**Operating Characteristics and Extensions of the FDR Procedure**

Christopher Genovese[1] and Larry Wasserman[2]
Carnegie Mellon University
March 2, 2001

We investigate the operating characteristics of the Benjamini-Hochberg false discovery rate (FDR) procedure for multiple testing. This is a distribution free method that controls the expected fraction of falsely rejected null hypotheses among those rejected. This paper provides a framework for understanding how and why this procedure works. We start by studying the special case where the p-values under the alternative have a common distribution, where we are able to obtain many insights into this new procedure. We first obtain bounds on the "deciding point" $D$ that determines the critical p-value. From this, we obtain explicit asymptotic expressions for a particular risk function. We introduce the dual notion of false non-rejections (FNR) and we consider a risk function that combines FDR and FNR. We also consider the optimal procedure with respect to a measure of conditional risk.

KEYWORDS: Multiple Testing, p-values, Risk, False Discovery Rate.

# 1 Introduction

In a breakthrough paper, Benjamini and Hochberg (BH 1995) introduced a new approach to multiple hypothesis testing that controls the false discovery rate (FDR), defined as the fraction of false rejections among those hypotheses rejected. The procedure is very appealing because it controls a quantity that is often of greater scientific relevance than the overall type I error rate. Also, the procedure is more powerful than the Bonferroni method. However, the reasons underlying the method's success are somewhat mysterious. In this

[1]Research supported by NSF Grant SES 9866147.
[2]Research supported by NIH Grant R01-CA54852-07 and NSF Grant DMS-98-03433

paper, we de-mystify FDR. Figure 1 is at the heart of our explanation. It shows how the BH method is intermediate between uncorrected testing and Bonferroni. The quantities $u^*$ and $\beta$ on the plot will be explained shortly.

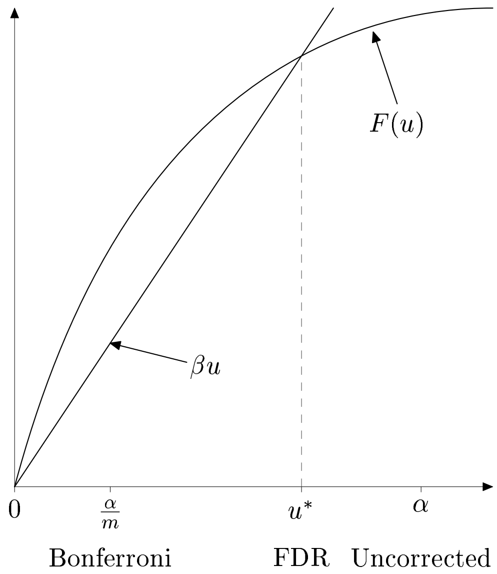


**Figure 1.** Geometric interpretation of the asymptotic behavior of the BH procedure. $F$ is the distribution of the p-value under the alternative. Asymptotically, the BH procedure rejects whenever the p-values less than $u^*$. The p-value thresholds for Bonferroni and uncorrected testing are also shown.

The broader purpose of this paper is to investigate the operating characteristics of the BH procedure as well as develop some extensions. Specifically, we approximate the marginal risk, defined as the total number of errors (false rejections and false non-rejections). The BH method was not designed to minimize this risk. Nevertheless, it is interesting to study this risk, first, because it gives insight into the procedure and second, because we have observed empirically that it often does give small marginal risk and so it is important to quantify when and how this small risk situation occurs. In particular, we compare the BH method explicitly with uncorrected (fixed level) testing, Bonferroni as well as the optimal Bayes classification rule. We introduce the dual notion of the False Non-discovery Rate (FNR), and we formulate and study a second risk that combines FDR and FNR. We also study this risk function from a conditional point of view. Conditional testing is discussed in a different context by Berger, Brown and Wolpert (1994). Other references on FDR include extensions of the BH method to dependent tests by Benjamini and Yekutieli (1999) and connections between FDR and minimax point estimation are made by Abramovich, Benjamini, Donoho and Johnstone (2000).

Let us first review the appropriate definitions. Consider a multiple testing situation in which $m$ tests are being performed. Suppose $m_0$ of the null hypotheses are true and $m_1 = m - m_0$ null hypotheses are false. We can categorize the $m$ tests in the following $2 \times 2$ table.

| | $H_0$ Not Rejected | $H_0$ Rejected | Total |
|---|---|---|---|
| $H_0$ True | $N_{0\|0}$ | $N_{1\|0}$ | $m_0$ |
| $H_0$ False | $N_{0\|1}$ | $N_{1\|1}$ | $m_1$ |
| Total | $m - R$ | $R$ | $m$ |

We define the False Discovery Rate (FDR) and the False Nondiscovery Rate (FNR) by

$$\mathrm{FDR} = \begin{cases} \dfrac{N_{1|0}}{R} & \text{if } R > 0 \\ 0, & \text{if } R = 0, \end{cases} \tag{1}$$

and

$$\mathrm{FNR} = \begin{cases} \dfrac{N_{0|1}}{m-R} & \text{if } R < m \\ 0 & \text{if } R = m. \end{cases} \tag{2}$$

The first is the proportion of rejections that are incorrect, and the second – the dual quantity – is the proportion of non-rejections that are incorrect.

The BH procedure works as follows. Suppose $P_1, \ldots, P_m$ are p-values for the $m$ tests. Let $P_{(1)} < \cdots < P_{(m)}$ denote the ordered p-values and let $\ell_i \equiv \ell(i) = i\alpha/m$. Define

$$D = \max\{i : P_{(i)} < \ell_i\}. \tag{3}$$

We refer to $D$ as the "deciding point." The procedure rejects all null hypotheses for which $P_i \leq P_{(D)}$; hence, for the BH procedure $R = D$. BH proved that this procedure ensures

$$\mathsf{E}\,(\mathrm{FDR}) \leq \frac{m_0}{m}\,\alpha \leq \alpha, \tag{4}$$

regardless of how many nulls are true and regardless of the distribution of the p-values under the alternatives.

In this paper, we examine the asymptotic behavior of the deciding point $D$, for both fixed and local alternatives. We then explore the marginal risk of this procedure, defined as the expected number of errors (false rejections and false non-rejections), and we examine how well the procedure controls the expected FNR. We obtain explicit, asymptotic expressions for the risk, and we compare these to some other procedures. In our results, we use an asymptotics in the the fraction of true nulls $A_0$ is kept fixed as the number of tests $m$ increases. The BH procedure is distribution free: the expected FDR is controlled regardless of $m_0$ and regardless of the distributions for the p-values corresponding to the non-null hypotheses. Currently, little is known about the operating characteristics of the BH procedure.

We begin by considering the "simple versus simple" case in which the p-values under the alternative have the same distribution. This may be regarded as a least favorable configuration for the non-null hypotheses. More importantly, by investigating this special case we develop some new insights into the procedure. Extensions to certain types of composite alternatives are discussed in the penultimate section of the paper.

Our investigation leads to the following conclusion. Asymptotically, the BH procedure corresponds to rejecting the null when the p-value is less than $u^*$ where $u^*$ is the solution to the equation $F(u) = \beta u$ and $\beta = (\frac{1}{\alpha} - A_0)/(1 - A_0)$. Here $F$ is the (common) distribution of the p-value under the alternative. Furthermore, as depicted in Figure 1, $\alpha/m \leq u^* \leq \alpha$ which shows that the BH method is intermediate between Bonferroni (corresponding to $\alpha/m$) and uncorrected testing (corresponding to $\alpha$).

An important step in our calculations involves bounding the random variable $D$. Sections 2 and 3 find such bounds in the fixed and local alternative cases, respectively. Section 4 contains risk calculations. In Section 5, we consider the False Nondiscovery Rate, and we give a heuristic for minimizing expected FNR subject to controlling expected FDR. In Section 6, we consider a different risk function and develop a procedure that minimizes the risk both unconditionally and in a conditional sense. We also make brief remarks about a Bayesian version of FDR. In Section 7, we consider extensions to composite alternatives. Section 8 contains some concluding remarks.

# 2 Bounds on Deciding Point: Fixed Alternative Case

Assume that we have $m$ tests. Let $S_0$ and $S_1$ be the set of indices $1, \ldots, m$ corresponding to true and false null hypotheses respectively. For $i = 0, 1$, let $m_i = \#S_i$ and $A_i = m_i/m$ be the corresponding number and proportions of each type. For now, we assume that all the false null hypothesis have a common alternative distribution. Let $F_\theta$ be the CDF of the p-value under the alternative and $f_\theta$ be the corresponding density.

Define $\ell_i = i\alpha/m$ and let $D = \max\{i : P_{(i)} < \ell_i\}$ be the "deciding point" for the FDR procedure. We are considering the asymptotic regime in which $m \to \infty$ while $A_0$ and $A_1$ stay fixed. In general, we will let the alternative $\theta$ depend on $m$, but we begin with the fixed alternative case in which $\theta_m = \theta$ for all $m$. Let $F \equiv F_\theta$ and $f \equiv f_\theta$ in this case.

THEOREM 1. *Assume that (i) $F$ is strictly concave and (ii) $F'(0) > \beta$, where*

$$\beta = \frac{\frac{1}{\alpha} - A_0}{1 - A_0} = \frac{1 - \alpha A_0}{\alpha A_1}.$$

*Let $u^*$ be the unique solution to the equation $F(u) = \beta u$. Define*

$$a_m = \frac{mu^*(1 - \epsilon_m)}{\alpha} \quad \text{and} \quad b_m = \frac{mu^*(1 + \epsilon_m)}{\alpha} \tag{5}$$

*where $\epsilon_m \to 0$ and $\epsilon_m \sqrt{m} \to \infty$ (e.g., $\epsilon_m = 1/\log\log m$). If $A_1 > 0$ then then, as $m \to \infty$,*

$$\mathsf{P}\{\, a_m \le D \le b_m \,\} \;\to\; 1, \tag{6}$$

*and*

$$\frac{b_m}{a_m} \;\to\; 1. \tag{7}$$

REMARK 1. The concavity condition on $F$ can be weakened. All that is needed is a unique solution $u^*$ such that (a) $F(u^*) = \beta u^*$ and (b) $F'(u^*) \neq \beta$. However, if the p-value is based on a test statistic whose density is eventually strictly decreasing, then concavity will hold, at least for small $\alpha$ which is all that is needed.

REMARK 2. If $A_1 = 0$ then all the nulls are true. In that case, the distribution of $D$ has a point mass of at least $1 - \alpha$ at 0.

PROOF. First, we record some useful properties of $u^*$. We know that $u^*$ exists and is unique because the conditions on $F$ and $\beta > 1$ imply that the line of slope $\beta$ crosses $F$ exactly once. Because $u^* = F(u^*)/\beta = \alpha A_1 F(u^*)/(1 - \alpha A_0)$, we can write the expressions for $a_m$ and $b_m$ as follows:

$$\begin{aligned} a_m &= \frac{A_1 F(u^*)\,(1-\epsilon_m)}{1-\alpha A_0}\, m \\ b_m &= \frac{A_1 F(u^*)\,(1+\epsilon_m)}{1-\alpha A_0}\, m. \end{aligned}$$

Notice that since $A_1 \leq 1 - \alpha A_0$, $1/\beta \leq \alpha$, so by the above, $u^* \leq \alpha F(u^*) \leq \alpha$, implying $F(u^*) \leq F(\alpha)$. Also note that $f(u^*) < \beta$ by the concavity of $F$.

Let $N_i = \#\{j : P_j \leq \ell_i\}$. Then, using the so called switching relation (see Abramovich, Benjamini, Donoho and Johnstone, 2000, for example) we note that

$$\{\, D > b_m \,\} = \bigcup_{k > b_m} \left\{\, P_{(k)} \leq \ell_k \,\right\} = \bigcup_{k > b_m} \{N_k \geq k\}.$$

Hence,

$$\mathsf{P}\{\, D > b_m \,\} \leq \sum_{k > b_m} \mathsf{P}\{\, N_k \geq k \,\}.$$

Define

$$\mu(t) = A_0 \alpha\, t + A_1 m F\left(\frac{\alpha t}{m}\right).$$

Notice that for $k$ integer, $\mu_k \equiv \mu(k) = \mathsf{E}\, N_k = \sum_i \mathsf{P}\{\, P_i \leq \ell_k \,\}$ where $N_k = \sum_{i=1}^m 1\{\, P_i \leq \ell_k \,\}$ . For $t > u^* m/\alpha$, $t - \mu(t)$ is increasing in $t$. To see this, first note that $f(tm/\alpha) < f(u^*) < \beta$ which implies that

$$\frac{d}{dt}(t - \mu(t)) = 1 - \alpha A_0 - A_1 m f(t/\alpha/m)(\alpha/m) > 1 - \alpha A_0 - \alpha A_1 \beta = 0$$

as claimed. Thus, $k - \mu_k \geq b_m - \mu_{b_m}$ for all $k > b_m$. Moreover,

$$\begin{aligned} b_m - \mu(b_m) &= (1-\alpha A_0)(1+\epsilon_m)\frac{u^* m}{\alpha} - m_1 F((1+\epsilon_m)u^*) \\ &= (1-\alpha A_0)(1+\epsilon_m)\frac{F(u^*) m}{\beta\alpha} - m_1 F((1+\epsilon_m)u^*) \\ &= m A_1 \left[(1+\epsilon_m)F(u^*) - F((1+\epsilon_m)u^*)\right] \\ &= m A_1 \left[\beta(1+\epsilon_m)u^* - (\beta + f(u^*)\epsilon_m)u^* + o(\epsilon_m)\right] \\ &= m A_1 \left[\epsilon_m u^*(\beta - f(u^*)) + o(\epsilon_m)\right]. \end{aligned} \tag{8}$$

As noted earlier, $f(u^*) < \beta$. It follows that

$$(\mu(b_m) - b_m)^2 = m^2\epsilon_m^2 \left[A_1 u^*(f(u^*) - \beta)\right]^2 + o\left(m^2\epsilon_m^2\right), \tag{9}$$

and therefore that $m^{-1}(\mu(b_m) - b_m)^2 \to \infty$.

By Hoeffding's inequality,

$$\begin{aligned} \sum_{k>b_m} \mathsf{P}\{ N_k \geq k \} &= \sum_{k>b_m} \mathsf{P}\left\{ \sum_{i=1}^m 1\{ P_i \leq \ell_k \} \geq k \right\} \\ &= \sum_{k>b_m} \mathsf{P}\left\{ \sum_{i=1}^m 1\{ P_i \leq \ell_k \} - \mu_k \geq k - \mu_k \right\} \\ &\leq \sum_{k>b_m} \exp\left\{ -\frac{2(k-\mu_k)^2}{m} \right\} \\ &\leq \sum_{k>b_m} \exp\left\{ -\frac{2(b_m-\mu_{b_m})^2}{m} \right\} \\ &\leq m \exp\left\{ -\frac{2(b_m-\mu_{b_m})^2}{m} \right\} \qquad (10) \\ &\to 0. \end{aligned}$$

Thus, $\mathsf{P}\{ D > b_m \} \to 0$ as $m \to \infty$.

Next we show that $\mathsf{P}\{ D < a_m \} \to 0$ as well. Notice that

$$\{D < a_m\} \subset \{P_{(a_m)} > \ell_{a_m}\} \subset \{N_{a_m} < a_m\}.$$

Hence,

$$\mathsf{P}\{ D < a_m \} \leq \mathsf{P}\{ N_{a_m} < a_m \}.$$

Recall that $a_m = (1-\epsilon_m)u^* m/\alpha$. Notice that $a_m/(1-\epsilon_m)$ (which is bigger than $a_m$) is a fixed point of $\mu(a)$. Using an argument like that used in (8), we have that

$$\mu(a_m) - a_m = mA_1 \left[\epsilon_m u^*(\beta - f(u^*)) + o(\epsilon_m)\right].$$

It follows that

$$(\mu(a_m) - a_m)^2 = m^2\epsilon_m^2 \left[A_1 u^*(f(u^*) - \beta)\right]^2 + o\left(m^2\epsilon_m^2\right),$$

and therefore that $m^{-1}(\mu(a_m) - a_m)^2 \to \infty$. By Hoeffding's inequality,

$$\begin{aligned} \mathsf{P}\{\, N_{a_m} < a_m \,\} &= \mathsf{P}\{\, N_{a_m} - \mu(a_m) < a_m - \mu(a_m) \,\} \\ &\leq \exp\left\{-2\,\frac{(\mu(a_m) - a_m)^2}{m}\right\} \qquad (11) \\ &\to 0. \end{aligned}$$

Equation (6) follows.

COROLLARY 1. *Under the conditions of Theorem 1,* $D/m \xrightarrow{p} u^*/\alpha$ *and hence,* $\ell(D) \xrightarrow{p} u^*$*.*

Thus, even though $u^*$ depends on $F$ and $A_0$, $D\alpha/m$ provides a distribution free, consistent estimator of $u^*$. Next we note that asymptotically, FDR is intermediate between uncorrected testing and Bonferroni.

LEMMA 1. *Under the conditions of Theorem 1, we have, for all large* $m$*, that* $\alpha/m \leq u^* \leq \alpha$*.*

PROOF. The upper bound was proved in the proof of Theorem 1. The lower bound is trivial.

# 3 Bounds on Deciding Point: Local Alternatives Case

We next investigate the risk when the alternative $\theta = \theta_m \to 0$ as $m \to \infty$. We now assume that the test statistics are Normal (0,1) under the null and N($\theta$,0) under the alternative. Hence, the cdf of the p-value (for one-sided testing) under the alternative is $F_\theta(u) = S(S^{-1}(u) - \theta)$ with density $f_\theta(u) = e^{-\theta^2/2}e^{\theta S^{-1}(u)}$ where $S(z) = 1 - \Phi(z)$. In what follows, we use the fact that, for all $u < .1$, $S^{-1}(u) = \{2\log(1/u) - \log\log(1/u) - r(u)\}^{1/2}$ where $0 \le r(u) \le 3$; for a proof of this, see, for example, Abramovich, Benjamini, Donoho and Johnstone (2000).

THEOREM 2. *Define $a_m$ and $b_m$ as in Theorem 1. Let*

$$\theta_m = c\left(\frac{\log m}{m}\right)^{1/4}$$

*where $c > 0$. Then Theorem 1 continues to hold. In this case, $a_m = O(\sqrt{m\log m})$ and $b_m = O(\sqrt{m\log m})$.*

PROOF. The proof is the same as for Theorem 1, however, we need to ensure that the term in (9), namely, $\sqrt{m}\epsilon_m u^*(\beta - f_m(u^*)) \to \infty$ to make sure that Hoeffding's inequality applies when bounding $a_m$. Here we have written $f_m = f_{\theta_m}$.

Recall that $u^*$ is defined by $F_{\theta_m}(u^*) = \beta u^*$. Note first that $u^* \to 0$ as $m \to \infty$. To see this, note that $F_{\theta_m}(u)$ converges uniformly to $u$ on compact subset of $(0, 1]$. Thus, since $\beta > 1$, eventually the solution of $F_{\theta_m}(u^*) = \beta u^*$ is not in $[a, 1]$ for any $a > 0$.

The equation $F_{\theta_m}(u^*) = \beta u^*$ can be rewritten as

$$S^{-1}(u^*) - S^{-1}(\beta u^*) = \theta_m. \tag{12}$$

Define $h_u(\beta) = S^{-1}(\beta u)$. Expanding around $\beta = 1$ we see that, for some $\tilde{\beta} \in [1, \beta]$ (depending on $u$),

$$\begin{aligned} h_u(\beta) &= h_u(1) + (\beta - 1)h'_u(\tilde{\beta}) \\ &= S^{-1}(u) - \frac{u(\beta-1)}{\phi(S^{-1}(\tilde{\beta}u))} \\ &= S^{-1}(u) - u(\beta - 1)\sqrt{2\pi}\exp\left\{\frac{1}{2}(S^{-1}(\tilde{\beta}u))^2\right\}. \end{aligned}$$

Hence,

$$\theta_m = h_{u^*}(1) - h_{u^*}(\beta) = u^*(\beta - 1)\sqrt{2\pi}\exp\left\{\frac{1}{2}(S^{-1}(\tilde{\beta}u^*))^2\right\}. \tag{13}$$

Substituting $S^{-1}(\tilde{\beta}u^*) = \left\{2\log(1/(\tilde{\beta}u^*)) - \log\log(1/(\tilde{\beta}u^*)) - r(\tilde{\beta}u^*)\right\}^{1/2}$ into (13) yields

$$\begin{aligned} u^* &= \log\left(\frac{1}{\tilde{\beta}u^*}\right)\frac{1}{2\pi}\frac{\tilde{\beta}}{(\beta-1)^2}e^r\theta_m^2 \\ &\geq \frac{1}{2\pi}\frac{1}{(\beta-1)^2}\theta_m^2. \end{aligned} \qquad (14)$$

From this it follows that

$$\log\left(\frac{1}{u^*}\right) \leq \log(2\pi(\beta-1)^2) + 2\log\left(\frac{1}{\theta}\right).$$

Hence,

$$\begin{aligned} \theta_m S^{-1}(\tilde{\beta}u^*) &\leq \theta_m S^{-1}(u^*) \\ &\leq \theta_m \left\{2\log(1/u^*)\right\}^{1/2} \\ &\leq \theta_m \left\{2\log(2\pi(\beta-1)^2) + 4\log\left(\frac{1}{\theta}\right)\right\}^{1/2} \\ &\to 0 \end{aligned}$$

as $\theta_m \to 0$. It follows that $\beta - f_m(u^*) \to \beta - 1$. From (14) it follows that $\sqrt{m}\epsilon_m u^*(\beta - f_m(u^*)) \to \infty$ and Hoeffding's inequality applies. Examining the formulae for $b_m$ and $a_m$, we see that both $b_m = \sqrt{m\log m}$ and $a_m = \sqrt{m\log m}$.

The interpretation of FDR as being intermediate between uncorrected testing and Bonferroni still holds in the local case.

LEMMA 2. *Under the conditions of Theorem 2, we have, for all large $m$, that $\alpha/m \leq u^* \leq \alpha$.*

PROOF. The proof for the upper bound is the same as before. The lower bound holds due to (14) and the form of $\theta_m$.

# 4 Asymptotic Marginal Risk

Let $\delta = (\delta_1, \ldots, \delta_m)$ where $\delta_i = 1$ if $H_{1i}$ is true and $\delta_i = 0$ if $H_{0i}$ is true. Let $\hat{\delta} = (\hat{\delta}_1, \ldots, \hat{\delta}_m)$ where $\hat{\delta}_i = 1$ if $H_{0i}$ is rejected and $\hat{\delta}_i = 0$ if $H_{0i}$ is accepted. Define the (marginal) average risk $R_m$ to be

$$R_m = \frac{1}{m}\mathsf{E}\left(\sum_{i=1}^{m} |\delta_i - \hat{\delta}_i|\right).$$

Thus, this risk is the average fraction of errors in both directions, i.e. false positives and false negatives.

THEOREM 3. *Let*

$$R_F = A_0 u^* + A_1\left[1 - F\left(u^*\right)\right] = A_0 u^* + A_1\left[1 - \beta u^*\right]. \tag{15}$$

*Then,* $R_m \sim R_F$ *as* $m \to \infty$.

PROOF. The procedure rejects a p-value $P_i$ if and only if $P_i \leq P_{(D)}$. Also, $P_i \leq P_{(D)}$ if and only if $P_i \leq \ell(D) = D\alpha/m$. Now,

$$\begin{aligned}
R_m &= \frac{1}{m}\mathsf{E}\left(\sum_{i=1}^{m} |\delta_i - \hat{\delta}_i|\right) \\
&= \frac{1}{m}\mathsf{E}\left(\sum_{i \in S_0} I(P_i < \ell(D)) + \sum_{i \in S_1} I(P_i > \ell(D))\right) \\
&= \frac{m_0}{m}\mathsf{P}\{\, P_0 < \ell(D)\,\} + \frac{m_1}{m}\mathsf{P}\{\, P_1 > \ell(D)\,\} \qquad (16) \\
&= A_0\mathsf{P}\{\, P_0 < \ell(D)\,\} + A_1\mathsf{P}\{\, P_1 > \ell(D)\,\} \qquad (17)
\end{aligned}$$

where $P_0$ represents a p-value under $H_0$ and $P_1$ represents a p-value under $H_1$. The result follows from Corollary 1.

## 4.1 Risk Comparisons

We now consider two other procedures. Uncorrected testing rejects whenever $P_i < \alpha$. The risk of uncorrected testing is

$$R_U = A_0\alpha + A_1(1 - F(\alpha)).$$

The well-known Bonferroni method rejects when $P_i < \alpha/m$. It has risk

$$R_B = A_0 \frac{\alpha}{m} + A_1 \left[ 1 - F\left( \frac{\alpha}{m} \right) \right].$$

We are interested in seeing when one method dominates. Indeed, we have noted empirically that $R_F$ seems to do quite well in many cases even though it was not designed expressly for this purpose. We also consider the optimal Bayes classification rule obtained by rejecting all $P_i < c$ and minimizing the risk $A_0 c + A_1(1 - F(c))$ over $c$. The solution is to choose $c$ to satisfy $f(c) = A_0/(1 - A_0)$. Unlike FDR, uncorrected testing and Bonferroni, this rule would require knowing $f$ and so is in a different spirit than the above rules. Nonetheless it makes an interesting baseline comparison.

Using the asymptotic formulae for risk with $m \to \infty$, we can see first that BH dominates Bonferroni, i.e. $R_B - R_F > 0$ precisely when

$$\begin{aligned} \frac{A_0}{1 - A_0} &< \frac{F(\alpha/m) - F(u^*)}{(\alpha/m) - u^*} \\ &\sim \frac{F(u^*)}{u^*} \\ &= \beta = \frac{(1/\alpha) - A_0}{1 - A_0} \end{aligned}$$

which corresponds to $A_0 < 1/(2\alpha)$. This always holds when $\alpha < 1/2$. Thus Bonferroni is not competitive and we do not consider it further.

Similarly, $R_U - R_F > 0$ when

$$\frac{A_0}{1 - A_0} > \frac{F(\alpha) - F(u^*)}{\alpha - u^*}$$

so neither dominates. In the local case, with $\theta_m \to 0$, $u^* \to 0$ so the condition becomes

$$\frac{A_0}{1 - A_0} > \frac{F(\alpha)}{\alpha}.$$

To better understand the comparisons, we now consider some examples. Figure 2 shows the risks of FDR, uncorrected testing and the Bayes risk in the Gaussian case for $\alpha = .05$ and various $\theta$'s. Note that the risk functions do not depend on $m$. We see that uncorrected testing dominates for small $\theta$ but that FDR nearly mimics the optimal Bayes rule for large $\theta$. This is rather remarkable given that FDR uses no knowledge of $F$. Figure 3 shows similar plots for the local case $\theta_m = 5(\log m/m)^{1/4}$ for various $m$'s. Here see that uncorrected and FDR are close, and for $A_0$ larger than some constant, both are close to the Bayes risk. To check the accuracy of the asymptotic approximations, we computed the FDR risk from the asymptotic formula and from simulation. Figure 4 shows some results for $m = 100$. We see that even at this $m$ the approximations are accurate. Similar results hold for the local case (not shown here).

# 5 The False Non-discovery Rate

Next we bound the expected False Negative Rate (FNR) which we defined in equation (2) as the proportion of false non-rejections among those tests whose null is not rejected.

THEOREM 4. *Under the conditions of Theorem 1,*

$$\mathsf{E}\,(\mathrm{FNR}) \sim A_1[1 - \beta u^*].$$

PROOF. From the definition,

$$\mathrm{FNR} = \frac{\sum_{i \in S_1} I(P_i > \ell(D))}{m - D}.$$

The result follows from an argument similar to that given in Theorem 3.

These definitions lead the the following natural question: find the procedure to minimize $\mathsf{E}\,(\mathrm{FNR})$ subject to $\mathsf{E}\,(\mathrm{FDR}) \leq \alpha$. Now we investigate

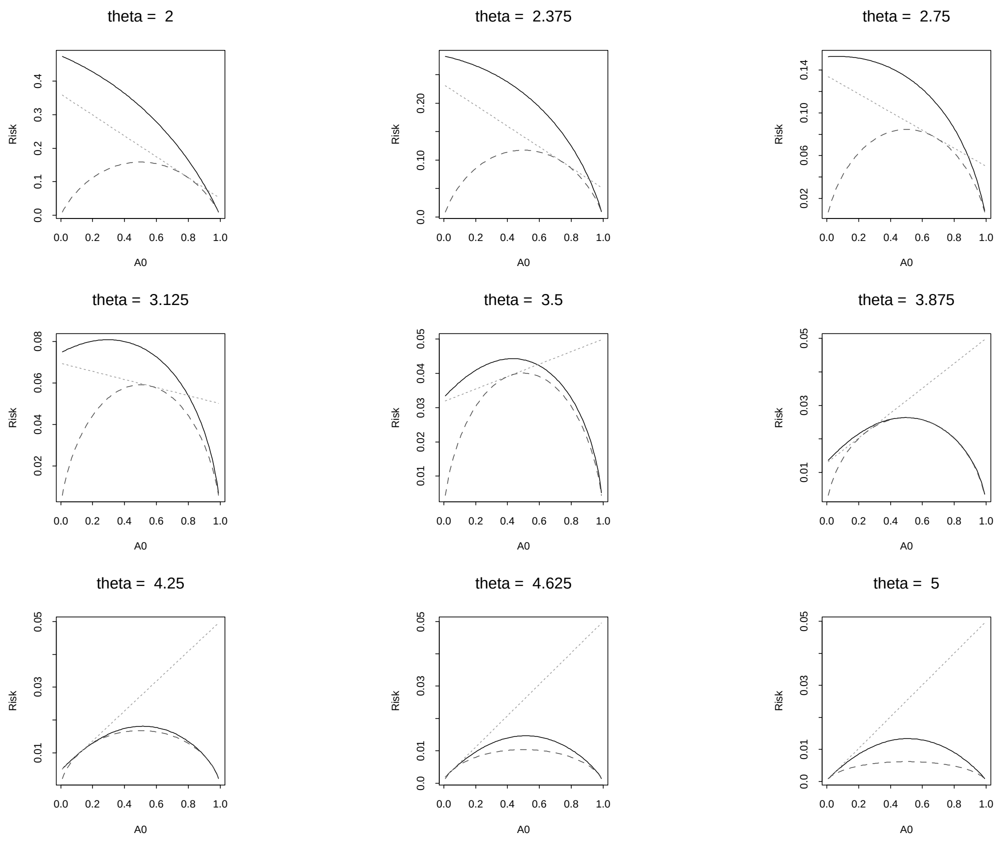


**Figure 2.** Marginal risk comparisons. Solid line is FDR, dotted is uncorrected and dashed is optimal Bayes risk.

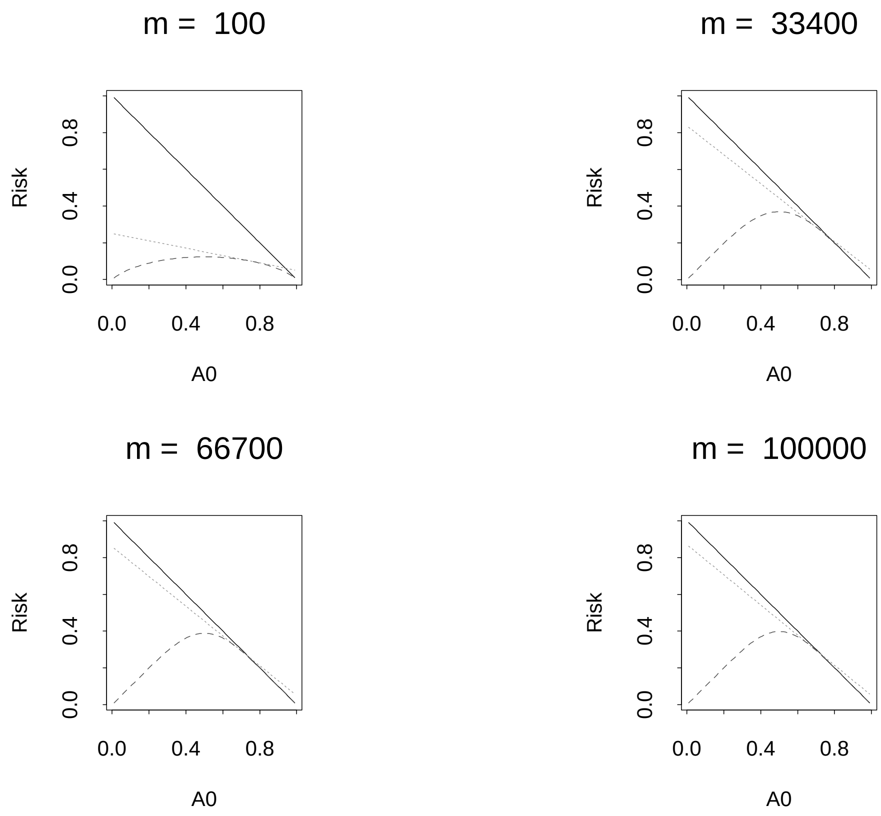


**Figure 3.** Marginal risk comparisons, local alternatives. Solid line is FDR, dotted is uncorrected and dashed is optimal Bayes risk.

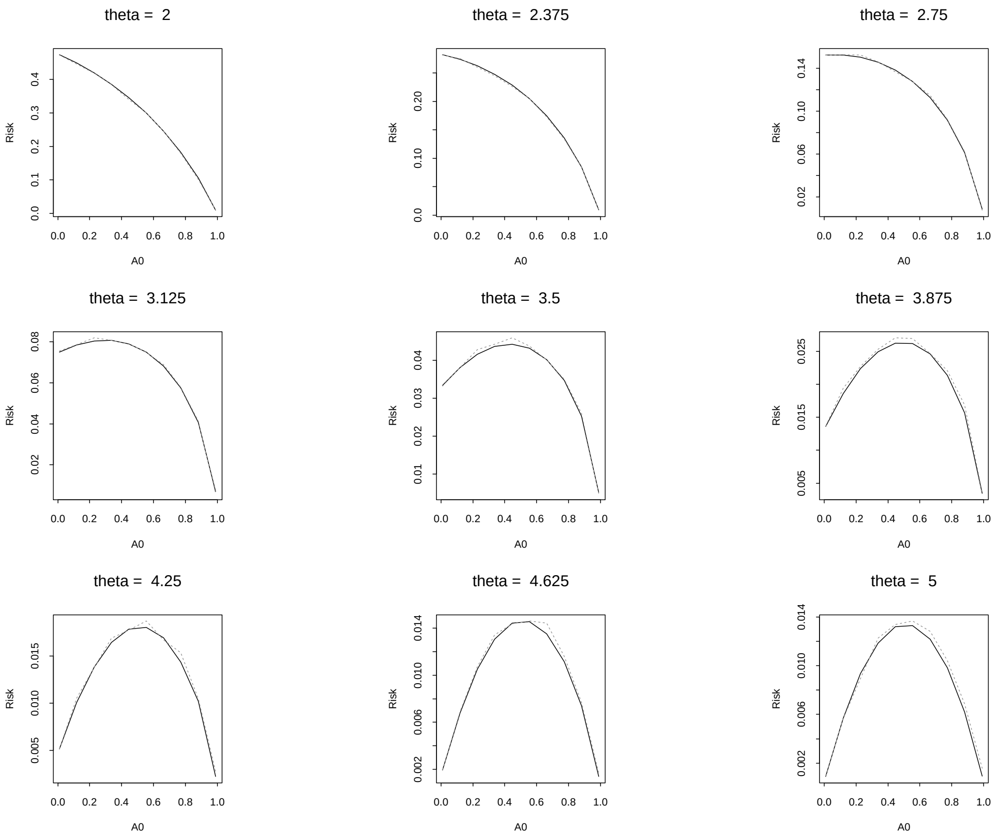


**Figure 4.** Comparison of asymptotic risk (for FDR) versus exact risk computed from 1000 simulations with m=100. Solid line is asymptotic, dotted line is exact.

this question using asymptotic approximations. We have seen that asymptotically, the BH procedure rejects p-values less than $u^*$. More generally, consider the procedure that rejects when $P_i < c$ for a constant $c$. Then,

$$\mathrm{FDR} = \frac{\sum_{i \in S_0} I(P_i < c)}{\sum_{i \in S_0} I(P_i < c) + \sum_{i \in S_1} I(P_i < c)}.$$

Now, $\sum_{i \in S_0} I(P_i < c) \sim \mathrm{Binomial}(m_0, c)$ and $\sum_{i \in S_1} I(P_i < c) \sim \mathrm{Binomial}(m_1, F(c))$. Hence,

$$\mathsf{E}\,(\mathrm{FDR}) = \frac{A_0 c}{A_0 c + A_1 F(c)} + O\left(\frac{1}{\sqrt{m}}\right).$$

Ignoring the approximation error, $c$ will satisfy $\mathsf{E}\,(\mathrm{FDR}) \leq \alpha$ if

$$\frac{F(c)}{c} \geq \frac{1-\alpha}{\alpha}\frac{A_0}{1-A_0} = \beta - \frac{1}{\alpha}. \tag{18}$$

It is easy to check that $u^*$ satisfies (18) for all $F$ and $A_0$ which confirms that the BH procedure satisfies that FDR property in a distribution free way. In fact, it is worth noting that

$$\frac{A_0 u^*}{A_0 u^* + A_1 F(u^*)} = \frac{m_0}{m}\alpha.$$

Now,

$$\mathrm{FNR} = \frac{\sum_{i \in S_1} I(P_i > c)}{\sum_{i \in S_1} I(P_i > c) + \sum_{i \in S_0} I(P_i > c)}$$

and hence

$$\mathsf{E}\,(\mathrm{FNR}) = \frac{A_1(1-F(c))}{A_1(1-F(c)) + A_0(1-c)} + O\left(\frac{1}{\sqrt{m}}\right).$$

Again, we ignore the $O(m^{-1/2})$ term in what follows. Now, the concavity of $F$ implies that $\mathsf{E}\,(\mathrm{FDR})$ is decreasing in $c$. Hence, minimizing $\mathsf{E}\,(\mathrm{FNR})$ subject to $\mathsf{E}\,(\mathrm{FDR}) \leq \alpha$ corresponds to choosing $c$ such that $\mathsf{E}\,(\mathrm{FDR}) = \alpha$ which implies that the optimal $c$, denoted by $c^*$ satisfies

$$\frac{F(c^*)}{c^*} = \beta - \frac{1}{\alpha}. \tag{19}$$

Note that

$$\frac{F(u^*)}{u^*} - \frac{F(c^*)}{c^*} = \beta$$

which thus represents, in some sense, the price paid for the distribution free property of the BH method. To measure this on the scale of $\mathsf{E}\,(\mathrm{FNR})$, let $\mathsf{E}\,(c)$ denote the expected FNR as a function of $c$. Figure 5 shows plots of $\mathsf{E}\,(u^*) - \mathsf{E}\,(c^*)$.

Clearly the BH method does not attain minimal $\mathsf{E}\,(\mathrm{FNR})$. It is an open question whether there exists another distribution free method that does attain lower $\mathsf{E}\,(\mathrm{FNR})$.

The BH procedure sets its decision point by using the last crossing (from the left) of the p-values under the line $\ell(t) = t\alpha/m$. Using the asymptotic approximation for $\mathsf{E}\,(\mathrm{FDR})$ above, we can give a heuristic exploration of other last-crossing procedures, each corresponding to a different function $\ell(t)$. Let $\ell(t) = r(t)/m$ for some function $r$ and define

$$D = \max\{i :\ P_{(i)} < \ell(i)\}.$$

The BH method corresponds to $r(t) = \alpha t$. The large deviation argument in Section 2 can applied with this more general $\ell(t)$ replacing the linear form used in Section 2. The bounds $a_m$ and $b_m$ take the same form but with $u^*$ now satisfying

$$\frac{F(u^*)}{u^*} = \frac{\frac{t}{r(t)} - A_0}{1 - A_0}$$

where $t$ satisfies $m_0 s(t) + m_1 F(s(t)) = t$. Hence $t$ implicitly depends on $F$. From the arguments before, the property $\mathsf{E}\,(\mathrm{FDR}) \leq \alpha$ requires that

$$\frac{F(u^*)}{u^*} \geq \frac{1-\alpha}{\alpha}\frac{A_0}{1-A_0}$$

which implies that

$$\frac{\frac{t}{r(t)} - A_0}{1 - A_0} \geq \frac{1-\alpha}{\alpha}\frac{A_0}{1-A_0}.$$

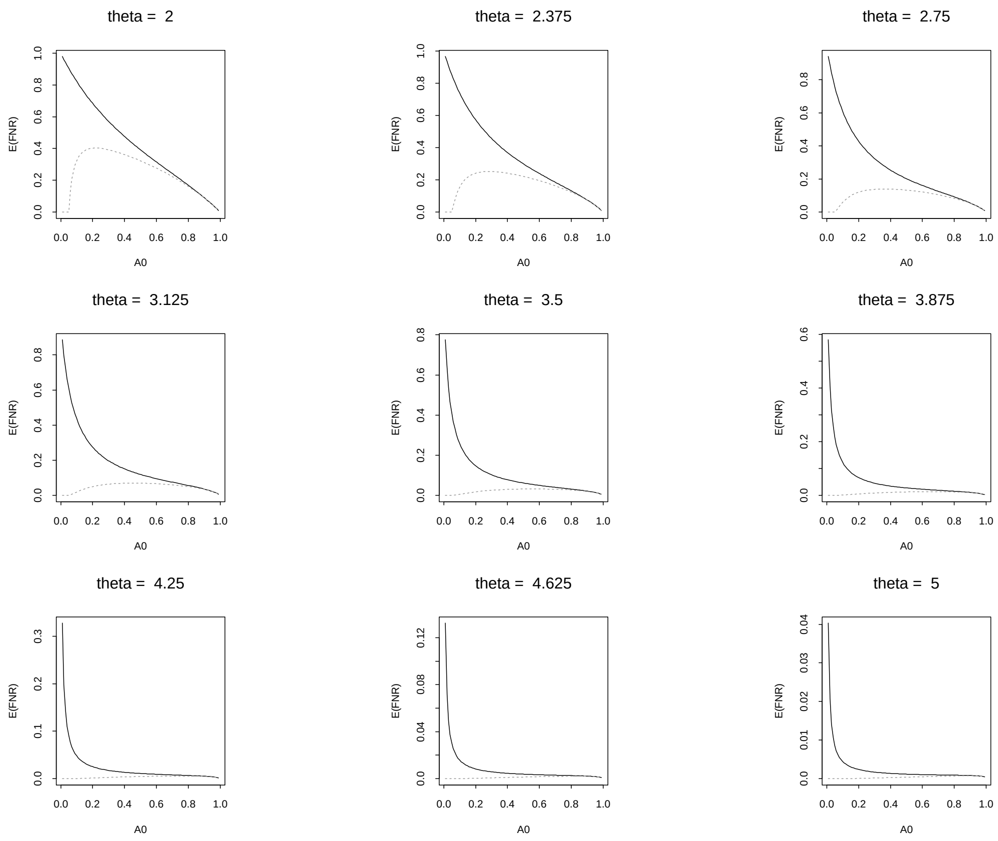


**Figure 5.** Expected false negative rate. Solid line is Benjamini-Hochberg method; dotted line is optimal.

This holds for all $A_0$ only if

$$\frac{r(t)}{t} \leq \alpha,$$

in other words, if $r(t) \leq t\alpha$. To minimize $\mathsf{E}\,(\text{FNR})$ requires moving $D$ to the right as much as possible which implies setting $r(t) = t\alpha$. Thus, it appears that among all last-crossing procedures, the BH method is optimal.

# 6 Combining FDR and FNR and Minimizing Conditional Risk

So far we have focused on a particular risk function and, separately, on $\mathsf{E}\,(\text{FNR})$. To gain further insight, we consider a different risk function in this section that combines FDR and FNR. The loss function is of the form FNR$+\lambda\,$FDR. The constant $\lambda$ is taken here to be user specified although one could consider data based methods for choosing $\lambda$. We will also introduce here an appealing notion of conditional risk and we will find a procedure that minimizes both conditional and unconditional risk. The calculations in this section are exact, finite sample results.

We maintain the assumption of the previous sections that both the null and alternative hypotheses are simple with p-value densities $f_0 \equiv 1_{[0,1]}$ and $f_1 \equiv f$. We also initially take $m_0$ as fixed and known. Under these assumptions, the decision problem in multiple testing can be viewed as that of assigning proper labels (null or alternative) to the sorted but unlabeled p-values. Loosely speaking, the p-values from the two groups of tests are generated and interleaved, but the group labels are unobserved. The group labeling is consequently random, even under strictly frequentist assumptions.

Let $\boldsymbol{p}$ denote the vector of sorted p-values. Let $G_i$ be the indicator that $p_i = P_{(i)}$ corresponds to a test with a false null and let $\boldsymbol{G}$ denote the vector

$(G_1, \ldots, G_m)$. We consider procedures that reject the null hypotheses with the smallest $r(\boldsymbol{p})$ p-values. The BH procedure is in this class. Notice that both FDR and FNR are determined by $\boldsymbol{G}$ and $r(\boldsymbol{p})$. Let $\mathsf{E}_r$ and $\mathsf{P}_r$ denote expected values and probabilities under the procedure $r$. Also, define the correct non-discovery rate $\mathrm{CNR} = 1 - \mathrm{FNR}$.

In this section, we consider a family of risk functions parametrized by $\lambda > 0$. Define the loss function $L_\lambda$ by

$$L_\lambda(\boldsymbol{G}, \boldsymbol{p}) = \mathrm{FNR}(\boldsymbol{G}, r(\boldsymbol{p})) + \lambda\, \mathrm{FDR}(\boldsymbol{G}, r(\boldsymbol{p})),$$

and corresponding risk

$$R_\lambda(m_0, r) = \mathsf{E}_r L_\lambda(\boldsymbol{G}, r(\boldsymbol{p})) \tag{20}$$

where the expectation is taken with respect to $\boldsymbol{G}$ and $\boldsymbol{p}$ under the procedure $r$. As noted earlier, we view $\lambda$ as user-specified. However, another motivation for this risk measure is as a means for solving the optimization problem of minimizing the $\mathsf{E}_r$FNR subject to the bound $\mathsf{E}_r\mathrm{FDR} \leq \alpha$. We can view $\lambda$ as the Lagrange multiplier in this problem maximizing the expected FNR subject to a bound on the expected FDR.

Now we define the conditional risk

$$R_\lambda(m_0, r \mid \boldsymbol{p}) = \mathsf{E}_r(L_\lambda(\boldsymbol{G}, r) \mid \boldsymbol{p}). \tag{21}$$

In this case, we can take $r$ to be an integer in $\{0, \ldots, m\}$. Note that we can write $p = (\boldsymbol{p}, \pi)$ where $\pi$ is the permutation that sorts the p-values. In the conditional risk, we are conditioning on $\boldsymbol{p}$ and the randomness is in $\pi$. This conditional risk is appealing in the usual sense of conditional inference (Kiefer 1977, Berger, Brown and Wolpert 1994) in which one conditions on meaningful functions of the data, here, the sorted p-values. It answers the question: what is the optimal procedure given the sorted p-values? Alternatively, minimizing the conditional risk may be seen simply as a strategy for minimizing the unconditional risk. That is, define an unconditional procedure $r_*(\boldsymbol{p};\lambda)$ to be the $r$ that minimizes the conditional risk for each $\boldsymbol{p}$. With this choice,

$$R_\lambda(m_0, r_* \mid \boldsymbol{p}) \leq R_\lambda(m_0, r \mid \boldsymbol{p}),$$

for any other $r$, so $r_*$ also minimizes the unconditional risk. We can easily choose $\lambda$ to bound $\mathsf{E}_r(\mathrm{FDR} \mid \boldsymbol{p})$ by $\alpha$ as tightly as possible, but the unconditional choice of $\lambda$ is more complicated.

We now derive the conditional risk minimizing procedure. Using the notation in Table 1, we can write

$$\begin{aligned} N_{0|1} &= m_1 - N_{1|1} \\ &= m_1 - R + N_{1|0} \\ &= m - R - (m_0 - N_{1|0}). \end{aligned}$$

Hence, when $R < m$,

$$\begin{aligned} \mathrm{FNR} &= \frac{N_{0|1}}{m-R} \\ &= 1 - \frac{m_0 - N_{1|0}}{m-R} \\ &= 1 - \frac{A_0 - (1/m)\, N_{1|0}}{1 - R/m} \\ &= 1 - \frac{A_0 - (R/m)\,\mathrm{FDR}}{1 - R/m}. \end{aligned}$$

Working under the procedure $r(\boldsymbol{p})$ and conditionally on $\boldsymbol{p}$ enables us to write $\mathsf{E}_r(\mathrm{CNR} \mid \boldsymbol{p}) = 1$ if $r = m$ and

$$\mathsf{E}_r(\mathrm{CNR} \mid \boldsymbol{p}) = \frac{A_0 - (r/m)\,\mathsf{E}_r(\mathrm{FDR} \mid \boldsymbol{p})}{1 - r/m} = \frac{A_0 - (1/m)\,\mathsf{E}_r(N_{1|0} \mid \boldsymbol{p})}{1 - r/m}, \quad (22)$$

if $r < m$. The $m+1$ quantities $\mathsf{E}_r(N_{1|0} \mid \boldsymbol{p})$ over $r = 0, \ldots, m$ determine the conditional risk since we can now write

$$R_\lambda(m_0, r|\boldsymbol{p}) = 1 - \frac{A_0 - \frac{1}{m}\mathsf{E}_r(N_{1|0}|\boldsymbol{p})}{1 - \frac{r}{m}} + \frac{\lambda}{r}\mathsf{E}_r(N_{1|0}|\boldsymbol{p}), \tag{23}$$

for $0 < r < m$. Hence, minimizing the risk only requires computing $\mathsf{E}_r(N_{1|0} \mid \boldsymbol{p})$.

Let $\mathcal{S}$ be the set of labeling vectors (vectors of ones and zeroes) of length $m$ with $m_0$ 0s, and let $\mathcal{S}_{r,k}$ be the subset of $\mathcal{S}$ with exactly $k$ 0s in the first $r \equiv r(\boldsymbol{p})$ positions. Then,

$$\begin{aligned}
\mathsf{P}_r\Big\{ N_{1|0} = k \mid \boldsymbol{p} \Big\} &= \sum_{s \in \mathcal{S}_{r,k}} \mathsf{P}_r\{ \boldsymbol{G} = s \mid \boldsymbol{p}, m_0 \} && (24)\\
&= \frac{\sum_{s \in \mathcal{S}_{r,k}} m_0! \prod_{i\,:\,s_i=0} f_0(p_i)\; m_1! \prod_{j\,:\,s_j=1} f_1(p_j)}{\sum_{s \in \mathcal{S}} m_0! \prod_{i\,:\,s_i=0} f_0(p_i)\; m_1! \prod_{j\,:\,s_j=1} f_1(p_j)} && (25)\\
&= \frac{\sum_{s \in \mathcal{S}_{r,k}} \prod_{j\,:\,s_j=1} f(p_j)}{\sum_{s \in \mathcal{S}} \prod_{j\,:\,s_j=1} f(p_j)} && (26)
\end{aligned}$$

Evaluating this expression is cumbersome. Now we develop formula that is much easier to evaluate.

Define the following generating functions:

$$\begin{aligned}
V_r(z) &= \prod_{i=1}^{r} (z + f(p_i)) \\
W_r(z) &= \prod_{i=r+1}^{m} (z + f(p_i)) \\
U(z) &= V_r(z)\, W_r(z),
\end{aligned}$$

where $f \equiv f_1$ is the p-value density under the alternative. (Note that $z + f(u) = zf_0(u) + f_1(u)$ since $f_0(u) \equiv 1$.) For a polynomial $P$ in $z$, let $\Big[P(z)\Big]_{z^k}$ denote the coefficient of $z^k$.

THEOREM 5. *For each $r$ we have*

$$\mathsf{E}_r(N_{1|0} \mid \boldsymbol{p}) = \Big[U(z)\Big]_{z^{m_0}}^{-1} \sum_{k=0}^{m_0-1} \Big[U(z)\Big]_{z^k} (-1)^{m_0-k-1} \sum_{i=1}^{r} (f(p_i))^{-(m_0-k)}. \quad (27)$$

PROOF. We have that

$$\frac{\displaystyle\sum_{s\in\mathcal{S}_{r,k}} \prod_{j\,:\,s_j=1} f(p_j)}{\displaystyle\sum_{s\in\mathcal{S}} \prod_{j\,:\,s_j=1} f(p_j)} = \frac{\Big[V_r(z)\Big]_{z^k} \cdot \Big[W_r(z)\Big]_{z^{m_0-k}}}{\Big[U(z)\Big]_{z^{m_0}}} \quad (28)$$

$$= \frac{\Big[V_r(z)\Big]_{z^k} \cdot \Big[W_r(z)\Big]_{z^{m_0-k}}}{\displaystyle\sum_{k=0}^{m} \Big[V_r(z)\Big]_{z^k} \cdot \Big[W_r(z)\Big]_{z^{m_0-k}}}. \quad (29)$$

Hence,

$$\begin{aligned}
\mathsf{E}_r(N_{1|0} \mid \boldsymbol{p}, m_0) &= \sum_{k=0}^{m} k \frac{\Big[V_r(z)\Big]_{z^k} \cdot \Big[W_r(z)\Big]_{z^{m_0-k}}}{\Big[U(z)\Big]_{z^{m_0}}} \\
&= \frac{\displaystyle\sum_{k=0}^{m} k \, \Big[V_r(z)\Big]_{z^k} \cdot \Big[W_r(z)\Big]_{z^{m_0-k}}}{\displaystyle\sum_{k=0}^{m} \Big[V_r(z)\Big]_{z^k} \cdot \Big[W_r(z)\Big]_{z^{m_0-k}}} \\
&= \frac{\Big[zV_r'(z)W_r(z)\Big]_{z^{m_0}}}{\Big[V_r(z)W_r(z)\Big]_{z^{m_0}}}.
\end{aligned}$$

The last equality follows from the fact that $\Big[zV_r'(z)\Big]_{z^k} = k \, \Big[V_r(z)\Big]_{z^k}$ and by the fact that the sum in the numerator is a convolution of generating functions. Notice that

$$zV_r'(z)W_r(z) = \sum_{i=1}^{r} z \prod_{\substack{j=1 \\ i\neq j}}^{m} (z + f(p_j)) \equiv \sum_{i=1}^{r} U_i(z).$$

Using the additivity of the coefficient operator, we have

$$\mathsf{E}_r(N_{1|0} \mid \boldsymbol{p}) = \frac{\displaystyle\sum_{i=1}^{r} \Big[U_i(z)\Big]_{z^{m_0}}}{\Big[U(z)\Big]_{z^{m_0}}}. \quad (30)$$

Now,

$$\Big[U_i(z)\Big]_{z^{m_0}} = \left[\frac{z/f(p_i)}{1+z/f(p_i)}\,U(z)\right]_{z^{m_0}} = -\sum_{k=0}^{m_0-1}\Big[U(z)\Big]_{z^k}\left(\frac{-1}{f(p_i)}\right)^{m_0-k},$$

and hence,

$$\mathsf{E}_r(N_{1|0}\mid \boldsymbol{p}) = \Big[U(z)\Big]_{z^{m_0}}^{-1}\sum_{k=0}^{m_0-1}\Big[U(z)\Big]_{z^k}(-1)^{m_0-k-1}\sum_{i=1}^{r}(f(p_i))^{-(m_0-k)}. \quad (31)$$

There are two possible paths towards solving the optimization problem. First, we can compute $\Big[U_i(z)\Big]_{z^{m_0}}$ for each $i = 1,\ldots,m-1$ and construct $\mathsf{E}_r(N_{1|0}\mid \boldsymbol{p}, m_0)$ for each $r$ accordingly, using equation (30). Second, we could use (31) which reduces the problem to finding the first $m_0{+}1$ coefficients of $U$. For small $m$, we can find the coefficients of $U$ by direct termwise recurrence. For very large $m$, it might be more effective to compute the coefficients with Gauss-Legendre quadrature of an appropriate order.

Let $g(r) = \mathsf{E}_r(N_{1|0}\mid \boldsymbol{p})$ for $0 \le r \le m$. Then, for $r = 0,\ldots,m$,

$$R_\lambda(m_0, r\mid \boldsymbol{p}) = \begin{cases} A_1 & \text{if } r = 0\\ 1-\frac{m_0}{m-r}+\left(\frac{1}{m-r}+\frac{\lambda}{r}\right)g(r) & \text{if } 0<r<m\\ \lambda A_0 & \text{if } r = m.\end{cases}$$

The comparison between the BH and $R_\lambda$-minimizing procedures is revealing. We consider here the simple case where $f_\theta$ is a $N(\theta,1)$, and for each $m_0$ and $\theta$, we simulate the sorted p-values from $m = 100$ tests. Figure 6 shows the risks as a function of $A_0$ for $\lambda = 1$. Also included in the comparison is the "naive" rule that rejects the $m_1$ hypotheses corresponding to the smallest p-values. Figure 7 shows the mean values of $D/m$ for the three procedures. The differences among the procedures is striking and is reminiscent of the comparisons above between BH and the Bayes classifier in Figure 3. Note in particular that the BH procedure performs nearly optimally for large $A_0$.

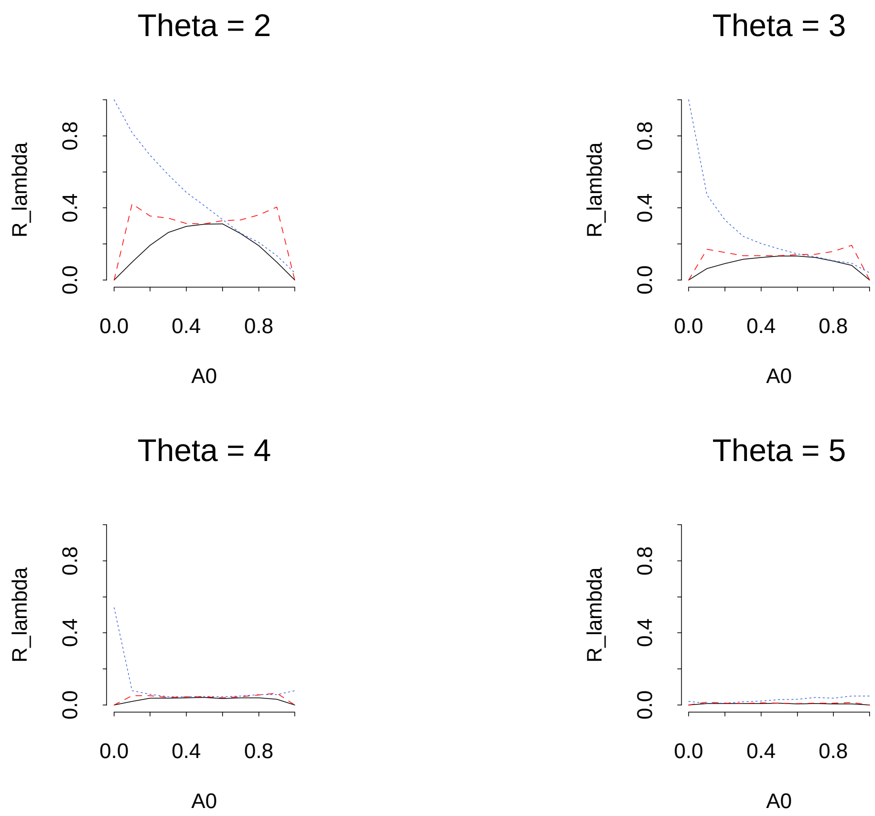


**Figure 6.** $R_\lambda$ profiles for three methods: (Solid line) $R_\lambda$-minimizer; (Dotted line) BH method; (Dashed line) the naive rule of rejecting $m_1$ hypotheses. Plots generated from simulated data with $m = 100$ over a range of $m_0$ and $\theta$.

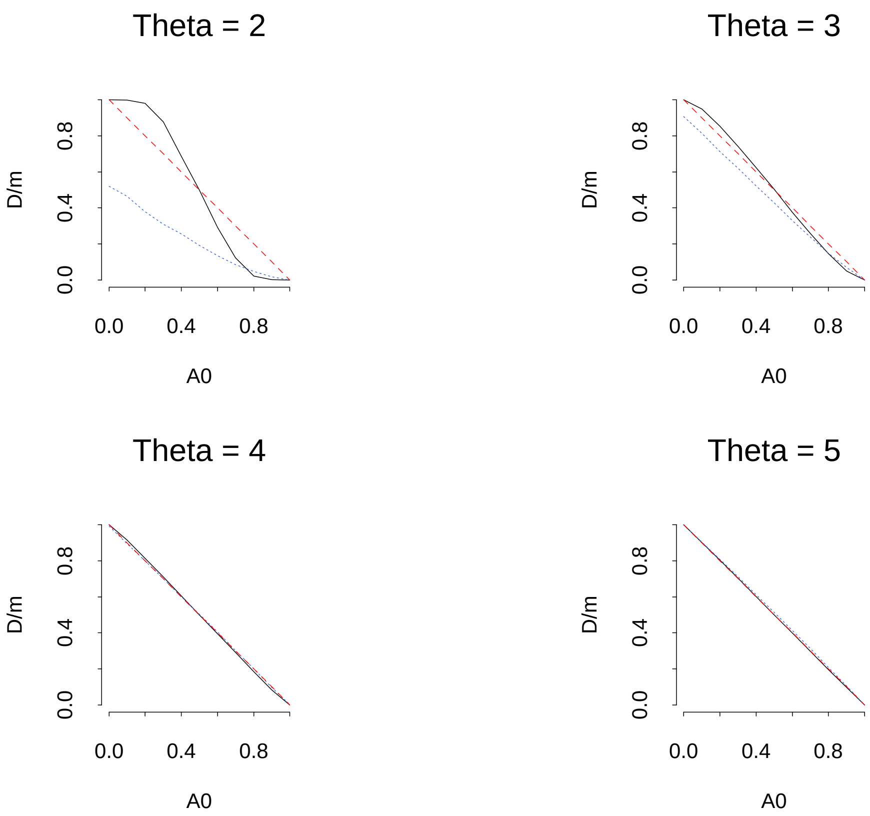


**Figure 7.** Mean $D/m$ for three methods: (Solid line) $R_\lambda$-minimizer; (Dotted line) BH method; (Dashed line) the naive rule of rejecting $m_1$ hypotheses. Plots generated from simulated data with $m = 100$ over a range of $m_0$ and $\theta$.

The systematic deviation of deciding point for the $R_\lambda$ minimizer from $m_1$ accounts for the improved risk over $A_0$ away from 0, 0.5, or 1.

It is also interesting to look at the marginal expected FDR from the $R_\lambda$-minimizing procedure as a function of $\lambda$. As shown in Figure 8, this decreases as $\lambda$ increases because the $\mathsf{E}\,(\mathrm{FDR})$ component gets greater weight.

If $m_0$ is not known but we have a prior distribution $\pi$ on $m_0$, then we can construct $r_*(\boldsymbol{p}, m_0)$ to be the minimizer of conditional Bayes risk $\mathsf{E}_\pi R_\lambda(m_0, r \mid \boldsymbol{p}, m_0)$. The calculations above all carry through for each $m_0$ where all the conditionals now include $m_0$ as well as $\boldsymbol{p}$. The risk function is obtained by averaging the risk functions for each $m_0$ with respect to the prior $\pi$. The conditionally optimal choice $r_*$ again minimizes the unconditional Bayes risk, which is $\mathsf{E}_\pi \mathsf{E}_r R_\lambda(m_0, r \mid \boldsymbol{p}, m_0)$ in this case. Figure 9 compares the risk functions given the true $m_0 = 80$ (i.e., a prior that is a point mass at 80) and given a prior that is uniform over the range of $m_0$ from 50 to 100. The latter is a reasonable default prior since in practice one usually has a reasonable idea of whether or not $A_0 > 1/2$. Although the risk for the uniform prior is less sharply peaked, it nonetheless has a minimum near the minimum for the point-mass prior.

We now show how these calculations are useful from a Bayesian perspective. Let $H_r$ be the hypotheses corresponding to the $r$ smallest p-values. Let $\mathrm{FDR}(H_r)$ be the *realized* FDR for this set and define $\mathrm{FNR}(H_r)$ similarly. From a Bayesian perspective, since $\mathrm{FDR}(H_r)$ and $\mathrm{FNR}(H_r)$ are functions of the random $\boldsymbol{G}$, there is a posterior distribution $\pi(\mathrm{FDR}(H_r) \mid \boldsymbol{p}, m_0)$ and $\pi(\mathrm{FNR}(H_r) \mid \boldsymbol{p}, m_0)$. One could approximate these posteriors by simulation. However, our previous calculations allow for exact calculations. For example,

$$E_\pi(FDR(H_r) \mid \boldsymbol{p}, m_0) \quad = \quad E(N_{1|0} \mid \boldsymbol{p}, m_0)/r$$

and

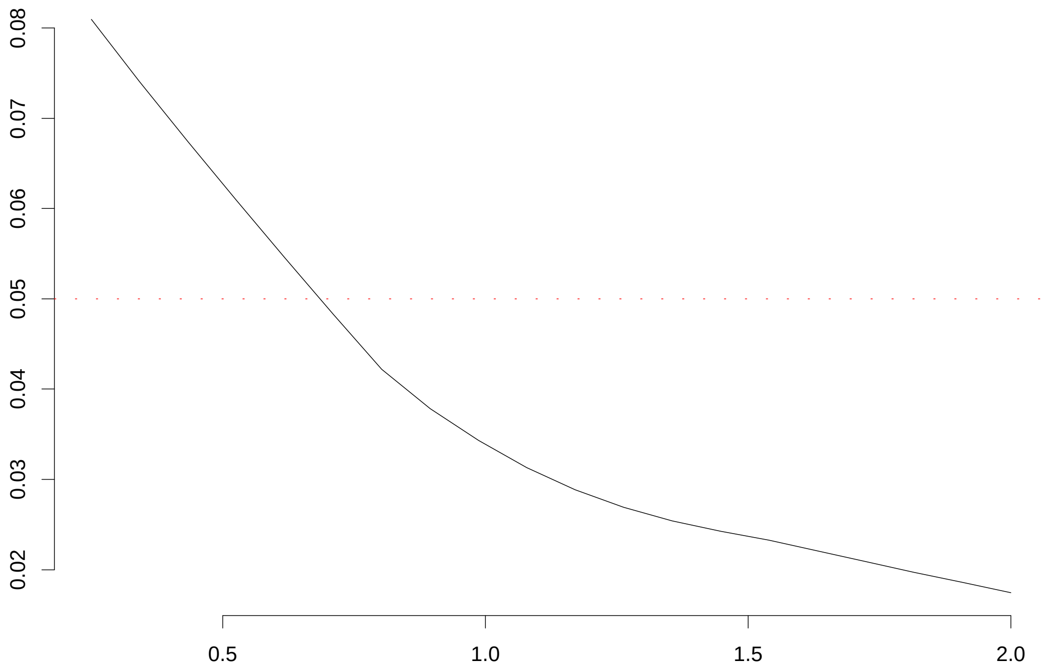


**Figure 8.** $\mathsf{E}\,(FDR)$ as a function of $\lambda$ for the $R_\lambda$-minimizing procedure. Based on Normal data with $m = 100$, $m_0 = 80$, and $\theta = 3$.

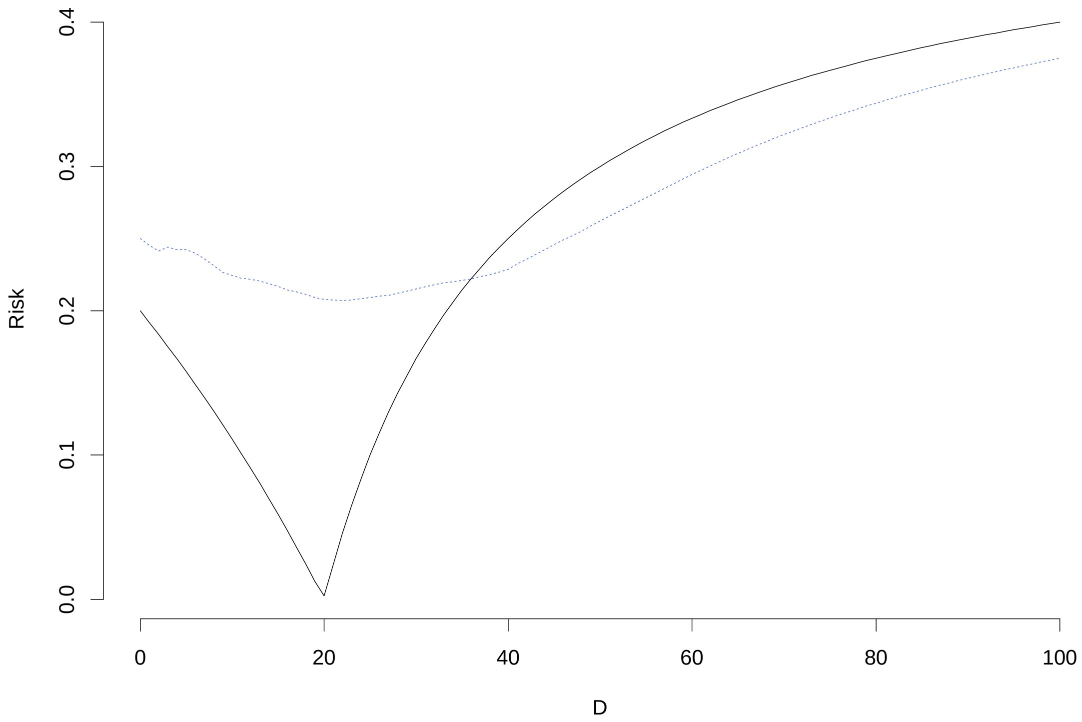


**Figure 9.** Comparison of conditional Bayes risk under two priors on $m_0$: (Solid Line) Point mass prior at the truth; (Dotted Line) Uniform from 50 to 100. The data were generated in the Normal case with $\theta = 5$ and the true $m_0 = 80$.

$$E_\pi(\mathrm{FNR}(H_r) \mid \boldsymbol{p}, m_0) \;=\; 1 - \frac{A_0 - \frac{1}{m}E(N_{1|0} \mid \boldsymbol{p}, m_0)}{1 - \frac{r}{m}}.$$

These expressions can be integrated with respect to a prior on $m_0$ and $f$ to obtain a fully Bayesian posterior estimate. The generating function approach immediately extends to compute factorial moments from which any higher moments can be determined. We can then derive the full posterior for $N_{1|0}$ by inverting the moments. Alternatively, the first two moments can be used to get a normal approximation, or we can get the posterior cdf of $N_{1|0}$ by direct calculation with the generating functions. A full development of a Bayesian version of FDR is beyond the scope of this paper; we shall report on it elsewhere.

# 7 Composite Alternatives

So far we have focused on the "simple versus simple" case in which, the p-values corresponding to hypotheses in which the null does not hold have a common distribution $F$. We now discuss extensions to more general alternatives. When studying operating characteristics like power for a single hypothesis test, it is not fruitful to consider all possible alternatives simultaneously; we can then get arbitrarily close to the null leading to trivial results. Similarly, we need some restriction of the class $\mathcal{F}$ of distributions under the alternative to get nontrivial results. Here we discuss two.

First suppose that $\mathcal{F}$ is bounded away from the null. (This is like testing $\theta = 0$ versus $\theta > \theta_1$ for some $\theta_1 > 0$.) More precisely, define the lower envelope $H(u) = \inf_{F \in \mathcal{F}} F(u)$ and assume that $H$ satisfies the conditions of Theorem 1. Define $u^*$ to satisfy $H(u^*) = \beta u^*$. The following extension of Theorem 3 holds. The proof is similar to other calculations in the paper and is omitted.

THEOREM 6. *With $u^*$ defined as above, we have*

$$A_0 u^* \;\leq\; \liminf_{m\to\infty} R_m \;\leq\; \limsup_{m\to\infty} R_m \;\leq\; A_0\alpha + A_1[1-\beta u^*].$$

A second extension is a "random effects" version. Suppose that $F$ is parameterized as $F_\theta$ and that p-values drawn from the alternative follow a two-stage model $\theta_i \sim G$ and $P_j|\theta_j \sim F_{\theta_j}$. Define $F(u) = \int F_\theta(u)dG(\theta)$. If we interpret all expectations marginally the we have:

THEOREM 7. *Theorems 1 and 3 hold with $u^*$ defined in terms of $F(u) = \int F_\theta(u)dG(\theta)$.*

# 8 Discussion

The Benjamini-Hochberg FDR procedure and its extensions in Benjamini and Yekuteli (1999) open up a whole new way of thinking about multiple testing. But unlike standard testing methods, the operating characteristics of the BH method have not been fully explored. Abramovich, Benjamini, Donoho and Johnstone (2000) showed that, in the context of minimax point estimation, the method has certain optimality properties. We have taken a first step in understanding the properties of the method, in the context of testing.

We have characterized the asymptotic behavior of the BH rejection cutoff. Specifically, we provide large-sample bounds on the cutoff for both the fixed and local alternatives. Asymptotically, the BH procedure corresponds to rejecting for p-values less than a quantity $u^*$. One way to think about the procedure is that it provides a consistent estimator of $u^*$. This quantifies the sense in which the BH procedure is intermediate between uncorrected testing and the Bonferroni method.

We have introduced two measures of risk under which multiple testing procedures can be evaluated. We also introduced the idea of False Nondiscovery Rate (FNR) which is a practically interesting dual to the FDR that we might wish to control. In the simple-simple case where the null and alternative distributions are known, it is possible to construct procedures that dominate the BH procedure. However, at least among a subclass of FDR-controlling procedures, the BH procedure is the optimal distribution free method. It is an open question whether there exist other distribution free methods with better risk properties.

The multiple testing problem can be thought of as a missing labels problem, and it can be profitable to consider procedures that work conditionally on the observed (but unlabelled) p-values. For any risk function that is linear in the FDR and FNR, the conditionally-optimal procedures will be unconditionally optimal, but there are practical differences when the dependence is nonlinear. We gave an explicit algorithm for minimizing conditional risk in the simple-simple case.

There remain many open questions, both about the BH procedure in particular and about the existence of alternative procedures with similar properties. Our results extend beyond the simple-simple case to more general alternative distributions, but there is room for further generalization. Moreover, we studied the local alternative case only for Normal distributions. We conjecture that these results hold for non-normal distributions though the rate at which $\theta_m$ must approach 0 may change.

We believe that the BH procedure will have a great impact on how multiple testing is done in practice, both among statisticians and scientists. We look forward to continued developments in this area.

# REFERENCES

Abramovich, F., Benjamini, Y., Donoho, D. and Johnstone, I. (2000). Adapting to unknown sparsity by controlling the false discovery rate. Technical report 2000-19. Department of Statistics. Stanford University.

Benjamini, Y. and Hochberg, Y. (1995). Controlling the false discovery rate: A practical and powerful approach to multiple testing. *Journal of the Royal Statistical Society, Series B*, 57, 289-300.

Benjamini, Y. and Yekutieli, D. (1999). The control of the false discovery rate in multiple testing under dependency. Technical report. Department of Statistics. Tel Aviv University.

Berger, J., Brown, L. and Wolpert, R. (1994). A unified conditional frequentist and Bayesian test for fixed and sequential hypothesis testing. *Ann. Statist.* 22, 1787-1807.

Kiefer, J. (1977). Conditional confidence statements and confidence estimators (with discussion). *Journal of the American Statistical Association*, 72, 789-827.