

The effect of correlation in false discovery rate estimation

BY ARMIN SCHWARTZMAN AND XIHONG LIN

*Department of Biostatistics, Harvard School of Public Health, Boston,
Massachusetts 02115, U.S.A.*

armins@hsph.harvard.edu xlin@hsph.harvard.edu

SUMMARY

The objective of this paper is to quantify the effect of correlation in false discovery rate analysis. Specifically, we derive approximations for the mean, variance, distribution and quantiles of the standard false discovery rate estimator for arbitrarily correlated data. This is achieved using a negative binomial model for the number of false discoveries, where the parameters are found empirically from the data. We show that correlation may increase the bias and variance of the estimator substantially with respect to the independent case, and that in some cases, such as an exchangeable correlation structure, the estimator fails to be consistent as the number of tests becomes large.

Some key words: High-dimensional data; Microarray data; Multiple testing; Negative binomial.

1. INTRODUCTION

Large-scale multiple testing is a common statistical problem in the analysis of high-dimensional data, particularly in genomics and medical imaging. An increasingly popular global error measure in these applications is the false discovery rate (Benjamini & Hochberg, 1995), defined as the expected proportion of false discoveries among the total number of discoveries. High-dimensional data are often highly correlated; in the microarray data example below, the raw pairwise correlations range from -0.9 to 0.96 . Correlation may greatly inflate the variance of both the number of false discoveries (Owen, 2005) and common false discovery rate estimators (Qiu & Yakovlev, 2006). While there exist procedures that control the false discovery rate under arbitrary dependence (Benjamini & Yekutieli, 2001), they have substantially less power than procedures that assume independence (Farcomeni, 2008) and the latter are often preferred. Because of the current widespread use of the false discovery rate, it is important to understand the effect of correlation in the analysis as it is typically done in practice, both for correct inference using current methods and as a guide for developing new methods for correlated data.

The goal of this paper is to quantify the effect of correlation in false discovery rate analysis. As a benchmark, we use the false discovery rate estimator of Storey et al. (2004) and Genovese & Wasserman (2004), originally proposed by Yekutieli & Benjamini (1999) and Benjamini & Hochberg (2000). This estimator is appealing because it provides estimates of the false discovery rate at all thresholds simultaneously. Furthermore, thresholding of this estimator is equivalent to the original false discovery rate controlling algorithm (Benjamini & Hochberg, 1995) under independence (Benjamini & Hochberg, 2000), and under specific forms of dependence such as positive regression dependence (Benjamini & Yekutieli, 2001) and weak dependence such as dependence in finite blocks (Storey et al., 2004). However, the generality of the estimator makes it conservative under correlation (Yekutieli & Benjamini,

1999) and it can perform poorly in genomic data (Qiu & Yakovlev, 2006). We show that correlation increases both the bias and variance of the estimator substantially compared with the independent case, but less so for small error levels. From a theoretical point of view, we show that in some cases, such as an exchangeable correlation structure, the estimator fails to be consistent as the number of tests increases.

As related work, Efron (2007a) and Schwartzman (2008) estimated the variance of the false discovery rate estimator by the delta method but did not consider the bias and skewness of the distribution. Owen (2005) quantified the variance of the number of discoveries given an arbitrary correlation structure, but did not provide results about the false discovery rate. Qiu & Yakovlev (2006) showed the strong effect of correlation on the false discovery rate estimator via simulations only.

Our contributions are as follows. First, we provide approximations for the mean, variance, distribution and quantiles of the false discovery rate estimator given an arbitrary correlation structure. This is achieved by modelling the number of discoveries with a negative binomial distribution whose parameters are estimated from the data based on the empirical distribution of the pairwise correlations. Our results are derived for test statistics whose marginal distribution is either normal or χ^2 . Secondly, we identify a necessary condition for consistency of the false discovery rate estimator as the number of tests increases and show that it is violated in situations such as exchangeable correlation.

2. THEORY

2.1. The false discovery rate estimator

Let H_1, \dots, H_m be m null hypotheses with associated test statistics T_1, \dots, T_m . The test statistics are assumed to have marginal distributions $T_j \sim F_0$ if H_j is true and $T_j \sim F_j$ if H_j is false, where F_0 is a common distribution under the null hypothesis and F_1, F_2, \dots are specific alternative distributions for each test. The test statistics may be dependent with $\text{corr}(T_i, T_j) = \rho_{ij}$. In particular, if the test statistics are z -scores, then this is the same as the correlation between the original observations.

The fraction $p_0 = m_0/m$ of tests where the null is true is called the null proportion. The complete null model is the one where $p_0 = 1$ and $T_j \sim F_0$ for all j . Without loss of generality, we focus on one-sided tests, where for each hypothesis H_j and a threshold u , the decision rule is $D_j(u) = I(T_j > u)$. Two-sided tests may be incorporated, for example, by defining $T_j = \tilde{T}_j^2$ or $T_j = |\tilde{T}_j|$, where \tilde{T}_j is a two-sided test statistic.

The false discovery rate is the expectation, under the true model, of the proportion of false positives among rejected null hypothesis or discoveries, i.e.

$$\text{FDR}(u) = E \left\{ \frac{V_m(u)}{R_m(u) \vee 1} \right\}, \quad (1)$$

where $V_m(u) = \sum_{j=1}^m D_j(u) I(H_j \text{ is true})$ is the number of false positives and $R_m(u) = \sum_{j=1}^m D_j(u)$ is the number of rejected null hypotheses or discoveries (Benjamini & Hochberg, 1995). The maximum operator \vee sets the fraction inside the expectation to zero when $R_m(u) = 0$.

When m is large, $\text{FDR}(u)$ is estimated by

$$\hat{\text{FDR}}_m(u) = \frac{\hat{p}_0 m \alpha(u)}{R_m(u) \vee 1}, \quad (2)$$

where \hat{p}_0 is an estimate of the null proportion p_0 , and $\alpha(u)$ is the marginal type-I-error level $\alpha(u) = E\{D_j(u)\} = \text{pr}(T_j > u)$, computed under the assumption that H_j is true. This estimator was proposed by Yekutieli & Benjamini (1999) and its asymptotic properties were studied by Storey et al. (2004) and Genovese & Wasserman (2004). A heuristic argument for this estimator is that the expectation of the false discovery proportion numerator, $E\{V_m(u)\}$, is equal to $p_0 m \alpha(u)$ under the true model. There are several ways to estimate p_0 (Genovese & Wasserman, 2004; Storey et al., 2004; Efron, 2007b; Jin & Cai, 2007). In applications p_0 is often close to 1 and setting $\hat{p}_0 = 1$ biases the estimate only slightly and in a conservative fashion (Efron, 2004). In this article, we focus on the effect that correlation has on the estimator (2) via the number of discoveries $R_m(u)$.

2.2. The number of discoveries

As a stepping stone towards studying the estimator (2), we first study the number of rejected null hypotheses or discoveries $R_m(u)$. Under the complete null model, and if the tests are independent, $R_m(u)$ is binomial with number of trials m and success probability $\alpha(u)$. In general, under the true model and allowing dependence, we have

$$\begin{aligned} E\{R_m(u)\} &= \sum_{i=1}^m E\{D_i(u)\} = \sum_{i=1}^m \beta_i(u), \\ \text{var}\{R_m(u)\} &= \sum_{i=1}^m \sum_{j=1}^m \text{cov}\{D_i(u), D_j(u)\} = \sum_{i=1}^m \sum_{j=1}^m \psi_{ij}(u), \end{aligned} \quad (3)$$

where

$$\beta_i(u) = \text{pr}_i(T_i > u), \quad \psi_{ij}(u) = \text{pr}(T_i > u, T_j > u) - \text{pr}(T_i > u)\text{pr}(T_j > u). \quad (4)$$

The quantity $\beta_i(u)$ is the per-test power. For those tests where the null hypothesis is true, this is equal to the marginal type-I-error level $\alpha(u)$. Summing over the diagonals in (3) reveals the mean-variance structure

$$E\{R_m(u)\} = m\bar{\beta}(u), \quad \text{var}\{R_m(u)\} = m\{\bar{\beta}(u) - \bar{\beta}^2(u)\} + m(m-1)\Psi_m(u), \quad (5)$$

where

$$\bar{\beta}(u) = \frac{1}{m} \sum_{i=1}^m \beta_i(u), \quad \bar{\beta}^2(u) = \frac{1}{m} \sum_{i=1}^m \beta_i^2(u), \quad \Psi_m(u) = \frac{2}{m(m-1)} \sum_{i < j} \psi_{ij}(u). \quad (6)$$

The quantities $\bar{\beta}(u)$ and $\bar{\beta}^2(u)$ are empirical moments of the power, while $\Psi_m(u)$ is the average covariance of the decisions $D_i(u)$ and $D_j(u)$ for $i \neq j$, a function of the pairwise correlations $\{\rho_{ij}\}$.

The dependence between the test statistics does not affect the mean of $R_m(u)$ but does affect its variance. It does so by adding the overdispersion term $m(m-1)\Psi_m(u)$ to the independent-case variance $m\{\bar{\beta}(u) - \bar{\beta}^2(u)\}$. Expression (5) under the complete null is an unconditional version of the conditional variance in expression (8) of Owen (2005). Similar expressions have appeared before in estimation problems with correlated binary data (Crowder, 1985; Prentice, 1986).

Another observation from (4) is that $\psi_{ij}(u)$ vanishes asymptotically as $u \rightarrow \infty$, so the effect of the correlation becomes negligible for very high thresholds. We will see in § 2.4 that the rate of decay is a quadratic exponential times a polynomial.

2.3. *Inconsistency of the false discovery rate estimator*

The overdispersion of the number of discoveries $R_m(u)$ has implications for the behaviour of the estimator (2). We consider the asymptotic case $m \rightarrow \infty$ in this section and treat the finite m case in § 2.4 and 2.5.

As $m \rightarrow \infty$, a sufficient condition for consistency of the estimator (2) is that the test statistics are independent or weakly dependent, e.g. dependent in finite blocks (Storey et al., 2004). On the other hand, taking m in (2) to the denominator, we see that a necessary condition for consistency is that the fraction of discoveries $R_m(u)/m$ has asymptotically vanishing variance. By (5), the variance of $R_m(u)/m$ is

$$\text{var} \left\{ \frac{R_m(u)}{m} \right\} = \frac{\bar{\beta}(u) - \bar{\beta}^2(u)}{m} + \left(1 - \frac{1}{m} \right) \Psi_m(u) \quad (7)$$

and is asymptotically zero if and only if the overdispersion Ψ_m is asymptotically zero, provided that $\bar{\beta}$ and $\bar{\beta}^2$ grow slower than linearly with m .

THEOREM 1. Assume $\{\bar{\beta}(u) - \bar{\beta}^2(u)\}/m \rightarrow 0$ as $m \rightarrow \infty$. Fix an ordering of the test statistics T_1, \dots, T_m and assume the autocovariance sequence $\psi_{i,i+k}(u) = \psi_{i+k,i}(u)$ in (7) has a limit $\Psi_\infty(u) \geq 0$ as $k \rightarrow \infty$ for every i . Then, as $m \rightarrow \infty$, $\text{var}\{R_m(u)/m\} \rightarrow \Psi_\infty(u) \geq 0$.

Example 1. Exchangeable correlation. Suppose the test statistics T_i have an exchangeable correlation structure so that $\rho_{ij} = \rho > 0$ is a constant for all $i \neq j$. For any fixed threshold, u , $\psi_{ij}(u) = \psi(u) > 0$ in (6) does not depend on i, j . As a consequence, $\Psi_m(u) = \psi(u) = \Psi_\infty(u) > 0$, and the estimator (2) is inconsistent. The case $\rho < 0$ is asymptotically moot because positive definiteness of the covariance of the T_i s requires $\rho > -1/(m-1)$, so ρ cannot be negative in the limit $m \rightarrow \infty$.

Example 2. Stationary ergodic covariance. Suppose the index i represents time or position along a sequence, and suppose the test statistic sequence T_i is a stationary ergodic process, e.g., M-dependent or autoregressive moving average, with Toeplitz correlation $\rho_{ij} = \rho_{i-j}$. Then $\rho_{i,i+k} = \rho_k \rightarrow 0$ for all i as $k \rightarrow \infty$, and so $\Psi_\infty(u) = 0$, pointwise for all u . By Theorem 1, the variance (7) converges to zero as $m \rightarrow \infty$.

Example 3. Finite blocks. Suppose the test statistics T_i are dependent in finite blocks, so that they are correlated within blocks but independent between blocks. Suppose the largest block has size K . If K increases with m such that $K/m \rightarrow 0$ as $m \rightarrow \infty$ then $\Psi_\infty(u) = 0$ for all u . On the other hand, if $K/m \rightarrow \gamma > 0$ then $\Psi_\infty(u) > 0$ and the estimator (2) is inconsistent.

Example 4. Strong mixing. Suppose the test statistics T_i have a strong mixing or α -mixing dependence; that is, the supremum over k of $|\text{pr}(A \cap B) - \text{pr}(A)\text{pr}(B)|$, where A is in the σ -field generated by T_1, \dots, T_k and B is in the σ -field generated by T_{k+1}, T_{k+2}, \dots , tends to 0 as $m \rightarrow \infty$ (Zhou & Liang, 2000). By taking $A = \{T_i > u\}$ and $B = \{T_j > u\}$ in (4), we have that $\Psi_\infty = 0$ and therefore, by Theorem 1, the variance (7) converges to zero as $m \rightarrow \infty$.

Example 5. Sparse dependence. Suppose the test statistics T_i are pairwise independent except for M_1 out of the $M = m(m-1)/2$ distinct pairs. If the dependence is asymptotically sparse in the sense that $M_1/M \rightarrow 0$ as $m \rightarrow \infty$, then $\Psi_\infty(u) = 0$ and the variance (7) goes to zero. However, if $M_1/M \rightarrow \gamma > 0$, then the result will depend on the dependence structure among the M_1 dependent pairs. If this dependence goes away asymptotically as in one of the cases above,

then the variance (7) goes to zero. Otherwise, if the dependence persists such that $\Psi_\infty(u) > 0$, the estimator (2) is inconsistent.

2.4. Quantifying overdispersion for finite m

We are interested in estimating the distributional properties of the false discovery rate estimator. This requires estimating the overdispersion $\Psi_m(u)$ in (5) and (7). This quantity is easy to write in terms of the pairwise correlations between the decision rules, as in (3), but not necessarily as a function of the pairwise test statistic correlations ρ_{ij} . In this section we provide expressions for $\Psi_m(u)$ for finite m assuming a specific bivariate probability model for every pair of test statistics, but without the need to assume a higher order correlation structure. We consider commonly used z and χ^2 tests.

Suppose first that every pair of test statistics (T_i, T_j) has the bivariate normal density with marginals $N(\mu_i, 1)$ and $N(\mu_j, 1)$, and $\text{corr}(T_i, T_j) = \rho_{ij}$. Denote by $\phi(t)$ and $\phi_2(t_i, t_j; \rho_{ij})$ the univariate and bivariate standard normal densities. Mehler's formula (Patel & Read, 1996; Kotz et al., 2000) states that the joint density $f_{ij}(t_i, t_j) = \phi_2(t_i - \mu_i, t_j - \mu_j; \rho_{ij})$ of (T_i, T_j) can be written as

$$f_{ij}(t_i, t_j) = \phi(t_i - \mu_i)\phi(t_j - \mu_j) \sum_{k=0}^{\infty} \frac{\rho_{ij}^k}{k!} \mathcal{H}_k(t_i - \mu_i) \mathcal{H}_k(t_j - \mu_j), \quad (8)$$

where $\mathcal{H}_k(t)$ are the Hermite polynomials: $\mathcal{H}_0(t) = 1$, $\mathcal{H}_1(t) = t$, $\mathcal{H}_2(t) = t^2 - 1$ and so forth.

THEOREM 2. Let $\mu = (\mu_1, \dots, \mu_m)^\top$. Under the bivariate normal model (8), the overdispersion term in (6) is

$$\Psi_m(u; \mu) = \sum_{k=1}^{\infty} \frac{\rho_k^*(u; \mu)}{k!}, \quad (9)$$

where

$$\rho_k^*(u; \mu) = \frac{2}{m(m-1)} \sum_{i < j} \rho_{ij}^k \phi(u - \mu_i) \phi(u - \mu_j) \mathcal{H}_{k-1}(u - \mu_i) \mathcal{H}_{k-1}(u - \mu_j). \quad (10)$$

Under the complete null model $\mu = 0$, (9) reduces to

$$\Psi_m(u) = \Psi_m(u; 0) = \phi^2(u) \sum_{k=1}^{\infty} \frac{\bar{\rho}^k}{k!} \mathcal{H}_{k-1}^2(u), \quad (11)$$

where $\bar{\rho}^k$ ($k = 1, 2, \dots$) are the empirical moments of the $m(m-1)/2$ correlations ρ_{ij} ($i < j$).

Another common null distribution in multiple testing problems is the χ^2 distribution. Let $f_\nu(t)$ and $F_\nu(t)$ be the density and distribution functions of the $\chi^2(\nu)$ distribution with ν degrees of freedom. Under the complete null, the pair of test statistics (T_i, T_j) admits a Lancaster bivariate model where T_i and T_j have the same marginal density f_ν , their correlation is ρ_{ij} , and their joint density is

$$f_\nu(t_i, t_j; \rho_{ij}) = f_\nu(t_i) f_\nu(t_j) \sum_{k=0}^{\infty} \frac{\rho_{ij}^k}{k!} \frac{\Gamma(\nu/2)}{\Gamma(\nu/2 + k)} \mathcal{L}_k^{(\nu/2-1)}\left(\frac{t_i}{2}\right) \mathcal{L}_k^{(\nu/2-1)}\left(\frac{t_j}{2}\right), \quad (12)$$

where $\mathcal{L}_k^{(v/2-1)}(t)$ are the generalized Laguerre polynomials of degree $v/2 - 1$: $\mathcal{L}_0^{(v/2-1)}(t) = 1$, $\mathcal{L}_1^{(v/2-1)}(t) = -t + v/2$, $\mathcal{L}_2^{(v/2-1)}(t) = t^2 - 2(v/2 + 1)t + (v/2)(v/2 + 1)$ and so forth (Koudou, 1998). The χ^2 distribution is not a location family with respect to the non-centrality parameter, so the Lancaster expansion does not hold in the non-central case.

THEOREM 3. *Assume the complete null. Under the bivariate χ^2 model (12), the overdispersion term in (6) is*

$$\Psi_m(u) = f_{v+2}^2(u) \sum_{k=1}^{\infty} \frac{\bar{\rho}^k}{k!} \frac{v^2 \Gamma(v/2)}{k^2 \Gamma(v/2 + k)} \left\{ L_{k-1}^{(v/2)} \left(\frac{u}{2} \right) \right\}^2, \quad (13)$$

where $\bar{\rho}^k$ ($k = 1, 2, \dots$) are the empirical moments of the $m(m-1)/2$ correlations ρ_{ij} ($i < j$).

Theorems 2 and 3 show that the overdispersion is a function of modified empirical moments of the pairwise correlations ρ_{ij} and depends on m only through these moments. This provides an efficient way to evaluate $\Psi_m(u)$ for a given set of pairwise correlations, as the expansions (11) and (13) may be approximated evaluating only the first few terms. This is in contrast to direct computation of (6), which requires $m(m-1)/2$ evaluations of the bivariate probabilities in (4), one for each value of ρ_{ij} . In addition, as a function of u , both (11) and (13) decay as a quadratic exponential times a polynomial, implying that the effect of correlation quickly becomes negligible for high thresholds. Finally, under the complete null, (11) and (13) indicate that the overdispersion, and thus also the variance (7), vanish asymptotically as $m \rightarrow \infty$ for all u if and only if $\bar{\rho}^k \rightarrow 0$ as $m \rightarrow \infty$ for all $k = 1, 2, \dots$

2.5. Distributional properties of the false discovery rate estimator

We now apply the above results towards quantifying the distributional properties of the false discovery rate estimator (2). Our strategy is to model $R_m(u)$ as a negative binomial variable and derive the properties of the estimator based on this model.

Seen as a function of m , the mean-variance structure (5) resembles that of the beta-binomial distribution, which has been used to model overdispersed binomial data (Prentice, 1986; McCullagh & Nelder, 1989). Because the beta-binomial distribution is difficult to work with, we propose instead to model $R_m(u)$ using a negative binomial distribution. The rationale is as follows. A binomial distribution with number of trials n and success probability p can be approximated by a Poisson distribution with mean parameter np when n is large and p is small. If p has a beta distribution with mean μ , then when n is large and μ is small, the distribution of np can be approximated by a gamma distribution with mean $n\mu$. Therefore, the beta-binomial mixture model can be approximated by a gamma-Poisson mixture model, which is equivalent to a negative binomial distribution (Hilbe, 2007).

It is convenient to parameterize the negative binomial distribution with parameters $\lambda \geq 0$ and $\omega \geq 0$, such that the mean and variance of $R \sim NB(\lambda, \omega)$ are

$$E(R) = \lambda, \quad \text{var}(R) = \lambda + \omega\lambda^2. \quad (14)$$

See the proof of Theorem 4 below for details. Here λ is a mean parameter, whereas ω controls the overdispersion with respect to the Poisson distribution. When $\omega \rightarrow 0$ with λ constant, the negative binomial distribution becomes Poisson with mean λ ; when $\lambda = 0$, it becomes a point mass at $R = 0$.

We describe how to estimate λ and ω from data in § 2.6. Given λ and ω , the moments, distribution and quantiles of the estimator (2) can be obtained directly from the negative binomial model as follows. For simplicity of notation, we omit the indices m and u .

THEOREM 4. Suppose $R \sim NB(\lambda, \omega)$ if $\lambda \geq 0$, $\omega > 0$, or $R \sim \text{Po}(\lambda)$ if $\lambda \geq 0$, $\omega = 0$, with moments (14), cumulative distribution function $F(k) = \text{pr}(R \leq k)$, and quantiles $F^{-1}(q) = \inf\{x : \text{pr}(R \leq x) \geq q\}$. Assume p_0 is known and define the probability of discovery to be

$$\gamma(\lambda, \omega) = \text{pr}(R > 0) = \begin{cases} 1 - (1 + \omega\lambda)^{-1/\omega} & (\omega > 0), \\ 1 - \exp(-\lambda) & (\omega = 0). \end{cases}$$

The mean and variance of $F\hat{\text{DR}} = p_0 m \alpha / (R \vee 1)$ are

$$\begin{aligned} E(F\hat{\text{DR}}) &= p_0 m \alpha \{1 - \gamma(\lambda, \omega) + \gamma(\lambda, \omega) \zeta(\lambda, \omega)\}, \\ \text{var}(F\hat{\text{DR}}) &= (p_0 m \alpha)^2 \{1 - \gamma(\lambda, \omega) + \gamma(\lambda, \omega) \zeta_2(\lambda, \omega)\} - E^2(F\hat{\text{DR}}), \end{aligned}$$

where

$$\begin{aligned} \zeta(\lambda, \omega) &= E\left(\frac{1}{R} \middle| R > 0\right) = \int_0^\lambda \frac{\gamma^{-1}(\lambda, \omega) - 1}{\gamma^{-1}(x, \omega) - 1} \cdot \frac{dx}{x(1 + \omega x)}, \\ \zeta_2(\lambda, \omega) &= E\left(\frac{1}{R^2} \middle| R > 0\right) = \int_0^\lambda \frac{\gamma^{-1}(\lambda, \omega) - 1}{\gamma^{-1}(x, \omega) - 1} \cdot \frac{\zeta(x, \omega) dx}{x(1 + \omega x)}. \end{aligned}$$

The distribution and quantiles of $F\hat{\text{DR}} = p_0 m \alpha / (R \vee 1)$ are

$$\begin{aligned} \text{pr}(F\hat{\text{DR}} \leq x) &= \begin{cases} 1 - F(k) & (a_{k+1} \leq x < a_k; \ k = 1, 2, \dots), \\ 1 & (a_1 \leq x), \end{cases} \\ \inf\{x : \text{pr}(F\hat{\text{DR}} \leq x) \geq q\} &= \begin{cases} a_{F^{-1}(1-q)+1} & (q \leq 1 - F(1)), \\ a_1 & (q > 1 - F(1)), \end{cases} \end{aligned}$$

where $a_k = p_0 m \alpha / k$.

As a notational choice, the functions $\zeta(\lambda, \omega)$ and $\zeta_2(\lambda, \omega)$ above were defined so that both are bounded, tend to 0 as $\lambda \rightarrow 0$, and tend to 1 as $\lambda \rightarrow \infty$. When $\omega = 0$, they are the same as (27) in Schwartzman (2008) divided by λ .

2.6. Estimation of the distributional properties of the false discovery rate estimator

We estimate the parameters λ and ω of the negative binomial model for $R_m(u)$ by the method of moments. Based on (5) but respecting the form (14), we use the approximations $E\{R_m(u)\} \sim m\alpha(u)$ and $\text{var}\{R_m(u)\} \sim m\alpha(u) + m^2\Psi_m(u; 0)$ for large m under the complete null. The form of the variance ensures that it is greater or equal to the mean as required by the negative binomial model. In the normal case, using (11), the overdispersion term $\Psi_m(u; 0)$ is estimated by

$$\hat{\Psi}_m(u; 0) = \phi^2(u) \sum_{k=1}^K \frac{\bar{\rho}^k}{k!} \mathcal{H}_{k-1}^2(u),$$

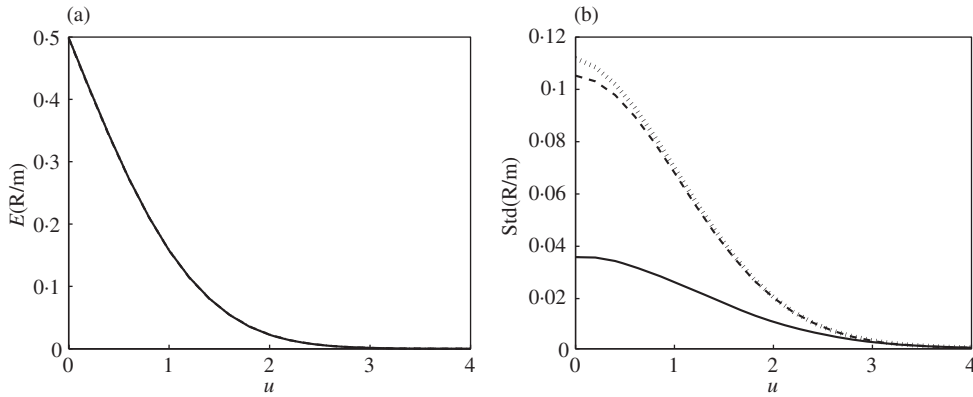


Fig. 1. Simulation results for Example 6: number of discoveries under the complete null, sparse correlation, $m = 100$. Expectation (a) and standard deviation (b) of $R_m(u)/m$. Plotted in both panels: the true value under independence (solid), the true value under dependence (dashed) and the value calculated using the polynomial expansion (11) (dotted).

where $\hat{\rho}^k$ ($k = 1, \dots, K$) denote the empirical moments of the $m(m-1)/2$ empirical pairwise correlations $\hat{\rho}_{ij}$ ($i < j$), after correction for sampling variability. In practice, $K = 3$ suffices. The overdispersion (13) is estimated similarly in the χ^2 case.

Matching the estimated moments of $R_m(u)$ to those of the negative binomial model (14) leads to the parameter estimates

$$\hat{\lambda}_m(u) = m\alpha(u), \quad \hat{\omega}_m(u) = \frac{\hat{\Psi}_m(u; 0)}{\alpha^2(u)}. \quad (15)$$

Finally, the moments and quantiles of \hat{FDR} are estimated using the formulas in Theorem 4 by plugging in the parameter estimates (15).

3. NUMERICAL STUDIES AND DATA EXAMPLES

3.1. Numerical studies

The following simulations illustrate the effect of correlation on the estimator (2) and show the accuracy of the negative binomial model in quantifying that effect. In the following simulations, $N = 10\,000$ datasets were drawn at random from the model $Y_j = \mu_j + Z_j$ ($j = 1, \dots, m$), where each Z_j has marginal distribution $N(0, 1)$. The tests were set up as $H_0 : \mu_j = 0$ vs. $H_j : \mu_j > 0$. The test statistics were taken as the z-scores $T_j = Y_j$, the signal-to-noise ratio being controlled by the strength of the signal μ_j . The true false discovery rate was computed according to (1) where the expectation was replaced by an average over the N simulated datasets. Similarly, the true moments and distribution of $R_m(u)$ and $\hat{FDR}_m(u)$ (2) were obtained from the N simulated datasets.

Example 6. Sparse correlation, complete null. A sparse correlation matrix $\Sigma = \{\rho_{ij}\}$ was generated as follows. First, a lower triangular matrix A was generated with diagonal entries equal to 1 and 10% of its lower off-diagonal entries randomly set to 0.6, zero otherwise. Then, Σ was computed as the covariance matrix AA' , normalized to be a correlation matrix. This process guarantees that Σ is positive definite. The resulting matrix Σ had 68% of its off-diagonal entries

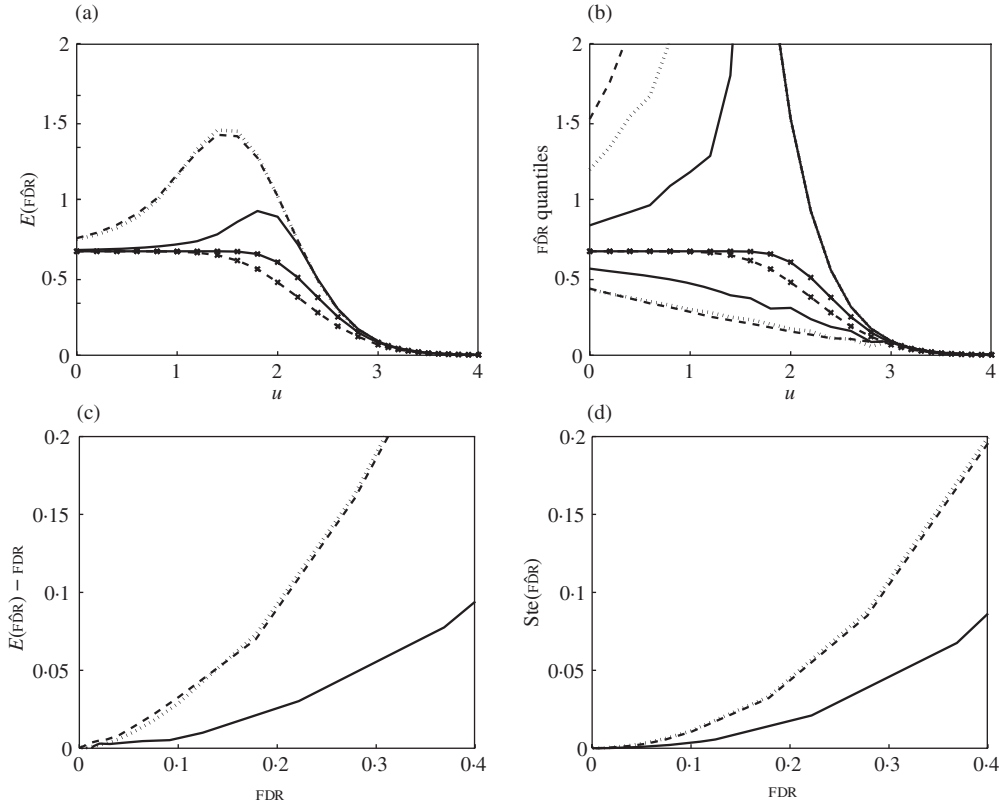


Fig. 2. Simulation results for Example 6: false discovery rate estimator under the complete null, sparse correlation, $m = 100$. (a) and (b) Expectation and quantiles 0.05 and 0.95 of \hat{FDR} as a function of threshold under independence (solid plain), under dependence (dashed plain) and estimated by the negative binomial model (dotted). Also plotted: true false discovery rate under independence (solid crossed) and under dependence (dashed crossed). (c) and (d) Bias and standard error of \hat{FDR} as a function of the true false discovery rate under independence (solid), under dependence (dashed) and estimated by the negative binomial model (dotted).

equal to zero, and empirical moments $\bar{\rho} = 0.0588$, $\bar{\rho}^2 = 0.0134$, $\bar{\rho}^3 = 0.0037$. The correlated observations $Z = (Z_1, \dots, Z_m)'$ were generated as $Z = \Sigma^{1/2}\varepsilon$, where $\varepsilon \sim N(0, I_m)$. The signal was set to $\mu_j = 0$ for all $j = 1, \dots, m$.

Figure 1 shows the effect of dependence on the discovery rate $R_m(u)/m$ for $m = 100$ under the complete null. According to (5), dependence does not affect the mean, as all the lines overlap in Figure 1(a), but affects the variance substantially for low thresholds. The estimated negative binomial mean and standard deviation coincide with the true values by design because the moments were matched, except that only three terms were used in the polynomial expansion (11) for $\Psi_m(u)$. The first of these terms is the largest by far.

Figure 2 shows the effect of dependence on the false discovery rate estimator (2) for $m = 100$ under the complete null. Figure 2(a) shows that the estimator is always positively biased. It may even be greater than 1 if the denominator is smaller than the numerator. Dependence causes the true false discovery rate to go down, further increasing the bias. Dependence also causes the variability of the estimator to increase dramatically, as indicated by the 0.05 and 0.95 quantiles in Fig. 2(b). Here, the ragged and non-monotone lines are a consequence of the discrete nature of

the distribution. In both panels (a) and (b), the expectation and quantiles of \hat{FDR} under dependency are reasonably well approximated by the negative binomial model.

The bias and standard error of the false discovery rate estimator (2) are easier to assess when plotted against the true false discovery rate in each case, as shown in Fig. 2(c) and (d). At practical false discovery rate levels such as 0.2, the bias of the estimator under independence is about 10% while under correlation is 25%, about 2.5 times larger. Similarly, the standard error under independence is about 9% while under correlation is 14%, about 1.7 times larger. The negative binomial model gives slight overestimates.

Other simulations for increasing m from $m = 10$ to $m = 10\,000$ indicate that, when plotted against the true false discovery rate as in Fig. 2(c) and (d), the bias and standard error of the estimator increase slightly as a function of m both under independence and correlation. This is because increasing m increases the expectation and variance of the estimator for any fixed threshold, but the threshold required for controlling false discovery rate at a given level also increases accordingly.

Example 7. Sparse correlation, non-complete null. Keeping the same correlation structure as before with $m = 100$, the signal was set to $\mu_j = 1$ ($j = 1, \dots, 5$), providing a null fraction $p_0 = 0.95$. Figure 3 shows the effect of dependence on the false discovery rate estimator (2). Figure 3(a) shows that the bias up persists as in the complete null case, while the overshoot has been diminished. Figure 3(b) shows that correlation increases the variability of the estimator. This is visible in this panel mostly in terms of the 0.05 quantile, as the 0.95 quantile lines coincide.

When plotted against the true false discovery rate, Fig. 3(c) and (d) shows that the bias and variance of the estimator are higher than in the complete null case, about twice as much at a false discovery rate level of 0.2. This is a consequence of the increased variance of $R_m(u)/m$ at medium-high thresholds, itself a consequence of the increased mean of $R_m(u)/m$ according to (5). In this regard, the effect of the signal is stronger than that of dependence. This explains why the negative binomial line, which was fitted assuming the complete null, does not match the simulations as well as when no signal is present.

3.2. Data example

We use the genetic microarray dataset from Mootha et al. (2003), which has $m = 10\,983$ expression levels measured among diabetes patients. For the purposes of this article, standard two-sample t -statistics were computed at each gene between the group with Type 2 diabetes mellitus, $n_1 = 17$, and the group with normal glucose tolerance, $n_2 = 17$. The t -statistics, having 32 degrees of freedom, were converted to the normal scale by a bijective quantile transformation (Efron, 2004, 2007a). The histogram of the $m = 10\,983$ test statistics in Fig. 4(a) is almost standard normal; its mean and standard deviation are 0.059 and 0.977. Thus the effect of correlation on the null distribution (Efron, 2007a) is minimal.

Figure 4(b) shows a histogram of 499 500 sample pairwise correlations computed from 1000 randomly sampled genes out of $m = 10\,983$. The pairwise correlations were computed between the gene expression levels across all 34 subjects after subtracting the means of both groups separately. These are approximately the same as the correlations between the z -scores given the moderate sample size of 34. The first three empirical moments of the sample pairwise correlations, obtained from a random sample of 2000 genes, were 0.0044, 0.0846 and 0.0020. To correct for sampling variability, empirical Bayes shrinking of the sample correlations towards zero (Owen, 2005; Efron, 2007a) resulted in the superimposed black histogram, with first three empirical moments 0.0029, 0.0586 and 0.0012.

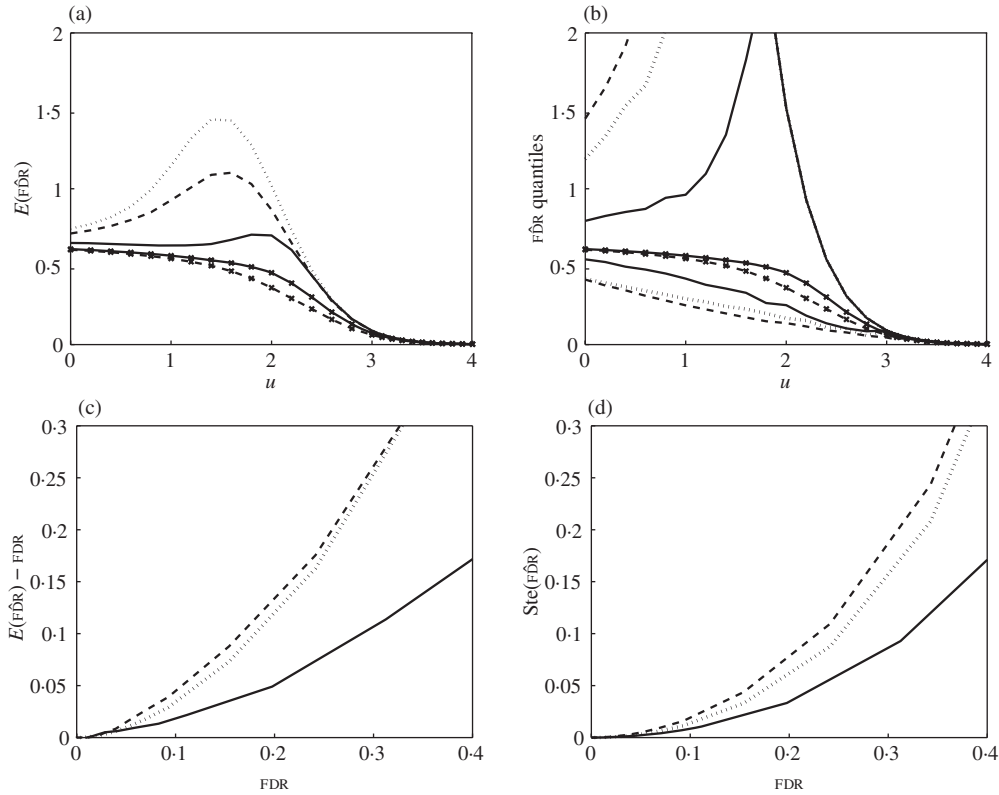


Fig. 3. Simulation results for Example 7: false discovery rate estimator with 5% signal at $\mu = 1$, sparse correlation, $m = 100$. Same quantities plotted as in Fig. 2. The negative binomial line (dotted) was fitted assuming the complete null.

Figure 4(c) shows the false discovery rate estimator (2) as a function of the threshold u . Superimposed are the 0.05 and 0.95 quantiles of the estimator according to the negative binomial model under the complete null. The 0.05 quantile line is much lower when correlation is taken into account, while the 0.95 quantile lines coincide. At $u = 4$, the estimated false discovery rate is 0.17, but the bands under correlation indicate that with 90% probability it could have been as low as 0.12 or as high as 0.35. For reference, we superimposed the 0.05 and 0.95 quantiles of the empirical distribution of the estimator obtained by permuting the subject labels between the two groups, while keeping genes belonging to the same subject together. We see that the permutation estimates closely resemble those of the negative binomial model under correlation.

4. DISCUSSION

The distribution of the false discovery rate estimator (2) is discrete and highly skewed, so rather than standard errors, as in Efron (2007a) and Schwartzman (2008), we chose to indicate the variability of the estimator by its quantiles. The 0.05 quantile indicates how deceptively low the estimates can be even when there is no signal. The 0.95 quantile is almost always equal to $m\alpha(u)$, the Bonferroni-adjusted level, showing that the estimator can be sometimes as conservative as the Bonferroni method. We emphasize that the negative binomial bands between the 0.05 and

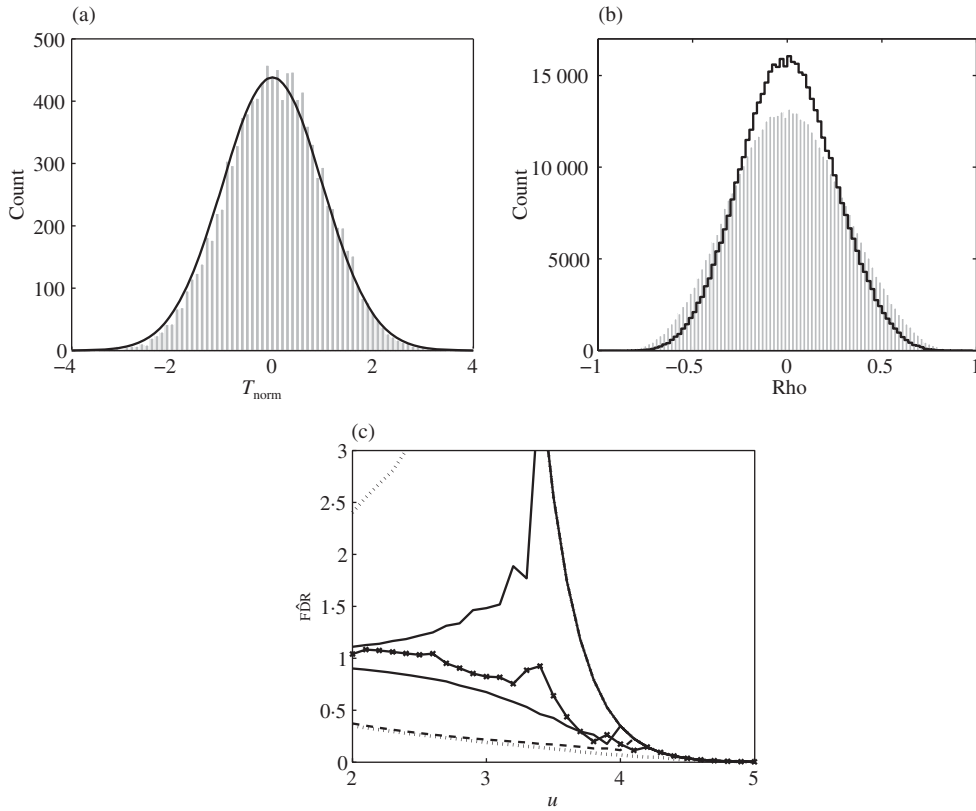


Fig. 4. The diabetes microarray data. (a) Histogram of the $m = 10\,983$ test statistics converted to normal scale; superimposed is the $N(0, 1)$ density times m . (b) Histogram of 499 500 pairwise sample correlations from 1000 randomly sampled genes; superimposed in black: histogram corrected for sampling variability. (c) False discovery rate curves: $\widehat{\text{FDR}}$ (solid crossed); quantiles 0.05 and 0.95 estimated by the negative binomial model assuming independence (solid) and assuming the pairwise correlations from the data (dashed); 0.05 and 0.95 quantiles estimated by permutations (dotted). All the 0.95 quantile lines overlap at the right edge.

0.95 quantiles describe the behavior under the complete null and are not 90% confidence bands for the true false discovery rate. It is interesting that, in our data example, correlation played an important role but did not cause a departure from the null distribution, as may have been predicted by Efron (2007a).

The negative binomial parameters can be computed via (5) and (15) for any pairwise distribution of the test statistics. The main motivation for assuming the normal and χ^2 models was to reduce computations, as the Mehler and Lancaster expansions of § 2.4 allowed reducing the correlation structure to the first few empirical moments of the pairwise correlations. A similar reduction was used by Owen (2005) and Efron (2007a). As shown in the data example, t -statistics can be handled easily by a quantile transformation to z -scores (Efron, 2007a). Similarly, F statistics can be transformed to χ^2 -scores (Schwartzman, 2008).

Summarizing the effect of the pairwise correlations by their empirical moments allows a gross comparison between different correlation structures. For example, the simulated scenario in Examples 6 and 7 of a sparse correlation structure with 68% pairwise correlations equal to

zero has nearly the same first moment but higher second moment as a fully exchangeable correlation with $\rho = 0.06$. Simulations show that both correlation structures have similar effects on the false discovery rate estimator.

Further work is required for the estimation of $\lambda_m(u)$ and $\omega_m(u)$ in the negative binomial model, not assuming the complete null. This is difficult but important because the effect of the signal may be stronger than that of the correlation. Since $\lambda_m(u)$ is the expected number of discoveries, one could estimate it by $R_m(u)$, but this estimate is too noisy. In terms of $\omega_m(u)$, one could estimate the required overdispersion term using Theorem 2, but the provided expression depends on the true signal, which is unknown. Estimating the signal as null may be appropriate if the signal is weak and/or p_0 is close to 1. Other options include using a highly regularized estimator of the signal vector μ . From the form of (10), it is expected that more weight would be given to pairwise correlations where the signal is large.

The negative binomial model was based on an approximation to a beta-binomial model when m is large and α is small. For larger α , the approximation may continue to hold as both distributions approach normality. Another approach sometimes used when strong dependency is suspected is to first transform the data to weak dependency via linear combination and then apply the false discovery rate procedure (Klebanov et al., 2008). This approach does not test the original marginal means but linear combinations of them. We chose not to do this so that the inference would continue to be about the original hypotheses.

The false discovery rate estimator is inherently biased and highly variable. It has been recognized before that under high correlation, the estimator ‘might become either downward biased or grossly upward biased’ (Yekutieli & Benjamini, 1999). The exchangeable correlation structure is a special case of positive regression dependence and therefore thresholding of the estimator guarantees control of the false discovery rate control (Benjamini & Yekutieli, 2001). The conservativeness of the thresholding procedure is reflected in the positive bias of the estimator, a consequence of overdispersion in the number of discoveries. In general, however, a positive bias is not enough to guarantee false discovery rate control as it makes the thresholding procedure conservative only on average, not uniformly over thresholds (Yekutieli & Benjamini, 1999). A correlation structure that produced underdispersion and negative bias might not provide control of the false discovery rate. Because the false discovery rate estimator was derived from the point of view of control, it is possible that better estimators might be found in the future by approaching the problem directly as an estimation problem.

ACKNOWLEDGMENT

This work was partially supported by a U.S. National Institutes of Health grant, the Claudia Adams Barr Program in Cancer Research, and the William F. Milton Fund.

APPENDIX

Technical details

Proof of Theorem 1. Define a mapping of the matrix $\{\psi_{ij}\}$ to the unit square $[0, 1]^2$ by the function

$$g_m(x, y) = \sum_{i=1}^m \sum_{j=1}^m \psi_{ij} 1\left(\frac{i-1}{m} \leq x < \frac{i}{m}, \frac{j-1}{m} \leq y < \frac{j}{m}\right).$$

Then the variance (3) is equal to the integral $\int_0^1 \int_0^1 g_m(x, y) dx dy$. The limit assumptions on ψ_{ij} imply that as $m \rightarrow \infty$, $g_m(x, y) \rightarrow \Psi_\infty$ pointwise for every (x, y) . Therefore, by the bounded convergence theorem, the integral converges to $\int_0^1 \int_0^1 \Psi_\infty dx dy = \Psi_\infty$. \square

Proof of Theorem 2. Let $\Phi^+(u)$ be the standard normal survival function and $\Phi_2^+(u, v; \rho)$ be the bivariate standard normal survival function with marginals $N(0, 1)$ and correlation ρ . Integrating (8) over the quadrant $[t_i - \mu_i, \infty] \times [t_j - \mu_j, \infty]$ gives

$$\begin{aligned} \Phi_2^+(t_i - \mu_i, t_j - \mu_j; \rho_{ij}) &= \Phi^+(t_i - \mu_i) \Phi^+(t_j - \mu_j) \\ &\quad + \phi(t_i - \mu_i) \phi(t_j - \mu_j) \sum_{k=1}^{\infty} \frac{\rho_{ij}^k}{k!} \mathcal{H}_{k-1}(t_i - \mu_i) \mathcal{H}_{k-1}(t_j - \mu_j), \end{aligned}$$

where we have used the property that the integral of $\phi(t) \mathcal{H}_k(t)$ over $[u, \infty)$ is $-\phi(u) \mathcal{H}_{k-1}(u)$ for $k \geq 1$. Replacing in (4) and then (6) gives the overdispersion term (11). \square

Proof of Theorem 3. Let $F_v(u, v; \rho)$ denote the bivariate χ^2 survival function with v degrees of freedom corresponding to the density (12). Integrating (12) over the quadrant $[u, \infty] \times [v, \infty]$ gives

$$F_v(u, v; \rho_{ij}) = F_v(u) F_v(v) + f_{v+2}(u) f_{v+2}(v) \sum_{k=1}^{\infty} \frac{\rho_{ij}^k}{k!} \frac{v^2 \Gamma(v/2)}{k^2 \Gamma(v/2 + k)} \mathcal{L}_{k-1}^{(v/2)}\left(\frac{u}{2}\right) \mathcal{L}_{k-1}^{(v/2)}\left(\frac{v}{2}\right),$$

where we have used the property that the integral of $f_v(t) \mathcal{L}_k^{(v/2-1)}(t/2)$ over $[u, \infty)$ is $(v/k) f_{v+2}(u) \mathcal{L}_{k-1}^{(v/2)}(u/2)$ for $k \geq 1$. Replacing in (4) and then (6) gives the overdispersion term (13). \square

Proof of Theorem 4. Let $R \sim NB(r, p)$ denote the common parameterization of the negative binomial distribution with probability mass function

$$\text{pr}(R = k) = \frac{\Gamma(k + r)}{k! \Gamma(r)} p^r (1 - p)^k \quad (k = 0, 1, 2, \dots), \quad (\text{A1})$$

where $0 < p < 1$ and $r \geq 0$. This parameterization is related to ours by

$$\lambda = r \frac{1 - p}{p} \geq 0, \quad \omega = \frac{1}{r} > 0 \quad \Longleftrightarrow \quad r = \frac{1}{\omega}, \quad p = \frac{1}{1 + \omega \lambda}. \quad (\text{A2})$$

The case $\omega = 0$, $R \sim \text{Po}(\lambda)$, is obtained as the limit of the above negative binomial distribution as $\omega \rightarrow 0$ such that λ remains constant. The same is true for the moments of R , which are continuous functions of ω .

From (A1), $\gamma(\lambda, \omega) = \text{pr}(R > 0) = 1 - p^r = 1 - (1 + \omega \lambda)^{-1/\omega}$, which becomes $1 - \exp(-\lambda)$ in the limit as $\omega \rightarrow 0$. For the mean, we have that

$$E(\text{FDR}) = E\left(\frac{p_0 m \alpha}{R \vee 1}\right) = p_0 m \alpha \left\{ \text{pr}(R = 0) + E\left(\frac{1}{R} \mid R > 0\right) \text{pr}(R > 0) \right\},$$

where $\text{pr}(R > 0) = \gamma(\lambda, \omega)$ and $\text{pr}(R = 0) = 1 - \gamma(\lambda, \omega)$ by definition. All that remains is the conditional expectation, which for $\omega > 0$ is equal to

$$\begin{aligned} E\left(\frac{1}{R} \mid R > 0\right) &= \frac{1}{1 - p^r} \sum_{k=1}^{\infty} \frac{1}{k} \frac{\Gamma(k + r)}{k! \Gamma(r)} p^r (1 - p)^k = \frac{p^r}{1 - p^r} \sum_{k=1}^{\infty} \frac{\Gamma(k + r)}{k! \Gamma(r)} \int_p^1 (1 - t)^{k-1} dt \\ &= \frac{p^r}{1 - p^r} \int_p^1 \frac{dt}{(1 - t)t^r} \sum_{k=1}^{\infty} \frac{\Gamma(k + r)}{k! \Gamma(r)} t^r (1 - t)^k = \frac{p^r}{1 - p^r} \int_p^1 \frac{1 - t^r}{(1 - t)t^r} dt. \end{aligned}$$

Replacing (A2) and making the change of variable $t = 1/(1 + \omega x)$ gives the expression for $\zeta(\lambda, \omega)$. The case $\omega = 0$ is obtained by similar calculations for the Poisson distribution or by taking the limit as $\omega \rightarrow 0$.

For the second moment, we have

$$E(\widehat{\text{FDR}}^2) = E\left\{\left(\frac{p_0 m \alpha}{R \vee 1}\right)^2\right\} = (p_0 m \alpha)^2 \left\{\text{pr}(R = 0) + E\left(\frac{1}{R^2} \mid R > 0\right) \text{pr}(R > 0)\right\}.$$

Again, we only need the conditional expectation, which for $\omega > 0$ is equal to

$$\begin{aligned} E\left(\frac{1}{R^2} \mid R > 0\right) &= \frac{1}{1 - p^r} \sum_{k=1}^{\infty} \frac{1}{k^2} \frac{\Gamma(k+r)}{k! \Gamma(r)} p^r (1-p)^k = \frac{p^r}{1 - p^r} \sum_{k=1}^{\infty} \frac{\Gamma(k+r)}{k! \Gamma(r)} \int_p^1 \frac{(1-t)^{k-1}}{k} dt \\ &= \frac{p^r}{1 - p^r} \int_p^1 \frac{dt}{(1-t)t^r} \sum_{k=1}^{\infty} \frac{1}{k} \frac{\Gamma(k+r)}{k! \Gamma(r)} t^r (1-t)^k. \end{aligned}$$

Following the argument in part (ii) of the proof, the last sum above is equal to $(1 - t^r)E\{(1/S)|S > 0\}$, where $S \sim NB(r, t)$. Then the change of variable $t = 1/(1 + \omega x)$ and replacing (A2) gives the expression for $\zeta_2(\lambda, \omega)$.

For the distribution, $\widehat{\text{FDR}}$ takes values the discrete values $a_k = p_0 m \alpha / k$ ($k = 1, 2, \dots$). If $a_{k+1} \leq x < a_k$ ($k = 1, 2, \dots$), then

$$\text{pr}(\widehat{\text{FDR}} \leq x) = P\left(\frac{p_0 m \alpha}{R \vee 1} \leq \frac{p_0 m \alpha}{k+1}\right) = \text{pr}(R \geq k+1) = 1 - \text{pr}(R \leq k) = 1 - F(k).$$

If $x \geq a_1 = p_0 m \alpha$, then x is greater than the largest value that $\widehat{\text{FDR}}$ can take, so $\text{pr}(\widehat{\text{FDR}} \leq x) = 1$.

For the quantiles, let $\varepsilon > 0$. If $q \leq 1 - F(1)$, $k = F^{-1}(1 - q)$, then

$$\begin{aligned} \text{pr}(\widehat{\text{FDR}} \leq a_{k+1}) &= \text{pr}\left(\widehat{\text{FDR}} \leq \frac{p_0 m \alpha}{k+1}\right) = \text{pr}(R \geq k+1) = 1 - F(k) \geq q, \\ \text{pr}(\widehat{\text{FDR}} \leq a_{k+1} - \varepsilon) &= \text{pr}\left(\widehat{\text{FDR}} \leq \frac{p_0 m \alpha}{k+2}\right) = \text{pr}(R \geq k+2) = 1 - F(k+1) < q. \end{aligned}$$

If $q > 1 - F(1)$,

$$\begin{aligned} \text{pr}(\widehat{\text{FDR}} \leq a_1) &= \text{pr}(\widehat{\text{FDR}} \leq p_0 m \alpha) = \text{pr}(R \vee 1 \geq 1) = 1 \geq q, \\ \text{pr}(\widehat{\text{FDR}} \leq a_1 - \varepsilon) &= \text{pr}\left(\widehat{\text{FDR}} \leq \frac{p_0 m \alpha}{2}\right) = \text{pr}(R \geq 2) = 1 - F(1) < q. \end{aligned}$$

□

REFERENCES

- 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). On the adaptive control of the false discovery rate in multiple testing with independent statistics. *J. Behav. Educ. Statist.* **25**, 60–83.
- BENJAMINI, Y. & YEKUTIELI, D. (2001). The control of the false discovery rate in multiple testing under dependency. *Ann. Statist.* **29**, 1165–88.
- CROWDER, M. (1985). Gaussian estimation for correlated binomial data. *J. R. Statist. Soc. B* **47**, 229–37.
- EFRON, B. (2004). Large-scale simultaneous hypothesis testing: the choice of a null hypothesis. *J. Amer. Statist. Assoc.* **99**, 96–104.
- EFRON, B. (2007a). Correlation and large-scale simultaneous hypothesis testing. *J. Amer. Statist. Assoc.* **102**, 93–103.
- EFRON, B. (2007b). Size, power and false discovery rates. *Ann. Statist.* **35**, 1351–77.
- FARCOMENI, A. (2008). A review of modern multiple hypothesis testing, with particular attention to the false discovery proportion. *Stat. Methods. Med. Res.* **17**, 347–88.
- GENOVESE, C. R. & WASSERMAN, L. (2004). A stochastic process approach to false discovery control. *Ann. Statist.* **32**, 1035–61.
- HILBE, J. M. (2007). *Negative Binomial Regression*. Cambridge, UK: Cambridge University Press.

- JIN, J. & CAI, T. T. (2007). Estimating the null and the proportion of nonnull effects in large-scale multiple comparisons. *J. Amer. Statist. Assoc.* **102**, 495–506.
- KLEBANOV L., QIU, X. & YAKOVLEV, A. (2008). Testing differential expression in nonoverlapping gene pairs: a new perspective for the empirical Bayes method. *J. Bioinform. Comput. Biol.* **6**, 301–16.
- KOTZ, S., BALAKRISHNAN, N. & JOHNSON, N. L. (2000). Bivariate and trivariate normal distributions. In *Continuous Multivariate Distributions, Vol. 1: Models and Applications*. pp. 251–348. New York: John Wiley & Sons.
- KOUDOU, A. E. (1998). Lancaster bivariate probability distributions with Poisson, negative binomial and gamma margins. *Test* **7**, 95–110.
- MCCULLAGH, P. & NELDER, J. A. (1989). *Generalized Linear Models*, 2nd ed. Boca Raton, FL: Chapman & Hall/CRC.
- MOOTHA, V. K., LINDGREN, C. M., ERIKSSON, F. K., SUBRAMANIAN, A., SIHAG, S., LEHAR, J., PUIGSERVER, P., CARLSSON, E., RIDDERSTRAALE, M., LAURILA, E., ET AL. (2003). PGC-1-responsive genes involved in oxidative phosphorylation are coordinately downregulated in human diabetes. *Nature Genetics* **34**, 267–73.
- OWEN, A. B. (2005). Variance of the number of false discoveries. *J. R. Statist. Soc. B* **67**, 411–26.
- PATEL, J. K. & READ, C. B. (1996). *Handbook of the Normal Distribution*, 2nd ed. New York: Marcel Dekker Inc.
- PRENTICE, R. L. (1986). Binary regression using an extended beta-binomial distribution, with discussion of correlation induced by covariate measurement errors. *J. Amer. Statist. Assoc.* **81**, 321–27.
- QIU, X. & YAKOVLEV, A. (2006). Some comments on instability of false discovery rate estimation. *J. Bioinform. Comput. Biol.* **4**, 1057–68.
- SCHWARTZMAN, A. (2008). Empirical null and false discovery rate inference for exponential families. *Ann. Appl. Statist.* **2**, 1332–59.
- 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.
- YEKUTIELI, D. & BENJAMINI, Y. (1999). Resampling-based false discovery rate controlling multiple test procedures for correlated test statistics. *J. Stat. Plan. Inf.* **82**, 171–96.
- ZHOU, Y. & LIANG, H. (2000). Asymptotic normality for L_1 norm kernel estimator of conditional median under α -mixing dependence. *J. Mult. Anal.* **73**, 136–54.

[Received June 2009. Revised August 2010]