Biometrika (2009), **96**, 4, *pp*. 1012–1018
doi: 10.1093/biomet/asp048
Advance Access publication 12 October 2009

A note on adaptive Bonferroni and Holm procedures under dependence

BY WENGE GUO

Department of Mathematical Sciences, New Jersey Institute of Technology, University Heights, New Jersey 07102-1982, U.S.A.

wenge.guo@gmail.com

SUMMARY

Hochberg & Benjamini (1990) first presented adaptive procedures for controlling familywise error rate. However, until now, it has not been proved that these procedures control the familywise error rate. We introduce a simplified version of Hochberg & Benjamini's adaptive Bonferroni and Holm procedures. Assuming a conditional dependence model, we prove that the former procedure controls the familywise error rate in finite samples while the latter controls it approximately.

Some key words: Bonferroni procedure; Conditional dependence; Familywise error rate; Holm procedure; Multiple testing; Step-down procedure.

1. INTRODUCTION

We consider the problem of simultaneously testing a finite number of null hypotheses H_i ($i = 1, \ldots, m$). A main concern in multiple testing is the multiplicity problem, namely, that the probability of committing at least one Type I error sharply increases with the number of the hypotheses tested at a prespecified level. There are two approaches to solving this problem. One approach is to control the familywise error rate, which is the probability of one or more false rejections, and the other is to control the false discovery rate, which is the expected proportion of Type I errors among the rejected hypotheses (Benjamini & Hochberg, 1995). The former approach works well for traditional small-scale multiple comparisons while the latter is more suitable for modern large-scale multiple-testing problems.

Several procedures have been proposed for controlling the familywise error rate, including proposals by Holm (1979) and Hochberg (1988). A well-known procedure for controlling the false discovery rate is the linear step-up procedure of Benjamini & Hochberg (1995). When some null hypotheses are false, these procedures are often conservative by a factor given by the proportion of the true null hypotheses among all null hypotheses. By exploiting knowledge of this proportion, Hochberg & Benjamini (1990) introduced adaptive Bonferroni, Holm and Hochberg procedures for controlling the familywise error rate. These adaptive procedures estimate the proportion and then use it to derive more powerful testing procedures. Until now, however, no one has proven that these adaptive procedures control the familywise error rate.

Recently, other adaptive procedures that control the false discovery rate have been introduced; e.g. by Storey et al. (2004), Genovese & Wasserman (2004), Benjamini et al. (2006), Sarkar (2006), Benjamini & Heller (2007), Gavrilov et al. (2009) and Sarkar & Guo (2009). In finite samples, however, all the existing procedures have been shown to control the false discovery rate only when the underlying test statistics are independent. Using a simulation study, Benjamini et al. (2006) demonstrated that some adaptive procedures, which control the false discovery rate under independence, may fail to control it under dependence. Therefore, it is important to study whether adaptive procedures control the familywise error rate or false discovery rate for dependent test statistics.

We introduce adaptive Bonferroni and Holm procedures, similar to those described in Hochberg & Benjamini (1990) and discuss their control of the familywise error rate under dependence. In our proposed adaptive procedures, the proportion of true nulls is estimated using an estimator of Storey et al. (2004), a simplified version of that used in Hochberg & Benjamini (1990). The dependence is described using a conditional model, a generalization of the random effects model introduced by Wu (2008). Assuming this dependence structure, we prove that the adaptive Bonferroni procedure controls the familywise error rate in finite samples while the adaptive Holm procedure controls it approximately. In addition, we prove that, even in finite samples, the adaptive Holm procedure can control the familywise error rate at a level slightly larger than the prespecified level. These results offer a partial answer to Hochberg & Benjamini's open problem. Finally, through a small simulation study, we illustrate that the adaptive Bonferroni and Holm procedures can be more powerful than the corresponding conventional procedures.

2. Main results

Given m null hypotheses $H_1, \ldots, H_m$, consider testing if $H_i = 0$, true, or $H_i = 1$, false, simultaneously for $i = 1, \ldots, m$, based on their respective p-values $P_1, \ldots, P_m$. Assume that H_i $(i = 1, \ldots, m)$, are Bernoulli random variables with $\mathrm{pr}(H = 0) = \pi_0 = 1 - \mathrm{pr}(H = 1)$, and the corresponding p-values P_i can be expressed as

$$P_i = (1 - H_i)U_i + H_i G_i^{-1}(U_i), \tag{1}$$

where U_i $(i = 1, \ldots, m)$ are independent and identically distributed uniform(0, 1) random variables that are independent of all H_i; G_i is some cumulative distribution function on (0, 1) and $G_i^{-1}(u)$ is the inverse of G_i. This mixture model was proposed by Wu (2008). The P_is are conditionally independent given H_i $(i = 1, \ldots, m)$, but H_is may be dependent. If the H_is are independent, then (1) reduces to the conventional random effect model (Storey, 2002, 2003; Genovese & Wasserman, 2004).

If V is the number of true null hypotheses rejected, then the familywise error rate is defined to be the probability of one or more false rejections, i.e. $\text{FWER} = \mathrm{pr}\{V > 0\}$. Let $P_{1:m} \leqslant \cdots \leqslant P_{m:m}$ be the ordered values of $P_1, \ldots, P_m$ and $H_{(1)}, \ldots, H_{(m)}$ be the corresponding null hypotheses. The Bonferroni procedure controls the familywise error rate at level $\pi_0\alpha$ for test statistics with arbitrary dependence by rejecting H_i whenever $P_i \leqslant \alpha/m$. Holm (1979) proposed a step-down version of the Bonferroni procedure, which controls the familywise error rate at α. Let $\alpha_i = \alpha/(m - i + 1)$ $(i = 1, \ldots, m)$ and r be the largest i such that $P_{1:m} \leqslant \alpha_1, \ldots, P_{i:m} \leqslant \alpha_i$, then under the Holm procedure, we reject the hypotheses $H_{(1)}, \ldots, H_{(r)}$. If r is not defined, then no hypothesis is rejected.

Because the above Bonferroni-type procedures are conservative by the factor π_0, knowledge of π_0 can be useful for improving the performance of Bonferroni and Holm's procedures. Several estimators of π_0 have been introduced; see Schweder & Spjøtvoll (1982), Storey et al. (2004), Meinshausen & Rice (2006), and Jin & Cai (2007), among others. We use Storey et al.'s simple estimator:

$$\hat{\pi}_0(\lambda) = \frac{m - R(\lambda) + 1}{(1 - \lambda)m}, \tag{2}$$

where λ is a prespecified constant, $R(\lambda) = \sum_{i=1}^{m} I(P_i \leqslant \lambda)$ is the number of p-values less than or equal to λ, and $I()$ is an indicator function. When $R(\lambda)$ is a fixed constant j, we use $\hat{\pi}_0(j)$ to denote $\hat{\pi}_0(\lambda)$. Storey et al.'s estimator is a simplified version of Schweder and Spjøtvoll's estimator, which was used in the adaptive procedures of Hochberg & Benjamini (1990) and Benjamini & Hochberg (2000).

Based on $\hat{\pi}_0(\lambda)$, an adaptive Bonferroni procedure is defined as follows.

Definition 1. *The level α adaptive Bonferroni procedure.*

1. *Given a fixed $\lambda \in (0, 1)$, find $R(\lambda) = \sum_{i=1}^{m} I(P_i \leqslant \lambda)$ and then calculate $\hat{\pi}_0$ based on* (2).
2. *Reject $H_{(1)}, \ldots, H_{(\hat{r})}$, where*

$$\hat{r} = \max\left\{i = 1, \ldots, R(\lambda) : P_{i:m} \leqslant \frac{\alpha}{\hat{\pi}_0 m}\right\}.$$

If the maximum does not exist, reject no hypothesis.

For the adaptive Bonferroni procedure, the following conclusion holds.

THEOREM 1. *In the conditional dependence model, the adaptive Bonferroni procedure controls the familywise error rate at level* α.

Remark 1. Theorem 1 strengthens a result of Sarkar (2006), who demonstrated that the adaptive Bonferroni procedure controls the false discovery rate under independence. Sarkar (2006) also considered a more general mixture model that does not require conditional independence of the p-values based on given null hypotheses.

Similar to Hochberg & Benjamini (1990), an adaptive Holm procedure based on $\hat{\pi}_0(\lambda)$ is as follows.

DEFINITION 2. *The level* α *adaptive Holm procedure.*

1. *Given a fixed* $\lambda \in (0, 1)$, *find* $\hat{\pi}_0$ *based on* (2) *and then calculate* $\hat{m}_0 = \hat{\pi}_0 m$ *and* $\hat{m}_j = \#\{P_i > \alpha/\hat{m}_{j-1}\}$ *for* $j = 1, \ldots, m$.
2. *Let* $\hat{k} = \max\left\{j = 0, \ldots, m : \hat{m}_{j+1} \leqslant \hat{m}_j \leqslant \hat{m}_0\right\}$, *if the maximum exists, otherwise let* $\hat{k} = 0$. *Reject* $H_{(1)}, \ldots, H_{(\hat{r})}$, *where*

$$\hat{r} = \max\left\{i = 1, \ldots, R(\lambda) : P_{i:m} \leqslant \frac{\alpha}{\hat{m}_{\hat{k}}}\right\}.$$

The adaptive Holm procedure is equivalent to the conventional Holm procedure when $\hat{\pi}_0 = 1$.

THEOREM 2. *In the conditional dependence model, let* $m_0 = \sum_{i=1}^{m} I(H_i = 0)$ *and* $A(\lambda m_0, m_0 - 1) = \mathrm{pr}(X = \lambda m_0)$, *where* $X \sim \mathrm{Bin}(m_0 - 1, \lambda)$. *Under the adaptive Holm procedure, we have:*

(i) *if* $\lim_{m\to\infty} m_0 = \infty$, *then* $\limsup_{m\to\infty} \text{FWER} \leqslant \alpha$. *That is, the adaptive Holm procedure approximately controls the familywise error rate at level* α;
(ii) *if* $\mathrm{pr}\{m_0 \geqslant n\} = 1$, *then the adaptive Holm procedure controls the familywise error rate at level* $(1 + \lambda d)\alpha$, *where* $d = \max_{m_0 \geqslant n} A(\lambda m_0, m_0 - 1)$ *and* n *is some positive integer.*

Remark 2. When n in Theorem 2(ii) is moderately large, $1 + \lambda d$ is only slightly larger than one. Therefore, the adaptive Holm procedure is slightly liberal at the most for finite samples.

3. A SIMULATION STUDY

We performed a small simulation study to compare the familywise error rate of our suggested procedures with that of the Bonferroni and Holm procedures. In Figs. 1(a) and (b) we compared the estimated familywise error rates with respect to the number of true null hypotheses and the common correlation, respectively. Each estimated familywise error rate was obtained by (i) generating $m = 200$ dependent normal random variables $N(\mu_i, 1)$ $(i = 1, \ldots, m)$, with a common correlation ρ and with m_0 of the 200 μ_is being equal to 0 and the remaining being equal to 6; (ii) applying these four procedures to test $H_i : \mu_i = 0$ against $K_i : \mu_i \neq 0$ simultaneously for $i = 1, \ldots, 200$ at level $\alpha = 0{\cdot}05$; and (iii) repeating steps (i) and (ii) 1000 times before observing the proportion of simulations where at least one true null hypothesis is falsely rejected. In Fig. 1(a) we set $\alpha = 0{\cdot}05$, $\lambda = 0{\cdot}2$ and $\rho = 0{\cdot}5$, and in Fig. 1(b) we set $\alpha = 0{\cdot}05$, $\lambda = 0{\cdot}2$ and $m_0 = 150$. As seen from Fig. 1(a), the estimated familywise error rates of our suggested adaptive procedures are much closer to the prespecified level than that of the conventional Bonferroni and Holm procedures. With an increasing number of true null hypotheses, the estimated familywise error rates decrease slightly. Also, as seen from Fig. 1(b), with the increasing common correlation among the test statistics, the estimated familywise error rates of the adaptive procedures change only slightly. By contrast, the estimated familywise error rates of the conventional procedures decrease to zero.

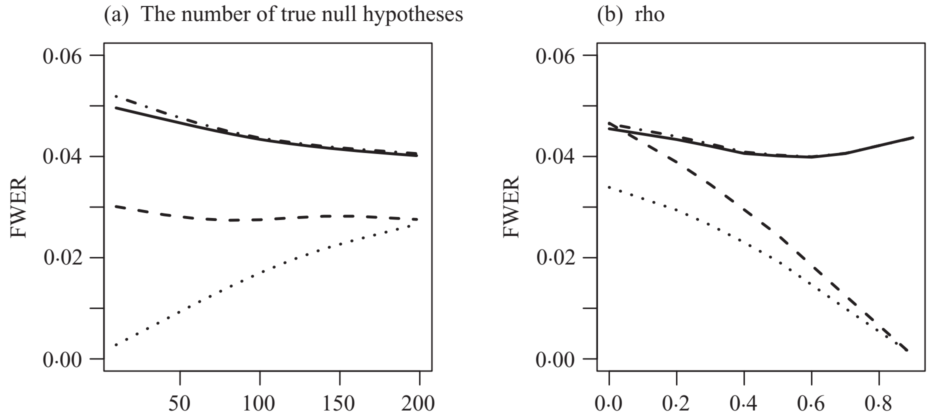

Fig. 1. Comparison of familywise error rates of four procedures: Bonferroni (dotted), adaptive Bonferroni (solid), Holm (dashed), and adaptive Holm (dot-dashed), with parameters $m = 200$, $\alpha = 0{\cdot}05$, $\lambda = 0{\cdot}2$. Specifically, (a) $\rho = 0{\cdot}5$ and (b) $m_0 = 150$.

Acknowledgement

This research was supported by the Intramural Research Program of the NIH, National Institute of Environmental Health Sciences. The author thanks Shyamal D. Peddada, David M. Umbach, the editor and an anonymous referee for their useful comments.

Appendix

Proof of Theorem 1

Using an argument similar to that of Sarkar & Guo (2009), we first consider conditional familywise error rate of the adaptive Bonferroni procedure, $\text{pr}\{V > 0 \mid P_{j:m} \leqslant \lambda < P_{j+1:m}, (H_j)_{j=1}^m\}$. For notational convenience, let $I = \{1, \ldots, m\}$ denote the index set of all null hypotheses and $m_0 = \sum_{i=1}^m I(H_i = 0)$ denote the number of true null hypotheses for given $(H_i)_{i=1}^m$. Let A_j, $A_j^{(-i)}$ and B_i respectively denote the events of $P_{j:m} \leqslant \lambda < P_{j+1:m}$, $P_{j-1:m-1}^{(-i)} \leqslant \lambda < P_{j:m-1}^{(-i)}$ and $P_i \leqslant \lambda$ for $i = 1, \ldots, m$ and $j = 0, \ldots, m$, where $P_{0:m} = 0$, $P_{m+1:m} = 1$ and $P_{1:m-1}^{(-i)} \leqslant \cdots \leqslant P_{m-1:m-1}^{(-i)}$ are the ordered p-values of $P_1, \ldots, P_m$ excluding P_i. Let $V(\lambda) = \sum_{i=1}^m I(H_i = 0, P_i \leqslant \lambda)$. Note that

$$
\begin{aligned}
\text{pr}(V > 0 \mid P_{j:m} &\leqslant \lambda < P_{j+1:m}, H_1, \ldots, H_m) \\
&= \text{pr}\left[\bigcup_{i\in I}\left\{P_i \leqslant \lambda,\, P_i \leqslant \frac{\alpha}{\tilde{\pi}_0(\lambda)m},\, H_i = 0\right\} \,\middle|\, A_j, H_1, \ldots, H_m\right] \\
&\leqslant \sum_{i=1}^m \text{pr}\left\{P_i \leqslant \lambda,\, P_i \leqslant \frac{\alpha}{\tilde{\pi}_0(j)m},\, H_i = 0 \,\middle|\, A_j, H_1, \ldots, H_m\right\} \\
&= \sum_{i=1}^m \frac{\text{pr}\left\{P_i \leqslant \frac{\alpha}{\tilde{\pi}_0(j)m}, B_i, A_j^{(-i)}, H_i = 0, (H_j)_{j=1}^m\right\}}{\text{pr}\left\{A_j, (H_j)_{j=1}^m\right\}}. \qquad \text{(A1)}
\end{aligned}
$$

In the last equality of (A1),

$$
\begin{aligned}
\text{pr}\{P_i &\leqslant \frac{\alpha}{\tilde{\pi}_0(j)m}, B_i, A_j^{(-i)}, H_i = 0, (H_j)_{j=1}^m\} \\
&= \text{pr}\left\{P_i \leqslant \frac{\alpha}{\tilde{\pi}_0(j)m}, B_i, A_j^{(-i)}, H_i \mid H_i = 0, (H_j)_{j\neq i}\right\} \text{pr}\{H_i = 0, (H_j)_{j\neq i}\}
\end{aligned}
$$

and this equals

$$\text{pr}\left\{P_i \leqslant \frac{\alpha}{\tilde{\pi}_0(j)m}, B_i \mid H_i = 0\right\} \text{pr}\{A_j^{(-i)}, H_i \mid H_i = 0, (H_j)_{j\neq i}\}\text{pr}\{H_i = 0, (H_j)_{j\neq i}\}$$

$$= \text{pr}\left\{P_i \leqslant \frac{\alpha}{\tilde{\pi}_0(j)m} \,\middle|\, B_i, H_i = 0\right\} \text{pr}\,(B_i \mid H_i = 0)$$

$$\otimes\, \text{pr}\{A_j^{(-i)}, H_i \mid H_i = 0, (H_j)_{j\neq i}\}\text{pr}\{H_i = 0, (H_j)_{j\neq i}\}$$

$$= \frac{\alpha}{\lambda\tilde{\pi}_0(j)m}\text{pr}\{B_i, A_j^{(-i)}, H_i = 0, (H_j)_{j=1}^m\}. \tag{A2}$$

Combining (A1) and (A2), we have

$$\text{pr}(V > 0 \mid A_j, H_1, \ldots, H_m) \leqslant \sum_{i=1}^{m} \frac{\alpha}{\lambda\tilde{\pi}_0(j)m}\text{pr}\left\{B_i, H_i = 0 \,\middle|\, A_j, (H_j)_{j=1}^m\right\}$$

$$= \frac{\alpha}{\lambda\tilde{\pi}_0(j)m} E\left\{V(\lambda) \,\middle|\, A_j, (H_j)_{j=1}^m\right\}. \tag{A3}$$

Thus,

$$\text{FWER} = \text{pr}\,(V > 0) = E\left[\sum_{j=0}^{m} I\left\{V > 0, A_j, (H_i)_{i=1}^m\right\}\right]$$

$$= E\left[\sum_{j=0}^{m} \text{pr}\left\{V > 0 \mid A_j, (H_i)_{i=1}^m\right\} I\left\{A_j, (H_i)_{i=1}^m\right\}\right]$$

$$\leqslant \frac{\alpha}{\lambda m} E\left\{\frac{V(\lambda)}{\hat{\pi}_0(\lambda)}\right\}$$

$$= \frac{(1-\lambda)\alpha}{\lambda} E\left[E\left\{\frac{V(\lambda)}{m - R(\lambda) + 1} \,\middle|\, (H_j)_{j=1}^m\right\}\right]. \tag{A4}$$

The inequality in (A4) follows from (A3) and the last equality follows from (2).

Under $(H_i)_{i=1}^m$, $m - R(\lambda) \geqslant m_0 - V(\lambda)$ and $m_0 - V(\lambda) \sim$ Bin $(m_0, 1-\lambda)$, so that the right side of (A4) is less than or equal to

$$\frac{(1-\lambda)\alpha}{\lambda} E\left[E\left\{\frac{V(\lambda)}{m_0 - V(\lambda) + 1} \,\middle|\, (H_j)_{j=1}^m\right\}\right] = \frac{(1-\lambda)\alpha}{\lambda} E\left[E\left\{\frac{m_0 + 1}{m_0 - V(\lambda) + 1} \,\middle|\, (H_j)_{j=1}^m\right\} - 1\right]$$

$$< \frac{(1-\lambda)\alpha}{\lambda}\left(\frac{1}{1-\lambda} - 1\right) = \alpha, \tag{A5}$$

which completes the proof. In (A5), we use the following lemma due to Benjamini et al. (2006).

LEMMA 1. *If* $Y \sim$ Bin (N, p) *then* $E\{(Y+1)^{-1}\} < \{(N+1)p\}^{-1}$.

Proof of Theorem 2

Similar to the arguments in the proof of Theorem 1, we first consider the conditional familywise error rate of the adaptive Holm procedure, $\text{pr}\left\{V > 0 \mid A_j, (H_j)_{j=1}^m\right\}$.

For given $(H_j)_{j=1}^m$, let $\hat{q}_{1:m_0} \leqslant \cdots \leqslant \hat{q}_{m_0:m_0}$ denote the ordered p-values corresponding to the true null hypotheses. Under the adaptive Holm procedure, we have

$$\text{pr}\,(V > 0, \hat{k} = 0 \mid A_j, H_1, \ldots, H_m) \leqslant \text{pr}\left\{\hat{q}_{1:m_0} \leqslant \lambda, \hat{q}_{1:m_0} \leqslant \frac{\alpha}{\hat{m}_0}, \hat{k} = 0 \mid A_j, (H_j)_{j=1}^m\right\} \tag{A6}$$

and for $k > 0$,

$$\text{pr}(V > 0, \hat{k} = k \mid A_j, H_1, \ldots, H_m) \leqslant \text{pr}\left\{\hat{q}_{1:m_0} \leqslant \lambda, \hat{q}_{1:m_0} \leqslant \frac{\alpha}{\hat{m}_k}, \hat{k} = k \mid A_j, (H_j)_{j=1}^m\right\}. \tag{A7}$$

For any $\hat{q}_{1:m_0}$ satisfying $\alpha/\hat{m}_0 < \hat{q}_{1:m_0} \leqslant \alpha/\hat{m}_k$, suppose that $\hat{q}_{1:m_0} \in (\alpha/\hat{m}_{j-1}, \alpha/\hat{m}_j]$ for some $j = 1, \ldots, k$, then by the definition of $\hat{m}_j$, we have $m_0 \leqslant \hat{m}_j$. That is, $\alpha/\hat{m}_0 < \hat{q}_{1:m_0} \leqslant \alpha/\hat{m}_k$ implies $\alpha/\hat{m}_0 < \hat{q}_{1:m_0} \leqslant \alpha/m_0$. Therefore, the right side of (A7) is less than or equal to

$$\text{pr}\left\{\hat{q}_{1:m_0} \leqslant \lambda, \hat{q}_{1:m_0} \leqslant \frac{\alpha}{\hat{m}_0}, \hat{k} = k \mid A_j, (H_j)_{j=1}^m\right\} + \text{pr}\left\{\hat{q}_{1:m_0} \leqslant \lambda, \frac{\alpha}{\hat{m}_0} < \hat{q}_{1:m_0} \leqslant \frac{\alpha}{m_0}, \hat{k} = k \mid A_j, (H_j)_{j=1}^m\right\}$$
$$= \text{pr}\left\{\hat{q}_{1:m_0} \leqslant \lambda, \hat{q}_{1:m_0} \leqslant \frac{\alpha}{m_0}, \hat{k} = k \mid A_j, (H_j)_{j=1}^m\right\}. \tag{A8}$$

Combining (A6)–(A8), we have

$$\begin{aligned}
\text{pr}(V > 0 \mid A_j, H_1, \ldots, H_m) &\leqslant \text{pr}\left\{\hat{q}_{1:m_0} \leqslant \lambda, \hat{q}_{1:m_0} \leqslant \frac{\alpha}{\min(m_0, \hat{m}_0)} \,\middle|\, A_j, (H_j)_{j=1}^m\right\} \\
&= \text{pr}\left(\bigcup_{i \in I}\left[P_i \leqslant \lambda, P_i \leqslant \frac{\alpha}{\min\{m_0, \tilde{m}_0(j)\}}, H_i = 0\right] \,\middle|\, A_j, (H_j)_{j=1}^m\right) \\
&\leqslant \sum_{i=1}^m \text{pr}\left[P_i \leqslant \lambda, P_i \leqslant \frac{\alpha}{\min\{m_0, \tilde{m}_0(j)\}}, H_i = 0 \mid A_j, (H_j)_{j=1}^m\right],
\end{aligned}$$

where $\tilde{m}_0(j)$ is the value of $\hat{m}_0(\lambda)$ when $R(\lambda) = j$. Using the arguments similar to those used in (A1)–(A3), we have

$$\text{pr}(V > 0 \mid A_j, H_1, \ldots, H_m) \leqslant \frac{\alpha}{\lambda \min\{m_0, \tilde{m}_0(j)\}} E\left\{V(\lambda) \mid A_j, (H_j)_{j=1}^m\right\}. \tag{A9}$$

Therefore, by (A9) and using the arguments similar to those used in (A4) and (A5), we have

$$\begin{aligned}
\text{FWER} &= \text{pr}\{V > 0\} \\
&= E\left[\sum_{j=1}^m pr\left\{V > 0 \mid A_j, (H_i)_{i=1}^m\right\} I\left\{A_j, (H_i)_{i=1}^m\right\}\right] \\
&\leqslant \frac{\alpha}{\lambda} E\left[\frac{V(\lambda)}{\min\{m_0, \hat{m}_0\}}\right] \\
&\leqslant \frac{(1-\lambda)\alpha}{\lambda} E\left[E\left\{\frac{V(\lambda)}{\min\{(1-\lambda)m_0, m_0 - V(\lambda) + 1\}} \,\middle|\, (H_j)_{j=1}^m\right\}\right] \\
&= \frac{(1-\lambda)\alpha}{\lambda} E\left[E\left\{\frac{V(\lambda) I\{V(\lambda) \leqslant \lambda m_0 + 1\}}{(1-\lambda)m_0} \,\middle|\, (H_j)_{j=1}^m\right\}\right] \\
&\quad + \frac{(1-\lambda)\alpha}{\lambda} E\left[E\left\{\frac{V(\lambda) I\{V(\lambda) > \lambda m_0 + 1\}}{m_0 - V(\lambda) + 1} \,\middle|\, (H_j)_{j=1}^m\right\}\right].
\end{aligned} \tag{A10}$$

Let $B(\cdot, N)$ denote the cumulative distribution function of a binomial random variable $X \sim \text{Bin}(N, p)$. Given $(H_j)_{j=1}^m$, $V(\lambda) \sim \text{Bin}(m_0, \lambda)$ and $m_0 - V(\lambda) \sim \text{Bin}(m_0, 1 - \lambda)$. Therefore,

$$E\left[\frac{V(\lambda) I\{V(\lambda) \leqslant \lambda m_0 + 1\}}{(1-\lambda)m_0} \,\middle|\, (H_j)_{j=1}^m\right] = \frac{\lambda B(\lambda m_0, m_0 - 1)}{1 - \lambda} \tag{A11}$$

and

$$E\left[\frac{V(\lambda)I\{V(\lambda)>\lambda m_0+1\}}{m_0-V(\lambda)+1}\,\middle|\,(H_j)_{j=1}^m\right] = \frac{\lambda}{1-\lambda}\sum_{\lambda m_0+1}^{m_0-1}\binom{m_0}{j}\lambda^j(1-\lambda)^{m_0-j}$$
$$\leqslant \frac{\lambda}{1-\lambda}\{1-B(\lambda m_0,m_0)\}$$
$$=\frac{\lambda}{1-\lambda}\{1-\lambda B(\lambda m_0-1,m_0-1)-(1-\lambda)B(\lambda m_0,m_0-1)\}. \tag{A12}$$

Combining (A10)–(A12), we have

$$\text{FWER} \leqslant \alpha E\left[1+\lambda\left\{B(\lambda m_0,m_0-1)-B(\lambda m_0-1,m_0-1)\right\}\right]=\alpha E\{1+\lambda A(\lambda m_0,m_0-1)\}, \tag{A13}$$

which yields the second claim. By (A13) and $\lim_{m_0\to\infty} A(\lambda m_0, m_0-1)=0$, the first claim follows immediately.

REFERENCES

BENJAMINI, Y. & HELLER, R. (2007). False discovery rate for spatial signals. *J. Am. Statist. Assoc.* **102**, 1272–81.
BENJAMINI, Y. & HOCHBERG, Y. (1995). Controlling the false discovery rate: a practical and powerful approach to multiple testing. *J. R. Statist. Soc.* B **57**, 289–300.
BENJAMINI, Y. & HOCHBERG, Y. (2000). The adaptive control of the false discovery rate in multiple hypothesis testing with independent statistics. *J. Educ. Behav. Statist.* **25**, 60–83.
BENJAMINI, Y., KRIEGER, K. & YEKUTIELI, D. (2006). Adaptive linear step-up procedures that control the false discovery rate. *Biometrika* **93**, 491–507.
GAVRILOV, Y., BENJAMINI, Y. & SARKAR, S. K. (2009). An adaptive step-down procedures with proven FDR control under independence. *Ann. Statist.* **37**, 619–29.
GENOVESE, C. & WASSERMAN, L. (2004). A stochastic process approach to false discovery control. *Ann. Statist.* **32**, 1035–1061.
HOCHBERG, Y. (1988). A sharper Bonferroni procedure for multiple testing of significance. *Biometrika* **75**, 800–2.
HOCHBERG, Y. & BENJAMINI, Y. (1990). More powerful procedures for multiple significance testing. *Statist. Med.* **9**, 811–8.
HOLM, S. (1979). A simple sequentially rejective multiple testing procedure. *Scand. J. Statist.* **6**, 65–70.
JIN, J. & CAI, T. (2007). Estimating the null and the proportion of nonnull effects in large-scale multiple comparisons. *J. Am. Statist. Assoc.* **102**, 495–506.
MEINSHAUSEN, N. & RICE, J. (2006). Estimating the proportion of false null hypotheses among a large number of independently tested hypotheses. *Ann. Statist.* **34**, 1850–73.
SARKAR, S. K. (2006). False discovery and false non-discovery rates in single-step multiple testing procedures. *Scand. Ann. Statist.* **34**, 394–415.
SARKAR, S. K. & GUO, W. (2009). On a generalized false discovery rate. *Ann. Statist.*, **37**, 337–63.
SCHWEDER, T. & SPJØTVOLL, E. (1982). Plots of p-values to evaluate many tests simultaneously. *Biometrika* **69**, 493–502.
STOREY, J. D. (2002). A direct approach to false discovery rates. *J. R. Statist. Soc.* B **64**, 479–98.
STOREY, J. D. (2003). The positive false discovery rate: a Bayesian interpretation and the q-value. *Ann. Statist.* **31**, 2013–35.
STOREY, J. D., TAYLOR, J. E. & SIEGMUND, D. (2004). Strong control, conservative point estimation and simultaneous conservative consistency of false discovery rates: a unified approach. *J. R. Statist. Soc.* B, **66**, 187–205.
WU, W. (2008). On false discovery control under dependence. *Ann. Statist.* **36**, 364–80.

[*Received March* 2008. *Revised March* 2009]