2012

Shrinkage in adaptive procedures for false discovery rate estimation in multiple testing: structure and synthesis

Debashis Ghosh, *Penn State University*

Shrinkage in adaptive procedures for false discovery rate estimation in multiple testing: structure and synthesis

Debashis Ghosh

Department of Statistics, Penn State University,

514A Wartik Lab, University Park, PA, 16802

E-mail: ghoshd@psu.edu

Abstract

There has been much interest in the study of adaptive estimation procedures for controlling the false discovery rate (FDR). In this article, we take the q-value approach of Storey (2002) and show how it can reexpressed as a particular type of shrinkage estimator. This representation leads to natural conditions on finite-sample FDR control for a general class of shrinkage estimators. In addition, many previous proposals from the literature can be unified under this framework for which finite-sample FDR results can be developed. Some asymptotic results are also provided.
Key words and phrases: Empirical Processes; Hypothesis Testing; Multiple Comparisons; Simultaneous Inference.

1 Introduction

Because of technological developments in fields such as neuroimaging and high-throughput genomics, analysts are dealing with experiments in which thousands of hypotheses are simultaneously tested. A fairly common problem is to identify interesting signals for further study. One school of thought has been to treat this signal detection problem as a multiple testing problem and to treat rejected null hypotheses as evidence for signal in the dataset. Classically, the usual quantity that has been controlled in this setting is the familywise error rate (FWER). More recently, authors have argued that control of the FWER is too stringent. Focus has shifted on methods for controlling the false discovery rate (FDR). Perhaps the most popular method for FDR control is the Benjamini-Hochberg (B-H) procedure (Benjamini and Hochberg, 1995). The B-H procedure proceeds as follows: based on the sorted p-values $p_{(1)} \le p_{(2)} \le \cdots \le p_{(n)}$, compute

$$\hat{k} = \max\{1 \le i \le n : p_{(i)} \le \alpha i/n\}.$$

Define $\hat{k}$ to be zero if the set is empty. If $\hat{k}$ is nonzero, then reject null hypotheses corresponding to $p_{(1)} \leq \cdots \leq p_{(\hat{k})}$; otherwise, reject nothing.

One finding of Benjamini and Hochberg (1995) was in the setting where the test statistics for testing n hypotheses are independent. If n_0 of the hypotheses correspond to true null hypotheses, then the B-H procedure controls the FDR at level $n_0\alpha/n$, where α is the target FDR level. This finding suggests the possibility of estimating n_0 from the data to improve power in the B-H procedure. A large amount of literature has been devoted to developing estimators of π_0 or equivalently, m_0. A nice overview of estimators of this quantity can be found in Langaas et al. (2005).

Another line of research for extending the B-H procedure and FDR procedures more generally has been to accommodate dependence. Recent work in this area has been intense. One track of research focuses on attempting to model the dependence explicitly (e.g., Pacifico Petrone et al., 2004, Leek and Storey, 2008; Sun and Cai, 2009; Efron, 2010, Schwartzman and Lin, 2011), while a second approach has been to argue for the insensitivity to dependence of the B-H and related procedures from an asymptotic point of view (Genovese and Wasserman, 2004; Storey et al., 2004; Ferreira and Zwinderman, 2006). In Section 4, we will adopt the latter viewpoint.

This note aims to characterize adaptive FDR-controlling procedures using a shrinkage framework. While there has been some discussion in this direction (Ghosh, 2006), the framework being described here is fairly general. We come to the following conclusions:

1. One interpretation of direct estimators of the false discovery rates (Storey, 2002) is as a shrinkage estimator of the unadjusted p-value, where the shrinkage factor depends on the excess number of rejections relative to the expected number assuming a binomial model;

2. There is a shrinkage phenomenon for the adaptive FDR procedures that differs substantially from James-Stein shrinkage estimation (James and Stein, 1961);

3. The structure of the shrinkage estimator allows for development of results regarding FDR control for many existing methods in the literature in finite samples as well as asymptotically.

4. A crucial ingredient in understanding FDR control is characterization of the monotonicity

properties of the FDR.

The structure of the paper is as follows. We review the multiple testing framework and the q-value estimation procedure of Storey (2002) in Section 2. The first part of Section 3 provides equivalence of the q-value with a certain shrinkage estimator. This representation, given in (2), provides the setup for develop new technical results about finite-sample control of q-value based FDR controlling procedures. We then review many previous proposals and demonstrate their finite-sample FDR control. Asymptotic results regarding FDR control are given in Section 4. Some discussion concludes Section 5.

2 Background and Preliminaries

2.1 Multiple Testing Procedures

We wish to test a set of n hypotheses, where n_0 of them correspond to true null hypotheses. The null hypotheses $H_0^1, \ldots, H_0^n$ are tested using p-values $p_1, \ldots, p_n$. A very useful table for defining the multiple testing error rates is given in Table (1)

Table 1: Outcomes of m tests of hypotheses

	Accept	Reject	Total
True Null	U	V	n_0
True Alternative	T	S	n_1
	W	R	n

Based on Table 1, the FWER is defined as $P(V \geq 1)$. By contrast, the FDR is given by

$$FDR \equiv E\left[\frac{V}{R} \mid R > 0\right] P(R > 0)$$

or $FDR = E[V/(R \vee 1)]$, where $a \vee b$ is the maximum of a and b. Asymptotically, the FDR should be equivalent to the so-called positive false discovery rate (pFDR), defined as

$$E\left[\frac{V}{R} \mid R > 0\right].$$

An important concept in the false discovery rate literature was proposed by Storey (2002) and is termed the q-value. Roughly speaking, for an individual p-value, the q-value is the smallest pFDR level at which one would reject the null hypothesis for that given p-value. The role the q-value plays in multiple testing is akin to that played by the p-value in the situation where there is one hypothesis to be tested. We will use the following algorithm for q-value estimation, proposed in Storey et al. (2004):

1. Select a value $\lambda \in (0, 1)$.

2. Estimate the proportion of true null hypotheses π_0 as

$$\hat{\pi}_0 \equiv \hat{\pi}_0(\lambda) = \frac{W(\lambda) + 1}{n(1-\lambda)},$$

   where $W(\lambda) = n - R(\lambda)$, and $R(\lambda) = \sum_{i=1}^{n} I(p_i \leq \lambda)$.

3. Define the q-value q_i corresponding to p_i, $i = 1, \ldots, n$ as $q_i = \hat{\pi}_0 p_i \{\sum_{i=1}^{n} I(p_i \leq \lambda) \vee 1\}^{-1}$.

The rule for rejecting hypotheses based on the above approach is to reject null hypotheses where $q_i \leq \alpha$, $i = 1, \ldots, n$. An equivalent procedure is to directly estimate the FDR (Storey, 2002). In particular, one fixes $0 < t < 1$. We then estimate the FDR as

$$\widehat{FDR}_\lambda(t) = \begin{cases} \frac{n\hat{\pi}_0(\lambda)t}{R(\lambda)\vee 1}, & \text{if } t \leq \lambda \\ 1, & \text{if } t > \lambda \end{cases}$$

and determine the threshold $t_\alpha \equiv t_\alpha(\widehat{FDR}) = \sup\{0 \leq t \leq 1 : \widehat{FDR}_\lambda(t) \leq \alpha\}$. Note that we have suppressed dependence of $t_\alpha(\widehat{FDR})$ on λ. All p-values smaller than this threshold are rejected.

Define the FDR of a procedure wherein all p-values less than a threshold t are rejected by

$$FDR(t) = E\left[\frac{V(t)}{R(t) \vee 1}\right], \tag{1}$$

where $R(t)$ are the number of p-values less than t, and $V(t)$ are the number of rejected p-values that correspond to the null being true. The notation is akin to what is given in Table (1) except for the fact that we now introduce a dependence on t. Storey et al. (2004) proved the following two results, which we state here as propositions.

Proposition 1: (Lemma 2 in Storey et al., 2004) The B-H procedure with n replaced by $n\hat{\pi}_0$ is equivalent to the procedure of rejecting null hypotheses with $p_i \leq t_\alpha(\widehat{FDR})$, $i = 1, \ldots, n$.

Proposition 2: (Theorem 3 in Storey et al., 2004) If the p-values corresponding to the true null hypotheses are independent, then for $\lambda > 0$,

$$FDR(t_\alpha) \leq \alpha.$$

3 Proposed approach

3.1 Shrinkage estimation and the q-value

We now make the following observation about the quantity $q_i, (i = 1, \ldots, n)$ from the previous section, namely that its numerator can be reexpressed as

$$\begin{aligned} \hat{\pi}_0 p_i &= \left(\frac{n - R(\lambda) + 1}{n(1-\lambda)} \right) p_i \\ &= \left(\frac{n(1-\lambda) - [R(\lambda) - n\lambda] + 1}{n(1-\lambda)} \right) p_i \\ &= \left(1 - \frac{R(\lambda) - n\lambda + 1}{n(1-\lambda)} \right) p_i \\ &= \{1 - G(\lambda)\} p_i, \end{aligned} \tag{2}$$

where

$$G(\lambda) = \frac{R(\lambda) - n\lambda + 1}{n(1-\lambda)}.$$

This is precisely the numerator of $\widehat{FDR}_\lambda(p_i)$. One can thus interpret the FDR evaluated at p_i as taking the unadjusted p-value and shrinking it towards zero, where the shrinkage factor is $G(\lambda)$. We term $G(\lambda)$ the shrinkage operator. In the hypothesis testing setting, zero would represent the strongest evidence against the null hypothesis. Note that the numerator of $G(\lambda)$ is the observed number of rejections minus the expected number, assuming that $R(\lambda)$ is distributed as binomial(n, λ). Such a model would hold under the assumption that the p-values are a random sample from a Uniform$(0, 1)$ distribution. Thus, the shrinkage factor depends on the excess number of rejections relative to the expected number assuming a binomial model for $R(\lambda)$.

Remark 1. Ghosh (2006, 2009) attempted to perform shrinkage on p-values and test statistics directly based on the theory of James-Stein estimation (James and Stein, 1961). However, that theory was originally developed in the *estimation* setting and not in the *hypothesis* testing setting.

The decomposition shown in (2) shows that the type of shrinkage correction that is needed does not follow directly from James-Stein estimators.

Remark 2. All of the discussion so far has been on control of the FDR. However, there is recent interest in the use of adaptive estimators for improved power in the setting where investigators might wish to control FWER (e.g., Finner and Gontscharuk, 2009). The decomposition (2) is still relevant to that setting as well so that results presented here should be easily modifiable to the FWER setting.

Let us continue to assume that $R(\lambda)$ is distributed Binomial(n, λ). Then we can write $G(\lambda)$ in one of two ways:

$$G(\lambda) = \frac{R(\lambda) - E\{R(\lambda)\} + 1}{n - E\{R(\lambda)\}} \quad (3)$$

$$= \frac{R(\lambda) - E\{R(\lambda)\} + 1}{\lambda^{-1}\mathrm{Var}\{R(\lambda)\}}. \quad (4)$$

We will refer to (3) as the "mean-only" shrinkage estimator for the q-value and (4) as the "variance-adjusted" shrinkage estimator for the q-value. What the structure of (3) and (4) suggest is a general recipe for constructing q-values:

(a). Formulate a model for $R(\lambda)$.

(b). Calculate the mean and the variance of $R(\lambda)$.

(c). Compute $G(\lambda)$ using either formula (3) or (4).

(d). Based on the calculated $G(\lambda)$, estimate the q-values using (2).

This formulation will allow us to entertain different forms of models for $R(\lambda)$. Note that formula (3) does not easily allow for incorporation of parameters pertaining to the variance of $R(\lambda)$, while formula (4) will allow us to do so. This will be expanded upon in the next section.

3.2 Finite-sample results

In this section, we will assume that λ is fixed. For two random variables X and Y, define $X < Y$ to mean that X is stochastically smaller than Y, or equivalently, that $P(Y > x) \geq P(X > x)$ for all real-valued x. We note that $t_\alpha(F) = \sup\{0 \le t \le 1 : F(t) \le \alpha\}$ is the thresholding function used to define the FDR-controlling threshold. We first show the following lemma:

Lemma 1: Let X and Y be random variables defined on $[0, 1]$ with cumulative distribution functions F_1 and F_2, respectively. If $Y > X$, then $t_\alpha(F_2) \le t_\alpha(F_1)$.

Proof: By assumption of stochastic ordering, $F_2(x) \le F_1(x)$ for all $x \in [0, 1]$. This implies that

$$\{0 \le t \le 1 : F_2(t) \le \alpha\} \subseteq \{0 \le t \le 1 : F_1(t) \le \alpha\}.$$

Taking supremums on both sides with respect to t yields the desired result. Lemma 1 implies the following corollary.

Corollary 1: Suppose two procedures yield estimated false discovery rates of $\widehat{FDR}^1_\lambda(t)$ and $\widehat{FDR}^2_\lambda(t)$, respectively, with $\widehat{FDR}^1_\lambda(t) < \widehat{FDR}^2_\lambda(t) \quad \forall t$. If $FDR(t)$ as defined in (1) is a monotone increasing function of t, then $FDR(t^2) \le FDR(t^1)$, where $t^j = t_\alpha\{\widehat{FDR}^j(t)\}$, $j = 1, 2$.

Proof: Lemma 1 implies that $t^1 \le t^2$. Applying the FDR function to both sides of the inequality along with the assumption that it is monotone increasing implies that $FDR(t^1) \le FDR(t^2)$.

The implication of Corollary 1 is that if $FDR(t^1)$ is controlled at a level α, so too will $FDR(t^2)$.

We defer discussion of the monotone increasing assumption about $FDR(t)$ to Section 3.4. We will work with this assumption throughout the rest of the paper. With this assumption, we see that one can make assumptions on the shrinkage operator that will guarantee FDR control. They are given in the following theorem.

Theorem 1: Assume that $FDR(t)$ in (1) is a monotone increasing function of t. If $K(\lambda) < G(\lambda)$ and the p-values corresponding to the true null hypotheses are independent, then the FDR estimation procedure with $K(\lambda)$ replacing $G(\lambda)$ will control the FDR at desired level.

Proof: Going in the reverse order of the derivation of (2) will yield the following inequality: for all t

$$\{1 - K(\lambda)\}t \ge \{1 - G(\lambda)\}t \tag{5}$$

Conditional on the number of rejections, the left-hand side of (5) is the numerator of estimated FDR defined using $K(\lambda)$, while the right-hand side is the estimated FDR as defined by Storey et al. (2004) for $t \leq \lambda$. This is equivalent to the fact that the estimated FDR based on $K(\lambda)$ is stochastically larger than the corresponding quantity using $G(\lambda)$. Thus by Corollary 1, for each t,

$$t_\alpha(\widehat{FDR}_\lambda^K(t)) \leq t_\alpha(\widehat{FDR}_\lambda(t)),$$

and application of the FDR functional to both sides yields the desired result.

The intuition behind Theorem 1 is fairly obvious. The condition on $K(\lambda)$ implies that the shrinkage of the p-value towards zero will be less than the shrinkage using $G(\lambda)$. Because of this, we are building an added conservatism into our multiple testing procedure relative to the procedure of Storey et al. (2004). This is how FDR control is maintained.

3.3 Synthesis of Prior proposals

We now describe various models that satisfy the assumptions for Theorem 1 to be satisfied. While FDR has been modelled using these models, nothing had been proven regarding their ability to control FDR in a finite-sample setting.

Beta-Binomial model: An alternative to the binomial model that was used to model $R(\lambda)$ in Section 3 is the beta-binomial. Thus, suppose that $R(\lambda)$ has the following probability mass function

$$f_r(r) = \binom{n}{r} \frac{B(r+\alpha, n-r+\beta)}{B(\alpha, \beta)},$$

where α and β are real-valued parameters and $B(a, b)$ is the beta function. Because the beta-binomial can be represented as a mixture of a binomial distribution and a Beta distribution, for this model, we can easily show using the law of iterated conditional expections that the variance of $R(\lambda)$ will be greater than the variance of $R(\lambda)$ under the binomial model from Section 3.1. This implies that q-values using the Beta-binomial model will satisfy the condition for Theorem 1 to hold. Hunt et al. (2009) used this model for FDR estimation.

Negative Binomial Model: Schwartzman and Lin (2011) proposed the following mean and variance models for $R(\lambda)$:

$$E\{R(\lambda)\} = n\beta(\lambda)$$

$$\text{Var}\{R(\lambda)\} = n[\beta(\lambda) - \beta^2(\lambda)] + n(n-1)\overline{\Psi(\lambda)},$$

where $\beta(\lambda) = \sum_{i=1}^{n} P(p_i \leq \lambda)$, $\beta^2(\lambda) = \sum_{i=1}^{n} P(p_i \leq \lambda)^2$ and

$$\overline{\Psi(\lambda)} = \frac{2}{n(n-1)} \sum_{i<j} \{P(p_i \leq \lambda, p_j \leq \lambda) - P(p_i \leq \lambda)P(p_j \leq \lambda)\}. \tag{6}$$

This corresponds to using a negative binomial model for the number of rejections. Again, as with the beta-binomial model, the variance under this model is greater than than predicted under the independent binomial model. In particular, the first term on the right-hand side of (6) corresponds to the variance under the binomial model case. The second term on the right-hand side is an added overdispersion term. Thus, for fixed λ and model parameters, q-values from this procedure will satisfy FDR control as per Theorem 1. We note that much of the results that are given in Schwartzman and Lin (2011) are asymptotic in nature. Here, we are in fact proving finite-sample control of FDR for their procedure.

Generalized Binomial Models: Altham (1978) proposed two generalizations of the binomial distribution. The first was termed an additive binomial model. For $R \equiv R(\lambda)$, the probability mass function of this model is given by

$$Pr(R = r) = \binom{n}{r} p^r (1-p)^{n-r} \left(\frac{p_{11} r(r-1)}{2p^2} + \frac{p_{10} r(n-r)}{p(1-p)} + \frac{p_{00}(n-r)(n-r-1)}{2p^2} - \frac{n(n-1)}{2} + 1 \right),$$

where $p = Pr(p_i \leq \lambda)$, $p_{11} = Pr(p_i \leq \lambda, p_j \leq \lambda)$, and $p_{00} = p_1 - p_{00}$. Interestingly, Altham shows that the mean and variance of this model are $E(R) = np$ and

$$\text{Var}(R) = np(1-p) + n(n-1)[p_{11} - p^2].$$

These moments coincide with those in the negative binomial model, so the same method of moments estimation approach of Schwartzman and Lin (2011) can be used here to estimate the mean and variance. As with their approach, the variance of R consists of two terms. The first term is the variance for the usual binomial distribution, while the second term is nonnegative. This would imply that the sufficiency condition needed for Theorem 1 to hold is in effect. Consequently, we have finite-sample control of the FDR.

The second binomial model that Altham (1978) proposed is a so-called multiplicative binomial model. For this model, the probability mass function of $R \equiv R(\lambda)$ is given by

$$Pr(R = r) = \frac{\binom{n}{r} p^r (1-p)^{n-r} \theta^{r(n-r)}}{f(p, \theta, r)},$$

where $p = Pr(p_i \leq \lambda)$, and

$$f(p, \theta, r) \equiv \sum_{j=0}^{r} \binom{r}{j} p^j (1-p)^{r-j} \theta^{j(r-j)}$$

is the normalizing constant needed to guarantee that $P(R = r)$ is a proper probability mass function that sums to one. The parameter θ is a measure of the association (dependence) between the random variables; by definition, $\theta \geq 0$. Values of $\theta > 1$ correspond to negative association among the random variables $I(p \leq \lambda)$, while the reverse holds for $\theta < 1$. The model reduces to the ordinary binomial model when $\theta = 1$. For values of $\theta < 1$, we have that the model will generate FDR estimates that will achieve FDR control under Theorems 1 and 2. This is because of the fact that $\theta > 1$ will lead to overdispersion relative to the binomial model. Such a model was utilized in a copy number variant problem by Lynch et al. (2007).

Poisson models An alternative to the binomial distribution and its variants that have been considered so far would be to use a Poisson sampling model. When the event of interest is rare, it is well-known that the binomial distribution can be approximated by a Poisson random variable. We thus take $R(\lambda)$ to be Poisson(λ), where $\lambda = np$. It is obvious that the numerator of $G(\lambda)$ does not change. The denominator of $G(\lambda)$, using the (4) representation, gives n, which is greater than $n(1 - \lambda)$. Thus, using this model would also make the shrinkage factor stochastically smaller so that the result of Theorem 1 would apply.

Remark 3: Everything we describe in this section presumes that λ is not estimated and is known. In practice, this parameter has to be estimated. The finite-sample results would no longer apply, and instead we would need recourse to the results in Section 4 that rely on asymptotic arguments.

3.4 Monotonicity of FDR(t)

A crucial assumption we have made throughout the article is that $FDR(t)$ as defined in (1) is a monotone increasing function of t. The range of t is from zero to one. This assumption states that as we increase t, then the false discovery rate either stays equal or gets larger. In the setting of a genomics dataset, increasing t leads to selecting more genes as statistically significant. The assumption on (1) in Theorem 1 means that the FDR cannot decrease as the list of significant genes gets larger.

The question of when $FDR(t)$ is a monotonically increasing function of t has been explored recently by Zeisel et al. (2011). They showed for a simple example that rejecting more p-values does not necessarily correspond to an increased FDR. In addition, they provided a set of sufficient conditions for the FDR to be a monotonically increasing function of t. This is given by the following theorem (Theorem 4.1 from Zeisel et al., (2011)).

Theorem 2: Let $\mathbf{p}$ denote an $n-$dimensional vector of p-values, and let t_1 and t_2 denote two thresholding procedures for selecting statistically significant hypotheses. Define $R^{(i)}(\mathbf{p})$ as the number of rejections associated with t_i, $i = 1, 2$. We assume the following conditions:

1. The p-values are iid:

$$p_1, \ldots, p_n \overset{iid}{\sim} \pi_0 + (1 - \pi_0) f_1(p),$$

   where $f_1(p)$ is monotonically nonincreasing and differentiable.

2. $R^{(1)}(\mathbf{p}) \leq R^{(2)}(\mathbf{p})$.

Then $FDR(t_1) \leq FDR(t_2)$.

As discussed by Zeisel et al. (2011), it does not appear possible to relax the condition on $f_1(p)$ or on dependence to guarantee the monotonicity of the FDR. Intuitively, this seems reasonable, as both $V(t)$ and $R(t)$ are nondecreasing functions of t so that it is hard to guarantee anything about the mean of the ratio of the two quantities without stronger assumptions.

We note in passing that if we calculate the local false discovery rate of Efron et al. (2001) using the mixture model in Theorem 2, it is given by

$$\text{locfdr}(p) = \frac{\pi_0}{\pi_0 + (1 - \pi_0) f_1(p)}.$$

Under the assumption that $f_1(p)$ is monotonically nonincreasing, then it can be easily proven that the ranking of the p-values will coincide with the ranking of the local false discovery rates. Thus, the thresholding rules based on rejecting p-values less than a certain threshold and rejecting local false discovery rates less than a threshold will reject the same hypotheses, provided that the thresholds for the two procedures are specified so as to reject the same number of hypotheses.

One possible strategy that avoids dealing with a ratio of random variables as in the definition of FDR is to focus instead on the conditional FDR (Tsai et al., 2003). In effect, this would aim at controlling $E(V/R|R = r)$ so that the denominator quantity is treated as a constant. This strategy is beyond the scope of the current paper.

4 Asymptotic Results

For the sake of completeness, we study the asymptotic behavior of the proposed q-value estimators as the number of hypotheses goes to infinity. We define the binary latent variables $H_i, \ldots, H_n$, where $H_i = 0$ if the ith null hypothesis is true, and 1 if it is false. We make the following assumptions:

A1. n_0 and n_1 go to infinity such that $n_1/n \to \pi_0 \in (0, 1)$.

A2. $n_0^{-1} \sum_{i=1} I(p_i \leq t, H_i = 0)$ and $n_1^{-1} \sum_{i=1} I(p_i \leq t, H_i = 1)$ satisfy a Glivenko-Cantelli condition, e.g., there exist cumulative distributive functions F_0 and F_1 such that

$$\sup_{-\infty<t<\infty} \|n_0^{-1} \sum_{i=1} I(p_i \leq t, H_i = 0) - F_0(t)\| \to 0$$

and

$$\sup_{-\infty<t<\infty} \|n_1^{-1} \sum_{i=1} I(p_i \leq t, H_i = 1) - F_1(t)\| \to 0$$

almost surely.

A3. $F_0(t) \leq t$ for all t.

Conditions (A1) - (A3) correspond exactly to conditions (7)-(9) in Storey et al. (2004). Condition (A1) guarantees that the limiting fraction of true null hypotheses is nondegenerate. The second condition (A2) guarantees the convergence of the cumulative distribution functions for the p-values corresponding to true nulls and true alternatives, respectively. Finally, condition (A3) is needed to ensure that the cdf for the true null hypotheses is stochastically larger than that of a Uniform(0,1) random variable.

Our approach has been to describe probability models for $R(\lambda)$ in this paper, which induce shrinkage operators for q-value estimation. So far, we have treated all parameters in these models as known. In practice, we will have to estimate these parameters from the data. Thus, we will need the following additional assumptions:

A4. Let θ denote the parameters in the model for $R(\lambda)$. We assume that the estimates of θ, $\hat{\theta}$, converge in probability to θ^*, the least false parameter (Hjort, 1992).

A5. The shrinkage operator K corresponding to the model for $R(\lambda)$ is differentiable in θ.

Before stating the relevant theorem, we make some more definitions. Define $\hat{q}_\lambda^K(t)$ denote the q-value based on $K(\hat{\theta})$, and let $\hat{q}_\lambda(t)$ be the q-value as defined by Storey et al. (2002). Based on these assumptions, we have the following theorem.

Theorem 3: Assume that conditions (A1)-(A5) hold. Suppose that $K(\theta^*) < G(\lambda)$. Then for any $\epsilon \geq 0$,

$$\lim_{n\to\infty} \inf_{t\geq\epsilon} \{\hat{q}_\lambda^K(t) - \text{q-value}(t)\} \geq 0$$

almost surely.

Proof: By assumptions (A4) and (A5) $K(\hat{\theta})$ will converge almost surely to $K(\theta^*)$ by an application of the continuous mapping theorem. Because $K(\theta^*) < G(\lambda)$ also by assumption, this means that $\hat{q}_\lambda^K(t) > \hat{q}_\lambda(t)$ for all t. This implies that

$$\hat{q}_\lambda^K(t) - \text{q-value}(t) > \hat{q}_\lambda(t) - \text{q-value}(t)$$

for all t. Taking the infimum and limits and $n \to \infty$, we have that

$$\lim_{n\to\infty} \inf_{t\geq\epsilon} \{\hat{q}_\lambda^K(t) - \text{q-value}(t)\} \geq \lim_{n\to\infty} \inf_{t\geq\epsilon} \{\hat{q}_\lambda(t) - \text{q-value}(t)\}. \tag{7}$$

By Theorem 7 of Storey et al. (2004), the right-hand side of (7) is nonnegative almost surely. This will imply the desired result.

5 Discussion

The take-home message of this note is that models for the number of rejections that lead to overdispersion relative to the binomial model will be able to provide proper control of the false discovery rate in many situations. This insight can be used to demonstrate that many procedures proposed in the literature will have proper finite-sample and asymptotic control of FDR.

Central to the proposed development in the paper is the reinterpretation of the direct FDR estimator of Storey et al. (2004) in terms of a shrinkage operator. Characterization of the shrinkage operator then allowed for development of finite-sample results about FDR control. This framework is quite general; it is important to note that each shrinkage operator will lead to a new estimator for the proportion of true null hypotheses as well as q-values.

Acknowledgments

The author would like to thank Tom Nichols for helpful discussions. This research is supported by the National Institutes of Health.

References

Altham, P. M. E. 1978. Two generalizations of the binomial distribution. *Journal of the Royal Statistical Society. Series C* **27**, 162-167.

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

Benjamini, Y. & Yekutieli, D. 2001. The control of the false discovery rate in multiple testing under dependency. *Annals of Statistics* **29**, 1165–1188.

Efron, B. 2010. Correlated z-Values and the accuracy of large-scale statistical estimates. *Journal of the American Statistical Association* **105**, 1042 – 1055.

Efron, B., Tibshirani, R., Storey, J. D., & Tusher, V. 2001, Empirical Bayes analysis of a microarray experiment, *Journal of the American Statistical Association* **96**, 1151 – 1160.

Ferreira, J. A. & Zwinderman, A. H. 2006. On the Benjamini- Hochberg method. *Annals of Statistics* **34**, 1827 – 1849.

Finner, H. & Gontscharuk, V. 2009. Controlling the familywise error rate with plug-in estimator for the proportion of true null hypotheses. *Journal of the Royal Statistical Society B* **71**, 10311048.

Genovese, C. R. & Wasserman, L. 2004. A stochastic process approach to false discovery control. *Annals of Statistics* **66**, 1035 – 1061.

Ghosh, D. 2006. Shrunken p-values for assessing differential expression, with applications to genomic data analysis. *Biometrics* **62** , 1099 – 1106.

Ghosh, D. 2009. Empirical Bayes methods for estimation and confidence intervals in high-dimensional problems. *Statistica Sinica* **19**, 125 – 143.

Hjort, N. L. 1992. On inference in parametric survival data models. *International Statistical Review* **60**, 355 – 387.

Hunt, D. L., Cheng, C. & Pounds, S. 2009. The beta-binomial distribution for estimating the number of false rejections in microarray gene expression studies. *Computational Statistics and Data Analysis* **53**, 1688 – 1700.

James, W. & Stein, C. 1961. Estimation with quadratic loss. *Proceedings of the 4th Berkeley Symposium in Mathematical Statistics and Probability 1*, 361 - 380, Univ. California Press, Berkeley.

Langaas, M., Lindqvist, B. H. & Ferkingstad, E. 2005. Estimating the proportion of true null hypotheses, with application to DNA microarray data. *Journal of the Royal Statistical Society Series B* **67**, 555 – 572.

Leek, J. T. & Storey, J. D. 2008 A general framework for multiple testing dependence. *Proceedings of the National Academy of Sciences* **105**, 18718-18723.

Lynch, A. G., Marioni, J. C. & Tavaré, S. 2007. Numbers of copy-number variations and false-negative rates will be underestimated if we do no account for dependence between repeated experiments. *American Journal of Human Genetics* **81**, 418 – 420.

Pacifico Petrone, M., Genovese, C., Verdinelli, I. & Wasserman, L. 2004 False discovery control for random elds. *Journal of the American Statistical Association* **99**,10021014.

Schwartzman, A. & Lin, X. 2011 The effect of correlation on false discovery rate estimation. *Biometrika* **98**, 199 – 214.

Storey, J. D. 2002. A direct approach to false discovery rates. *Journal of the Royal Statistical Society Series B* **64**, 479 – 498.

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. *Journal of the Royal Statistical Society Series B* **66**, 187 – 205.

Sun, W. & Cai, T. 2009. Large-scale multiple testing under dependency. *Journal of the Royal Statistical Society Series B* **71**, 393-424.

Tsai, C. A., Hsueh, H.-M. & Chen, J. J. 2003. Estimation of false discovery rates in multiple testing: application to gene microarray data. *Biometrics* **59**, 1071 – 1081.

Zeisel, A., Zuk, O. & Domany, E. 2011. FDR control with adaptive procedures and FDR monotonicity. *Annals of Applied Statistics* **5**, 943 – 968.