymptotic form of the threshold c_B^2 for the sequence itself will in fact satisfy (8.92) with $D_1 = 2(1-\Phi(\sqrt{C}))$ (i.e., it will satisfy (8.86)). □

Proof of Theorem 5.1:

First we prove that for optimality to hold conditions (5.39) and (5.40) on s_t (as defined in (5.38)) are sufficient. Note that (5.39) guarantees that (8.84) and (8.85) hold. According to Lemma 8.3, the threshold of the BFDR control rule can thus be written as

$$c_B^2 = \log v + \tilde{z}_t$$

with

$$\tilde{z}_t = -\log v + 2\log\left(\frac{f}{r_\alpha}\right) - \log\left(2\log\left(\frac{f}{r_\alpha}\right)\right) + C_1 + o_t \ .$$

According to Theorem 4.1, such a rule is asymptotically optimal if

$$\log v - 2\log\left(\frac{f}{r_\alpha}\right) + \log\left(2\log\left(\frac{f}{r_\alpha}\right)\right) = o(\log v) \tag{8.93}$$

and

$$\log v - 2\log\left(\frac{f}{r_\alpha}\right) + \log\left(2\log\left(\frac{f}{r_\alpha}\right)\right) - 2\log\log v \to -\infty \ . \tag{8.94}$$

Since (8.84) holds, condition (8.93) is fulfilled if and only if

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

i.e. if and only if condition (5.39) holds. From (5.38), we obtain

$$\log v = 2(1+s_t)\log(f/r_\alpha)$$

and simple computations show the equivalence of (8.94) and (5.40).

Let us now prove that for optimality to hold, it is also necessary that conditions (5.39) and (5.40) are satisfied by s_t. The optimality of a BFDR control rule implies, from Lemma 8.2, that both (8.84) and (8.85) hold. We may thus use the asymptotic approximation to the threshold given in (8.86). Since the rule is optimal, Theorem (4.1) implies that (8.93) and (8.94) must hold, which have been shown to be equivalent to (5.39) and (5.40).

8.4. *Proofs of Theorems 6.1 and 6.3.* First we will prove two lemmas needed for the proof of Theorem 6.3.

Proof of Lemma 6.1

PROOF. Given the condition $K = k$, there are $(m-k)$ true nulls. Let the corresponding ordered p-values be $\tilde{p}_{(1)} \leq ... \leq \tilde{p}_{(m-k)}$. 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

$$\tilde{k} = argmax_i \left\{ \tilde{p}_{(i)} \leq \frac{\alpha(i+k)}{m} \right\} . \tag{8.95}$$

Let $\tilde{V_1}$ be the corresponding number of rejections. Then $E(V|K=k) \leq E(\tilde{V_1})$, since the number of false rejections for the original BH, V, is not larger than $\tilde{V_1}$. Now, consider m i.i.d p-values $q_1, \ldots, q_m$ from the total null (i.e, each of the m nulls is true), which are independent of the given original p-values. Let $\tilde{q}_{(1)} \leq \ldots \leq \tilde{q}_{(m-k)}$ be the ordered values from the subsequence $q_1, \ldots, q_{m-k}$. Then $\tilde{q}_{(1)}, \ldots, \tilde{q}_{(m-k)}$ and $\tilde{p}_{(1)}, ..., \tilde{p}_{(m-k)}$ have exactly the same distribution. Let V_1 and V_2 be the number of rejections of null when the procedure (8.95) is applied to the first $(m-k)$ or m q's respectively. Then $E(V|K=k) \leq E(\tilde{V_1}) = E(V_1) \leq E(V_2)$.

Now the bound on k (see the assumption of Lemma 6.1) guarantees that the right hand side of (8.95) is smaller than 1 for all possible i. We can thus apply Lemma 4.2 of [16] 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 6.1

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

□

Proof of Lemma 6.2

PROOF. Define $C_2 := \frac{1}{\alpha_\infty} - 1$ and $m_0 := \min(m, C_2 m)$. The following holds

$$E(V) \leq \sum_{k=0}^{m_0} E(V|K=k)P(K=k) + mP(K > C_2 m) . \tag{8.96}$$

The first term can be bounded for m large enough using Lemma 6.1 :

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

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

$$\sum_{k=0}^{m_0} E(V|K=k)P(K=k) < C_1 m\alpha_m p_m \ ,$$

for any constant C_1 satisfying the assumption of Lemma 6.2.

Finally, note that the second term of (8.96) vanishes for $\alpha_\infty < 0.5$. On the other hand, for $\alpha_\infty \in [0.5, 1)$, Lemma 7.1 of [1] yields

$$mP(K > C_2 m) \leq m \exp(-\frac{1}{4} m p_m h(C_2/p_m)) \ ,$$

where $h(x) = \min(|x-1|, |x-1|^2)$. If $p_m \to 0$, then for any constant $C_3 \in (0, C_2)$ and sufficiently large m, the right hand side is bounded from above by $m\exp(-C_3 m) \to 0$. Now, from the assumptions $mp_m \to s > 0$ and $\alpha_m \to \alpha_\infty > 0.5$, it follows that for any constant $c > 0$ and sufficiently large m, the second term of (8.96) is smaller than $c\alpha_m m p_m$ and Lemma 6.2 follows. □

Proof of Theorem 6.3

The proof of Theorem 6.3 is based on a sequence of four lemmas. The first of these lemmas states that when $p_m \to 0$ such that $mp_m \to \infty$, the tail probability $(1 - F(c_{GW}))$ (of $\left|\frac{X_i}{\sigma}\right|$ being greater than the asymptotically optimal threshold c_{GW}) can be very well approximated by its estimate $(1 - \hat{F}_m(c_{GW}))$ based on the empirical distribution function.

LEMMA 8.4. *Suppose Assumption (A) is true and $p_m \to 0$. Consider the multiple testing rule based on the GW threshold c_{GW}, defined in (5.54), where the level α_m is chosen in such a way that this rule is asymptotically optimal (i.e the condition in Theorem 5.1 is satisfied). For any constant $\xi \in (0,1)$,*

(8.97)
$$P\left(\frac{1-\hat{F}_m(c_{GW})}{1-F(c_{GW})} > (1+\xi)\right) \leq \exp\left\{-\frac{1}{4} m p_m D(r_{\alpha_m}+1)\xi^2(1+o_m)\right\}$$

and

(8.98)
$$P\left(\frac{1-\hat{F}_m(c_{GW})}{1-F(c_{GW})} < (1-\xi)\right) \leq \exp\left\{-\frac{1}{4} m p_m D(r_{\alpha_m}+1)\xi^2(1+o_m)\right\} \ .$$

PROOF. Let $1 - F(c_{GW}) = (1 - p_m)t_1(c_{GW}) + p_m(1 - t_2(c_{GW}))$, where $t_1(c_{GW})$ and $t_2(c_{GW})$ denote the probability of type I and type II errors for the procedure based on the GW threshold. Theorem 5.2 implies that (5.56) holds and hence $t_1(c_{GW})$ and $t_2(c_{GW})$ are exactly of the same asymptotic form as the corresponding probabilities of type I and type II errors for the rule controlling the BFDR at the same level. Therefore, using Theorem 5.1, we obtain

$$1 - F(c_{GW}) = p_m D(r_{\alpha_m} + 1)(1 + o_m) \quad . \tag{8.99}$$

Now observe that $Y = m(1 - \hat{F}_m(c_{GW}))$ is a Binomial $B(m, 1 - F(c_{GW}))$ random variable. Therefore, from Bennett's inequality (e.g. see [30], page 440), it follows that

$$P(Y > m(1 - F(c_{GW}))(1 + \xi)) \le \exp\left\{-\frac{1}{4} m p_m D(r_{\alpha_m} + 1)\xi^2(1 + o_m)\right\}$$

and

$$P(Y < m(1 - F(c_{GW}))(1 - \xi)) \le \exp\left\{-\frac{1}{4} m p_m D(r_{\alpha_m} + 1)\xi^2(1 + o_m)\right\}$$

and the proof of Lemma 8.4 is complete. □

The next lemma shows that with a large probability the random threshold of BH is bounded from above by a sequence which converges to c_{GW}.

LEMMA 8.5. *Suppose Assumption (A) holds. Assume that α_m and p_m satisfy the assumptions of Lemma 8.4. Let c_{BH} be the BH threshold at level α_m and let $\tilde{c}_1 = \tilde{c}_{1m}$ be the GW threshold (5.54) at level $\alpha_{1m} = \alpha_m(1 - \xi_m)$, where $\xi_m \to 0$ and $\xi_m = o(1 - \alpha_m)$ as $m \to 0$. It follows that*

$$\tilde{c}_1 = c_{GW} + o_m \tag{8.100}$$

and

$$P(c_{BH} \ge \tilde{c}_1) \le \exp\left\{-\frac{1}{4} m p_m D(r_{\alpha_m} + 1)\xi_m^2(1 + o_m)\right\} \quad .$$

PROOF. Note that since $\xi_m = o(1 - \alpha_m)$ and $\xi_m \to 0$, it follows that $r_{\alpha_{1m}} = r_{\alpha_m}(1 + o_m)$. Thus, similarly to c_{GW}, $\tilde{c}_1$ satisfies equation (5.56) and (8.100) follows.

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

$$\begin{aligned} P(c_{BH} < \tilde{c}_1) \quad &\ge \quad P\left(\frac{2(1 - \Phi(\tilde{c}_1))}{(1 - \hat{F}_m(\tilde{c}_1))} \le \alpha_m\right) \\ &= \quad P\left(\frac{2(1 - \Phi(\tilde{c}_1))}{1 - F(\tilde{c}_1)} \frac{1 - F(\tilde{c}_1)}{1 - \hat{F}_m(\tilde{c}_1)} \le \alpha_m\right) = P\left(\frac{1 - \hat{F}_m(\tilde{c}_1)}{1 - F(\tilde{c}_1)} \ge 1 - \xi_m\right) \quad . \end{aligned}$$

Now the proof follows from (8.100) and invoking the arguments of Lemma 8.4. □

Based on Lemmas 8.4 and 8.5, we can now prove Lemma 8.6 which provides an upper bound on c_{BH} in terms of c_{GW}.

LEMMA 8.6. *Suppose that Assumption (A) holds and that p_m and α_m satisfy the assumptions of Theorem 6.3. Let c_{BH} and c_{GW} denote the thresholds of BH and their Genovese-Wasserman approximation (5.54) corresponding to FDR level α_m. It follows that for every $\epsilon > 0$, every constant $\beta_u > 0$ and for sufficiently large m, we have*

$$P(c_{BH} > c_{GW} + \epsilon) \leq m^{-\beta_u} \quad . \tag{8.101}$$

PROOF. We first observe that when $p_m \to 0$ and assumptions (6.64), (6.65) and (6.71) hold, one can invoke Lemma 8.5 by choosing $\xi_m = (\log m)^{c_\xi}$, where $c_\xi \in ((1-\beta_p)/2, 0)$. By doing this, we conclude that for every constant $\beta_u > 0$ and sufficiently large m

$$P(c_{BH} \geq \tilde{c}_1) \leq m^{-\beta_u}, \tag{8.102}$$

since by (6.62) $mp_m > \log^{\beta_p} m$. Thus (8.101) follows from (8.100). □

To finish the proof of Theorem 6.3, we show that with a very large probability c_{BH} can be bounded from below by a sequence which asymptotically converges to c_{GW}. The proof of this is substantially more complicated than the previous proof, since for $c_{BH} > \tilde{c}_2$ to hold, it is necessary that $\frac{2(1-\Phi(y))}{1-\hat{F}_m(y)} > \alpha$ for all $y < \tilde{c}_2$.

LEMMA 8.7. *Suppose that Assumption (A) holds and that p_m and α_m satisfy the assumptions of Theorem 6.3. Let c_{BH} and c_{GW} denote the thresholds of BH and their Genovese-Wasserman approximation (5.54) corresponding to FDR level α_m. For every $\epsilon > 0$, every constant $\beta_u > 0$ and sufficiently large m, the following holds*

$$P(c_{BH} < c_{GW} - \epsilon) \leq m^{-\beta_u} \quad . \tag{8.103}$$

PROOF. Let $\tilde{c}_2 = \tilde{c}_{2m}$ be the GW threshold (5.54) at the level $\alpha_{2m} = \alpha_m(1+\xi_m)$, where $\xi_m = (\log m)^{c_\xi}$, with $c_\xi \in ((1-\beta_p)/2, 0)$. Since $\xi_m \to 0$, it follows that $r_{\alpha_{2m}} = r_{\alpha_m}(1+o_m)$ and $\tilde{c}_2$ satisfies equation (5.56) with $\alpha = \alpha_m$. Thus it easily follows that $\tilde{c}_2$ is asymptotically optimal and satisfies

$$\tilde{c}_2 = c_{GW} + o_m \quad , \tag{8.104}$$

where c_{GW} is given by (5.54) with $\alpha = \alpha_m$.

Now observe that from the monotonicity of $\frac{(1-\Phi(c))}{1-F(c)}$

$$c_{BH} < \tilde{c}_2 \text{ if and only if } \frac{2(1-\Phi(c_{BH}))}{1-F(c_{BH})} > \alpha_m(1+\xi_m) \quad . \tag{8.105}$$

Moreover, from the definition of c_{BH},

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

Thus, the event $\{c_{BH} < \tilde{c}_2\}$ implies that

$$\frac{1-\hat{F}_m(c_{BH})}{1-F(c_{BH})} > (1+\xi_m)$$

and in consequence

$$P(c_{BH} < \tilde{c}_2) \leq P\left(\sup_{c\in[0,\tilde{c}_2)} \frac{1-\hat{F}_m(c)}{1-F(c)} > (1+\xi_m)\right) \quad . \tag{8.106}$$

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

$$P(c_{BH} \in [0,\tilde{c}_2)) \leq P\left(\sup_{t\in[0,z_{1m})} \frac{1-\hat{G}_m(t)}{1-t} \geq (1+\xi_m)\right) \quad ,$$

where $z_{1m} = F(\tilde{c}_2) = 1 - p_m D(r_{\alpha_m}+1)(1+o_m)$ and $\hat{G}_m(t)$ is the empirical cdf for m independent observations from the uniform distribution on $[0,1]$.

Let $u_i = \frac{i}{m}$ and $k_{1m} = \lceil m(1-C_1 p_m)\rceil$, where $C_1 \in (0,D)$. From the monotonicity of $G_m(t)$ and t, it follows that for sufficiently large m

$$P\left(\sup_{t\in[0,z_{1m})} \frac{(1-\hat{G}_m(t))}{1-t} > (1+\xi_m)\right) \leq$$

$$\sum_{i=0}^{k_{1m}} P\left(1-\hat{G}_m(u_i) > (1-u_i-\frac{1}{m})(1+\xi_m)\right) \quad . \tag{8.107}$$

Note that for $i=0$ and sufficiently large m,

$$P\left(1-\hat{G}_m(u_i) > (1-u_i-\frac{1}{m})(1+\xi_m)\right) = 0 \quad .$$

We now consider the case $i \in \{1,\ldots,k_{1m}\}$.

Observe that over this range of i

$$1 - u_i \geq C_1 \frac{\log^{\beta_p} m}{m} - \frac{1}{m}$$

and therefore

$$(1 - u_i - \frac{1}{m}) = (1 - u_i)(1 - t_m) \ ,$$

where $t_m = O([\log m]^{-\beta_p}) = o(\xi_m)$.

Now using Bennett's inequality, we obtain that for every $i \in \{1, \ldots, k_{1m}\}$

$$\begin{aligned} P(1 - \hat{G}_m(u_i) &> (1 - u_i)(1 + \xi_m)(1 - t_m)) \\ &\leq \exp\left(-\frac{1}{4} m(1 - u_i)\xi_m^2(1 + o_m)\right) \\ &\leq \exp\left(-\frac{C_2}{4} (\log m)^{\beta_p} \xi_m^2 (1 + o_m)\right) \end{aligned}$$

for any constant $C_2 \in (0, C_1)$. Thus for any constant $\beta_u > 0$, sufficiently large m and any $i \in \{1, \ldots, k_{1m}\}$, the following holds

$$mP(1 - \hat{G}_m(u_i) \geq (1 - u_i)(1 + \xi_m)(1 - t_m)) < m^{-\beta_u}$$

and Lemma 8.7 follows by invoking (8.107) and (8.104). □

The proof of Theorem 6.3 results from combining Lemma 8.6 and Lemma 8.7.

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., 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.

[4] 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* 211–230. Beachwood Ohio.

[5] Cai, T. and Jin, J. (2009). Optimal rates of convergence for estimating the null and proportion of non-null effects in large-scale multiple testing. *Ann. Statist.* to appear.

[6] Cai, T., Jin, J. and Low, M. (2007). Estimation and confidence sets for sparse normal mixtures. *Ann. Statist.***35**, 2421-24.

[7] Carvalho, C.M., Polson N. and Scott J. G. (2008). The horseshoe estimator for sparse signals. *Duke University Department of Statistical Science Technical Report 2008-31.*

[8] Do K.-A, Müller P. and Tang F. (2005). A Bayesian Mixture Model for Differential Gene Expression. *Applied Statistics***54** 627-644.

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

[10] Donoho, D.L. and Jin, J. 2006. Asymptotic minimaxity of false discovery rate thresholding for sparse exponential data. *Ann. Statist.* **34** 2980–3018.

[11] 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.

[12] Donoho, D.L. and Johnstone, I. M. 1995. Adapting to unknown smoothness via wavelet shrinkage. *J. Amer. Statist. Assoc.* **90** 1200–1224.

[13] Efron, B. (2008). Microarrays, Empirical Bayes and the Two-Groups Model. *Statist. Sci.* **23** 1–22. MR2431866

[14] Efron, B. and Tibshirani, R. (2002). Empirical bayes methods and false discovery rates for microarrays. *Genetic Epidemiology.* **23** 70–86. MR1946571

[15] 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.

[16] Finner H. and Roters M. (2002). Multiple hypotheses testing and expected number of type I errors. *Ann. Statist.* **30** 220-238. MR1892662

[17] 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

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

[19] Jeng, X. (2009). Covariance Adaptation and Regularization in Large-Scale Hypothesis Testing and High-Dimensional Regression. PhD thesis, Department of Statistics, Purdue University.

[20] Jin, J. and Cai, T. C.. Estimating the null and the proportion of non-null effects in large-scale multiple comparisons. *J. Amer. Statist. Assoc.* **102** 495-506.

[21] Lehmann, E. L. 1957. A theory of some multiple decision problems, I. *Ann. Math. Stat.* **28** 1–25.

[22] Lehmann, E. L. 1957. A theory of some multiple decision problems, II. *Ann. Math. Stat.* **28** 547–572.

[23] Lehmann, E. L. and Romano, J. P. (2005). Generalizations of the familywise error rate. *Ann. Statist.* **33** 1138–1154. MR2195631

[24] 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

[25] 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

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

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

[28] Scott, J. G. and Berger, J. O. (2008). Bayes and empirical-Bayes multiplicity adjustment in the variable-selection problem. Duke University Department of Statistical Science, Discussion Paper 2008-10.

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

[30] Shorack, G. R. and Wellner, J. A. (1986). *Empirical processes with applications to Statistics.* Wiley Series in Probability and Mathematical Statistics. MR0838963

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

[32] Storey, J. D. (2002). A direct approach to false discovery rates. *J. R. Stat. Soc. Ser. B.* **64** 479–498. MR1924302

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

[34] Storey, J. D. (2007). The optimal discovery procedure: a new approach to simultaneous significance testing. *J. R. Statist. Soc. B.* **69** 347–368. MR2323757

[35] 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.* **66** 187–205. MR2035766

[36] Sun, W. and Cai, T. C.. (2007) Oracle and adaptive compound decision rules for false discovery rate control. *J. Amer. Statist. Assoc.* **102** 901-912.

[37] Tai, Y. C. and Speed, T. P. (2006). A multivariate empirical Bayes statistic for replicated microarray time course data. *Ann. Statist.* **34** 2387-2412.

Institute of Mathematics and Computer Science
Wrocław University of Technology
ul. Janiszewskiego 14a
50-370 Wroclaw, Poland
E-mail: Malgorzata.Bogdan@pwr.wroc.pl

150 North University Street
Department of Statistics
Purdue University
West Lafayette, IN-47907, USA
ghosh@purdue.edu, bogdanm@stat.purdue.edu

203 B.T.Road
Bayesian and Interdisciplinary Research Unit
Indian Statistical Institute
Kolkata 700109, West Bengal,India
arc@isical.ac.in, jayanta@isical.ac.in

Brünnerstrasse 72
Department of Statistics and Decision Support Systems
University of Vienna
1210 Vienna, Austria
Florian.Frommlet@univie.ac.at