Dynamic adaptive procedures that control the false discovery rate

Peter W. MacDonald and Kun Liang

University of Waterloo
200 University Ave W, Waterloo ON
N2L 3G1, Canada
e-mail: pwmacdon@uwaterloo.ca*;* kun.liang@uwaterloo.ca

Arnold Janssen

Heinrich-Heine University Düsseldorf
Mathematical Institute
Universitätststraße 1
40225 Düsseldorf, Germany
e-mail: Arnold.Janssen@uni-duesseldorf.de

Abstract: In the multiple testing problem with independent tests, the classical linear step-up procedure controls the false discovery rate (FDR) at level $\pi_0 \alpha$, where π_0 is the proportion of true null hypotheses and α is the target FDR level. Adaptive procedures can improve power by incorporating estimates of π_0, which typically rely on a tuning parameter. Fixed adaptive procedures set their tuning parameters before seeing the data and can be shown to control the FDR in finite samples. We develop theoretical results for dynamic adaptive procedures whose tuning parameters are determined by the data. We show that, if the tuning parameter is chosen according to a stopping time rule, the corresponding dynamic adaptive procedure controls the FDR in finite samples. Examples include the recently proposed right-boundary procedure and the widely used lowest-slope procedure, among others. Simulation results show that the right-boundary procedure is more powerful than other dynamic adaptive procedures under independence and mild dependence conditions. The right-boundary procedure is implemented in the Bioconductor R package `calm`.

MSC 2010 subject classifications: Primary 62G10; secondary 62F03, 62M02.
Keywords and phrases: False discovery rate, multiple hypothesis testing, simultaneous inference, martingale, stopping time.

Contents

1. Introduction

Powerful modern computers have introduced large data sets to diverse fields of research, and testing of hundreds or even thousands of hypotheses simultaneously has become commonplace in statistical applications such as genetics, neuroscience, and astronomy. Since its inception in Benjamini and Hochberg [2], the false discovery rate (FDR), the expected proportion of false positives, has been widely adopted as an error measure for such large-scale problems. Much research effort has been made to improve Benjamini and Hochberg's initial method, in particular developing efficient estimators of the FDR that lead to powerful procedures which maintain FDR control. In this paper, we provide the proof of finite sample FDR control for a large class of data-adaptive procedures. First, we briefly review the literature.

Consider the classical problem of testing m independent simultaneous null hypotheses, of which m_0 are true and $m_1 = m - m_0$ are false. Denote the associated p-values by $p_1, p_2, ..., p_m$ and the ordered p-values by $p_{(1)} \leq \cdots \leq p_{(m)}$. For $t \in [0, 1]$, define the following empirical processes [19]:

$$
\begin{aligned}
V(t) &= \#\{\text{true null } p_i : p_i \leq t\}, \\
S(t) &= \#\{\text{false null } p_i : p_i \leq t\}, \\
R(t) &= V(t) + S(t).
\end{aligned}
$$

Then the FDR at a p-value cut-off $t \in (0, 1]$ is defined as

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

For a fixed FDR threshold α, Benjamini and Hochberg [2] proposed a linear step-up FDR controlling procedure (the BH procedure) which sets the p-value cut off at $p_{(k)}$, where $k = \max\{i : p_{(i)} \leq i\alpha/m\}$. The procedure has been shown to control the FDR conservatively at level $\pi_0 \alpha$ under independence, where $\pi_0 = m_0/m$ is the proportion of true nulls [5]. To tighten the FDR control, we

could use the *adaptive procedure* that applies the BH procedure at the threshold of $\alpha/\hat{\pi}_0$, where $\hat{\pi}_0$ is preferably a conservative estimate of π_0.

Instead of finding a rejection region to control the FDR, Storey [18] proposed to estimate the FDR for a fixed rejection region. When $R(t) > 0$ and under the usual assumptions that true null p-values are independent and uniformly distributed on $(0, 1)$, a natural estimator for $FDR(t)$ arises as

$$\widehat{FDR}(t) = \frac{\hat{E}[V(t)]}{R(t)} = \frac{m\hat{\pi}_0 t}{R(t)}.$$

The FDR control and FDR estimation approaches are intricately connected. With $\hat{\pi}_0 = 1$, the BH procedure can be viewed as finding the largest p-value whose FDR estimate is below or equal to α.

For a fixed tuning parameter $\lambda \in [0, 1)$, Storey [18] proposed a widely used π_0-estimator as

$$\hat{\pi}_0(\lambda) = \frac{m - R(\lambda)}{(1 - \lambda)m}.$$

Using $\hat{\pi}_0(\lambda)$ in $\widehat{FDR}$ leads to

$$\widehat{FDR}_\lambda(t) = \frac{m\hat{\pi}_0(\lambda)t}{R(t) \vee 1},$$

and Liang and Nettleton [15] showed that $\widehat{FDR}_\lambda(t)$ is a conservative estimator of $\widehat{FDR}(t)$, i.e.,

$$E[\widehat{FDR}_\lambda(t)] \geq FDR(t).$$

To control the FDR in the adaptive procedure, it is a good practice to bound $\hat{\pi}_0$ away from zero, and Storey et al. [19] proposed an asymptotically equivalent estimator:

$$\hat{\pi}_0^*(\lambda) = \frac{m - R(\lambda) + 1}{(1 - \lambda)m}.$$

Because $\hat{\pi}_0^*(\lambda) \geq \hat{\pi}_0(\lambda)$, using $\hat{\pi}_0^*(\lambda)$ in $\widehat{FDR}$ leads to conservative estimation of the FDR. On the other hand, Storey et al. [19] showed that the adaptive procedure with $\hat{\pi}_0^*(\lambda)$ controls the FDR. Furthermore, if we use π_0-estimators that are more conservative than $\hat{\pi}_0^*(\lambda)$ in the adaptive procedures, the FDR control can also be guaranteed [15]. Such examples include the two-stage procedure of Benjamini et al. [4] and the one-stage and two-stage procedures of Blanchard and Roquain [6]. We will refer to the adaptive procedures that use fixed λ parameters as *fixed adaptive* procedures. In summary, it is well established in the literature that for fixed adaptive procedures, conservative FDR estimation and FDR control are closely related.

In practice, the selection of λ amounts to a trade-off between the bias and variance of $\hat{\pi}_0^*(\lambda)$ and should depend on the data at hand. We will refer to the adaptive procedures that use data to select λ as the *dynamic adaptive* procedures. Interestingly, Liang and Nettleton [15] showed that if λ is chosen according to

a certain stopping time rule, then conservative π_0 and FDR estimation can still be guaranteed. Examples include the lowest-slope procedure of Benjamini and Hochberg [3] and the right-boundary procedure of Liang and Nettleton [15]. In spite of their conservative estimation, it is unclear whether such procedures will still maintain FDR control. Recently, Heesen and Janssen [12] have proposed a class of weighted Storey π_0-estimators with data-dependent weights and showed that the corresponding dynamic adaptive procedures control the FDR in finite samples. However, the weight measurability condition required by Heesen and Janssen [12] is not compatible with the stopping time condition required in the lowest-slope and right-boundary procedures, for which a proof of FDR control remains elusive.

In this paper, we strive to prove the FDR control for a large class of dynamic adaptive procedures, which include the right-boundary and lowest-slope procedures as special cases. The lowest-slope procedure is historically important in the field of multiple testing and especially in the FDR literature. The lowest-slope π_0-estimator was first proposed in Hochberg and Benjamini [13] to control familywise error rate (FWER), and its idea can be traced back to Schweder and Spjøtvoll [17]. According to Benjamini [1], Benjamini and Hochberg attempted but could not show that the least-slope procedure controls the FDR and presented the non-adaptive BH procedure in Benjamini and Hochberg [2] as a result. As the earliest adaptive FDR procedure, the lowest-slope procedure is widely used, but its control of the FDR has not been theoretically established.

We organize the rest of the paper as follows. In Section 2, we show finite sample FDR control for a very general class of dynamic adaptive procedures and give specific examples of possible λ selection rules. In Section 3, we conduct simulation studies to demonstrate the advantages of dynamic adaptive procedures. Finally, we discuss the issues of identifiability, dependence, and discrete p-values, and conclude Section 4. Technical proofs are postponed until Appendix A.

2. Dynamic adaptive procedures

Throughout this paper, we will assume that the true null p-values are independent and identically distributed as $\mathrm{Unif}(0, 1)$ random variables, and are independent of the false null p-values. Under this model, arbitrary dependence is allowed among the false null p-values. This is the same condition adopted by Benjamini and Hochberg [2], Storey et al. [19], and Liang and Nettleton [15], who call it the *null independence model*. Notice that under this model, the number of true nulls m_0 is fixed. A more general model with possibly random m_0 is termed as the *basic independence model* by Heesen and Janssen [11]; note that results in the fixed m_0 model can be easily extended to the random case by conditioning on m_0 and integrating. We begin by presenting our main theoretical result.

2.1. FDR control

In this section, we will show the control of the FDR for the same class of stopping time rules used by Liang and Nettleton [15] to establish the conservative FDR

estimation. Similar as in Heesen and Janssen [11], for $0 < \kappa < 1$, we divide the unit interval into a rejection region $[0, \kappa]$ and an estimation region $[\kappa, 1]$. We will first use the p-values in the estimation region to determine the tuning parameter λ and the corresponding $\hat{\pi}_0^*(\lambda)$, then we decide the p-value rejection threshold in the rejection region. It may appear restrictive to limit the rejection threshold to be no greater than κ. In practice, we can set κ not too small, say $\kappa = \alpha$, and it will be unlikely that the above restriction will affect the final rejection threshold. We refer readers to Remark 1 of Storey et al. [19] for a more detailed justification.

We require the definition of the (forward) p-value filtration $\{\mathcal{F}_t\}_{t \in [\kappa, 1)}$, where $\mathcal{F}_t = \sigma(R(s) : \kappa < s \leq t)$. The σ-algebra $\mathcal{F}_t$ can be thought of as the information given by all the p-values located in the interval $(\kappa, t]$ plus the number of p-values no larger than κ. The λ selection rules considered are those such that λ is a stopping time with respect to $\{\mathcal{F}_t\}_{t \in [\kappa, 1)}$.

We define the following FDR estimator

$$\widehat{\mathrm{FDR}}_\lambda^*(t) = \begin{cases} \frac{m \hat{\pi}_0^*(\lambda) t}{R(t) \vee 1} & t \leq \kappa, \\ 1 & t > \kappa. \end{cases} \tag{1}$$

Furthermore, for any function $F : [0, 1] \to \mathbb{R}$, define the α-level thresholding functional by

$$t_\alpha(F) = \sup\{0 \leq t \leq 1 : F(t) \leq \alpha\}.$$

Then $t_\alpha(\widehat{\mathrm{FDR}}_\lambda^*)$ is the rejection threshold for the dynamic adaptive procedure based on λ.

As our main theoretical result, we show that the dynamic adaptive procedures control the FDR.

Theorem 1. *Under the null independence model, suppose λ is a stopping time with respect to $\{\mathcal{F}_t\}_{t \in [\kappa, 1)}$, and satisfies $0 < \kappa \leq \lambda < 1$ almost surely for a fixed constant κ. Then*

$$\mathrm{FDR}\{t_\alpha(\widehat{\mathrm{FDR}}_\lambda^*)\} \leq \alpha.$$

The proof of Theorem 1 and its required lemmas are presented in the Appendix A. Briefly, the proof of Theorem 1 relies heavily on Lemma 1, whose proof follows that of Proposition 1 of Heesen and Janssen [12]. We then construct a supermartingale (Lemma 3) and invoke the optional stopping theorem to bound the FDR below α.

The stopping time rules required in Theorem 1 form a very general class, but it is not clear how they should be constructed in practice. For illustration purpose, we will analyze existing stopping time rules in the literature and show that they can be easily modified to satisfy the conditions of Theorem 1. Through this analysis, we will also draw insight and motivate new rules.

2.2. Histogram-based λ selection rules

As an example of the stopping time λ selection rule, we begin by formally defining the *right-boundary* procedure [15]. For $k \geq 1$, consider a fixed and

finite λ candidate set $\Lambda = \{\lambda_1, \ldots, \lambda_k\}$ that divides the interval $(0, 1]$ into $k + 1$ bins with boundaries at $\lambda_0 \equiv 0 < \lambda_1 < \ldots < \lambda_k < \lambda_{k+1} \equiv 1$ such that the ith bin is $(\lambda_{i-1}, \lambda_i]$ for $i = 1, \ldots, k + 1$. This partition resembles the construction of a histogram of the p-values. Then the right-boundary procedure chooses the tuning parameter $\lambda = \lambda_j$, where

$$j = \min\{1 \le i \le k : \hat{\pi}_0^*(\lambda_i) \ge \hat{\pi}_0^*(\lambda_{i-1})\} \tag{2}$$

if this set is non-empty, and otherwise chooses $j = k$. That is, we choose λ as the right boundary of the first bin where the π_0 estimate at its right boundary is larger or equal to that at its left boundary. To ensure $\lambda \ge \kappa$, we can simply set $\lambda_1 = \kappa$, or we can require that $\lambda_i \ge \kappa$ in addition to the condition $\hat{\pi}_0^*(\lambda_i) \ge \hat{\pi}_0^*(\lambda_{i-1})$ in (2). Then, it is clear that λ is a stopping time with respect to $\{\mathcal{F}_t\}_{t \in [\kappa, 1)}$. While in this definition we define a stopping rule using the more conservative estimator $\hat{\pi}_0^*(\lambda)$, λ is still a stopping time if we substitute $\hat{\pi}_0(\lambda)$ in (2), as in the original right boundary procedure of [15]. Such a substitution will only affect the λ selection rule, and Theorem 1 will still give finite sample control of the FDR as long as the thresholding procedure uses the estimator $\widehat{\text{FDR}}_\lambda^*(t)$ defined in (1).

It is straightforward to show that the right-boundary procedure chooses the first bin whose p-value density is less or equal to its tail average. Typically, the overall p-value density shows a decreasing trend, and we want to choose a λ not too small (large) to avoid high bias (variance). By design, the right-boundary procedure is likely to stop at a bin when the expected reduction in bias is comparable to the variation of $\hat{\pi}_0^*(\lambda)$. In summary, the main idea behind the right-boundary procedure is to identify a λ that would balance the bias and variance of the corresponding π_0-estimator.

The smaller the number of λ candidates, the less sensitive the right-boundary procedure is to the change in p-value density. In the extreme case, if $k = 1$, then the right-boundary procedure reduces to choosing a fixed $\lambda = \lambda_1$. On the other hand, if we set k too large, then we risk stopping too early and choosing a small λ and its associated high positive bias in π_0 estimation. This is because at each λ candidate, there is a positive probability the procedure could stop, and checking the stopping condition too frequently will likely lead to early stop. Past simulation studies [15, 16] suggest that an equal-distance 20-bin setup is a reasonable choice for the number of tests m in the thousands.

2.3. Quantile-based λ selection rules

The histogram-based rules require the explicit specification of the λ candidates beforehand. Alternatively, we can let the data to determine the candidates by choosing λ among p-value quantiles. For example, Benjamini et al. [4] proposed the k-quantile procedure which selects $\lambda = p_{(k)}$ for some prespecified $1 \le k \le m$. They recommended that $k = \lfloor \frac{m}{2} \rfloor$ such that λ is approximately the median of the p-values.

Benjamini and Hochberg [3] proposed the lowest-slope procedure to control the FDR, which chooses the tuning parameter $\lambda = p_{(j)}$, where

$$j = \min\{2 \leq i \leq m : \hat{\pi}_0^*(p_{(i)}) > \hat{\pi}_0^*(p_{(i-1)})\}. \tag{3}$$

Comparing to (2), it is easy to see that the right-boundary and lowest-slope procedures are essentially identical except that they use different λ candidate sets. Similar as our minor modification to the right-boundary procedure so that $\lambda \geq \kappa$, we can require that $p_{(i)} \geq \kappa$ in addition to the condition $\hat{\pi}_0^*(p_{(i)}) > \hat{\pi}_0^*(p_{(i-1)})$ in (3).

Because the lowest-slope procedure checks its stopping condition at every realized p-value, it tends to stop too early and suffer high positive bias in estimating π_0. Not surprisingly, simulation studies in the literature have shown that the lowest-slope procedure is one of the most conservative and least powerful adaptive procedures, for example, see [15] and [16], among many others. In a sense, the lowest-slope procedure is penalized by the same multiplicity it tries to address. This undesirable result can be easily remedied by considering fewer stopping points, similar in spirit to choosing a reasonable number of bins for the right-boundary procedure. We analogously define a *right-boundary quantile* procedure which applies the original right-boundary procedure to an arbitrary grid of fixed quantiles of the p-value distribution. In the simulations to follow we will show that the right-boundary and right-boundary quantile procedures provide the best performance among dynamic adaptive procedures known to control the FDR in finite samples.

3. Simulations

We conducted simulation studies to evaluate the FDR control, power and m_0 estimation properties of the dynamic adaptive procedures in the literature. The candidate procedures are

– `BH`, the original step-up procedure of Benjamini and Hochberg [2];
– `ORC`, the oracle procedure by applying the BH procedure at level α/π_0, assuming known π_0;
– `RB20`, the right-boundary procedure with $\Lambda = \{0.05, 0.1, ..., 0.95\}$;
– `LSL`, the modified lowest-slope procedure;
– `RB20q`, the right-boundary procedure that considers the 20 evenly spaced p-value quantiles. More specifically, $\Lambda = \{q_{0.05}, q_{0.1}, ..., q_{0.95}\}$ where q_γ denote the γ quantile of the p-values;
– `HJW`, the weight shifting method of Heesen and Janssen [12].

The simulation settings are similar to those in Liang and Nettleton [15]. When true null p-values are independent, all the procedures considered are established to control the FDR in finite samples at level α. `BH` controls the FDR conservatively at level $\pi_0\alpha$. The finite sample control of `RB20`, `LSL` and `RB20q` all follow from Theorem 1. `HJW` is a particular example from a class of dynamic

adaptive procedures shown to control the FDR[12]. We implement HJW as described in Section 5 of [12], with fixed tuning parameters $\epsilon = 0.05$, $k = 6$, and $(\lambda_i)_{i=0}^{k} = (0.5, 0.6, 0.7, 0.8, 0.9, 0.95)$.

Simulations are based on $J = 10000$ replications, and the nominal FDR level is $\alpha = 0.05$. For each replication, $m = 10000$ one-sided tests of $H_0 : \mu = 0$ are performed, with standard normal true null statistics, and false null statistics having $N(\mu, 1)$ distribution. Effect sizes μ are set to $0.5, 1, 2$ and 4. For effect sizes larger than 4, the false null p-values are well separated from the true null p-values, and all procedures achieve full power relative to ORC.

3.1. Independent tests

Simulation results for independent test statistics are reported in Figure 1. The first row plots average realized FDR, the second the power relative to ORC, and the third the log mean-squared error (MSE) of $\hat{m}_{0,j} = \hat{\pi}_{0,j} m$, where $\{\hat{\pi}_{0,j}\}_{j=1}^{J}$ are the π_0 estimates from each of the J replications, and MSE is defined as

$$\text{MSE} = \frac{1}{J} \sum_{j=1}^{J} (\hat{m}_{0,j} - m_0)^2.$$

All procedures control the FDR below the nominal level 0.05 and see an increase in the FDR and relative power as the signal strength μ increases. RB20 and RB20q provide the greatest relative power in all settings, and this is because they have the smallest MSE of $\hat{m}_0$. When the signal strength is larger, and the optimal λ may be smaller than $\lambda_1 = 0.05$, the minimal possible value from RB20, in which case the quantile-based bins of RB20q can provide a marginal improvement over RB20 by considering smaller stopping points, similar to the RB20* procedure in Liang and Nettleton [15]. HJW, although similar in spirit to RB20, cannot achieve the same power performance since it restricts its estimation region to $[0.5, 1]$, and its right-to-left measurability condition forces it to sometimes over-weight the influence of smaller p-values in the estimation of π_0. Since it is known that ORC controls the FDR at exactly level α [5], all average realized FDR levels are corrected by the difference between the FDR of ORC and the target FDR level α.

3.2. Dependent tests

We also performed a simulation study with dependent test statistics. In particular, statistics have block auto-regressive order 1 correlation structure with block size 50 and correlation $\rho^{|i-j|}$ between the ith and jth elements in any block, and correlation coefficient $\rho = -0.9$. Block structure such as this has been used by Liang and Nettleton [15], among others, to recreate the varying positive and negative correlations expected among genes in the same biological pathway. Results are reported in Figure 2. As above, all procedures control the FDR below the nominal level 0.05 and increase in the FDR and relative power as the signal
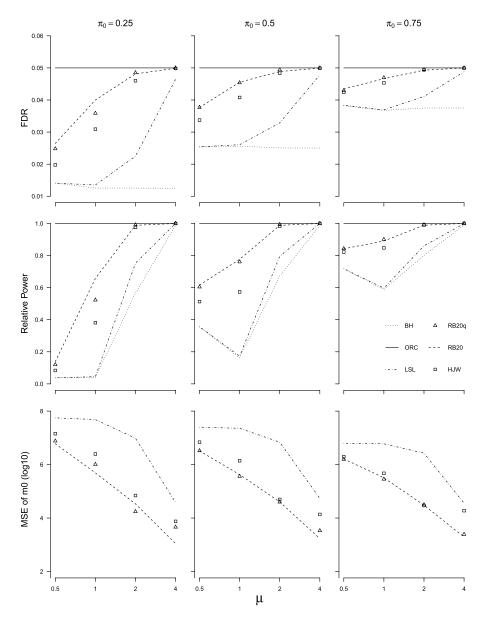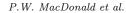
Fig 1. *Simulation results for independent test statistics.*

strength increases. `RB20` and `RB20q` remain the best in terms of power. There is some evidence that all procedures, including `ORC`, become conservative in the weak signal case, due to the dependence among the test statistics.

Liang and Nettleton [15] show that, under weak dependence conditions, the dynamic adaptive procedures working with fixed grids (e.g., the right-boundary
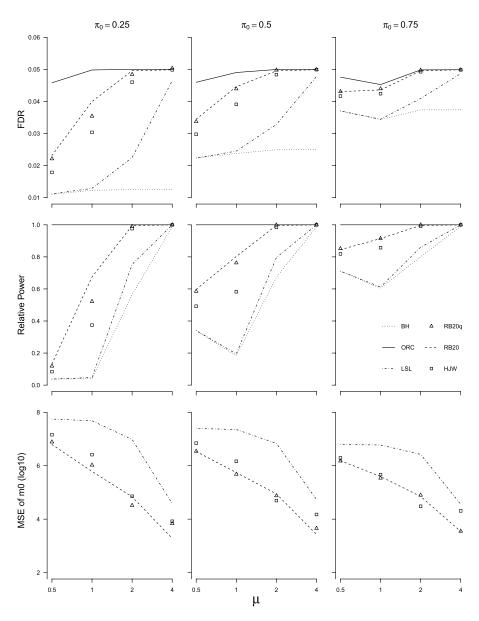
FIG 2. *Simulation results for correlated test statistics,* $\rho = -0.9$.

procedure) provide simultaneously conservative estimation and control of the FDR asymptotically. For details, see Theorems 5 and 6 of Liang and Nettleton [15]. Such theory explains why the FDR is under control in our dependent simulation setting. If we limit block sizes to be a constant or below a certain threshold and let the number of tests (m) increases, the weak dependence con-

ditions are likely to hold. This is because although tests within the same block are correlated, we will have more and more independent blocks as m increases.

4. Discussion and conclusions

4.1. Identifiability

All of the results proven in this paper give only conservative control and estimation, rather than exact control or estimation. Benjamini and Yekutieli [5], among others, have shown that the original BH procedure has the FDR exactly equal to $\pi_0\alpha$, but in the adaptive case in which we incorporate an estimate of π_0, identifiability issues manifest themselves, as discussed in Section 3.1 of Genovese and Wasserman [10].

Under the commonly used *two-group model* [9] where all p-values are independent, with random $M_0 \sim BIN(m, \pi_0)$, and when λ is selected using a stopping time rule from a fixed candidate set Λ, it can be shown that

$$\text{FDR}(t_\alpha(\widehat{\text{FDR}}_\lambda^*)) \leq \alpha \cdot \sup_{\lambda \in \Lambda} P(i\text{th null hypothesis is true} \mid p_i > \lambda).$$

There may in fact be no λ for which $P(i\text{th null hypothesis is true} \mid p_i > \lambda) = 1$, a result of the false null p-value distribution having a non-zero uniform component. This is termed *impurity* by Genovese and Wasserman [10]. Such purity issues are the reason that we cannot, without further assumptions on F_1, find an unbiased Storey-type estimator for π_0, and can only conclude conservatism.

4.2. Dependence

The results of this paper are proven under the classical null independence model, but prior FDR control literature has considered estimation and control properties under dependence assumptions on the true null p-values, in particular, the *positive regression dependence on a subset* (PRDS) condition in Benjamini and Yekutieli [5] and the *reverse martingale dependence* (RMD) condition in Heesen and Janssen [11].

Proposition 6.2 of Heesen and Janssen [11] implies that finite sample control will not hold under every PRDS or RMD model, even for fixed adaptive procedures like those described by Storey et al. [19]. Nonetheless, it may still be possible to further limit the class of models or describe an alternative dependence model such that finite sample control can be proven for adaptive or even dynamic adaptive procedures. In particular, simulation studies in this paper and Liang and Nettleton [15] motivate that finite sample control may hold under certain types of block or autoregressive dependence.

4.3. Discrete p-values

As most papers in the FDR literature, we have assumed that the true null *p*-values follow Unif(0,1). In many practical applications, discrete *p*-values are observed, and we will discuss the homogeneous and heterogeneous discrete cases separately.

In many applications, the discrete *p*-values have a set of common support points, and we call such setting as the homogeneous discrete *p*-values setting. For example, in high-throughput genetic experiments, *p*-values obtained through permutation tests have a set of identical support points. The common support provides natural grid candidates, and Liang [14] proposed the discrete right-boundary procedure that applies the idea of the right-boundary procedure to this setting and showed its conservative π_0 and FDR estimation.

For heterogeneous discrete *p*-values setting, which is also common in practice, many methods have been developed, but few have been shown to be powerful and control the FDR in finite samples. Recently, Döhler et al. [7] proved that several new procedures control the FDR and demonstrated their power in simulation studies. Adaptive procedures that incorporate π_0 estimates are suggested, but their control of the FDR has not been established and will be interesting future work.

4.4. Conclusions

For the adaptive procedures with $\hat{\pi}_0^*(\lambda)$ estimator, we show that if λ is a forward stopping time, then the FDR is controlled. We demonstrated through simulation that the right-boundary procedure (RB20) and quantile-based right-boundary procedure (RB20q) outperform the competing dynamic adaptive procedures in terms of power and estimation accuracy of π_0 while maintaining FDR control at the nominal level. In similar simulation settings in Liang and Nettleton [15], the right-boundary procedure, with estimator $\hat{\pi}_0(\lambda)$ and a different candidate set Λ than RB20, was shown to be more powerful than many other adaptive procedures, such as $\lambda = 0.5$, the median adaptive procedure and the two-stage procedures of Benjamini et al. [4], and the two-stage procedure of Blanchard and Roquain [6]. The simulation results thus far show that the right-boundary procedure is one of the most powerful adaptive procedures that controls the FDR.

Our results strengthen the connection between the FDR estimation approach and the FDR control approach. With a conservative FDR estimator, we can use the step-up procedure to find the largest *p*-value whose FDR estimate is below the target FDR level and controls the FDR as a result. This connection is the most evident for fixed adaptive procedures through the work of Storey et al. [19] and Liang and Nettleton [15]. It is further studied for certain dynamic adaptive procedures by Heesen and Janssen [11]. We extend this connection to still more dynamic adaptive procedures in this paper. The FDR estimation approach is more direct, and conservative FDR estimation much easier to establish than finite sample control of the FDR. Such insight could be useful in the future to design and evaluate new FDR controlling procedures.

Acknowledgements

Appendix A: Proofs

A.1. Lemmas for Theorem 1

We require the following three lemmas.

Lemma 1. *Under the conditions of Theorem 1, the dynamic adaptive procedure with the stopping time tuning parameter $1 > \lambda \geq \kappa > 0$ has*

$$\mathrm{FDR}\{t_\alpha(\widehat{\mathrm{FDR}}_\lambda^*)\} \leq \frac{\alpha}{\kappa} E\left[\frac{V(\kappa)}{m \cdot \hat{\pi}_0^*(\lambda)}\right].$$

Lemma 1 follows immediately from Proposition 1 of Heesen and Janssen [12]. Although their Proposition 1 is established under the basic independence model, their proof still works under our null independence model.

Alternatively, Lemma 1 can be viewed as a special case of Lemma 6.1 in Heesen and Janssen [11] by noting that our stopping time condition satisfies their condition (A1) and the null independence model is a special case of their reverse martingale model.

Lemma 2. *Suppose $X \sim BIN(n,p)$. Then*

$$E\left[\frac{1}{n - X + 1}\right] \leq \frac{1}{(n+1)(1-p)}.$$

Lemma 2 is given as Lemma 1 in [4].

Similar to the definition of the forward p-value filtration in Section 2.1, define the forward true null filtration $\{\mathcal{G}_t\}_{t \in [0,1]}$, where

$$\mathcal{G}_t = \sigma(V(s) : 0 \leq s \leq t),$$

and the σ-algebra generated by the false null p-values,

$$\mathcal{S} = \sigma(S(t) : 0 \leq t \leq 1).$$

Lemma 3. *Define the filtration $\{\mathcal{H}_t\}_{t \in [0,1]}$ by*

$$\mathcal{H}_t = \sigma(\mathcal{G}_t, \mathcal{S}).$$

Then the process

$$M(t) = \left\{\frac{1-t}{m_0 - V(t) + 1}\right\}_{t \in [0,1]}$$

is a supermartingale with respect to $\{\mathcal{H}_t\}_{t \in [0,1]}$.

Proof. Note that $0 \leq M(t) \leq 1$ a.s. for all $t \in [0,1]$, so certainly $M(t)$ is a collection of integrable random variables. Now fix $0 \leq s \leq t$. If $t = 1$, then

$$E\left[M(t)\,\Big|\,\mathcal{H}_s\right] = 0 \leq M(s).$$

Otherwise, $s \leq t < 1$, and

$$
\begin{aligned}
E\left[M(t)\,\Big|\,\mathcal{H}_s\right] &= E\left[M(t)\,\Big|\,V(s)\right] \\
&= E\left[\frac{1-t}{m_0 - V(t) + 1}\,\Big|\,V(s)\right] \\
&= E\left[\frac{1-t}{(m_0 - V(s)) - (V(t) - V(s)) + 1}\,\Big|\,V(s)\right],
\end{aligned}
$$

where the first equality follows since $\mathcal{S}$ is independent of $M(t)$ by null independence. Also by null independence, conditional on $V(s)$,

$$V(t) - V(s) \sim \mathrm{BIN}\left(m_0 - V(s), \frac{t-s}{1-s}\right).$$

Thus by Lemma 2,

$$
\begin{aligned}
E\left[M(t)\,\Big|\,\mathcal{H}_s\right] &\leq \frac{1-t}{(m_0 - V(s) + 1)(1 - \frac{t-s}{1-s})} \\
&= \frac{1-s}{m_0 - V(s) + 1} \\
&= M(s). \quad \square
\end{aligned}
$$

A.2. Proof of Theorem 1

Proof. By Lemma 1, it follows that

$$\mathrm{FDR}\{t_\alpha(\widehat{\mathrm{FDR}}_\lambda^*)\} \leq \alpha E\left[\frac{V(\kappa)}{\kappa m \hat{\pi}_0^*(\lambda)}\right],$$

and hence it suffices to show that

$$E\left[\frac{V(\kappa)}{\kappa m \hat{\pi}_0^*(\lambda)}\right] \leq 1.$$

Since $m - R(\lambda) \geq m_0 - V(\lambda)$,

$$
\begin{aligned}
E\left[\frac{V(\kappa)}{\kappa m \hat{\pi}_0^*(\lambda)}\right] &= E\left[\frac{1-\lambda}{m - R(\lambda) + 1} \cdot \frac{V(\kappa)}{\kappa}\right] \\
&\leq E\left[\frac{1-\lambda}{m_0 - V(\lambda) + 1} \cdot \frac{V(\kappa)}{\kappa}\right].
\end{aligned}
$$

$$\begin{aligned}
&= E\left\{\frac{V(\kappa)}{\kappa} \cdot E\left[\frac{1-\lambda}{m_0 - V(\lambda) + 1}\bigg|\mathcal{H}_\kappa\right]\right\} \\
&\leq E\left[\frac{1-\kappa}{m_0 - V(\kappa) + 1} \cdot \frac{V(\kappa)}{\kappa}\right] \\
&\leq 1 - \kappa^{m_0} \\
&\leq 1.
\end{aligned}$$

The third to last inequality follows from Lemma 3 and the optional stopping theorem [8]. Note that λ is a stopping time with respect to $\{\mathcal{F}_t\}_{t\in[\kappa,1)}$, and thus is also a stopping time with respect to the larger filtration $\{\mathcal{H}_t\}_{t\in[0,1]}$ to which the supermartingale in Lemma 3 is adapted. The second to last inequality follows from the binomial argument of Storey et al. [19], Theorem 3 (since κ is a fixed constant). □

References

[1] BENJAMINI, Y. (2010). Discovering the false discovery rate. *Journal of the Royal Statistical Society, Series B* **72** 405–416. MR2758522

[2] BENJAMINI, Y. and HOCHBERG, Y. (1995). Controlling the false discovery rate: a practical and powerful approach to multiple testing. *Journal of the Royal Statistical Society. Series B* **57** 289–300. MR1325392

[3] BENJAMINI, Y. and HOCHBERG, Y. (2000). On the adaptive control of the false discovery rate in multiple testing with independent statistics. *Journal of Educational and Behavioral Statistics* **25** 60–83. MR1869245

[4] BENJAMINI, Y., KRIEGER, A. M. and YEKUTIELI, D. (2006). Adaptive linear step-up procedures that control the false discovery rate. *Biometrika* **93** 491–507. MR2261438

[5] BENJAMINI, Y. and YEKUTIELI, D. (2001). The control of the false discovery rate in multiple testing under dependency. *Annals of Statistics* **29** 1165–1188. MR1869245

[6] BLANCHARD, G. and ROQUAIN, E. (2009). Adaptive false discovery rate control under independence and dependence. *Journal of Machine Learning Research* **10** 2837–2871. MR2579914

[7] DÖHLER, S., DURAND, G., ROQUAIN, E. et al. (2018). New FDR bounds for discrete and heterogeneous tests. *Electronic Journal of Statistics* **12** 1867–1900. MR3813600

[8] DOOB, J. (1953). *Stochastic Processes*. New York: Wiley. MR0058896

[9] EFRON, B., TIBSHIRANI, R., STOREY, J. D. and TUSHER, V. (2001). Empirical Bayes analysis of a microarray experiment. *Journal of the American Statistical Association* **96** 1151–1160. MR1946571

[10] GENOVESE, C. and WASSERMAN, L. (2004). A stochastic process approach to false discovery control. *Annals of Statistics* **32** 1035–1061. MR2065197

[11] HEESEN, P. and JANSSEN, A. (2015). Inequalities for the false discovery rate (FDR) under dependence. *Electronic Journal of Statistics* **9** 679–716. MR3331854

[12] HEESEN, P. and JANSSEN, A. (2016). Dynamic adaptive multiple tests with finite sample FDR control. *Journal of Statistical Planning and Inference* **168** 38–51. MR3412220

[13] HOCHBERG, Y. and BENJAMINI, Y. (1990). More powerful procedures for multiple significance testing. *Statistics in Medicine* **9** 811–818.

[14] LIANG, K. (2016). False discovery rate estimation for large-scale homogeneous discrete p-values. *Biometrics* **72** 639–648. MR3515790

[15] LIANG, K. and NETTLETON, D. (2012). Adaptive and Dynamic Adaptive Procedures for False Discovery Rate Control and Estimation. *Journal of the Royal Statistical Society, Series B* **74** 163–182. MR2885844

[16] NETTLETON, D., HWANG, J. T. G., CALDO, R. A. and WISE, R. P. (2006). Estimating the number of true null hypotheses from a histogram of p values. *Journal of Agricultural, Biological, and Environmental Statistics* **11** 337–356.

[17] SCHWEDER, T. and SPJØTVOLL, E. (1982). Plots of p-values to evaluate many tests simultaneously. *Biometrika* **69** 493–502.

[18] STOREY, J. D. (2002). A direct approach to false discovery rates. *Journal of the Royal Statistical Society, Series B* **64** 479–498. MR1924302

[19] STOREY, J. D., TAYLOR, J. E. and SIEGMUND, D. (2004). Strong control, conservative point estimation and simultaneous conservative consistency of false discovery rates: a unified approach. *Journal of the Royal Statistical Society, Series B* **66** 187–205. MR2035766