GENERALIZED STEPWISE PROCEDURES FOR
CONTROLLING THE FALSE DISCOVERY RATE

A Dissertation in
Statistics
by
Scott Roths

© 2011 Scott Roths

Submitted in Partial Fulfillment
of the Requirements
for the Degree of

Doctor of Philosophy

August 2011
The dissertation of Scott Roths was reviewed and approved* by the following:

Jogesh Babu  
Professor of Statistics  
Dissertation Advisor  
Chair of Committee  

Debashis Ghosh  
Professor of Statistics and Public Health Sciences  
Co-Chair of Committee  

Bing Li  
Professor of Statistics  

Stephen Simpson  
Professor of Mathematics  

Bruce Lindsay  
Professor of Statistics  
Department Head  

*Signatures are on file in the Graduate School.
Abstract

Among the most popular procedures used to control the false discovery rate (FDR) in large-scale multiple testing are the stepwise ones, where the marginal $p$-values are ordered and compared with specified cutoffs, according to a stopping rule. Starting with the most significant $p$-value, the stepdown procedure rejects each null hypothesis as long as the corresponding cutoff is not exceeded. The stepup procedure starts with the least significant $p$-value and proceeds in the opposite direction and accepts each null hypothesis, provided the $p$-value does not exceed its cutoff. These procedures have been shown to control the FDR under certain types of dependencies, including the kind relevant to multiple comparisons with a control. A generalization of these stepwise procedures allows the stepdown procedure to continue rejecting as long as the fraction of $p$-values not exceeding their cutoffs is sufficiently large is proposed. The stepup procedure is similarly generalized. For appropriate choices of this fraction bound, increased power may be obtained. The proposed method is illustrated with simulated and real data.
Table of Contents

List of Figures vi

Chapter 1

Literature review of multiple testing procedures 1
1.1 Introduction ............................................. 1
1.2 Error rates for multiple testing .......................... 3
  1.2.1 Traditional error rates to control .................. 4
  1.2.2 The false discovery rate ........................... 5
  1.2.3 The false discovery proportion ...................... 7
  1.2.4 Generalized error rates ............................. 8
1.3 Procedures in multiple testing ........................... 10
  1.3.1 Step-up procedures ................................ 11
  1.3.2 Step-down procedures .............................. 14
  1.3.3 Storey’s pFDR procedure .......................... 16
1.4 Asymptotic considerations ............................... 17
1.5 Outline of the dissertation .............................. 18

Chapter 2

Generalized stepwise procedures 20
2.1 Introduction ............................................. 20
2.2 Preliminaries ............................................. 21
2.3 Limiting thresholds ..................................... 24
  2.3.1 Limiting threshold of the BHP ..................... 24
  2.3.2 Limiting threshold of the proposed procedure ..... 27
2.4 FDR control .............................................. 30
  2.4.1 Recommendations for choice of γ ................. 30
    2.4.1.1 Moderate non-null distribution ............... 31
2.4.1.2 Extreme non-null distribution 32
2.4.2 Modification of $\alpha$ to control the FDR 34
2.5 Conclusions and other considerations 37

Chapter 3
Simulations 40
3.1 Introduction 40
3.2 Referential description of simulated procedures 41
3.3 Generating the data 42
3.4 Results 45
3.4.1 Small number of tests ($m = 10$) 46
3.4.2 Moderate number of tests ($m = 500$) 47
3.4.3 Large number of tests ($m = 5000$) 48

Chapter 4
Applications 60
4.1 Introduction 60
4.2 Astronomical source detection 60
4.2.1 The role of radio surveys in astronomy 61
4.2.2 Bump-hunting in radio datacubes 63
4.2.3 Statistical application 64
4.2.4 Comments 68
4.3 Genetic expression 69
4.3.1 Normalization 70
4.3.2 Comparing gene expressions among groups 72
4.4 Conclusions 72

Chapter 5
Final comments and future work 75

Appendix A
A.1 Proof of the theorem 77
A.2 Genovese and Wasserman FDP Control 81

Appendix B
Computer code 84
B.1 R code for simulations 84
B.2 R code for applications 89

Bibliography 92
# List of Figures

<table>
<thead>
<tr>
<th>Figure</th>
<th>Description</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>1.1</td>
<td>Summary of error types for a single hypothesis test</td>
<td>2</td>
</tr>
<tr>
<td>1.2</td>
<td>Summary of outcomes for $m$ hypothesis tests</td>
<td>5</td>
</tr>
<tr>
<td>1.3</td>
<td>The first row gives ranges for the index $i$, and the second row gives the relationship between $P(i)$ and its cutoff $\alpha_i$. Note that this relationship for $i \in [i_D, i_U - 1]$ may be in either direction.</td>
<td>11</td>
</tr>
<tr>
<td>1.4</td>
<td>The relationship between $\frac{1}{m-x+1}$ (curved line) and $\frac{1}{m}$ (straight line) reveals the increased power of the BHP over the Hochberg procedure.</td>
<td>13</td>
</tr>
<tr>
<td>1.5</td>
<td>The curved line represents the critical values of the Benjamini and Lui procedure, and straight line represents those of the BHP over the Hochberg procedure.</td>
<td>15</td>
</tr>
<tr>
<td>2.1</td>
<td>The solution to (2.3) $u^* \approx .028$ in this case, which is at the intersection of $F(u)$ and $\beta u$. The dashed line is the null distribution function for reference.</td>
<td>25</td>
</tr>
<tr>
<td>2.2</td>
<td>A plot of the $p$-value threshold versus number of tests $m$ for the proposed procedure. The non-null distribution is $F(u) = \sqrt{u}$, $\alpha = .1$, and $\gamma = .2$. The horizontal lines are $u'$, the limiting threshold by the theorem.</td>
<td>29</td>
</tr>
<tr>
<td>2.3</td>
<td>a) $\gamma^<em>$ versus $\alpha$ with $A_0 = .9$; the horizontal line is $1 - A_0 = .1$. b) $\gamma^</em>$ versus $A_0$ with $\alpha = .3$; the straight line is $1 - A_0$.</td>
<td>32</td>
</tr>
<tr>
<td>2.4</td>
<td>The most extreme non-null distribution $G^<em>(u) = A_0 u + 1 - A_0$ is plotted alongside other non-null distributions. The steep line is $\frac{u}{\alpha}$, and the intersections represent $u^</em>$ values. $A_0 = .8$, and $\alpha = .05$ for both cases.</td>
<td>34</td>
</tr>
<tr>
<td>2.5</td>
<td>$\alpha^*$ versus $\alpha$ (slightly curved line) bounded below by $(1 - \gamma)\alpha$ (straight line).</td>
<td>37</td>
</tr>
<tr>
<td>3.1</td>
<td>Brief description of simulated procedures by labels.</td>
<td>42</td>
</tr>
</tbody>
</table>
3.2 Estimated density plots of generated $p$-values with moderate departure from null ($\mu = 2$) and extreme departure from null ($\mu = 4$). $A_0 = .8$ and $m = 500$ in both case. 44

3.3 Results for $m = 10$ tests, independent $p$-values, $\mu = 2$, and $A_0 = .8$ 48

3.4 Results for $m = 10$ tests, independent $p$-values, $\mu = 2$, and $A_0 = .95$ 48

3.5 Results for $m = 10$ tests, independent $p$-values, $\mu = 4$, and $A_0 = .8$ 49

3.6 Results for $m = 10$ tests, independent $p$-values, $\mu = 4$, and $A_0 = .95$ 49

3.7 Results for $m = 10$ tests, dependent $p$-values, $\mu = 2$, and $A_0 = .8$ 50

3.8 Results for $m = 10$ tests, dependent $p$-values, $\mu = 2$, and $A_0 = .95$ 50

3.9 Results for $m = 10$ tests, dependent $p$-values, $\mu = 4$, and $A_0 = .8$ 51

3.10 Results for $m = 10$ tests, dependent $p$-values, $\mu = 4$, and $A_0 = .95$ 51

3.11 Results for $m = 500$ tests, independent $p$-values, $\mu = 2$, and $A_0 = .8$ 52

3.12 Results for $m = 500$ tests, independent $p$-values, $\mu = 2$, and $A_0 = .95$ 52

3.13 Results for $m = 500$ tests, independent $p$-values, $\mu = 4$, and $A_0 = .8$ 53

3.14 Results for $m = 500$ tests, independent $p$-values, $\mu = 4$, and $A_0 = .95$ 53

3.15 Results for $m = 500$ tests, dependent $p$-values, $\mu = 2$, and $A_0 = .8$ 54

3.16 Results for $m = 500$ tests, dependent $p$-values, $\mu = 2$, and $A_0 = .95$ 54

3.17 Results for $m = 500$ tests, dependent $p$-values, $\mu = 4$, and $A_0 = .8$ 55

3.18 Results for $m = 500$ tests, dependent $p$-values, $\mu = 4$, and $A_0 = .95$ 55

3.19 Results for $m = 500$ tests, dependent $p$-values, $\mu = 4$, and $A_0 = .3$ 56

3.20 Results for $m = 500$ tests, independent $p$-values, $\mu = 2$, and $A_0 = .3$ 56

3.21 Results for $m = 5000$ tests, independent $p$-values, $\mu = 2$, and $A_0 = .8$ 56

3.22 Results for $m = 5000$ tests, independent $p$-values, $\mu = 2$, and $A_0 = .95$ 57

3.23 Results for $m = 5000$ tests, independent $p$-values, $\mu = 4$, and $A_0 = .8$ 57

3.24 Results for $m = 5000$ tests, independent $p$-values, $\mu = 4$, and $A_0 = .95$ 57
3.25 Results for $m = 5000$ tests, dependent $p$-values, $\mu = 2$, and
$A_0 = .8$ .................................................. 58
3.26 Results for $m = 5000$ tests, dependent $p$-values, $\mu = 2$, and
$A_0 = .95$ .................................................. 58
3.27 Results for $m = 5000$ tests, dependent $p$-values, $\mu = 4$, and
$A_0 = .8$ .................................................. 58
3.28 Results for $m = 5000$ tests, dependent $p$-values, $\mu = 4$, and
$A_0 = .95$ .................................................. 59

4.1 Visual detection of source signals in a single channel of a datacube 61
4.2 Structure of a three-dimensional datacube ...................... 63
4.3 a) Approximate normal density of noise in Figure 4.1 with mean
$\approx 0$ and standard deviation $\approx 0.0223$ b) Datacube replaced
with $p$-values according to the normal noise in a) ............... 66
4.4 $p$-value thresholds for the datacube image in Figure 4.1 with
$\alpha = .05$ .................................................. 66
4.5 $p$-values in the datacube, discretized according to rejection for
the BenHoch (a) and the gam.su (b) ......................... 67
4.6 Sub-cube of Figure 4.1 containing the sources and the $p$-values
transformed similar to those in the first step above .......... 67
4.7 $p$-value thresholds for the data sub-cube image in Figure 4.6
with $\alpha = .05$ .................................................. 68
4.8 $p$-values in the data sub-cube, discretized according to rejection
for the gam.su procedure (a) and the BenLui procedure (b) 68
4.9 Diagnostic plots before normalization .......................... 70
4.10 Residual plots after linear regression ........................ 71
4.11 Residual plots after lowess regression ......................... 71
4.12 Residuals and histogram of studentized residuals of ANOVA
model for a representative gene ............................ 72
4.13 $p$-value thresholds for the microarray with $\alpha = .1$ ....... 73
4.14 $p$-values for the ANOVA at each gene location ............ 73
Chapter 1

Literature review of multiple testing procedures

When performing a single statistical test of two competing hypotheses, the probability of rejecting a true null hypothesis (false discovery) is traditionally controlled, and a testing procedure is distinguished by its power in rejecting the null hypothesis when it is false. When multiple tests are performed simultaneously, the quantity to control is not so easily defined, particularly when the number of tests is very large. Benjamini and Hochberg [1] define a new quantity, the false discovery rate, to control and argue that many times, it is of more scientific interest than the traditional quantity. This chapter provides a brief literature review of some of the procedures designed to control the false discovery rate.

1.1 Introduction

In a single hypothesis test, the default situation is stated as the null hypothesis \( H_0 \), and the burden of “proof” is on the alternative or non-null hypothesis \( H_a \). These hypotheses are mutually exclusive and may be stated in a variety of ways, depending on the experiment. A common example is \( H_0 : \mu = \mu_0 \) versus \( H_a : \mu \neq \mu_0 \), where \( \mu \) is the unknown mean of the population of interest,
and $\mu_0$ is some specified (known) value. From sample data representing the population, a statistic known as the *p-value* is computed and measures the strength of the evidence against $H_0$. In the entirety of this thesis, we will refer to a *p*-value corresponding to a test where $H_0$ is true by a *null p-value*. For a test where $H_0$ is false, we will use the term *non-null p-value* in similar fashion.

As a random variable—that is, in consideration before the data is actually sampled—a null *p*-value is constructed to be uniformly distributed between zero and one (or approximately uniform, depending on the sample size). A non-null *p*-value is stochastically no larger than a null *p*-value. That is, if $P^t$ and $P^f$ are a null *p*-value and a non-null *p*-value, respectively, then $\mathcal{P}(P^t \leq u) = u$, and $\mathcal{P}(P^f \leq u) \geq u$ for any $u \in [0, 1]$. Thus, the smaller an observed *p*-value is, the more “evidence” it carries that $H_0$ is false, and the problem of deciding whether to reject or to accept $H_0$ reduces to establishing a threshold, or *rejection region*, for the *p*-value.

Since the truth of $H_0$ is obviously unknown in practice, it is possible to reject $H_0$ in favor of $H_a$ when $H_0$ is actually true. This situation is called a false discovery or a *type I error*. The other possible error, where $H_0$ is accepted when it is actually false, is called a *type II error*. It should be noted that if the null hypothesis is true, then a type II error is not even defined for that test. If $H_0$ is rejected, then a type I error is made; otherwise, no error is made. These are the only two outcomes. Likewise, when $H_0$ is false, a type I error is not defined. Figure 1.1 provides a summary of the outcomes, depending on the truth of $H_0$ and whether it is rejected.

<table>
<thead>
<tr>
<th></th>
<th>$H_0$ accepted</th>
<th>$H_0$ rejected</th>
</tr>
</thead>
<tbody>
<tr>
<td>$H_0$ true</td>
<td>No error</td>
<td>Type I error (false positive)</td>
</tr>
<tr>
<td>$H_0$ false</td>
<td>Type II error</td>
<td>No error</td>
</tr>
</tbody>
</table>

*Figure 1.1. Summary of error types for a single hypothesis test*
Each of these errors carries its own consequences, and ideally, both would have negligible probabilities or rates of occurrence. Unfortunately, for a fixed sample size it is impossible to control both error rates. Controlling the type I error rate requires reducing the \( p \)-value threshold to inhibit rejection of \( H_0 \), which in turn increases the probability of accepting \( H_0 \) when it is false, the type II error. The convention is to bound the type I error rate at a level \( \alpha \) and to minimize the type II error rate subject to that constraint. The effectiveness of a hypothesis test to reject \( H_0 \) when it is false is referred to as its power. Since \( \mathcal{P}(P^t \leq u) \leq \alpha \) for \( u \in [0, \alpha] \), and \( \mathcal{P}(P^f \leq u) \) is increasing in \( u \), the solution is to set the \( p \)-value threshold to \( \alpha \). That is, \( H_0 \) is rejected in favor of \( H_a \) if the \( p \)-value is in the range \([0, \alpha]\). Otherwise, \( H_0 \) is accepted.

### 1.2 Error rates for multiple testing

When multiple hypothesis tests are considered simultaneously, the problem of error rate control is not so straightforward. Consider \( m \) hypothesis tests giving rise to \( m_0 \) null \( p \)-values and \( m - m_0 \) non-null \( p \)-values. Let \( R_i = 1 \) if the \( i \)th null hypothesis is rejected and \( R_i = 0 \) otherwise. Without loss of generality, let \( 1 \leq i \leq m_0 \) index the null \( p \)-values, and let \( m_0 + 1 \leq i \leq m \) index the non-null \( p \)-values. For a specific test, the per comparison error rate (PCER) is the probability of making a type I error (rejecting the null hypothesis when it is true). That is, for each \( 1 \leq i \leq m_0 \),

\[
\text{PCER} = \mathcal{P}(R_i = 1).
\]

This error rate applies only to a single test, and does not directly apply to the overall experiment-wise conclusion. Without correction for multiplicity, one significant result is less meaningful as more attempts are made. For example, if ten independent throws of a dart at a board are made, each with probability \( \frac{1}{10} \) of hitting the center, then one would expect a dart to hit the center. In the same way, if an experiment involving ten hypothesis tests is
performed, with each test having a PCER of $\frac{1}{10}$, then at least one null hypothesis is expected to be rejected—even if all null hypotheses are true. Thus, the probability of any false discovery certainly exceeds the PCER when all ten tests are considered simultaneously as a single experiment. Hence, the following error rates are considered.

### 1.2.1 Traditional error rates to control

Traditionally, when conducting multiple tests simultaneously, the preferred error rate to control is the family-wise error rate (FWER), the probability of making at least one false discovery among all tests. It is defined as

$$\text{FWER} = \mathbb{P}(\max\{R_1, \ldots, R_{m_0}\} = 1).$$  \hspace{1cm} (1.1)

Note the distinction between the PCER and the FWER. The former applies individually to the $i$th $p$-value, provided $1 \leq i \leq m_0$. The latter applies to the family of tests conducted simultaneously. Controlling the FWER allows individual conclusions to be aggregated to an overall conclusion, which is appropriate when the collection of tests is considered as a family. The FWER is necessarily no smaller than the PCER, which can be seen explicitly by writing for any $1 \leq i \leq m_0$,

$$\text{PCER} = \mathbb{P}(R_i = 1) \leq \mathbb{P}(\max\{R_1, \ldots, R_{m_0}\} = 1) = \text{FWER}. \hspace{1cm} (1.2)$$

So, if the FWER for a family of tests is to equal (or bounded above by) a level $\alpha$, then each individual test should be conducted at a PCER less than $\alpha$. This is like requiring the dart thrower above to stand farther away from the board if he is allowed ten throws. Then, if he hits the center in any throw at this increased distance, his accomplishment is comparable to a successful throw with only one attempt at the original distance.

As mentioned, the FWER has desirable properties and is commonly employed when the number of individual tests $m$ is small. However, the FWER
gets quite large (approaching one) as \( m_0 \leq m \) increases to infinity, which makes controlling the FWER, while maintaining reasonable power, restrictively difficult when the number of tests is large. Also, note from 1.1 that the FWER does not take into account how many non-null hypotheses are correctly rejected. It seems reasonable that a multiple testing procedure should allow many correct individual decisions to offset a limited few incorrect decisions. With this in mind, consider the following error rate.

### 1.2.2 The false discovery rate

In what follows, it is convenient to define 

\[
V = \sum_{i=1}^{m_0} R_i \quad \text{and} \quad R = \sum_{i=1}^{m} R_i
\]

as the number of false discoveries (true null hypotheses rejected) and the total number of discoveries (total rejections), respectively. According to the notation of [1], which is used often in the literature, we can summarize all the relevant quantities in a multiple-testing situation in Figure 1.2. In comparison with Figure 1.1, we have, in addition to \( V \) and \( R \) described above, that \( U \) is the number of true null hypotheses correctly accepted, \( T \) is the number of false null hypotheses incorrectly accepted, \( S \) is the number of false null hypotheses correctly rejected, and \( m_1 = m - m_0 \) is the number of false null hypotheses.

Benjamini and Hochberg [1] introduce the false discovery rate (FDR) as the expected proportion of false discoveries among all discoveries, which can be expressed in the following equivalent ways:

\[
FDR = \mathbb{E} \left[ \frac{V}{R \lor 1} \right] = \mathbb{E} \left[ \frac{V}{R} \mid R > 0 \right] P(R > 0) = \mathbb{E} \left[ \frac{V}{R} \right],
\]

where the expectation is with respect to the joint distribution of all \( p \)-values.
and the last equality makes use of the convention that \( \frac{0}{0} \equiv 0 \) in the event that no rejection is made. Benjamini and Hochberg (BH) show in [1] that the FDR is never larger than the FWER and can be much smaller if a small proportion of the tests are actually null. To see this, we can write

\[
\text{FWER} = \mathbb{E}\left[\frac{V}{\bar{V}}\right] \geq \mathbb{E}\left[\frac{V}{R}\right] = \text{FDR},
\]

where the event in the first expectation is \( \frac{1}{1} = 1 \) if any null hypothesis is rejected and \( \frac{0}{0} = 0 \) otherwise, reducing to 1.1. Thus, we have the relationship FWER \( \leq \) FDR. Although it is tempting to assume that PCER \( \leq \) FDR, this is not always the case. If \( m_0 \) is sufficiently small, and the non-null \( p \)-values are sufficiently likely to be rejected, we may have FDR < PCER. However, if all the null hypotheses are true \( (m_0 = m) \), then FDR = FWER, and so PCER \( \leq \) FDR by (1.2).

The fact that FWER = FDR when all the null hypotheses are true is of some importance. It allows FDR-controlling procedures to be used to test the intersection hypothesis: \( H_0^* \) :“all individual null hypotheses are true”, where \( H_0^* \) is rejected if and only if at least one individual null hypothesis is rejected. See, for example, [8] or [27]. In fact, the Simes test in [30], another example of the intersection test, is adopted by the authors in [1] as their FDR-controlling procedure. Not only does this procedure control the FDR when all the null hypotheses are true, which is called weak control, but it also maintains FDR control in the presence of false null hypotheses, which is called strong control. Discussion of Benjamini and Hochberg’s application of Simes’ method to FDR control occurs throughout this thesis.

As can be seen when \( m_0 < m \), the FDR tends to be smaller and therefore easier to control than the FWER, which allows for more rejections and higher power. The tradeoff, of course, is that true null hypotheses are more likely to be rejected as well, and controlling the FWER may still be preferred when the number of tests is small or when the conclusion of the whole experiment depends on the validity of every individual result. This is not always the
case, however. For example, a treatment may be compared with a control in terms of many characteristics. If the treatment excels in the vast majority of individual tests, one would like to conclude that the treatment is better even if a small proportion of the individual results are erroneous [1]. Thus, if one can make meaningful conclusions for an experiment despite rejecting some true null hypotheses, then controlling the FDR results in more power by allowing correct discoveries to outweigh false discoveries. Two applications to astronomical source detection and microarray gene expressions in Chapter 4 make use of this.

A slight variation on the FDR, originally due to [32], is to condition on the total number of rejections being strictly positive. This leads to the positive false discovery rate (pFDR):

\[ pFDR = \mathbb{E} \left[ \frac{V}{R} \Bigg| R > 0 \right]. \]

This definition is considered and summarily dismissed by Benjamini and Hochberg in [1] in favor of (1.3) because it is identically one when \( m_0 = m \) and therefore cannot be controlled—and does not reduce to the FWER—in that case. In general, we still have FDR \( \leq pFDR \). The authors in [32] argue, however, that when \( m_0 = m \), an error rate identically one is appropriate. They also show how the pFDR lends itself to a Bayesian interpretation. Other work with the pFDR can be found in [31], [34], and [33].

1.2.3 The false discovery proportion

We see from (1.3) that FDR is the expectation of the proportion of false discoveries among all discoveries. It is interesting that more attention is paid to expectation of this random variable than to the random variable itself. This ratio is referred to as the false discovery proportion (FDP):

\[ FDP = \frac{V}{R}. \]
As a random variable, it is unique among the other error quantities discussed in this section, and as such, it cannot in general be directly bounded by \( \alpha \) as the other quantities can. Rather, for \( 0 \leq \delta < 1 \), an upper \( \delta \) percentile of its distribution is considered. The relevant quantity is \( \mathcal{P}(\text{FDP} > \delta) \), which we may call the FDP\(_\delta\). For small choices of \( \delta \), the FDR is smaller than the FDP\(_\delta\), but control of one can lead to control of the other. We have FDP\(_\delta\) < \( \frac{\text{FDR}}{\delta} \) for any \( 0 \leq \delta < 1 \) by Chebyshev’s Inequality, and to go the other direction, note

\[
\text{FDR} = \mathcal{E}(\text{FDP}) = \int_0^1 \mathcal{P}(\text{FDP} > x) \, dx < \delta + (\text{FDP}\_\delta)(1 - \delta).
\]

Since the FWER can be written as \( \mathcal{P}(\text{FDP} > 0) \), we have FDP\(_\delta\) ≤ FWER. Also, if \( m_0 = m \), then FDR = FDP\(_\delta\) for any \( \delta \), and both reduce to the FWER.

Since FDP\(_\delta\) ≤ FDP\(_0\) = FWER, any procedure controlling the FWER may be directly applied to control the FDP\(_0\). For \( \delta > 0 \), such a procedure can be modified in the following way. Given a procedure that controls FWER by rejecting \( R \) \( p \)-values, then increasing the rejections to \( R \frac{1}{1-\delta} \) provides FDP\(_\delta\) control. To see this, note that the modification adds at most \( \frac{\delta}{1-\delta}R \) null \( p \)-values to the rejected pool. So, the FDP\(_\delta\) for the modified procedure is

\[
\mathcal{P}(\text{FDP} > \delta) \leq P \left( \frac{V + \frac{\delta}{1-\delta}R}{R} > \delta \right) = \mathcal{P} \left( \frac{V}{R} > 0 \right) = \text{FWER}.
\]

Of course, \( \frac{R}{1-\delta} \) would need to be rounded down if it is not an integer.

### 1.2.4 Generalized error rates

We briefly mention several other error rates that generalize those listed above. The first, due to [2], allows preference for certain individual tests whose false discovery consequences are perhaps more important. An example in clinical trials is the multiple end-points problem with primary and secondary tests. This involves introducing weights to the \( p \)-values, and the relevant error rate
becomes the \textit{weighted false discovery rate (WFDR)}:

\[
WFDR = \mathcal{E} \left[ \frac{\sum_{i=1}^{m} w_i R_i}{\sum_{i=1}^{m} w_i R_i} \right]
\]

where \( w_i \) is the weight associated with the \( i \)th \( p \)-value, and \( \sum_{i=1}^{m} w_i = m \). When all weights are equal, the WFDR reduces to the FDR in (1.3).

Other generalizations are obtained by essentially allowing \( k \) false discoveries (type I errors) without penalty and limiting errors from there. Lehmann and Romano [16] introduce the \( k \)-FWER, which is the probability of rejecting at least \( k \) null \( p \)-values:

\[
k\text{-FWER} = \mathcal{P}(V \geq k).
\]

It is easily verified that the \( k \)-FWER reduces to the FWER when \( k = 1 \) and that the \( k \)-FWER is at most as large as the FWER for \( k > 1 \). This provides a less conservative error to control when the researcher is willing to tolerate \( k \) false discoveries. Moreover, any FWER-controlling procedure can be trivially modified to control the \( k \)-FWER by rejecting an additional \( k - 1 \) \( p \)-values; the converse of this strategy does not hold. For more literature on the \( k \)-FWER or on procedures that control the \( k \)-FWER, see [27], [24], or [26].

The FDR and FDP are similarly generalized. Introduced by Sarkar [26], the \( k \)-FDP is the random variable defined as

\[
k\text{-FDP} = \left[ \frac{V}{R} \right] I(V \geq k),
\]

where \( I \) is the indicator (characteristic) function, and the \( k \)-FDR is the expectation of the \( k \)-FDP. For given \( \delta \), we have \( k\text{-FDP}_\delta = \mathcal{P}(k\text{-FDP} > \delta) \). Using similar arguments to those above in comparing the FWER with the FDP, we can show that given a procedure that rejects \( R \) \( p \)-values and controls the \( k \)-FWER, we can reject \( \frac{R-k+1}{1-\delta} \) \( p \)-values to control the \( k \)-FDP. Many other results on \( k \)-FDR and \( k \)-FWER properties, as well as procedures controlling these quantities can be found in [27], [26], and [28].
1.3 Procedures in multiple testing

For \( m \) tests with \( P_1, \ldots, P_{m_0} \) null \( p \)-values and \( P_{m_0+1}, \ldots, P_m \) non-null \( p \)-values, let \( P(1) \leq \cdots \leq P(m) \) denote the ordered values, and let \( 0 < \alpha_1 \leq \cdots \leq \alpha_m < 1 \) be a set of non-decreasing constants. Essentially, a stepwise procedure replaces each \( P(i) \) with a one or zero according to the indicator function \( I(P(i) \leq \alpha_i) \), sets a threshold, and rejects all \( p \)-values falling below this value. The manner of choosing this threshold distinguishes the two types of stepwise procedures defined next. As a matter of convenience, we say that a \( p \)-value is rejected (accepted) to mean that the null hypothesis corresponding to that \( p \)-value is rejected (accepted).

**Definition 1.** For \( \alpha_i \) and \( P(i) \) defined as above, the step-up procedure (SUP) rejects \( P(1), \ldots, P(i_U) \), where \( i_U \) is given by

\[
i_U = \max\{i : P(i) \leq \alpha_i\},
\]

and the stepdown procedure (SDP) rejects \( P(1), \ldots, P(i_D) \), where

\[
i_D = \max\{i : P(\ell) \leq \alpha_\ell \text{ for all } 1 \leq \ell \leq i\}.
\]

In either case, no \( p \)-value is rejected if the set is empty.

To justify the terms “step-up” and “step-down”, consider arranging the ordered \( p \)-values vertically with the smallest \( P(1) \) at the top. Then, the SUP starts at the bottom (largest \( p \)-value) and steps up, sequentially comparing each \( P(i) \) with \( \alpha_i \) until the first instance where the \( p \)-value does not exceed its cutoff; this occurs at \( i_U \). Then, \( P(i_u) \) and all smaller \( p \)-values are rejected. Similarly, the SDP starts at the top and steps down, comparing the ordered \( p \)-values in similar fashion until the first one falls above its cutoff; all \( p \)-values evaluated up to this point are rejected. This is crudely summarized in Figure 1.3 and in Example 1. For a horizontal treatment of these procedures, see Example 1.
<table>
<thead>
<tr>
<th>$i$</th>
<th>$[1, i_D]$</th>
<th>$i_D + 1$</th>
<th>$[i_D + 2, i_U - 1]$</th>
<th>$i_U$</th>
<th>$[i_U + 1, m]$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$P_{(i)} \leq \alpha_i$</td>
<td>yes</td>
<td>no</td>
<td>either</td>
<td>yes</td>
<td>no</td>
</tr>
</tbody>
</table>

**Figure 1.3.** The first row gives ranges for the index $i$, and the second row gives the relationship between $P_{(i)}$ and its cutoff $\alpha_i$. Note that this relationship for $i \in [i_D, i_U - 1]$ may be in either direction.

**Example 1.** For a set of $m = 12$ tests, let $I_i = I(P_{(i)} \leq \alpha_i)$ for $i = 1, \ldots, 12$, and consider the following configuration of $I_1, \ldots, I_{12}$:

1 1 1 0 1 0 1 0 1 0 0 0

According to Definition 1, $i_D = 3$ and $i_U = 9$, where the difference is due to the extra condition in (1.5). It is typical for $i_D < i_U$ unless the smallest $p$-values are quite extreme. Note that once the $p$-values are replaced with these indicators, the stepwise methods simply step through them until a stopping point is reached.

It is worth noting that both the SUP and the SDP reduce to the well-known Bonferroni adjustment for multiplicity when each $\alpha_i = \frac{\alpha}{m}$. Also, the case where each $\alpha_i = 1 - (1 - \alpha)^{1/m}$ is known as the Sidak procedure. Holland and Copenhaver [11] show that this procedure controls the FWER under independence and a certain type of positive dependence.

### 1.3.1 Step-up procedures

Although Simes [30] introduced the following step-up procedure to control the FWER when all the null hypotheses are true (weak control), Benjamini and Hochberg [1] were the first to apply it to control the FDR for any configuration of false null hypotheses (strong control), and so throughout this thesis, it is credited to the authors in [1].

**Definition 2.** For a fixed number $\alpha \in (0, 1)$ and a family of $m$ hypothesis tests with $p$-values $P_1, \ldots, P_m$, let $P_{(i)}$ denote the $i$th smallest $p$-value. The
Benjamini and Hochberg procedure, or BHP, rejects $P_{(1)}, \ldots, P_{(i^*)}$, where

$$i^* = \max \left\{ i : P_{(i)} \leq \frac{i}{m} \alpha \right\}.$$ 

If the set is empty, then no $p$-value is rejected.

The BHP is the special case of the SUP with $\alpha_i = \frac{i}{m} \alpha$, but it is a special case particularly well-suited to FDR control as will be revealed throughout this thesis. When the $p$-values are independent, the authors in [1] show that their procedure satisfies $\text{FDR} \leq \frac{m_0}{m} \alpha$ and therefore controls the FDR at level $\alpha$, albeit conservative by the factor $\frac{m_0}{m}$. Although simple to use, this procedure is remarkably efficient. In the “simple versus simple” and under independence, Genovese and Wasserman [7] show that the BHP is asymptotically ($m$ tends to infinity with $\frac{m_0}{m}$ fixed) optimal among all procedures that reject the smallest $k$ (fixed) $p$-values in that it minimizes the proportion of false non-discoveries.

Obviously, test statistics are not always independent. The usual settings where FDR control would be applied involve positively correlated test statistics. Remarkably, Benjamini and Yekutieli [5] show that the BHP maintains control of the FDR, without modification, when the $p$-values exhibit certain forms of dependence, including positive regression dependence (PRD). Examples for which this type of dependence holds include multivariate normal test statistics with nonnegative covariances [5] and unidimensional latent distributions [12]. See, [15] or [25] for further discussion on positive dependence.

The PRD property does not include all types of positively associated $p$-values, however [15]. In particular, normal test statistics for pairwise comparisons are known not to have PRD. To combat this, many modifications to the BHP have been proposed, resulting in more generality with respect to dependency structures but at the expense of power. Benjamini and Yekutieli [5] prove that conducting the BHP with $\alpha' = \alpha(\sum_{i=1}^{m} i^{-1})^{-1}$ controls the FDR at level $\frac{m_0}{m} \alpha$ under arbitrary dependence.

In light of conservative adjustments made to account for general dependencies, Benjamini, Krieger, and Yekutieli [3] investigate an adaptive step-up
procedure to increase power. Recognizing the persistent presence of $m_0$ in the FDR bounds above, they use information from the—likely large—pool of $p$-values to estimate the number of true null hypotheses $m_0$. Essentially, they perform the BHP procedure twice, the first time to estimate $m_0$, and the second time to establish the $p$-value cutoff. This effectively tightens the upper bound on FDR and increases power. However, FDR control does not generally hold under PRD as it did for the single-stage version. The authors discuss how the bias and variance of the $m_0$ estimator can lead to loss of FDR control in some cases when $p$-values are positively dependent. Thus, the class of dependence for which the adaptive procedure controls the FDR remains to be established.

In addition to the BHP, another notable step-up procedure is the Hochberg procedure [10], whose critical values are $\alpha_i = \frac{\alpha}{m-i+1}$. Since $\frac{1}{m-x+1} \leq \frac{x}{m}$ for $1 \leq x \leq m$, the critical values for the Hochberg procedure are no larger than those for the BHP. This translates to the Hochberg procedure rejecting fewer $p$-values, making it less powerful than the BHP. Moreover, this disparity in critical values increases with $m$ as Figure 1.4 illustrates. This latter tendency is particularly relevant in the discussion of asymptotic properties in the next two chapters. In its defense, the Hochberg procedure controls the FWER under independence or PRD, while the BHP fails in these regards.

![Figure 1.4](image.png)

**Figure 1.4.** The relationship between $\frac{1}{m-x+1}$ (curved line) and $\frac{x}{m}$ (straight line) reveals the increased power of the BHP over the Hochberg procedure.
Finally, if the dependence structure is completely unknown, Benjamini and Yekutieli [5] show that dividing the critical constants by \( D = \sum_{i=1}^{m} \frac{1}{i} \) is a sufficiently conservative adjustment. Specifically, they modify the BHP constants, replacing them with \( \alpha_i = \frac{i}{m}D^{-1}\alpha \), and show that FDR control is satisfied under arbitrary dependence.

### 1.3.2 Step-down procedures

The extra condition in (1.5), which requires the \( p \)-values to be sufficiently small at each step until the threshold, effectively reduces the number of rejections compared with its step-up equivalent. The drawback to this, of course, is loss of power, but often this leads to control of the more restrictive FWER or the FDR in the presence of more general dependence.

Two notable examples of step-down procedures are the Benjamini and Lui procedure [4], one of the most powerful step-down procedures, and the Holm procedure [13], which is less powerful but enjoys FWER control under arbitrary dependence. The Holm procedure is the step-down version of the Hochberg procedure in that it uses the same set of critical constants \( \alpha_i = \frac{\alpha}{m-i+1} \). For their procedure, Benjamini and Lui use the critical constants

\[
\alpha_i = 1 - \left[ 1 - \min\left( 1, \frac{m}{m-i+1}\alpha \right) \right]^{\frac{1}{m-i+1}},
\]

which are optimized for their proof that FDR is controlled under independence. They argue further that the power for this procedure can exceed that of the BHP in certain situations. Sarkar [25] later proved that its FDR control holds for PRD as well. See Figure 1.5 for a comparison between the critical values of the Benjamini and Lui procedure and those of the BHP. As in the comparison with the Hochberg procedure, the Benjamini and Lui cutoffs become more conservative for smaller \( i \), but remain higher than those of the BHP for large \( i \). This is where the potential increase in power the authors mention comes from.
If FDR control under arbitrary dependence is desired, Guo and Rao [9] offer an adjustment for the step-down procedure that provides less conservative critical values $\alpha_i$ than those proposed by Benjamini and Yekutieli in [5]. Specifically, for $m_1 = m - m_0$, the Guo and Rao procedure uses

$$\alpha_i = \frac{i}{m} \left[ \max_{1 \leq m_0 \leq m} \frac{m_0 + 1}{m} \sum_{j=1}^{m_1+1} \frac{1}{j} + \frac{m_1}{m_1 + 1} - \frac{m_1}{m} \right]^{-1} \alpha.$$ 

We saw above that a FWER procedure that rejects $R$ $p$-values can be modified to $\frac{R}{1-\delta}$ to control the FDP$_\delta$. Using this idea, Genovese and Wasserman [8] modify a step-down procedure that controls FWER and use it to control the $\delta$FDP. Their step-down procedure uses the critical values $\alpha_i = B_{1,m+1-i}(\alpha)$, where $B_{a,b}$ is the well-known beta distribution with mean $\frac{a}{a+b}$. The justification for this as a FWER-controlling procedure follows from the beta distribution of the ordered $p$-values under $H_0$ and is also discussed in [22]. With this choice of $\alpha_i$, the authors in [8] define $i_D$ according to (1.5) and finally set $\frac{i_D}{1-\delta}$ as the number of rejections to control the $\delta$FDP. In fact, the authors in [8] use different arguments to obtain this result, and details explaining how their approach reduces to this $\frac{i_D}{1-\delta}$ are given in Appendix A.

Finally, Lehmann and Romano [16] provide a direct step-down procedure
that controls the δFDP for a certain class of dependent p-values. They use critical values
\[ \alpha_i = \frac{([\delta i] + 1)\alpha}{m + [\delta i] + 1 - i}, \]
where [δi] is the largest whole number equal to or less than δi.

1.3.3 Storey’s pFDR procedure

Continuing his approach to multiple hypothesis testing, Storey controls the FDR and his pFDR by first fixing a rejection region and conservatively estimating the resulting FDR and pFDR. Then, he reverses direction and chooses the (fixed) rejection region so that the FDR estimate is below a specified threshold. The end result is an FDR-controlling procedure.

In his [31] work, Storey derives a conservative estimator for FDR as a function of a p-value threshold t. That is, \( R(t) = \#p\text{-values} \leq t \), \( V(t) \) is defined similarly, and \( FDR(t) = \mathbb{E}\left[\frac{V(t)}{R(t)}\right] \). His estimator is
\[
\hat{FDR}_\lambda(t) = \frac{\hat{\pi}_0(\lambda)t}{\{R(t) \vee 1\}/m},
\]
where \( \hat{\pi}_0(\lambda) \) is an estimate of \( \frac{m_0}{m} \) for some tuning parameter \( \lambda \in [0, 1) \). More specifically,
\[
\hat{\pi}_0(\lambda) = \frac{m - R(\lambda)}{(1 - \lambda)m}.
\]
Since the null p-values are uniformly distributed over (0, 1), roughly \( \pi_0(1 - \lambda) \) of them should fall in the interval \( (\lambda, 1] \); this yields \( [m - R(\lambda)]/m \approx \pi_0(1 - \lambda) \) with \( \mathbb{E}[\hat{\pi}_0(\lambda)] \geq \pi_0 \), provided the null p-values are stochastically greater than the non-null p-values. The idea behind the conservatism is so that \( FDR(t) \leq \alpha \) when \( \hat{FDR}(t) \leq \alpha \), and the use of \( \lambda \) serves as a balancing mechanism between bias and variance.

Storey shows in his first theorem that for any fixed \( \lambda \in [0, 1) \), \( FDR(t) \leq \hat{FDR}_\lambda(t) \). Then, he finds the largest \( t \) such that \( \hat{FDR}_\lambda(t) \leq \alpha \) so that FDR is no greater than \( \alpha \). Remarkably, the cutoff he obtains equals the cutoff in
Benjamini and Hochberg [1].

1.4 Asymptotic considerations

Frequently, statistical procedures make use of large sample theory to justify approximations in the finite applications. In the case of multiple testing considered here, the sample sizes are already involved in computation of the \( p \)-values. These computations may or may not rely on large sample theory. The asymptotic setting considered is for large \( m \), the number of tests. As discussed above, the FDR is particularly suited to large-scale testing of this nature because it does not grow prohibitively large. Fortunately, as is not always the case with sample size asymptotics, the applications involving these FDR procedures commonly have sufficiently large \( m \) to justify asymptotic results.

Genovese and Wasserman [7] were among the first to evaluate the Benjamini and Hochberg procedure (BHP) in this asymptotic setting. Their main result is to show that the BHP asymptotically rejects all \( p \)-values below \( u^* \), and they give a closed-form expression for this limiting threshold \( u^* \). The conditions for this result include holding constant \( \frac{A_0}{m} \) = proportion of true hypotheses and \( \frac{A_1}{m} \) = proportion of false hypotheses and assuming that \( p \)-values under false null hypotheses have a common distribution. These conditions are not required to control the finite-sample FDR as shown in BH’s original paper. This approach and its results are of particular interest in this thesis and are discussed more in the following chapters.

In another point of this paper, Genovese and Wasserman [7] introduce the false negative rate (FNR), defined as the proportion of false non-rejections among those null hypotheses not rejected. This brings up the question of choosing the (fixed) cutoff that minimizes expected FNR subject to control of expected FDR asymptotically (as \( m \to \infty \)). Under the same assumptions as before, they show that the optimal cutoff requires knowledge of \( A_0 \) as well as \( G \), the distribution of the \( p \)-values as a mixture of true and false hypotheses. If these unknown quantities can be well-estimated from the data, the cutoff can
be improved over that used in the BH procedure. This point will also return to the discussion in Chapter 2.

1.5 Outline of the dissertation

Among the variety of ways to address the multiplicity problem in the literature, sequential or step-wise procedures are no doubt the most popular. They are easy to perform, and they can be modified—usually with only the proper choice of $\alpha_i$—to control many if not all the error rates mentioned for positive or arbitrarily dependent $p$-values. However, amid the various ways these step-up and step-down procedures have been modified and generalized, it is interesting to note one persistent characteristic: their stopping rule requires that, until the stopping point is reached, every evaluated $p$-value must be large (step-up version) or every evaluated $p$-value must be small (step-down). This dissertation investigates a generalization of this stopping rule paradigm by allowing the stepping-up (stepping-down) to continue, provided a sufficiently large fraction of the evaluated $p$-values are large (small).

In the next chapter, we define new procedures based on this idea of a generalized stopping rule. We see in an example how such a procedure falls intermediate between the SUP and the SDP. Also, we explore the limiting behavior of these procedures as the number of tests $m$ tends to infinity by following the approach of Genovese and Wasserman [7]. The procedure of Benjamini and Hochberg [1] serves as a reference. Since the new procedures involve an additional quantity, we discuss options for choosing this quantity, based on whether the distribution of the non-null $p$-values is moderate or extreme.

In Chapter 3, we provide an assortment of simulated results comparing the new procedures with some of the prominent existing ones. Among other quantities, we report the expected number of rejections and the FDR, two primary quantities when comparing multiple procedures. Additionally, we provide a measure of variability for these quantities of interest, a sometimes
overlooked property in the literature. In Chapter 4, we apply the procedures to two examples, one in astronomy and the other in genetics, and in the last chapter we make some conclusions and mention several ways from which this work can proceed in the future.
Chapter 2

Generalized stepwise procedures

2.1 Introduction

As discussed in the previous chapter, stepwise procedures have received much attention in multiple testing situations where the false discovery rate or the false discovery proportion is the error quantity to control. In both the step-up and step-down versions, the marginal \( p \)-values are ordered and compared with specified cutoffs, according to a stopping rule. Starting with the most significant \( p \)-value, the step-down procedure (SDP) rejects each null hypothesis as long as the corresponding cutoff is not exceeded. The step-up procedure (SUP) starts with the least significant \( p \)-value and proceeds in the opposite direction and accepts each null hypothesis, provided the \( p \)-value does not exceed its cutoff. In this chapter, the proposed procedures offer a generalization by allowing the SDP to continue rejecting as long as the fraction of \( p \)-values not exceeding their cutoffs is sufficiently large. The SUP is similarly generalized.

In the next section, we define the proposed procedures as generalizations of existing step-up and step-down procedures involving a new quantity \( \gamma \), and show by example how the new procedures can fall intermediate between these. In Section 2.3, we investigate the limiting threshold for \( p \)-value rejection by following the approach in [7]. In Section 2.4, we continue the approach of [7] in terms of the limiting FDR of the BHP and of the new procedures. As
expected, the FDR control of the new procedures depends on the choice of $\gamma$, and we investigate suitable choices for this quantity in for two cases of non-null distribution. Adaptation to $\gamma$ through the error bound $\alpha$ is also explored.

2.2 Preliminaries

Let the fixed constants $0 < \alpha_1 \leq \cdots \leq \alpha_m < 1$ serve as cutoff points, and consider the SDP that rejects $P_{(1)}, \ldots, P_{(i_D)}$, where $i_D$ is defined by

$$i_D = \max \{ i : P_{(\ell)} \leq \alpha_{\ell} \text{ for all } 1 \leq \ell \leq i \}.$$

If this set is empty, then take $i_D = 0$, and no $p$-value is rejected. Since this method requires that all $p$-values do not exceed their cutoffs, it is easy to verify that the following is an equivalent way to compute $i_D$:

$$i_D = \max \left\{ i : \frac{1}{i} \sum_{j=1}^{i} I(P_{(j)} \leq \alpha_{j}) \geq 1 \right\}.$$

The proposed procedures introduce a new quantity $\gamma \in [0, 1)$, which may be specified with $\alpha$ in advance of the experiment, and choice of this value is discussed in this chapter. It will serve as the cutoff for the fraction of $p$-values not exceeding their $\alpha_i$ thresholds, starting with the smallest (most significant) $p$-value. It may be applied in two different ways, depending on whether the fraction of small $p$-values is required to be large at every step; this distinction is similar to that which differentiates the traditional step-down and step-up procedures.

**Definition 3.** For constants $0 < \alpha_1 \leq \cdots \leq \alpha_m < 1$, the $\gamma$-step-down procedure, or $\gamma$SDP, rejects $P_{(1)}, \ldots, P_{(i_D(\gamma))}$, where $i_D(\gamma)$ is defined by

$$i_D(\gamma) = \max \left\{ i : \frac{1}{i} \sum_{j=1}^{i} I(P_{(j)} \leq \alpha_{j}) \geq 1 - \gamma \text{ for all } 1 \leq \ell \leq i \right\}. \quad (2.1)$$
The $\gamma$step-up procedure, or $\gamma$SUP, rejects $P_{(1)}, \ldots, P_{(i_U(\gamma))}$, where $i_U(\gamma)$ is defined by

$$i_U(\gamma) = \max \left\{ i : \frac{1}{i} \sum_{j=1}^{i} I(P_j \leq \alpha_j) \geq 1 - \gamma \right\}.$$  

(2.2)

In each case, no $p$-value is rejected if the set is empty.

An explanation of this naming choice is in order. Although the $\gamma$SDP generalizes the SDP, the $\gamma$SUP does not generalize the SUP. The $\gamma$SDP is referred to as a “step-down” procedure specifically because of the characteristic that the fraction of small $p$-values must be large at every step of the way, while the $\gamma$SUP does not require this. This choice of terms makes sense when each $p$-value $P_{(i)}$ is considered not in terms of the indicator $I(P_{(i)} \leq \alpha_i)$ but rather in terms of the indicator corresponding to whether the fraction of these indicators exceeds $1 - \gamma$ as the example below illustrates.

**Example 2.** Continuing Example 1, let $I_1, \ldots, I_{12}$ have the following configuration:

\[
\begin{array}{ccccccccccccc}
1 & 1 & 1 & 0 & 1 & 0 & 1 & 0 & 1 & 0 & 0 & 0 \\
\end{array}
\]

Recall from Example 1 that $i_D = 3$ and $i_U = 9$. The SDP starts at the left (smallest $p$-values) and steps right until the first 0 is reached at $I_4$; thus, the SDP rejects the smallest three $p$-values. The SUP starts at the right and steps left until the first 1 is reached at $I_9$; it, therefore, rejects the smallest nine $p$-values.

For $\gamma = .3$, the values $i_D(.3)$ and $i_U(.3)$ may also be found directly from the $I_i$. However, to justify the choice of the naming conventions, replace each $I_i$ with the indicator $I(\frac{1}{\gamma} \sum_{j=1}^{i} I_j \geq 1 - .7)$ to get

\[
\begin{array}{ccccccccccccc}
1 & 1 & 1 & 1 & 1 & 1 & 0 & 1 & 0 & 0 & 0 & 0 \\
\end{array}
\]

Now, the stopping rules for the traditional stepwise procedures may be employed. We find that $i_D(.3) = 5$, and $i_U(.3) = 7$.  

It is easily seen that $i_D \leq i_D(\gamma) \leq i_U(\gamma)$ with probability one and that
all three procedures become the traditional SDP when \( \gamma = 0 \). Obviously, the case \( \gamma = 1 \) is not appropriate since the resulting procedure rejects every null hypothesis, regardless of its \( p \)-value. In fact, most values of \( \gamma \) result in too many rejections, and the error rate control is compromised. However, under certain conditions, we find that the proposed methods perform suitably well. Simulation results in Chapter 3 suggest choosing \( \gamma = \alpha \) is reasonable.

As mentioned above, both of the proposed methods generalize the SDP, even though the \( \gamma \)SUP uses a fundamental characteristic of the SUP. This raises the question of what modification is necessary to generalize the SUP. One answer is to require that the fraction of \( p \)-values not exceeding their cutoffs is sufficiently large for all \( P_i \) not rejected. Note that \( i_U \), defined in 1.4 may be written as

\[
i_U = \min \left\{ i : \frac{1}{m-i} \sum_{j=i+1}^{m} I(P_{ij} \leq \alpha_j) \leq 0 \right\},
\]

and for \( \gamma \in [0, 1) \), this may be generalized to

\[
i_U'(\gamma) = \min \left\{ i : \frac{1}{m-i} \sum_{j=i+1}^{m} I(P_{ij} \leq \alpha_j) \leq \gamma \right\}
\]

or to, with the additional condition of requiring this property at each term in the sum,

\[
i_D'(\gamma) = \max \left\{ i : \frac{1}{m-\ell} \sum_{j=\ell+1}^{m} I(P_{ij} \leq \alpha_j) \leq \gamma \text{ for all } i \leq \ell \leq m \right\}
\]

While the generalization of the SDP yielded \( i_D \leq i_D(\gamma) \leq i_U(\gamma) \), the effect here is the reverse. We have \( i_U = i_U'(0) = i_D'(0) \geq i_D'(\gamma) \geq i_U'(\gamma) \). This translates into less power because the index of the largest rejected \( p \)-value can be no greater than that for SDP. Following Example 2, if the configuration of the indicators \( I_i \) is

\[
1\ 1\ 1\ 0\ 1\ 0\ 1\ 0\ 0\ 0\ 0,
\]
then \( i_U = 9, i'_D(.3) = 7, \) and \( i'_U(.3) = 5. \) These generalizations never yield more rejections than the SUP for a given \( \alpha, \) and they are quite conservative for large \( m. \) However, they perform well when \( m \) is small or when \( \gamma \) is chosen small. They are mentioned further in the next chapters, but for the remaining of this chapter, the “proposed procedure” will refer to the \( \gamma \)SUP, where the number of rejections is given in (2.2).

### 2.3 Limiting thresholds

In this section, we investigate some asymptotic properties of the cutoff value for the proposed procedure \( \gamma \)SUP. Since the sample size of each test is already incorporated into the \( p \)-values, the asymptotic results are with respect to \( m, \) the number of tests considered. This number is usually quite large, so that asymptotic results are of practical, as well as theoretical, interest. Since we follow the approach Genovese and Wasserman take in [7], it will be helpful to briefly review their work.

#### 2.3.1 Limiting threshold of the BHP

For \( m \) tests, \( m_0 \) of which are null, let the fraction \( \frac{m_0}{m} \) converge to \( A_0 \) as \( m \) tends to infinity, and let \( F(u) \) be the common distribution of each non-null \( p \)-value. Then, Genovese and Wasserman [7] show that the cutoff, or deciding point, for the original Benjamini and Hochberg procedure (BHP), which is the SUP discussed above with \( \alpha_i = \frac{1}{m} \alpha, \) converges to a point \( u^* \) satisfying

\[
\frac{F(u^*)}{u^*} = \frac{1/\alpha - A_0}{1 - A_0} \equiv \beta. \tag{2.3}
\]

The existence and uniqueness of \( u^* \) are not guaranteed in general but are satisfied given the following assumptions the authors make: \( F \) is strictly concave, \( F'(0) > \beta, \) and \( A_0 < 1. \) These qualities are illustrated in Figure 2.1. It is also important to note that \( F(u) < \beta u \) for \( u > u^*. \)
To shed some light on these three assumptions, consider \( \max\{\alpha, 1-A_0\} \leq \frac{1}{10} \), a typical range for these quantities. Then, \( \beta \geq 91 \), and \( F(u) > \beta u \) for small \( u \) requires that the non-null \( p \)-values be extremely small, relative to the null \( p \)-values, whose marginal distribution is the identity function. If \( F(u) < \beta u \) for all \( u \), then the authors’ result does not apply. Of course, for larger values of \( 1-A_0 \) and \( \alpha \), \( F \) need not be as extreme. The authors in [7] do address a situation where \( F \) resolves to uniformity in their section on local alternatives. They show that in this case, \( u^* \) tends to zero, which is naturally expected if \( F \) becomes indistinguishable from the null distribution. It is also interesting to note that in this asymptotic framework, the increased power of the BHP over its step-down version (the SDP with \( \alpha_i = \frac{i}{m} \alpha \)) is not fully realized. This is discussed further below.

Since the BHP is the SUP discussed above with \( \alpha_i = \frac{i}{m} \alpha \), recall that it rejects \( P_{(1)}, \ldots, P_{(i^*)} \), where

\[
i^* = \max \left\{ i : P_{(i)} \leq \frac{i}{m} \alpha \right\}.
\]

(2.4)

The authors’ first theorem states that this cutoff \( P_{(i^*)} \) converges to \( u^* \). To see
why this is true, it is convenient to work in terms of $G(u) = A_0 u + (1 - A_0) F(u)$, which may be thought of as a mixture of the null (uniform) distribution $u \mapsto u$ and the non-null distribution $F$, applied to null and non-null $p$-values alike. In this case, the qualitative behavior of $G(u)$ relative to $u \alpha$ is similar to that of $F(u)$ relative to $\beta u$, and we use both in the discussion below as is convenient.

From this mixture model perspective, the empirical distribution of each $p$-value is $G_m(u) = \frac{1}{m} \sum_{j=1}^{m} I(P_j \leq u)$, and we can write $P(i) = G_m^{-1}(\frac{i}{m})$ for $i = 1, \ldots, m - 1$, and $G_m^{-1}(1) = \max_{1 \leq j \leq m} P_j$, where for $0 < u < 1$,

$$G^{-1}(u) = \inf\{y : G(y) \geq u\}$$

Applying this to (2.4) gives

$$\frac{i^*}{m} = \max \left\{ \frac{i}{m} : G_m^{-1}\left(\frac{i}{m}\right) \leq \frac{i}{m} \right\}$$

$$\xrightarrow{P} \sup \left\{ v : G^{-1}(v) \leq v \alpha \right\} = \frac{1}{\alpha} \sup \left\{ u : \frac{u}{\alpha} \leq G(u) \right\} = \frac{u^*}{\alpha}. \quad (2.5)$$

The last equality follows from the monotonicity of $G$ and from $\frac{u}{\alpha} = G(u) \iff \frac{F(u)}{u} = \beta$. The details of this result are given in [7]. Finally, we see that the cutoff for the BHP indeed converges to $u^*$:

$$P(i^*) = G_m^{-1}\left(\frac{i^*}{m}\right) \xrightarrow{P} G^{-1}\left(\frac{u^*}{\alpha}\right) = u^*, \quad (2.6)$$

where $G^{-1}$ is the inverse function of $G$, which exists since $G$ is strictly increasing. We should mention here that, although the authors in [7] show that $\frac{i^*}{m} \xrightarrow{P} \frac{u^*}{\alpha}$ and that $P(i^*) \xrightarrow{P} u^*$, they do not work in terms of $G^{-1}$, and the expressions in (2.5) and (2.6) should be viewed as intuitive interpretations of their results.

We take a similar approach to investigate the limit of the cutoff for the proposed procedure.
2.3.2 Limiting threshold of the proposed procedure

Let $\alpha_i = \frac{1}{m} \alpha$. Recall that for a fixed number of tests $m$, the $\gamma$SUP rejects the smallest $i_U(\gamma)$ $p$-values, where

$$i_U(\gamma) = \max \left\{ i : \frac{1}{i} \sum_{j=1}^{i} I \left( P(j) \leq \frac{j}{m} \alpha \right) \geq 1 - \gamma \right\}$$

The following theorem establishes the limiting threshold for the $\gamma$SUP.

**Theorem 1.** Let the (common) distribution $F$ of the non-null $p$-values have the following properties:

1. $F$ is strictly concave,
2. $F'(0) > \beta$, where $\beta$ is defined in (2.3), and
3. $A_0 < 1$.

Let $u^*$ be the unique solution to $F(u^*) = \beta u^*$. Then, the cutoff for the $\gamma$SUP converges to $G^{-1}(\frac{u^*}{\alpha(1-\gamma)}) \equiv u'$. By the monotonicity of $G^{-1}$ and for fixed $\alpha$, $u' = G^{-1}(\frac{u^*}{\alpha(1-\gamma)}) \geq G^{-1}(\frac{u^*}{\alpha}) = u^*$ with equality if and only if $\gamma = 0$. This means the limiting threshold for the $\gamma$SUP is at least as large as that for the BHP and, therefore, that the $\gamma$SUP is at least as powerful as the BHP.

Before we prove this theorem, we consider an interesting implication. By the monotonicity of $G$, $u' \geq u^*$ with equality if and only if $\gamma = 0$. That is, the cutoff for the $\gamma$SUP is at least as large as that for the BHP under the specified assumptions, with equality when $\gamma = 0$. Thus, in terms of its limiting threshold and for the setting imposed by the conditions on $F$, the $\gamma$SUP generalizes the SUP. However, as already mentioned, the $\gamma$SUP also reduces to the SDP when $\gamma = 0$. In other words, while the SUP is generally more powerful than the SDP, for the current setting they are asymptotically equal, and the $\gamma$SUP reduces to their common limit when $\gamma = 0$.

The following argument is similar to that above for the BHP. Full details of the proof are given in Appendix A. Writing $P_{(i)} = G_m^{-1}(\frac{i}{m})$ and applying
this to $i_U(\gamma)$ gives

$$i_U(\gamma) = \frac{m}{m} \max \left\{ i : \frac{1}{i} \sum_{j=1}^{i} I \left( P(j) \leq \frac{j}{m} \alpha \right) \geq 1 - \gamma \right\}$$

$$= \max \left\{ i : \frac{1}{m} \sum_{j=1}^{i} I \left( \frac{G^{-1}(\frac{j}{m}) \leq \frac{j}{m} \alpha}{} \right) \geq 1 - \gamma \right\}$$

Now, under the same conditions as those imposed by the authors in [7], $\frac{i_U(\gamma)}{m}$ converges in probability to

$$\sup \left\{ u : \int_{0}^{u} I \left( G^{-1}(t) \leq ta \right) dt \geq 1 - \gamma \right\}.$$
tributions considered in the simulations in Chapter 3 and represents a typical choice in the literature. Example 3.

It is worth mentioning that for the most part, we work with \( \alpha_i = \frac{1}{m} \alpha \) to allow for easier comparison with the BHP, the most powerful FDR-controlling procedure discussed so far. However, the theorem can potentially accommodate procedures with other critical values, provided \( \alpha_i \) converges to a function of \( u \) as \( \frac{i}{m} \) tends to \( u \). For example, the critical values of the Benjamini and Lui procedure can be written as

\[
\alpha_i = 1 - \left[ 1 - \min \left( 1, \frac{1}{1 - \frac{i}{m} + \frac{1}{m}} \right) \right]^{\frac{1}{1/m - 1/(m + i/m)}},
\]

which would converge to zero as \( m \to \infty \) with \( \frac{i}{m} \to u \). Thus, if the proposed modification with \( \gamma \) is applied to the Benjamini and Lui procedure, we expect the limiting threshold to be \( G^{-1}(0) = 0 \). This is consistent with Figure 1.5, which shows the collapsing of the Benjamini and Lui critical values towards zero as \( m \) increases. The Hochberg step-up procedure with critical values \( \alpha_i = \frac{\alpha}{m-i+1} = [m(1 - \frac{i}{m} + \frac{1}{m})]^{-1} \alpha \) exhibits similar behavior.
2.4 FDR control

We have seen that the limiting threshold for the $\gamma_{\text{SUP}}$ is at least as large as that for the BHP and, therefore, that the $\gamma_{\text{SUP}}$ is at least as powerful as the BHP. Of course, FDR control must be satisfied as well, and to address this, we follow the approach of Genovese and Wasserman [7] again.

2.4.1 Recommendations for choice of $\gamma$

Let $P_1, \ldots, P_{m_0}$ be the null $p$-values, and let $P_{m_0+1}, \ldots, P_m$ be the non-null $p$-values. For a fixed threshold $c$, the number of null rejections (false discoveries) is $\sum_{i=1}^{m_0} I(P_i \leq c) \equiv V$, and the total number of rejections is $\sum_{i=1}^{m} I(P_i \leq c) \equiv R$. Then, $\mathcal{E}(V) = m_0c$, $\mathcal{E}(R) = m_0c + (m - m_0)F(c)$, and we can write by a first-order Taylor expansion [7],

$$\text{FDR} = \mathcal{E}\left[\frac{V}{R}\right] = \frac{m_0c}{m_0c + (m - m_0)F(c)} + O\left(\frac{1}{\sqrt{m}}\right) \to \frac{A_0c}{G(c)}.$$  \hspace{1cm} (2.7)

In the case of the BHP, the limiting cutoff is $u^* = G^{-1}(\frac{\alpha}{\alpha})$, and the expression above reduces to $A_0\alpha < \alpha$. In the case of the $\gamma_{\text{SUP}}$, (2.7) becomes

$$\frac{A_0G^{-1}(\frac{u^*}{\alpha(1-\gamma)})}{G^{-1}(\frac{u^*}{\alpha(1-\gamma)})} = A_0\alpha \left[\frac{G^{-1}(\frac{u^*}{\alpha(1-\gamma)})}{G^{-1}(\frac{u^*}{\alpha(1-\gamma)})}\right].$$ \hspace{1cm} (2.8)

Ideally, we could find $\gamma > 0$ for which $\text{FDR} \leq A_0\alpha$ and enjoy FDR control for any value $A_0 < 1$. However, this is impossible—at least, in this asymptotic framework. The quantity in the brackets is necessarily greater than unity by the assumptions on $F$. Moreover, the factor by which it exceeds unity depends on $G$. Recall from (2.3) and Figure 2.1 that $F(u) \leq \beta u \iff G^{-1}(\frac{\beta}{\alpha}) \geq u$ whenever $u \geq u^*$. Here, the argument of $G$ is $\frac{u^*(\omega)}{1-\gamma}$, which exceeds $u^*(\alpha)$ for any $\gamma > 0$. Nevertheless, $A_0$ is assumed strictly less than one in practically every application in the FDR literature, and we seek to take advantage of this.
2.4.1.1 Moderate non-null distribution

If (2.8) is to be bounded by \( \alpha \), we must have

\[
A_0 G^{-1} \left( \frac{u^*}{\alpha (1 - \gamma)} \right) \leq \frac{u^*}{1 - \gamma} \iff G(u^*) \leq (1 - \gamma) G \left( \frac{u^*}{A_0 (1 - \gamma)} \right).
\]

(2.9)

A solution to (2.9) exists for sufficiently small \( \gamma > 0 \) because the right side tends to \( G(\frac{u^*}{A_0}) < \frac{u^*}{\alpha A_0} = \frac{G(u^*)}{A_0} \), which is strictly larger than \( G(u^*) \). Looking at Figure 2.1, the desired quality of \( F \) is that after it crosses the line \( \beta u \), it remains sufficiently large for \( u^* < u \leq \frac{u^*}{A_0 (1 - \gamma)} \). This implies that \( F \) does not increase too quickly. Since \( F(1) = 1 \), and \( \beta \) is potentially much greater than one, the steeper the slope of \( F \) for small \( u \), the more quickly it must level off in order to reach one at \( u = 1 \). To achieve the desired quality, \( F \) must start out moderately above \( \beta u \), cross it early, and then gradually decrease in slope until \( u = 1 \). This property is to easier satisfy if the non-null \( p \)-values only moderately deviate from the null distribution.

**Example 4.** For \( F(u) = \sqrt{u} \), we can solve for the optimal \( \gamma \), say \( \gamma^* \), analytically. Of course, the solution depends on both \( A_0 \) and \( \alpha \), and Figure 2.3 illustrates the behavior between the \( \gamma^* \) and each of these variables with the other fixed. In Figure 2.3a, \( \gamma^* \) is plotted against \( \alpha \) for fixed \( A_0 = .9 \). The horizontal line is \( 1 - A_0 = .1 \) and bounds \( \gamma^* \) below. In Figure 2.3b, \( \gamma^* \) is plotted against \( A_0 \) for fixed \( \alpha = .3 \) — this is an unusually large value for \( \alpha \), but it does not alter the qualitative nature of the plot. Again, the straight line is \( 1 - A_0 \), which now depends on the abscissa.

An interesting property is revealed in both images. The optimal \( \gamma \) to satisfy (2.9) always lies above \( 1 - A_0 \). Since (2.9) will be satisfied for any \( 0 \leq \gamma \leq \gamma^* \), we may use \( \hat{\gamma} = 1 - \hat{A}_0 \) if an estimate \( \hat{A}_0 \) of \( A_0 \) is available. Since \( \alpha \) is usually small in practice (usually less than .1), the conservatism is marginal. To re-iterate the observation in the above paragraph, taking \( \gamma = 1 - \hat{A}_0 \) is recommended only if non-null \( p \)-values are expected to deviate
slightly from the null distribution.

![Graphs](image)

**Figure 2.3.** a) $\gamma^*$ versus $\alpha$ with $A_0 = .9$; the horizontal line is $1 - A_0 = .1$. b) $\gamma^*$ versus $A_0$ with $\alpha = .3$; the straight line is $1 - A_0$.

The behavior exhibited in Figure 2.3 is not always typical. Simulation results in Chapter 3 consider a non-null distribution for which $\gamma = 1 - A_0$ does not satisfy (2.9), and FDR is therefore not controlled at level $\alpha$. In order to obtain a conservative option these situations, we consider the most extreme $F$ possible, and investigate what this implies about $\gamma$ for FDR control. The reasoning is that if $\gamma$ is chosen assuming the non-null distribution is the most extreme it can be, then for the (unknown) non-null distribution at hand, the FDR will be conservatively controlled.

### 2.4.1.2 Extreme non-null distribution

Recall that the null $p$-value distribution is $u \mapsto u$ for any $u \in [0,1]$, and that the non-null distribution is $F(u) \geq u$ for $u \in [0,1]$. Extreme (smaller) non-null $p$-values tend to arise from a distribution $F$ for which $F(u)$ is much greater than $u$. Based on the closing comments in the above subsection and on the simulation results in the next chapter, the FDR control of the proposed
procedure is increasingly compromised for increasingly extreme non-null distributions. Since, we never know exactly how extreme it will be, we therefore address the most extreme case and recommend applying the resulting solution for $\gamma$ as a conservative option in general.

Although $F$ varies among applications, it is a probability distribution bounded above by one, so that the most extreme—although not attainable in practice—case is $F^*(u) = 1$ for $u \in (0, 1]$, resulting in $G^*(u) = A_0 u + 1 - A_0$. Even though the choice $F^*$ does not strictly satisfy the assumptions as stated in the conjecture, the authors mention in [7] that their convergence result requires only a unique solution $u^*$ satisfying

1. $F(u^*) = \beta u^*$ and
2. $F'(u^*) \neq \beta$,

and we work with these relaxed conditions as well. These conditions are satisfied for $F_\mu = \Phi_\mu \circ \Phi_0^{-1}$ for any $\mu > 0$, and $F_\mu(u)$ is increasing in $\mu$ for fixed $u$. So, $F^*$ is viewed as an approximation of $F_\mu$ for large $\mu$. Likewise, $G^*(u) \approx A_0 u + (1 - A_0) F_\mu(u)$ for large $\mu$, and Figure 2.4 illustrates the tendency of $A_0 u + (1 - A_0) F_\mu(u)$ toward $G^*(u)$. Since $G^*$ produces the most rejections, the $\gamma$ satisfying 2.9 as an equality with $G^*$ is expected to provide FDR for any other $G$ as well, albeit conservatively. This claim is supported in the simulation results. Fortunately, the simple form of $G^*$ allows for an easy solution in $\gamma$, and it is easy to verify that $\gamma^* = u^* = \frac{\alpha(1 - A_0)}{1 - \alpha A_0}$ satisfies $G^*(u^*) = \frac{u^*}{\alpha}$ and (2.9).

**Example 5.** To see, qualitatively, how conservative this adjustment can be when $G$ is not the most extreme case, consider Figure 2.4, where it is compared with two other non-null distributions $\Phi_3 \circ \Phi_0^{-1}$ and $\Phi_4 \circ \Phi_0^{-1}$. Other values are $A_0 = .8$ and $\alpha = .05$. In both cases, $G^*(u)$ lies slightly above the other with this disparity more pronounced when $\mu = 4$ in Figure 2.4b. The more steeply inclined line is $\frac{u}{\alpha}$, and the two intersections represent the $u^*$ solutions. The case $\mu = 4$ is considered in Chapter 3 as the extreme case, which seems reasonable given the apparent similarity to $G^*$. \hfill $\Box$
a) $G^*(u)$ compared with $.8 + .2\Phi_3 \circ \Phi_0^{-1}(u)$  

b) $G^*(u)$ compared with $.8 + .2\Phi_4 \circ \Phi_0^{-1}(u)$

**Figure 2.4.** The most extreme non-null distribution $G^*(u) = A_0 u + 1 - A_0$ is plotted alongside other non-null distributions. The steep line is $\frac{u}{\gamma}$, and the intersections represent $u^*$ values. $A_0 = .8$, and $\alpha = .05$ for both cases.

Thus, if the non-null distribution is expected to be quite extreme, then taking $\gamma = \frac{\alpha(1 - A_0)}{1 - \alpha A_0}$ is a suitable choice. Of course, this requires knowledge or an estimate of $A_0$. If $A_0$ is to be estimated, a multiple-stage procedure like that proposed by [3] may be utilized, where the first stage is used to estimate $A_0$ without consideration of $\gamma$, and the second stage is used to determine the number of $p$-value rejections with the introduction of $\hat{A}_0$. This is taken up in Chapter 3.

### 2.4.2 Modification of $\alpha$ to control the FDR

As the previous subsection points out, the proposed procedure is expected to provide FDR control for $\gamma > 0$ for certain forms of $F$ and where $A_0 < 1$. Although this is a reasonable assumption, it is worthwhile to investigate a modification to allow for $A_0$ arbitrarily close to one. This involves conducting the procedure at a reduced $\alpha$ level, say $\alpha’$, to compensate for the anti-conservatism induced by $\gamma$.

Let $\alpha$ (fixed) be the desired FDR bound a priori. Consider performing the $\gamma$SUP with $p$-value cutoffs $\alpha_i = \frac{i}{m} \alpha’$, where $\alpha’ < \alpha$. Smaller values of $\alpha$ inhibit
rejections and ultimately compromise power, but this modification is to provide FDR control for $A_0$ arbitrarily close to one. In order to distinguish between $u^*$ and the limiting threshold obtained with $\alpha'$, we write the aforementioned $u^*$ more explicitly as $u^*(\alpha)$, and for any $0 < a < 1$, we define $u^*(a)$ as the unique solution to

$$G(u^*(a)) = \frac{u^*(a)}{a}.$$  \hfill (2.10)

Thus, by the same arguments as those above, the limiting threshold of the BHP is $u^*(\alpha) = G^{-1}(\frac{u^*(\alpha)}{\alpha})$, and for any $\alpha' \in (0, 1)$, the limiting threshold of the $\gamma$SUP procedure is expected to be $G^{-1}(\frac{u^*(\alpha')}{\alpha'(1-\gamma)}).

The task at hand is to find the largest $\alpha'$ such that FDR is bounded by $\alpha$. To satisfy (2.7) and (2.8) for any value of $A_0$, we must have (for limiting $m$)

$$\frac{G^{-1}(\frac{u^*(\alpha')}{}^{\alpha'}}{G^{-1}(\frac{1}{\alpha})} \leq \alpha \iff \frac{G^{-1}(\frac{1}{\alpha})}{\alpha} \leq \alpha \iff G^{-1}(\frac{x}{\alpha}) \leq x,$$

where $x = \frac{\alpha}{\alpha'} u^*(\alpha')$. The assumptions on $F$ require that $G^{-1}(\frac{1}{\alpha}) \leq u \iff u \leq u^*(\alpha)$, and we have the following condition for FDR control:

$$\frac{\alpha}{\alpha'} \frac{u^*(\alpha')}{1-\gamma} \leq u^*(\alpha) \iff \frac{u^*(\alpha')}{u^*(\alpha)} \leq (1-\gamma)\frac{\alpha'}{\alpha}.$$ \hfill (2.11)

The above condition is viewed as a constraint on $\alpha'$ and $\gamma$; the desired FDR bound, $\alpha$, is assumed fixed beforehand.

It should also be noted that taking $A_0 = 1$ in the statement above represents the “worst-case scenario” when seeking FDR control in that any $\gamma > 0$ immediately leads (asymptotically) to loss of FDR control unless $\alpha' < \alpha$. Unlike in the above discussion, where the difference $1-A_0 > 0$ is explicitly taken advantage of, here we seek a remedy solely through modification of $\alpha'$. Even though, $A_0 < 1$ is rightly taken for the purpose of the example below—indeed, the asymptotics of this chapter rely on $A_0 < 1$—the point is that the solution we seek cannot explicitly take advantage of it.

To begin, it is interesting to note how the bound in (2.11) relates to the
limiting thresholds of these procedures. Recall from the previous subsection that the threshold for the $\gamma$SUP, when conducted with $\alpha'$ in place of $\alpha$, is expected to converge to $G^{-1}\left(\frac{u^*(\alpha')}{(1-\gamma)}\right)$. When comparing this to the limiting threshold of the BHP, $G^{-1}\left(\frac{u^*(\alpha)}{\alpha}\right)$, we find that any gain in power requires that:

$$\frac{u^*(\alpha')}{\alpha'(1-\gamma)} \geq \frac{u^*(\alpha)}{\alpha} \iff \frac{u^*(\alpha')}{u^*(\alpha)} \geq (1-\gamma)\frac{\alpha'}{\alpha}. \quad (2.12)$$

Thus, we seek equalities in (2.11) and (2.12), with priority on preserving the former if equality can’t be attained. The following example demonstrates that the bound in (2.11) may be attained.

**Example 6.** Let $A_0 = .9$, $\gamma = .2$, and let the non-null distribution be $F(u) = \sqrt{u}$. Denote by $\alpha^*(\alpha)$ the value of $\alpha'$ that satisfies equality in (2.11) as a function of $\alpha$. The plot in Figure 2.5 shows the behavior of $\alpha^*$ relative to $(1-\gamma)\alpha$. For any $\alpha$, $\alpha^*$ is bounded below by $(1-\gamma)\alpha$. Similar to the result in the above subsection, (2.11) is satisfied for any $0 < \alpha' \leq \alpha^*(\alpha)$ so that setting $\alpha' = (1-\gamma)\alpha$ is a reasonable modification in order to compensate for the anti-conservative behavior of the proposed procedure for the for non-null distributions similar to $F$.

Looking at Figure 2.1, the desired quality of $F$ is that before it crosses the line $\beta u$, it remains sufficiently small for $(1 - \gamma)^2 u^*(\alpha) \leq u \leq u^*(\alpha)$. Combining
Figure 2.5. $\alpha^*$ versus $\alpha$ (slightly curved line) bounded below by $(1 - \gamma)\alpha$ (straight line).

this property with (2.9), we expect a choice of non-null distributions $F(u)$ near $\beta u$ to fare well in the simulation studies. As noted in the previous subsection, this property pertains to non-null distributions that deviate only moderately from the null distribution.

2.5 Conclusions and other considerations

In this chapter, we focus mainly on comparisons between the proposed $\gamma$SUP and the step-up procedure of Benjamini and Hochberg [1] under the asymptotic framework of [7] for large number of tests $m$. As expected, we see that the power and FDR of the $\gamma$SUP depends on $\gamma$ and that both of these properties reduce to those of the BHP when $\gamma = 0$. This is somewhat surprising when considering that the $\gamma$SUP is equal to the step-down version of the BHP—for every $m$—when $\gamma = 0$. That is, the BHP and its step-down version become equivalent for sufficiently large $m$. Although the step-down version is more conservative in finite cases, its conservatism vanishes as $m$ tends to infinity, where it effectively becomes the BHP (step-up version).
Unfortunately for the γSUP, which takes advantage of increased power over the step-down version, this increased power is translated into loss of FDR control as \( m \) gets large. Thus, we investigate ways to remedy this. The first approach is to incorporate the quantity \( A_0 < 1 \), which we handle in two ways depending on whether the non-null distribution is moderate or extreme. However, both of these approaches require estimation of \( A_0 \), since it is not an observable quantity. We employ a two-stage process by which \( A_0 \) can be estimated and then be subsequently utilized to choose \( \gamma \), according to the expected extremity of the non-null distribution.

It is perhaps not surprising that the potential gain in power from the proposed generalization involving \( \gamma \) would be transferred into loss of FDR control without taking advantage of \( A_0 < 1 \), an unknown quantity, or conservatively modifying the value for \( \alpha \). The BHP, as the authors in [7] point out, is optimal among all “last crossing procedures” in that its limiting threshold is as large as can be to maintain FDR control for any \( A_0 \leq 1 \). This property reflects the intimate relationship between the BHP and the definition of the FDR itself. Nevertheless, the treatment of this proposed generalization beneficially raises many follow-up questions. Some of these are discussed in Chapter 5. One of these, we mention now.

Since the BHP is asymptotically optimal in the sense the authors [7] consider, perhaps the proposed generalization should be modified to force the limiting threshold to equal that of the BHP, while still enjoying increased power for finite \( m \). This suggests taking \( \gamma = \gamma_m \) and requiring \( \gamma_m \to 0 \) with increasing \( m \) in such a way that its limiting threshold equals that of the BHP. For appropriately chosen \( \gamma_m \), perhaps FDR control may be achieved with an increase in power for certain non-null distributions. After all, the BHP makes no assumption about the non-null distribution \( F \). In general, this is a desirable trait, but in specific situations where \( F \) may be reasonably ascertained, it is at a loss for potential improvement.

After the discussion here concerning the proposed γSUP, the BHP, and the various modifications. It is interesting to turn now to simulation results, where
these procedures and the modifications may be compared for many different situations.
Simulations

3.1 Introduction

As the literature demonstrates, the false discovery rate (FDR) lends itself well to large-scale multiple testing situations. Not only are these situations becoming increasingly common—in astronomy for example—but they also allow for asymptotic considerations with respect to the number of tests. In this chapter, we explore numerical results for a variety of cases. Although some rely on asymptotic properties, it is interesting to see how they behave in finite settings.

After summarizing the simulated procedures for easy reference in the next section, we describe at some length how the data are generated for simulation in Section 3.3. We also outline the various choices for the input quantities, such as $A_0$ and $\alpha$. In Section 3.4, we comment briefly on the numerical results obtained by the simulations. Recall from our discussion in the previous chapter that the recommended choice for $\gamma$ depends on the non-null distribution, and the effects of these choices are pointed out. Finally, the last section contains the tabulated simulation results.
3.2 Referential description of simulated procedures

With the many procedures discussed, both existing and introduced here, it is helpful to have an abbreviated description for reference when evaluating the performances in the simulations of this chapter. Although we cannot perform an exhaustive comparison among all the procedures and parametric values, we choose the following for discussion now. By “small” $p$-value, we mean an ordered $p$-value that falls below its cutoff: $P_{(i)} \leq \alpha_i$. The group of procedures considered is given in Figure 3.1.

For the proposed procedures introduced in this work, the default choice is $\gamma = \alpha$, which corresponds to the $\gamma$SUP labeled gam.su. Also, earlier discussion supports the adjustment $\alpha' = (1 - \gamma)\alpha$ for moderate non-null distributions, suggesting $\alpha' = (1 - \alpha)\alpha$ when $\gamma$ is set to $\alpha$. These conditions correspond to the gam.su.mod and gam.sd.mod procedures.

Note that the procedures, BenHoch.adp, gam.su.adp, and gam.su.adp2, addressed in the simulations are conducted in two stages. The first stage is used to estimate $A_0$, the fraction of null $p$-values, and the second stage incorporates this estimate $\hat{A}_0$ and uses it to adjust the FDR control. The second stage allows them to “adapt” to the situation presented to them; hence their suffixes. In both cases, $\hat{A}_0 = \frac{m-R}{m}$, the fraction of $p$-values not rejected, but each adaptive procedure utilizes this information in its own way.

The BenHoch.adp adaptive procedure replaces $\alpha$ with $\alpha \frac{m}{\hat{A}_0}$, which compromises for its conservative FDR control by $\alpha A_0$ otherwise. This strategy is discussed in [3]. The modification in gam.su.adp is to set $\gamma = 1 - \hat{A}_0$, and the modification in gam.su.adp2 is to set $\gamma = \frac{\alpha(1-\hat{A}_0)}{1-\alpha A_0}$ as motivated by the discussion in Section 2.4.1.2. As mentioned there, these adjustments depending on the non-null distribution. The BenHoch.adp is recommended for moderate non-null distributions ($\mu = 2$), and BenHoch.adp2 is recommended for extreme non-null distributions ($\mu = 4$).
<table>
<thead>
<tr>
<th>Label</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>BenHoch</td>
<td>Benjamini and Hochberg step-up procedure rejects up to the largest small $p$-value.</td>
</tr>
<tr>
<td>BenHoch.adp</td>
<td>Benjamini and Hochberg adaptive step-up procedure estimates $A_0$.</td>
</tr>
<tr>
<td>BenLui</td>
<td>Benjamini and Lui step-down procedure rejects up to the largest consecutive small $p$-value.</td>
</tr>
<tr>
<td>gam.su</td>
<td>Proposed procedure $\gamma$SUP requires $1 - \gamma$ fraction of the rejected $p$-values to be small.</td>
</tr>
<tr>
<td>gam.su.adp</td>
<td>Proposed $\gamma$SUP adaptive procedure estimates $A_0$ and sets $\hat{\gamma} = 1 - \hat{A}_0$, for moderate non-nulls.</td>
</tr>
<tr>
<td>gam.su.adp2</td>
<td>Proposed $\gamma$SUP adaptive procedure estimates $A_0$ and sets $\hat{\gamma} = \alpha(1 - \hat{A}_0)/(1 - \alpha A_0)$, for extreme non-nulls.</td>
</tr>
<tr>
<td>gam.sd</td>
<td>Proposed procedure $\gamma$SDP requires $1 - \gamma$ fraction of the rejected $p$-values to be consecutively small.</td>
</tr>
<tr>
<td>gam.su.mod</td>
<td>Proposed procedure $\gamma$SUP with conservative modification $\alpha' = (1 - \gamma)\alpha$ requires $1 - \gamma$ fraction of the rejected $p$-values to be small.</td>
</tr>
<tr>
<td>gam.sd.mod</td>
<td>Proposed procedure $\gamma$SUP with conservative modification $\alpha' = (1 - \gamma)\alpha$ requires $1 - \gamma$ fraction of the rejected $p$-values to be consecutively small.</td>
</tr>
</tbody>
</table>

**Figure 3.1.** Brief description of simulated procedures by labels.

### 3.3 Generating the data

For any hypothesis test, the $p$-value is a standardized statistic measuring the amount of evidence in favor of the alternative. It is interpreted as the probability, assuming the null hypothesis $H_0$ is true, of observing sample data at least as extreme as that observed. This seemingly circular description arises
from consideration of the sample data as both a random variable (before data is collected) and a non-random constant (after data is collected). For example, if $X$ represents the random data yet to be observed and $x$ represents the realization of $X$ after observing the data, then the realized $p$-value is $P_0(X < -|x|) + P(X > |x|)$, where $P_0$ is the probability assuming $H_0$ is true. If this value is small, we are more likely to reject $H_0$ in favor of $H_a$. In the sections that follow, we consider two different non-null distributions corresponding to extreme departure from $H_0$ and moderate departure from $H_0$.

We simulate both scenarios with the following framework.

Consider a normal random variable $X$ with mean $\mu$ and variance 1. We denote the distribution of $X$ by $\Phi_\mu$ and the inverse of this function by $\Phi_\mu^{-1}$. If the hypotheses are $H_0 : \mu = 0$ versus $H_a : \mu \neq 0$, then the $p$-value for the test is $2\Phi_0(-|X|)$ by the symmetry of the normal density. Also, the distribution function of the $p$-value is

$$P(2\Phi_0(-|X|) < u) = \Phi_\mu(\Phi_0^{-1}(\frac{u}{2})) + 1 - \Phi_\mu(\Phi_0^{-1}(1 - \frac{u}{2})),$$

where the last equality uses the symmetry property: $-\Phi^{-1}(\frac{u}{2}) = \Phi^{-1}(1 - \frac{u}{2})$.

Now, it can be shown that the above is bounded below by $u$ with equality if and only if $\mu = 0$, as claimed by $H_0$. To visualize why this is true consider that the normal density is uni-modal and symmetric around $\mu$. For fixed $u$, the probability above corresponds to the sum of two tail probabilities and is minimized when $\mu$ lies equidistant between the two boundaries. Thus, the $p$-value is uniformly distributed on $(0, 1)$ when $H_0$ is true—its distribution function is the identity on this interval—and stochastically smaller when $H_0$ is false.

To generate the $p$-values, we sample $X_1, \ldots, X_{m_0}$ random variables with distribution $\Phi_0$ and $X_{m_0+1}, \ldots, X_m$ with distribution $\Phi_\mu$, where $m_0$, $m$, and $\mu$ are all specified. Then, applying the transformation $2\Phi_0(-|X|)$ to each
\(X_1, \ldots, X_m\) provides a set of \(m_0\) (null) uniform random variables and a set of \(m - m_0\) (non-null) random variables that are stochastically smaller than uniform. These numerical tasks are easily performed with the R software package, and the specific lines of code are included in Appendix B. Moreover, this approach allows the sample \(X_1, \ldots, X_m\) to be generated jointly with correlation coefficient \(\rho\), also specified. This type of positive dependence satisfies the PRD property to which many of the discussed procedures apply [5]. To clarify, the correlation applies to the \(p\)-values jointly. Individual null \(p\)-values, however correlated with the non-null ones in the same sample, are still marginally uniformly distributed, and the non-null ones are stochastically smaller.

The choice of non-null distribution is distinguished by its departure from uniformity on \((0, 1)\). Given the way the data are simulated, this corresponds to choosing different values of \(\mu\), where the larger \(\mu\) is in absolute values, the smaller (more extreme) the non-null \(p\)-values will be. For these simulations, we use values \(\mu = 2\) and \(\mu = 4\) for the moderate and extreme departures from null, respectively. Figure 3.2 gives estimated sampling densities of the \(p\)-values for \(\mu = 2\) and \(\mu = 4\).

![Figure 3.2](image.png)

**Figure 3.2.** Estimated density plots of generated \(p\)-values with moderate departure from null (\(\mu = 2\)) and extreme departure from null (\(\mu = 4\)). \(A_0 = .8\) and \(m = 500\) in both cases.
With all the procedures and parameters discussed, one can imagine the vast number of resulting combinations to consider. However, to keep the results manageable, we basically reduce our choices to “low” and “high” values, with the exception being the number of tests \( m \). The correlation values are \( \rho = 0 \) and \( \rho = .5 \), corresponding to independent and positively dependent \( p \)-values, respectively, and the fraction of null \( p \)-values is either \( A_0 = .8 \) or \( A_0 = .95 \); for the case \( m = 10 \), \( A_0 = .9 \). The FDR bound is \( \alpha = .05 \) in all cases.

### 3.4 Results

The most important quantities we are interested in are the FDR, the number of rejections \( R \), and their standard deviations. We wish to reject as many \( p \)-values as possible while controlling the FDR (or another error rate) at the specified \( \alpha = .05 \) level. To this end, we generally prefer the procedure with the highest expected number of rejections subject to the FDR constraint. However, we cannot ignore the variability of the realizations of these quantities, and these are measured by their standard deviations. The following example provides an extreme case to reveal the importance of variability, which is sometimes overlooked in the FDR literature.

**Example 7.** Consider the (degenerate) multiple-testing procedure that ignores the values of the \( p \)-values altogether. With probability \( \alpha \), it rejects all of them, and with probability \( 1 - \alpha \) it accepts all of them. Obviously, this is a ridiculous approach, but is worth noting that it controls the FDR for any \( m_0 \) and \( \alpha \), for any configuration of non-null distributions, and under any dependence structure. Since the fraction of false positives \( \frac{V}{R} \) is either \( \frac{m_0}{m} \) with probability \( \alpha \) or zero with probability \( 1 - \alpha \), we have simply

\[
\text{FDR} = \mathcal{E}\left[\frac{V}{R}\right] = \frac{m_0}{m} \alpha.
\]

Moreover, the expected number of rejections \( \mathcal{E}(R) = \alpha m \) is quite impressive. Of course, the variability of this strategy is its downfall. We have \( \mathcal{V}(R) = \)
\(\alpha(1 - \alpha)m^2\), which becomes extremely high with large \(m\), and, although it is not as succinctly written here, \(V(\frac{V}{R})\) is also unacceptably high in comparison with the other procedures. Although this is a ridiculously extreme example, the point is that we cannot consider only the expected number of rejections and FDR in our assessment of multiple-testing procedures.

In all simulated cases, the output has the following format. The row labels pertain to the procedures, which are described in the previous section, and the column labels are as follows: \(E(V)\), \(E(R)\), FDR, and FDP are the expected number of null rejections \(V\), the expected number of rejections \(R\), the FDR, and the \(\delta\)FDP for \(\delta = \alpha\), according to \(N = 10000\) independent and identical simulations of \(p\)-values with distributions described above. Of course, these procedures are not designed to control the FDP for any particular \(\delta\), but it is still interesting to see how this rate varies among the procedures. Each of these four quantities is obtained by averaging their realizations for the \(N\) simulations. The last four columns are the estimated standard deviations of the quantities averaged in the first four.

The numerical results are provided below in Figure 3.3 through Figure 3.28. We divide our discussion of these results into three groups, determined by the number of tests \(m\).

### 3.4.1 Small number of tests \((m = 10)\)

Although the benefit of the FDR as an error bound in multiple testing is fully realized in large-scale situations, it may still be utilized in smaller cases as well. Examples of this arise in pairwise comparisons following an analysis of variance involving a small number of groups. If a control group is compared with each of the others, the induced dependence among these individual tests is of the PRD type [5].

When the non-null distribution is moderate \((\mu = 2)\), all procedures perform similarly. Each controls the FDR slightly conservatively with the conservatism roughly proportionate to \(A_0\). The BenLui procedure is better suited to smaller
and performs well here. When a positive dependence is introduced ($\rho = .5$), we find essentially the same results but in a more conservative way. In all cases, the BenHoch procedure performs well as expected.

3.4.2 Moderate number of tests ($m = 500$)

This case serves as a reasonably large-scale situation but somewhat before the asymptotic results necessarily apply.

More so here than in the previous case, the two proposed adaptive procedures distinguish themselves based on the extremity of the non-null distribution, as parameterized by $\mu$. As expected, the gam.su.adp procedure does well when the non-null distribution is moderate ($\mu = 2$), and the gam.su.adp2 procedure does well in the extreme case. The gam.su and gam.su tend to fall between the adaptive versions but are a more stable choice if the non-null distribution is unknown. The procedures with modified $\alpha' = (1 - \alpha)\alpha$ demonstrate a slightly conservative adjustment to the gam.su procedure but are otherwise unremarkable. The BenHoch procedures are both admirable.

The dependent case for this group reveals no major differences. The proposed adaptive procedures still perform well in their appropriate situations, and the gam.su procedure tends to fall somewhere in between.

In the interest of completeness, we would be remiss not to address the situation where $A_0$ is small. Therefore, we include two cases for $A_0 = .3$ in this middle group of $m = 500$. The first has independent $p$-values with $\mu = 2$, a moderate deviation from the null distribution. The second has dependent $p$-values with $\mu = 4$, an extreme deviation from the null. The first adaptive gam.su.adp procedure is noticeably extreme in this case, but recall that it is not recommended when the non-null distribution is extreme, as it is when $\mu = 4$. 
3.4.3 Large number of tests \((m = 5000)\)

Although \(m = 5000\) hardly covers the extremely large-scale applications, the results here represent the largest-\(m\) group considered in this work.

We now turn to the numerical results.

<table>
<thead>
<tr>
<th></th>
<th>(E(V))</th>
<th>(E(R))</th>
<th>FDR</th>
<th>FDP</th>
<th>(s(V))</th>
<th>(s(R))</th>
<th>(s(FDR))</th>
<th>(s(FDP))</th>
</tr>
</thead>
<tbody>
<tr>
<td>BenHoch</td>
<td>0.0713</td>
<td>0.5450</td>
<td>0.0418</td>
<td>0.0662</td>
<td>0.2768</td>
<td>0.7451</td>
<td>0.1701</td>
<td>0.2486</td>
</tr>
<tr>
<td>BenHoch.adp</td>
<td>0.0784</td>
<td>0.5528</td>
<td>0.0429</td>
<td>0.0708</td>
<td>0.2991</td>
<td>0.7764</td>
<td>0.1689</td>
<td>0.2565</td>
</tr>
<tr>
<td>BenLui</td>
<td>0.0447</td>
<td>0.4692</td>
<td>0.0342</td>
<td>0.0441</td>
<td>0.2095</td>
<td>0.6214</td>
<td>0.1682</td>
<td>0.2053</td>
</tr>
<tr>
<td>gam.su</td>
<td>0.0581</td>
<td>0.5037</td>
<td>0.0361</td>
<td>0.0551</td>
<td>0.2464</td>
<td>0.7061</td>
<td>0.1618</td>
<td>0.2282</td>
</tr>
<tr>
<td>gam.su.adp</td>
<td>0.0736</td>
<td>0.5267</td>
<td>0.0379</td>
<td>0.0556</td>
<td>0.3391</td>
<td>0.7818</td>
<td>0.1659</td>
<td>0.2292</td>
</tr>
<tr>
<td>gam.su.adp2</td>
<td>0.0581</td>
<td>0.5037</td>
<td>0.0361</td>
<td>0.0551</td>
<td>0.2464</td>
<td>0.7061</td>
<td>0.1618</td>
<td>0.2282</td>
</tr>
<tr>
<td>gam.sd</td>
<td>0.0581</td>
<td>0.5037</td>
<td>0.0361</td>
<td>0.0551</td>
<td>0.2464</td>
<td>0.7061</td>
<td>0.1618</td>
<td>0.2282</td>
</tr>
<tr>
<td>gam.su.mod</td>
<td>0.0580</td>
<td>0.5030</td>
<td>0.0361</td>
<td>0.0550</td>
<td>0.2463</td>
<td>0.7057</td>
<td>0.1619</td>
<td>0.2280</td>
</tr>
<tr>
<td>gam.sd.mod</td>
<td>0.0580</td>
<td>0.5030</td>
<td>0.0361</td>
<td>0.0550</td>
<td>0.2463</td>
<td>0.7057</td>
<td>0.1619</td>
<td>0.2280</td>
</tr>
</tbody>
</table>

Figure 3.3. Results for \(m = 10\) tests, independent \(p\)-values, \(\mu = 2\), and \(A_0 = .8\)

<table>
<thead>
<tr>
<th></th>
<th>(E(V))</th>
<th>(E(R))</th>
<th>FDR</th>
<th>FDP</th>
<th>(s(V))</th>
<th>(s(R))</th>
<th>(s(FDR))</th>
<th>(s(FDP))</th>
</tr>
</thead>
<tbody>
<tr>
<td>BenHoch</td>
<td>0.0586</td>
<td>0.2715</td>
<td>0.0433</td>
<td>0.0537</td>
<td>0.2561</td>
<td>0.5089</td>
<td>0.1897</td>
<td>0.2254</td>
</tr>
<tr>
<td>BenHoch.adp</td>
<td>0.0594</td>
<td>0.2681</td>
<td>0.0423</td>
<td>0.0533</td>
<td>0.2628</td>
<td>0.5148</td>
<td>0.1865</td>
<td>0.2246</td>
</tr>
<tr>
<td>BenLui</td>
<td>0.0467</td>
<td>0.2570</td>
<td>0.0404</td>
<td>0.0455</td>
<td>0.2175</td>
<td>0.4652</td>
<td>0.1901</td>
<td>0.2084</td>
</tr>
<tr>
<td>gam.su</td>
<td>0.0526</td>
<td>0.2585</td>
<td>0.0399</td>
<td>0.0486</td>
<td>0.2417</td>
<td>0.4926</td>
<td>0.1837</td>
<td>0.2150</td>
</tr>
<tr>
<td>gam.su.adp</td>
<td>0.0626</td>
<td>0.2794</td>
<td>0.0449</td>
<td>0.0555</td>
<td>0.2841</td>
<td>0.5284</td>
<td>0.1934</td>
<td>0.229</td>
</tr>
<tr>
<td>gam.su.adp2</td>
<td>0.0526</td>
<td>0.2585</td>
<td>0.0399</td>
<td>0.0486</td>
<td>0.2417</td>
<td>0.4926</td>
<td>0.1837</td>
<td>0.2150</td>
</tr>
<tr>
<td>gam.sd</td>
<td>0.0526</td>
<td>0.2585</td>
<td>0.0399</td>
<td>0.0486</td>
<td>0.2417</td>
<td>0.4926</td>
<td>0.1837</td>
<td>0.2150</td>
</tr>
<tr>
<td>gam.su.mod</td>
<td>0.0526</td>
<td>0.2582</td>
<td>0.0399</td>
<td>0.0486</td>
<td>0.2417</td>
<td>0.4925</td>
<td>0.1837</td>
<td>0.2150</td>
</tr>
<tr>
<td>gam.sd.mod</td>
<td>0.0526</td>
<td>0.2582</td>
<td>0.0399</td>
<td>0.0486</td>
<td>0.2417</td>
<td>0.4925</td>
<td>0.1837</td>
<td>0.2150</td>
</tr>
</tbody>
</table>

Figure 3.4. Results for \(m = 10\) tests, independent \(p\)-values, \(\mu = 2\), and \(A_0 = .95\)
<table>
<thead>
<tr>
<th>Method</th>
<th>E(V)</th>
<th>E(R)</th>
<th>FDR</th>
<th>FDP</th>
<th>s(V)</th>
<th>s(R)</th>
<th>s(FDR)</th>
<th>s(FDP)</th>
</tr>
</thead>
<tbody>
<tr>
<td>BenHoch</td>
<td>0.1166</td>
<td>1.9600</td>
<td>0.0386</td>
<td>0.1069</td>
<td>0.3513</td>
<td>0.5454</td>
<td>0.1139</td>
<td>0.3090</td>
</tr>
<tr>
<td>BenHoch.adp</td>
<td>0.1510</td>
<td>2.0025</td>
<td>0.0484</td>
<td>0.1321</td>
<td>0.4128</td>
<td>0.5888</td>
<td>0.1270</td>
<td>0.3386</td>
</tr>
<tr>
<td>BenLui</td>
<td>0.0482</td>
<td>1.8317</td>
<td>0.0174</td>
<td>0.0469</td>
<td>0.2202</td>
<td>0.4974</td>
<td>0.0807</td>
<td>0.2114</td>
</tr>
<tr>
<td>gam.su</td>
<td>0.1084</td>
<td>1.9428</td>
<td>0.0361</td>
<td>0.1001</td>
<td>0.3380</td>
<td>0.5460</td>
<td>0.1104</td>
<td>0.3001</td>
</tr>
<tr>
<td>gam.su.adp</td>
<td>0.2170</td>
<td>2.0637</td>
<td>0.0534</td>
<td>0.1037</td>
<td>0.6738</td>
<td>0.8116</td>
<td>0.1578</td>
<td>0.3049</td>
</tr>
<tr>
<td>gam.su.adp2</td>
<td>0.1084</td>
<td>1.9428</td>
<td>0.0361</td>
<td>0.1001</td>
<td>0.3380</td>
<td>0.5460</td>
<td>0.1104</td>
<td>0.3001</td>
</tr>
<tr>
<td>gam.sd</td>
<td>0.1084</td>
<td>1.9428</td>
<td>0.0361</td>
<td>0.1001</td>
<td>0.3380</td>
<td>0.5460</td>
<td>0.1104</td>
<td>0.3001</td>
</tr>
<tr>
<td>gam.su.mod</td>
<td>0.1082</td>
<td>1.9424</td>
<td>0.0361</td>
<td>0.1000</td>
<td>0.3375</td>
<td>0.5454</td>
<td>0.1104</td>
<td>0.3000</td>
</tr>
<tr>
<td>gam.sd.mod</td>
<td>0.1082</td>
<td>1.9424</td>
<td>0.0361</td>
<td>0.1000</td>
<td>0.3375</td>
<td>0.5454</td>
<td>0.1104</td>
<td>0.3000</td>
</tr>
</tbody>
</table>

**Figure 3.5.** Results for $m = 10$ tests, independent $p$-values, $\mu = 4$, and $A_0 = .8$

<table>
<thead>
<tr>
<th>Method</th>
<th>E(V)</th>
<th>E(R)</th>
<th>FDR</th>
<th>FDP</th>
<th>s(V)</th>
<th>s(R)</th>
<th>s(FDR)</th>
<th>s(FDP)</th>
</tr>
</thead>
<tbody>
<tr>
<td>BenHoch</td>
<td>0.0890</td>
<td>0.9735</td>
<td>0.0436</td>
<td>0.0822</td>
<td>0.3093</td>
<td>0.4605</td>
<td>0.1485</td>
<td>0.2747</td>
</tr>
<tr>
<td>BenHoch.adp</td>
<td>0.0998</td>
<td>0.9808</td>
<td>0.0475</td>
<td>0.0896</td>
<td>0.3359</td>
<td>0.4859</td>
<td>0.1542</td>
<td>0.2856</td>
</tr>
<tr>
<td>BenLui</td>
<td>0.0494</td>
<td>0.9324</td>
<td>0.0262</td>
<td>0.0482</td>
<td>0.2222</td>
<td>0.3953</td>
<td>0.1204</td>
<td>0.2142</td>
</tr>
<tr>
<td>gam.su</td>
<td>0.0809</td>
<td>0.9600</td>
<td>0.0400</td>
<td>0.0758</td>
<td>0.2918</td>
<td>0.4535</td>
<td>0.1423</td>
<td>0.2647</td>
</tr>
<tr>
<td>gam.su.adp</td>
<td>0.0957</td>
<td>0.9794</td>
<td>0.0448</td>
<td>0.0831</td>
<td>0.3638</td>
<td>0.5002</td>
<td>0.1526</td>
<td>0.276</td>
</tr>
<tr>
<td>gam.su.adp2</td>
<td>0.0809</td>
<td>0.9600</td>
<td>0.0400</td>
<td>0.0758</td>
<td>0.2918</td>
<td>0.4535</td>
<td>0.1423</td>
<td>0.2647</td>
</tr>
<tr>
<td>gam.sd</td>
<td>0.0809</td>
<td>0.9600</td>
<td>0.0400</td>
<td>0.0758</td>
<td>0.2918</td>
<td>0.4535</td>
<td>0.1423</td>
<td>0.2647</td>
</tr>
<tr>
<td>gam.su.mod</td>
<td>0.0805</td>
<td>0.9595</td>
<td>0.0398</td>
<td>0.0755</td>
<td>0.2909</td>
<td>0.4629</td>
<td>0.1420</td>
<td>0.2642</td>
</tr>
<tr>
<td>gam.sd.mod</td>
<td>0.0805</td>
<td>0.9595</td>
<td>0.0398</td>
<td>0.0755</td>
<td>0.2909</td>
<td>0.4629</td>
<td>0.1420</td>
<td>0.2642</td>
</tr>
</tbody>
</table>

**Figure 3.6.** Results for $m = 10$ tests, independent $p$-values, $\mu = 4$, and $A_0 = .95$
<table>
<thead>
<tr>
<th>E(V)</th>
<th>E(R)</th>
<th>FDR</th>
<th>FDP</th>
<th>s(V)</th>
<th>s(R)</th>
<th>s(FDR)</th>
<th>s(FDP)</th>
</tr>
</thead>
<tbody>
<tr>
<td>BenHoch</td>
<td>0.0943</td>
<td>0.5851</td>
<td>0.0334</td>
<td>0.0550</td>
<td>0.4864</td>
<td>0.9684</td>
<td>0.1529</td>
</tr>
<tr>
<td>BenHoch.adp</td>
<td>0.1251</td>
<td>0.6164</td>
<td>0.0366</td>
<td>0.0609</td>
<td>0.6405</td>
<td>1.1011</td>
<td>0.1578</td>
</tr>
<tr>
<td>BenLui</td>
<td>0.0484</td>
<td>0.4851</td>
<td>0.0252</td>
<td>0.0365</td>
<td>0.2974</td>
<td>0.7685</td>
<td>0.1411</td>
</tr>
<tr>
<td>gam.su</td>
<td>0.0814</td>
<td>0.5431</td>
<td>0.0298</td>
<td>0.0484</td>
<td>0.4514</td>
<td>0.9277</td>
<td>0.1455</td>
</tr>
<tr>
<td>gam.su.adp</td>
<td>0.1555</td>
<td>0.6252</td>
<td>0.0390</td>
<td>0.0538</td>
<td>0.8586</td>
<td>1.2797</td>
<td>0.1709</td>
</tr>
<tr>
<td>gam.su.adp2</td>
<td>0.0814</td>
<td>0.5431</td>
<td>0.0298</td>
<td>0.0484</td>
<td>0.4514</td>
<td>0.9277</td>
<td>0.1455</td>
</tr>
<tr>
<td>gam.sd</td>
<td>0.0814</td>
<td>0.5431</td>
<td>0.0298</td>
<td>0.0484</td>
<td>0.4514</td>
<td>0.9277</td>
<td>0.1455</td>
</tr>
<tr>
<td>gam.su.mod</td>
<td>0.0812</td>
<td>0.5424</td>
<td>0.0297</td>
<td>0.0483</td>
<td>0.4510</td>
<td>0.9275</td>
<td>0.1452</td>
</tr>
<tr>
<td>gam.sd.mod</td>
<td>0.0812</td>
<td>0.5424</td>
<td>0.0297</td>
<td>0.0483</td>
<td>0.4510</td>
<td>0.9275</td>
<td>0.1452</td>
</tr>
</tbody>
</table>

**Figure 3.7.** Results for \( m = 10 \) tests, dependent \( p \)-values, \( \mu = 2 \), and \( A_0 = .8 \)

<table>
<thead>
<tr>
<th>E(V)</th>
<th>E(R)</th>
<th>FDR</th>
<th>FDP</th>
<th>s(V)</th>
<th>s(R)</th>
<th>s(FDR)</th>
<th>s(FDP)</th>
</tr>
</thead>
<tbody>
<tr>
<td>BenHoch</td>
<td>0.1055</td>
<td>0.3210</td>
<td>0.0416</td>
<td>0.0536</td>
<td>0.5781</td>
<td>0.7550</td>
<td>0.1821</td>
</tr>
<tr>
<td>BenHoch.adp</td>
<td>0.1281</td>
<td>0.3410</td>
<td>0.0425</td>
<td>0.0548</td>
<td>0.7233</td>
<td>0.8871</td>
<td>0.1837</td>
</tr>
<tr>
<td>BenLui</td>
<td>0.0555</td>
<td>0.2701</td>
<td>0.0322</td>
<td>0.0395</td>
<td>0.3317</td>
<td>0.5468</td>
<td>0.1654</td>
</tr>
<tr>
<td>gam.su</td>
<td>0.0893</td>
<td>0.3003</td>
<td>0.0370</td>
<td>0.0478</td>
<td>0.5238</td>
<td>0.7074</td>
<td>0.1724</td>
</tr>
<tr>
<td>gam.su.adp</td>
<td>0.1074</td>
<td>0.3257</td>
<td>0.0372</td>
<td>0.0478</td>
<td>0.7001</td>
<td>0.8758</td>
<td>0.1729</td>
</tr>
<tr>
<td>gam.su.adp2</td>
<td>0.0893</td>
<td>0.3003</td>
<td>0.0370</td>
<td>0.0478</td>
<td>0.5238</td>
<td>0.7074</td>
<td>0.1724</td>
</tr>
<tr>
<td>gam.sd</td>
<td>0.0893</td>
<td>0.3003</td>
<td>0.0370</td>
<td>0.0478</td>
<td>0.5238</td>
<td>0.7074</td>
<td>0.1724</td>
</tr>
<tr>
<td>gam.su.mod</td>
<td>0.0892</td>
<td>0.3000</td>
<td>0.0370</td>
<td>0.0477</td>
<td>0.5237</td>
<td>0.7071</td>
<td>0.1723</td>
</tr>
<tr>
<td>gam.sd.mod</td>
<td>0.0892</td>
<td>0.3000</td>
<td>0.0370</td>
<td>0.0477</td>
<td>0.5237</td>
<td>0.7071</td>
<td>0.1723</td>
</tr>
</tbody>
</table>

**Figure 3.8.** Results for \( m = 10 \) tests, dependent \( p \)-values, \( \mu = 2 \), and \( A_0 = .95 \)
<table>
<thead>
<tr>
<th></th>
<th>E(V)</th>
<th>E(R)</th>
<th>FDR</th>
<th>FDP</th>
<th>s(V)</th>
<th>s(R)</th>
<th>s(FDR)</th>
<th>s(FDP)</th>
</tr>
</thead>
<tbody>
<tr>
<td>BenHoch</td>
<td>0.1315</td>
<td>1.9738</td>
<td>0.0371</td>
<td>0.0807</td>
<td>0.5457</td>
<td>0.6820</td>
<td>0.1348</td>
<td>0.2724</td>
</tr>
<tr>
<td>BenHoch.adp</td>
<td>0.1835</td>
<td>2.0342</td>
<td>0.0448</td>
<td>0.0958</td>
<td>0.7397</td>
<td>0.8499</td>
<td>0.1480</td>
<td>0.2943</td>
</tr>
<tr>
<td>BenLui</td>
<td>0.0518</td>
<td>1.8357</td>
<td>0.0193</td>
<td>0.0395</td>
<td>0.3045</td>
<td>0.5571</td>
<td>0.1058</td>
<td>0.1948</td>
</tr>
<tr>
<td>gam.su</td>
<td>0.1113</td>
<td>1.9390</td>
<td>0.0320</td>
<td>0.0705</td>
<td>0.4955</td>
<td>0.6696</td>
<td>0.1256</td>
<td>0.2560</td>
</tr>
<tr>
<td>gam.su.adp</td>
<td>0.2154</td>
<td>2.0518</td>
<td>0.0436</td>
<td>0.0743</td>
<td>0.9379</td>
<td>1.0508</td>
<td>0.1588</td>
<td>0.2623</td>
</tr>
<tr>
<td>gam.su.adp2</td>
<td>0.1113</td>
<td>1.9390</td>
<td>0.0320</td>
<td>0.0705</td>
<td>0.4955</td>
<td>0.6696</td>
<td>0.1256</td>
<td>0.2560</td>
</tr>
<tr>
<td>gam.sd</td>
<td>0.1113</td>
<td>1.9390</td>
<td>0.0320</td>
<td>0.0705</td>
<td>0.4955</td>
<td>0.6696</td>
<td>0.1256</td>
<td>0.2560</td>
</tr>
<tr>
<td>gam.su.mod</td>
<td>0.1108</td>
<td>1.9379</td>
<td>0.0319</td>
<td>0.0701</td>
<td>0.4949</td>
<td>0.6694</td>
<td>0.1254</td>
<td>0.2553</td>
</tr>
<tr>
<td>gam.sd.mod</td>
<td>0.1108</td>
<td>1.9379</td>
<td>0.0319</td>
<td>0.0701</td>
<td>0.4949</td>
<td>0.6694</td>
<td>0.1254</td>
<td>0.2553</td>
</tr>
</tbody>
</table>

Figure 3.9. Results for $m = 10$ tests, dependent $p$-values, $\mu = 4$, and $A_0 = .8$

<table>
<thead>
<tr>
<th></th>
<th>E(V)</th>
<th>E(R)</th>
<th>FDR</th>
<th>FDP</th>
<th>s(V)</th>
<th>s(R)</th>
<th>s(FDR)</th>
<th>s(FDP)</th>
</tr>
</thead>
<tbody>
<tr>
<td>BenHoch</td>
<td>0.1161</td>
<td>1.0026</td>
<td>0.0422</td>
<td>0.0671</td>
<td>0.5505</td>
<td>0.6315</td>
<td>0.1637</td>
<td>0.2502</td>
</tr>
<tr>
<td>BenHoch.adp</td>
<td>0.1432</td>
<td>1.0266</td>
<td>0.0452</td>
<td>0.0707</td>
<td>0.7008</td>
<td>0.7728</td>
<td>0.1700</td>
<td>0.2563</td>
</tr>
<tr>
<td>BenLui</td>
<td>0.0522</td>
<td>0.9348</td>
<td>0.0261</td>
<td>0.0404</td>
<td>0.2944</td>
<td>0.4192</td>
<td>0.1339</td>
<td>0.1969</td>
</tr>
<tr>
<td>gam.su</td>
<td>0.0991</td>
<td>0.9790</td>
<td>0.0370</td>
<td>0.0592</td>
<td>0.5009</td>
<td>0.5948</td>
<td>0.1534</td>
<td>0.2360</td>
</tr>
<tr>
<td>gam.su.adp</td>
<td>0.1296</td>
<td>1.0142</td>
<td>0.0383</td>
<td>0.0599</td>
<td>0.7591</td>
<td>0.8311</td>
<td>0.1586</td>
<td>0.2373</td>
</tr>
<tr>
<td>gam.su.adp2</td>
<td>0.0991</td>
<td>0.9790</td>
<td>0.0370</td>
<td>0.0592</td>
<td>0.5009</td>
<td>0.5948</td>
<td>0.1534</td>
<td>0.2360</td>
</tr>
<tr>
<td>gam.sd</td>
<td>0.0991</td>
<td>0.9790</td>
<td>0.0370</td>
<td>0.0592</td>
<td>0.5009</td>
<td>0.5948</td>
<td>0.1534</td>
<td>0.2360</td>
</tr>
<tr>
<td>gam.sd.mod</td>
<td>0.0988</td>
<td>0.9786</td>
<td>0.0369</td>
<td>0.0591</td>
<td>0.4993</td>
<td>0.5934</td>
<td>0.1533</td>
<td>0.2358</td>
</tr>
</tbody>
</table>

Figure 3.10. Results for $m = 10$ tests, dependent $p$-values, $\mu = 4$, and $A_0 = .95$
<table>
<thead>
<tr>
<th></th>
<th>E(V)</th>
<th>E(R)</th>
<th>FDR</th>
<th>FDP</th>
<th>s(V)</th>
<th>s(R)</th>
<th>s(FDR)</th>
<th>s(FDP)</th>
</tr>
</thead>
<tbody>
<tr>
<td>BenHoch</td>
<td>0.5238</td>
<td>11.4645</td>
<td>0.0393</td>
<td>0.3414</td>
<td>0.8014</td>
<td>0.0634</td>
<td>0.4742</td>
<td></td>
</tr>
<tr>
<td>BenHoch.adp</td>
<td>0.5474</td>
<td>11.7053</td>
<td>0.0400</td>
<td>0.3470</td>
<td>0.8250</td>
<td>0.0631</td>
<td>0.4760</td>
<td></td>
</tr>
<tr>
<td>BenLui</td>
<td>0.0363</td>
<td>3.0204</td>
<td>0.0117</td>
<td>0.0354</td>
<td>0.1918</td>
<td>0.0716</td>
<td>0.1888</td>
<td></td>
</tr>
<tr>
<td>gam.su</td>
<td>0.4153</td>
<td>9.8381</td>
<td>0.0335</td>
<td>0.2707</td>
<td>0.7311</td>
<td>0.0660</td>
<td>0.4443</td>
<td></td>
</tr>
<tr>
<td>gam.su.adp</td>
<td>0.4669</td>
<td>10.3932</td>
<td>0.0354</td>
<td>0.3042</td>
<td>0.7918</td>
<td>0.0622</td>
<td>0.4601</td>
<td></td>
</tr>
<tr>
<td>gam.su.adp2</td>
<td>0.4012</td>
<td>9.7169</td>
<td>0.0331</td>
<td>0.2761</td>
<td>0.7084</td>
<td>0.0659</td>
<td>0.4471</td>
<td></td>
</tr>
<tr>
<td>gam.sd</td>
<td>0.4133</td>
<td>9.8091</td>
<td>0.0334</td>
<td>0.2714</td>
<td>0.7284</td>
<td>0.0660</td>
<td>0.4447</td>
<td></td>
</tr>
<tr>
<td>gam.su.mod</td>
<td>0.4133</td>
<td>9.8123</td>
<td>0.0334</td>
<td>0.2867</td>
<td>0.7293</td>
<td>0.0661</td>
<td>0.4522</td>
<td></td>
</tr>
<tr>
<td>gam.sd.mod</td>
<td>0.4112</td>
<td>9.7834</td>
<td>0.0333</td>
<td>0.2861</td>
<td>0.7263</td>
<td>0.0661</td>
<td>0.4520</td>
<td></td>
</tr>
</tbody>
</table>

**Figure 3.11.** Results for $m = 500$ tests, independent $p$-values, $\mu = 2$, and $A_0 = .8$

<table>
<thead>
<tr>
<th></th>
<th>E(V)</th>
<th>E(R)</th>
<th>FDR</th>
<th>FDP</th>
<th>s(V)</th>
<th>s(R)</th>
<th>s(FDR)</th>
<th>s(FDP)</th>
</tr>
</thead>
<tbody>
<tr>
<td>BenHoch</td>
<td>0.1279</td>
<td>1.2894</td>
<td>0.0485</td>
<td>0.1089</td>
<td>0.3944</td>
<td>1.5620</td>
<td>0.1626</td>
<td>0.3115</td>
</tr>
<tr>
<td>BenHoch.adp</td>
<td>0.1259</td>
<td>1.2721</td>
<td>0.0473</td>
<td>0.1069</td>
<td>0.3925</td>
<td>1.5680</td>
<td>0.1607</td>
<td>0.3090</td>
</tr>
<tr>
<td>BenLui</td>
<td>0.0473</td>
<td>0.7832</td>
<td>0.0338</td>
<td>0.0461</td>
<td>0.2179</td>
<td>0.8757</td>
<td>0.1648</td>
<td>0.2097</td>
</tr>
<tr>
<td>gam.su</td>
<td>0.0953</td>
<td>1.0872</td>
<td>0.0402</td>
<td>0.0837</td>
<td>0.3350</td>
<td>1.4190</td>
<td>0.1553</td>
<td>0.2770</td>
</tr>
<tr>
<td>gam.su.adp</td>
<td>0.1071</td>
<td>1.1586</td>
<td>0.0420</td>
<td>0.0926</td>
<td>0.3589</td>
<td>1.4779</td>
<td>0.1536</td>
<td>0.2899</td>
</tr>
<tr>
<td>gam.su.adp2</td>
<td>0.0953</td>
<td>1.0872</td>
<td>0.0402</td>
<td>0.0837</td>
<td>0.3350</td>
<td>1.4190</td>
<td>0.1553</td>
<td>0.2770</td>
</tr>
<tr>
<td>gam.sd</td>
<td>0.0953</td>
<td>1.0872</td>
<td>0.0402</td>
<td>0.0837</td>
<td>0.3350</td>
<td>1.4190</td>
<td>0.1553</td>
<td>0.2770</td>
</tr>
<tr>
<td>gam.su.mod</td>
<td>0.0950</td>
<td>1.0849</td>
<td>0.0402</td>
<td>0.0834</td>
<td>0.3346</td>
<td>1.4160</td>
<td>0.1555</td>
<td>0.2765</td>
</tr>
<tr>
<td>gam.sd.mod</td>
<td>0.0960</td>
<td>1.0849</td>
<td>0.0402</td>
<td>0.0834</td>
<td>0.3346</td>
<td>1.4160</td>
<td>0.1555</td>
<td>0.2765</td>
</tr>
</tbody>
</table>

**Figure 3.12.** Results for $m = 500$ tests, independent $p$-values, $\mu = 2$, and $A_0 = .95$
\begin{tabular}{lcccccccc}
E(V) & E(R) & FDR & FDP & s(V) & s(R) & s(FDR) & s(FDP) \\
BenHoch & 3.8774 & 95.9692 & 0.0400 & 0.3157 & 2.0427 & 3.6613 & 0.0202 & 0.4648 \\
BenHoch.adp & 4.9052 & 98.1031 & 0.0494 & 0.4772 & 2.3569 & 3.7813 & 0.0225 & 0.4995 \\
BenLui & 0.0453 & 55.8623 & 0.0008 & 0.0000 & 0.2146 & 5.0537 & 0.0038 & 0.0000 \\
gam.su & 6.1003 & 100.0853 & 0.0603 & 0.6456 & 2.4842 & 4.0740 & 0.0229 & 0.4784 \\
gam.su.adp & 21.6928 & 119.6164 & 0.1797 & 1.0000 & 5.1380 & 5.9287 & 0.0343 & 0.0000 \\
gam.su.adp2 & 3.9819 & 96.0821 & 0.0409 & 0.3370 & 2.1404 & 4.0777 & 0.0210 & 0.4727 \\
gam.sd & 6.1003 & 100.0853 & 0.0603 & 0.6456 & 2.4842 & 4.0740 & 0.0229 & 0.4784 \\
gam.su.mod & 6.0856 & 100.0587 & 0.0602 & 0.7168 & 2.4839 & 4.0771 & 0.0229 & 0.4506 \\
gam.sd.mod & 6.0856 & 100.0587 & 0.0602 & 0.7168 & 2.4839 & 4.0771 & 0.0229 & 0.4506 \\
\end{tabular}

**Figure 3.13.** Results for $m = 500$ tests, independent $p$-values, $\mu = 4$, and $A_0 = .8$

\begin{tabular}{lcccccccc}
E(V) & E(R) & FDR & FDP & s(V) & s(R) & s(FDR) & s(FDP) \\
BenHoch & 1.0768 & 21.6708 & 0.0471 & 0.3435 & 1.1011 & 2.5294 & 0.0456 & 0.4749 \\
BenHoch.adp & 1.1304 & 21.8232 & 0.0491 & 0.3595 & 1.1327 & 2.5458 & 0.0465 & 0.4799 \\
BenLui & 0.0490 & 13.7412 & 0.0035 & 0.0458 & 0.2232 & 2.5179 & 0.0162 & 0.2091 \\
gam.su & 1.4084 & 22.3718 & 0.0598 & 0.4329 & 1.2134 & 2.7555 & 0.0482 & 0.4955 \\
gam.su.adp & 1.3813 & 22.2375 & 0.0581 & 0.4487 & 1.2790 & 2.9351 & 0.0505 & 0.4974 \\
gam.su.adp2 & 1.0129 & 21.4829 & 0.0447 & 0.3275 & 1.0654 & 2.5369 & 0.0446 & 0.4693 \\
gam.sd & 1.4081 & 22.3692 & 0.0598 & 0.4332 & 1.2134 & 2.7586 & 0.0482 & 0.4955 \\
gam.su.mod & 1.4041 & 22.3620 & 0.0596 & 0.5272 & 1.2099 & 2.7546 & 0.0481 & 0.4993 \\
gam.sd.mod & 1.4038 & 22.3594 & 0.0596 & 0.5271 & 1.2098 & 2.7578 & 0.0481 & 0.4993 \\
\end{tabular}

**Figure 3.14.** Results for $m = 500$ tests, independent $p$-values, $\mu = 4$, and $A_0 = .95$
<table>
<thead>
<tr>
<th></th>
<th>E(V)</th>
<th>E(R)</th>
<th>FDR</th>
<th>FDP</th>
<th>s(V)</th>
<th>s(R)</th>
<th>s(FDR)</th>
<th>s(FDP)</th>
</tr>
</thead>
<tbody>
<tr>
<td>BenHoch</td>
<td>2.6597</td>
<td>17.9273</td>
<td>0.0284</td>
<td>0.0974</td>
<td>15.9255</td>
<td>35.5825</td>
<td>0.1231</td>
<td>0.2965</td>
</tr>
<tr>
<td>BenHoch.adp</td>
<td>3.8286</td>
<td>19.7441</td>
<td>0.0317</td>
<td>0.1055</td>
<td>23.3483</td>
<td>42.2877</td>
<td>0.1284</td>
<td>0.3072</td>
</tr>
<tr>
<td>BenLui</td>
<td>0.0440</td>
<td>3.1417</td>
<td>0.0102</td>
<td>0.0129</td>
<td>0.5133</td>
<td>7.7945</td>
<td>0.0981</td>
<td>0.1128</td>
</tr>
<tr>
<td>gam.su</td>
<td>2.7639</td>
<td>17.3219</td>
<td>0.0270</td>
<td>0.0962</td>
<td>16.5865</td>
<td>36.6400</td>
<td>0.1159</td>
<td>0.2949</td>
</tr>
<tr>
<td>gam.su.adp</td>
<td>5.2988</td>
<td>20.9166</td>
<td>0.0350</td>
<td>0.1149</td>
<td>30.7189</td>
<td>49.9084</td>
<td>0.1323</td>
<td>0.3189</td>
</tr>
<tr>
<td>gam.su.adp2</td>
<td>2.5126</td>
<td>16.5648</td>
<td>0.0254</td>
<td>0.0886</td>
<td>15.8869</td>
<td>35.3721</td>
<td>0.1138</td>
<td>0.2842</td>
</tr>
<tr>
<td>gam.sd</td>
<td>2.7467</td>
<td>17.2152</td>
<td>0.0269</td>
<td>0.0963</td>
<td>16.5475</td>
<td>36.6313</td>
<td>0.1155</td>
<td>0.2950</td>
</tr>
<tr>
<td>gam.su.mod</td>
<td>2.7556</td>
<td>17.2901</td>
<td>0.0270</td>
<td>0.1003</td>
<td>16.5605</td>
<td>36.5998</td>
<td>0.1159</td>
<td>0.3004</td>
</tr>
<tr>
<td>gam.sd.mod</td>
<td>2.7387</td>
<td>17.1858</td>
<td>0.0268</td>
<td>0.1002</td>
<td>16.5229</td>
<td>36.5918</td>
<td>0.1155</td>
<td>0.3003</td>
</tr>
</tbody>
</table>

Figure 3.15. Results for $m = 500$ tests, dependent $p$-values, $\mu = 2$, and $A_0 = .8$

<table>
<thead>
<tr>
<th></th>
<th>E(V)</th>
<th>E(R)</th>
<th>FDR</th>
<th>FDP</th>
<th>s(V)</th>
<th>s(R)</th>
<th>s(FDR)</th>
<th>s(FDP)</th>
</tr>
</thead>
<tbody>
<tr>
<td>BenHoch</td>
<td>1.9596</td>
<td>3.8960</td>
<td>0.0334</td>
<td>0.0843</td>
<td>17.5150</td>
<td>19.5996</td>
<td>0.1484</td>
<td>0.2779</td>
</tr>
<tr>
<td>BenHoch.adp</td>
<td>2.5980</td>
<td>4.5480</td>
<td>0.0337</td>
<td>0.0845</td>
<td>23.6569</td>
<td>25.6934</td>
<td>0.1495</td>
<td>0.2782</td>
</tr>
<tr>
<td>BenLui</td>
<td>0.0674</td>
<td>0.8523</td>
<td>0.0138</td>
<td>0.0234</td>
<td>1.3952</td>
<td>2.6798</td>
<td>0.1094</td>
<td>0.1512</td>
</tr>
<tr>
<td>gam.su</td>
<td>1.8594</td>
<td>3.6057</td>
<td>0.0296</td>
<td>0.0750</td>
<td>17.5561</td>
<td>19.6326</td>
<td>0.1389</td>
<td>0.2634</td>
</tr>
<tr>
<td>gam.su.adp</td>
<td>2.5811</td>
<td>4.3115</td>
<td>0.0272</td>
<td>0.0697</td>
<td>27.0664</td>
<td>29.0804</td>
<td>0.1344</td>
<td>0.2547</td>
</tr>
<tr>
<td>gam.su.adp2</td>
<td>1.7386</td>
<td>3.4743</td>
<td>0.0286</td>
<td>0.0744</td>
<td>17.0128</td>
<td>19.0675</td>
<td>0.1359</td>
<td>0.2624</td>
</tr>
<tr>
<td>gam.sd</td>
<td>1.7991</td>
<td>3.5452</td>
<td>0.0290</td>
<td>0.0744</td>
<td>17.3733</td>
<td>19.4740</td>
<td>0.1368</td>
<td>0.2624</td>
</tr>
<tr>
<td>gam.su.mod</td>
<td>1.8555</td>
<td>3.5988</td>
<td>0.0296</td>
<td>0.0751</td>
<td>17.5295</td>
<td>19.6053</td>
<td>0.1388</td>
<td>0.2636</td>
</tr>
<tr>
<td>gam.sd.mod</td>
<td>1.7953</td>
<td>3.5382</td>
<td>0.0290</td>
<td>0.0745</td>
<td>17.3473</td>
<td>19.4470</td>
<td>0.1368</td>
<td>0.2626</td>
</tr>
</tbody>
</table>

Figure 3.16. Results for $m = 500$ tests, dependent $p$-values, $\mu = 2$, and $A_0 = .95$
<table>
<thead>
<tr>
<th></th>
<th>E(V)</th>
<th>E(R)</th>
<th>FDR</th>
<th>FDP</th>
<th>s(V)</th>
<th>s(R)</th>
<th>s(FDR)</th>
<th>s(FDP)</th>
</tr>
</thead>
<tbody>
<tr>
<td>BenHoch</td>
<td>4.2287</td>
<td>95.1681</td>
<td>0.0382</td>
<td>0.1613</td>
<td>17.0615</td>
<td>21.1004</td>
<td>0.1012</td>
<td>0.3678</td>
</tr>
<tr>
<td>BenHoch.adp</td>
<td>5.9898</td>
<td>98.0029</td>
<td>0.0459</td>
<td>0.1893</td>
<td>24.7436</td>
<td>26.9598</td>
<td>0.1114</td>
<td>0.3918</td>
</tr>
<tr>
<td>BenLui</td>
<td>0.0483</td>
<td>55.6134</td>
<td>0.0076</td>
<td>0.0123</td>
<td>0.5618</td>
<td>29.3861</td>
<td>0.0757</td>
<td>0.1102</td>
</tr>
<tr>
<td>gam.su</td>
<td>6.6878</td>
<td>98.8471</td>
<td>0.0606</td>
<td>0.3038</td>
<td>17.2905</td>
<td>23.0684</td>
<td>0.0964</td>
<td>0.4599</td>
</tr>
<tr>
<td>gam.su.adp</td>
<td>24.9358</td>
<td>119.4394</td>
<td>0.1875</td>
<td>0.9578</td>
<td>29.8861</td>
<td>34.8513</td>
<td>0.1040</td>
<td>0.2011</td>
</tr>
<tr>
<td>gam.su.adp2</td>
<td>4.1964</td>
<td>94.8363</td>
<td>0.0380</td>
<td>0.1608</td>
<td>16.6983</td>
<td>22.8591</td>
<td>0.0951</td>
<td>0.3674</td>
</tr>
<tr>
<td>gam.sd</td>
<td>6.6146</td>
<td>98.7137</td>
<td>0.0597</td>
<td>0.3020</td>
<td>17.1979</td>
<td>23.4237</td>
<td>0.0941</td>
<td>0.4591</td>
</tr>
<tr>
<td>gam.su.mod</td>
<td>6.6620</td>
<td>98.8035</td>
<td>0.0604</td>
<td>0.4470</td>
<td>17.2274</td>
<td>23.0889</td>
<td>0.0961</td>
<td>0.4972</td>
</tr>
<tr>
<td>gam.sd.mod</td>
<td>6.6000</td>
<td>98.6839</td>
<td>0.0596</td>
<td>0.4452</td>
<td>17.1660</td>
<td>23.4226</td>
<td>0.0941</td>
<td>0.4970</td>
</tr>
</tbody>
</table>

**Figure 3.17.** Results for $m = 500$ tests, dependent $p$-values, $\mu = 4$, and $A_0 = .8$

<table>
<thead>
<tr>
<th></th>
<th>E(V)</th>
<th>E(R)</th>
<th>FDR</th>
<th>FDP</th>
<th>s(V)</th>
<th>s(R)</th>
<th>s(FDR)</th>
<th>s(FDP)</th>
</tr>
</thead>
<tbody>
<tr>
<td>BenHoch</td>
<td>1.8073</td>
<td>21.9193</td>
<td>0.0331</td>
<td>0.1168</td>
<td>14.8624</td>
<td>15.9680</td>
<td>0.1219</td>
<td>0.3212</td>
</tr>
<tr>
<td>BenHoch.adp</td>
<td>2.3649</td>
<td>22.5511</td>
<td>0.0343</td>
<td>0.1194</td>
<td>20.5915</td>
<td>21.4455</td>
<td>0.1249</td>
<td>0.3243</td>
</tr>
<tr>
<td>BenLui</td>
<td>0.0398</td>
<td>13.9077</td>
<td>0.0088</td>
<td>0.0129</td>
<td>0.4879</td>
<td>7.5074</td>
<td>0.0860</td>
<td>0.1128</td>
</tr>
<tr>
<td>gam.su</td>
<td>2.0034</td>
<td>22.2628</td>
<td>0.0410</td>
<td>0.1470</td>
<td>14.8635</td>
<td>16.3472</td>
<td>0.1160</td>
<td>0.3541</td>
</tr>
<tr>
<td>gam.su.adp</td>
<td>3.1845</td>
<td>23.3321</td>
<td>0.0447</td>
<td>0.1571</td>
<td>27.4585</td>
<td>28.5387</td>
<td>0.1255</td>
<td>0.3639</td>
</tr>
<tr>
<td>gam.su.adp2</td>
<td>1.5631</td>
<td>21.4329</td>
<td>0.0282</td>
<td>0.1008</td>
<td>14.3504</td>
<td>15.8005</td>
<td>0.1134</td>
<td>0.3011</td>
</tr>
<tr>
<td>gam.sd</td>
<td>1.9135</td>
<td>22.1673</td>
<td>0.0406</td>
<td>0.1466</td>
<td>14.5026</td>
<td>16.1023</td>
<td>0.1144</td>
<td>0.3537</td>
</tr>
<tr>
<td>gam.su.mod</td>
<td>1.9983</td>
<td>22.2529</td>
<td>0.0409</td>
<td>0.1508</td>
<td>14.8357</td>
<td>16.3233</td>
<td>0.1158</td>
<td>0.3579</td>
</tr>
<tr>
<td>gam.sd.mod</td>
<td>1.9072</td>
<td>22.1556</td>
<td>0.0404</td>
<td>0.1503</td>
<td>14.4694</td>
<td>16.0755</td>
<td>0.1143</td>
<td>0.3574</td>
</tr>
</tbody>
</table>

**Figure 3.18.** Results for $m = 500$ tests, dependent $p$-values, $\mu = 4$, and $A_0 = .95$
Table 3.19. Results for $m = 500$ tests, dependent $p$-values, $\mu = 4$, and $A_0 = .3$

<table>
<thead>
<tr>
<th></th>
<th>$E(V)$</th>
<th>$E(R)$</th>
<th>FDR</th>
<th>FDP</th>
<th>$s(V)$</th>
<th>$s(R)$</th>
<th>$s(FDR)$</th>
<th>$s(FDP)$</th>
</tr>
</thead>
<tbody>
<tr>
<td>BenHoch</td>
<td>5.0792</td>
<td>342.9462</td>
<td>0.0162</td>
<td>0.0789</td>
<td>9.9585</td>
<td>28.6030</td>
<td>0.0368</td>
<td>0.2696</td>
</tr>
<tr>
<td>BenHoch-adp</td>
<td>17.6935</td>
<td>363.0352</td>
<td>0.0475</td>
<td>0.3006</td>
<td>20.8714</td>
<td>25.4406</td>
<td>0.0535</td>
<td>0.4585</td>
</tr>
<tr>
<td>BenLui</td>
<td>0.0439</td>
<td>207.0073</td>
<td>0.0021</td>
<td>0.0051</td>
<td>0.5400</td>
<td>107.7197</td>
<td>0.0370</td>
<td>0.0712</td>
</tr>
<tr>
<td>gam.su</td>
<td>18.4048</td>
<td>359.4244</td>
<td>0.0518</td>
<td>0.3776</td>
<td>10.8709</td>
<td>32.9098</td>
<td>0.0368</td>
<td>0.3645</td>
</tr>
<tr>
<td>gam.su.adp</td>
<td>18.4420</td>
<td>359.5490</td>
<td>0.0519</td>
<td>0.3783</td>
<td>10.8764</td>
<td>31.8030</td>
<td>0.0368</td>
<td>0.4585</td>
</tr>
<tr>
<td>gam.su.adp2</td>
<td>18.4184</td>
<td>359.4985</td>
<td>0.0518</td>
<td>0.5115</td>
<td>10.8714</td>
<td>31.8928</td>
<td>0.0348</td>
<td>0.4999</td>
</tr>
<tr>
<td>gam.sd</td>
<td>18.4048</td>
<td>359.4244</td>
<td>0.0518</td>
<td>0.3776</td>
<td>10.8709</td>
<td>32.9098</td>
<td>0.0368</td>
<td>0.3645</td>
</tr>
<tr>
<td>gam.su.mod</td>
<td>18.4184</td>
<td>359.4985</td>
<td>0.0518</td>
<td>0.5115</td>
<td>10.8714</td>
<td>31.8928</td>
<td>0.0348</td>
<td>0.4999</td>
</tr>
<tr>
<td>gam.sd.mod</td>
<td>18.3820</td>
<td>359.3764</td>
<td>0.0517</td>
<td>0.3776</td>
<td>10.8709</td>
<td>32.9723</td>
<td>0.0368</td>
<td>0.4848</td>
</tr>
</tbody>
</table>

Table 3.20. Results for $m = 500$ tests, independent $p$-values, $\mu = 2$, and $A_0 = .3$

<table>
<thead>
<tr>
<th></th>
<th>$E(V)$</th>
<th>$E(R)$</th>
<th>FDR</th>
<th>FDP</th>
<th>$s(V)$</th>
<th>$s(R)$</th>
<th>$s(FDR)$</th>
<th>$s(FDP)$</th>
</tr>
</thead>
<tbody>
<tr>
<td>BenHoch</td>
<td>4.3468</td>
<td>106.3367</td>
<td>0.0401</td>
<td>0.2848</td>
<td>0.0239</td>
<td>0.2007</td>
<td>2e-04</td>
<td>0.0045</td>
</tr>
<tr>
<td>BenHoch-adp</td>
<td>4.5316</td>
<td>108.6284</td>
<td>0.0409</td>
<td>0.2999</td>
<td>0.0247</td>
<td>0.2064</td>
<td>2e-04</td>
<td>0.0046</td>
</tr>
<tr>
<td>BenLui</td>
<td>0.0395</td>
<td>7.9797</td>
<td>0.0049</td>
<td>0.0391</td>
<td>0.0020</td>
<td>0.0283</td>
<td>3e-04</td>
<td>0.0019</td>
</tr>
<tr>
<td>gam.su</td>
<td>4.2621</td>
<td>105.0272</td>
<td>0.0397</td>
<td>0.2790</td>
<td>0.0239</td>
<td>0.2085</td>
<td>2e-04</td>
<td>0.0045</td>
</tr>
<tr>
<td>gam.su.adp</td>
<td>4.0113</td>
<td>101.7592</td>
<td>0.0385</td>
<td>0.2639</td>
<td>0.0232</td>
<td>0.2086</td>
<td>2e-04</td>
<td>0.0044</td>
</tr>
<tr>
<td>gam.su.adp2</td>
<td>3.8439</td>
<td>99.6688</td>
<td>0.0377</td>
<td>0.2517</td>
<td>0.0224</td>
<td>0.2005</td>
<td>2e-04</td>
<td>0.0043</td>
</tr>
<tr>
<td>gam.sd</td>
<td>4.2578</td>
<td>104.9534</td>
<td>0.0397</td>
<td>0.2788</td>
<td>0.0239</td>
<td>0.2105</td>
<td>2e-04</td>
<td>0.0045</td>
</tr>
<tr>
<td>gam.su.mod</td>
<td>4.2395</td>
<td>104.7378</td>
<td>0.0396</td>
<td>0.3243</td>
<td>0.0238</td>
<td>0.2083</td>
<td>2e-04</td>
<td>0.0047</td>
</tr>
<tr>
<td>gam.sd.mod</td>
<td>4.2347</td>
<td>104.6594</td>
<td>0.0396</td>
<td>0.3243</td>
<td>0.0238</td>
<td>0.2102</td>
<td>2e-04</td>
<td>0.0047</td>
</tr>
</tbody>
</table>

Table 3.21. Results for $m = 5000$ tests, independent $p$-values, $\mu = 2$, and $A_0 = .8$
<table>
<thead>
<tr>
<th></th>
<th>E(V)</th>
<th>E(R)</th>
<th>FDR</th>
<th>FDP</th>
<th>s(V)</th>
<th>s(R)</th>
<th>s(FDR)</th>
<th>s(FDP)</th>
</tr>
</thead>
<tbody>
<tr>
<td>BenHoch</td>
<td>0.4085</td>
<td>6.9440</td>
<td>0.0472</td>
<td>0.0073</td>
<td>0.0482</td>
<td>0.0009</td>
<td>0.0046</td>
<td></td>
</tr>
<tr>
<td>BenHoch.adp</td>
<td>0.4077</td>
<td>6.9255</td>
<td>0.0469</td>
<td>0.0073</td>
<td>0.0486</td>
<td>0.0009</td>
<td>0.0046</td>
<td></td>
</tr>
<tr>
<td>BenLui</td>
<td>0.0476</td>
<td>2.0335</td>
<td>0.0203</td>
<td>0.0022</td>
<td>0.0140</td>
<td>0.0011</td>
<td>0.0021</td>
<td></td>
</tr>
<tr>
<td>gam.su</td>
<td>0.2844</td>
<td>5.5159</td>
<td>0.0372</td>
<td>0.0061</td>
<td>0.0467</td>
<td>0.0009</td>
<td>0.0041</td>
<td></td>
</tr>
<tr>
<td>gam.su.adp</td>
<td>0.2813</td>
<td>5.4967</td>
<td>0.0371</td>
<td>0.0060</td>
<td>0.0467</td>
<td>0.0009</td>
<td>0.0041</td>
<td></td>
</tr>
<tr>
<td>gam.su.adp2</td>
<td>0.2813</td>
<td>5.4967</td>
<td>0.0371</td>
<td>0.0060</td>
<td>0.0462</td>
<td>0.0009</td>
<td>0.0041</td>
<td></td>
</tr>
<tr>
<td>gam.sd</td>
<td>0.2836</td>
<td>5.5105</td>
<td>0.0372</td>
<td>0.0061</td>
<td>0.0467</td>
<td>0.0009</td>
<td>0.0041</td>
<td></td>
</tr>
<tr>
<td>gam.su.mod</td>
<td>0.2829</td>
<td>5.4973</td>
<td>0.0371</td>
<td>0.0061</td>
<td>0.0466</td>
<td>0.0009</td>
<td>0.0041</td>
<td></td>
</tr>
<tr>
<td>gam.sd.mod</td>
<td>0.2821</td>
<td>5.4922</td>
<td>0.0370</td>
<td>0.0060</td>
<td>0.0465</td>
<td>0.0009</td>
<td>0.0041</td>
<td></td>
</tr>
</tbody>
</table>

**Figure 3.22.** Results for $m = 5000$ tests, independent $p$-values, $\mu = 2$, and $A_0 = .95$

<table>
<thead>
<tr>
<th></th>
<th>E(V)</th>
<th>E(R)</th>
<th>FDR</th>
<th>FDP</th>
<th>s(V)</th>
<th>s(R)</th>
<th>s(FDR)</th>
<th>s(FDP)</th>
</tr>
</thead>
<tbody>
<tr>
<td>BenHoch</td>
<td>38.4908</td>
<td>959.1931</td>
<td>0.0401</td>
<td>0.0671</td>
<td>0.0649</td>
<td>0.1175</td>
<td>1e-04</td>
<td>0.0025</td>
</tr>
<tr>
<td>BenHoch.adp</td>
<td>48.6057</td>
<td>980.7103</td>
<td>0.0495</td>
<td>0.4622</td>
<td>0.0744</td>
<td>0.1203</td>
<td>1e-04</td>
<td>0.0050</td>
</tr>
<tr>
<td>BenLui</td>
<td>0.0420</td>
<td>346.1785</td>
<td>0.0001</td>
<td>0.0000</td>
<td>0.0020</td>
<td>0.1536</td>
<td>0e+00</td>
<td>0.0000</td>
</tr>
<tr>
<td>gam.su</td>
<td>61.6824</td>
<td>1004.1486</td>
<td>0.0614</td>
<td>0.9477</td>
<td>0.0774</td>
<td>0.1239</td>
<td>1e-04</td>
<td>0.0022</td>
</tr>
<tr>
<td>gam.su.adp</td>
<td>213.2903</td>
<td>1193.8257</td>
<td>0.1785</td>
<td>1.0000</td>
<td>0.1599</td>
<td>0.1842</td>
<td>1e-04</td>
<td>0.0000</td>
</tr>
<tr>
<td>gam.su.adp2</td>
<td>40.6528</td>
<td>964.0538</td>
<td>0.0421</td>
<td>0.1169</td>
<td>0.0664</td>
<td>0.1213</td>
<td>1e-04</td>
<td>0.0032</td>
</tr>
<tr>
<td>gam.sd</td>
<td>61.6824</td>
<td>1004.1486</td>
<td>0.0614</td>
<td>0.9477</td>
<td>0.0774</td>
<td>0.1239</td>
<td>1e-04</td>
<td>0.0022</td>
</tr>
<tr>
<td>gam.su.mod</td>
<td>61.5314</td>
<td>1003.8971</td>
<td>0.0612</td>
<td>0.9755</td>
<td>0.0773</td>
<td>0.1238</td>
<td>1e-04</td>
<td>0.0015</td>
</tr>
<tr>
<td>gam.sd.mod</td>
<td>61.5314</td>
<td>1003.8971</td>
<td>0.0612</td>
<td>0.9755</td>
<td>0.0773</td>
<td>0.1238</td>
<td>1e-04</td>
<td>0.0015</td>
</tr>
</tbody>
</table>

**Figure 3.23.** Results for $m = 5000$ tests, independent $p$-values, $\mu = 4$, and $A_0 = .8$

<table>
<thead>
<tr>
<th></th>
<th>E(V)</th>
<th>E(R)</th>
<th>FDR</th>
<th>FDP</th>
<th>s(V)</th>
<th>s(R)</th>
<th>s(FDR)</th>
<th>s(FDP)</th>
</tr>
</thead>
<tbody>
<tr>
<td>BenHoch</td>
<td>10.3415</td>
<td>216.6230</td>
<td>0.0475</td>
<td>0.1414</td>
<td>0.0340</td>
<td>0.0781</td>
<td>1e-04</td>
<td>0.0049</td>
</tr>
<tr>
<td>BenHoch.adp</td>
<td>10.8862</td>
<td>218.1420</td>
<td>0.0496</td>
<td>0.4708</td>
<td>0.0351</td>
<td>0.0788</td>
<td>2e-04</td>
<td>0.0050</td>
</tr>
<tr>
<td>BenLui</td>
<td>0.0481</td>
<td>85.5618</td>
<td>0.0006</td>
<td>0.0000</td>
<td>0.0022</td>
<td>0.0757</td>
<td>0e+00</td>
<td>0.0000</td>
</tr>
<tr>
<td>gam.su</td>
<td>13.9984</td>
<td>225.7968</td>
<td>0.0617</td>
<td>0.7613</td>
<td>0.0386</td>
<td>0.0820</td>
<td>2e-04</td>
<td>0.0043</td>
</tr>
<tr>
<td>gam.su.adp</td>
<td>13.5014</td>
<td>224.6031</td>
<td>0.0598</td>
<td>0.7192</td>
<td>0.0391</td>
<td>0.0858</td>
<td>2e-04</td>
<td>0.0043</td>
</tr>
<tr>
<td>gam.su.adp2</td>
<td>9.7676</td>
<td>214.9320</td>
<td>0.0452</td>
<td>0.3568</td>
<td>0.0329</td>
<td>0.0781</td>
<td>1e-04</td>
<td>0.0048</td>
</tr>
<tr>
<td>gam.sd</td>
<td>13.9984</td>
<td>225.7968</td>
<td>0.0617</td>
<td>0.7613</td>
<td>0.0386</td>
<td>0.0820</td>
<td>2e-04</td>
<td>0.0043</td>
</tr>
<tr>
<td>gam.su.mod</td>
<td>13.9606</td>
<td>225.7087</td>
<td>0.0616</td>
<td>0.8134</td>
<td>0.0386</td>
<td>0.0821</td>
<td>2e-04</td>
<td>0.0039</td>
</tr>
<tr>
<td>gam.sd.mod</td>
<td>13.9606</td>
<td>225.7087</td>
<td>0.0616</td>
<td>0.8134</td>
<td>0.0386</td>
<td>0.0821</td>
<td>2e-04</td>
<td>0.0039</td>
</tr>
</tbody>
</table>

**Figure 3.24.** Results for $m = 5000$ tests, independent $p$-values, $\mu = 4$, and $A_0 = .95$
<table>
<thead>
<tr>
<th>E(V)</th>
<th>E(R)</th>
<th>FDR</th>
<th>FDP</th>
<th>s(V)</th>
<th>s(R)</th>
<th>s(FDR)</th>
<th>s(FDP)</th>
</tr>
</thead>
<tbody>
<tr>
<td>BenHoch</td>
<td>26.4569</td>
<td>171.7777</td>
<td>0.0261</td>
<td>0.0871</td>
<td>1.5415</td>
<td>3.4772</td>
<td>0.0011</td>
</tr>
<tr>
<td>BenHoch.adp</td>
<td>38.2364</td>
<td>190.3472</td>
<td>0.0295</td>
<td>0.0959</td>
<td>2.2574</td>
<td>4.1252</td>
<td>0.0012</td>
</tr>
<tr>
<td>BenLui</td>
<td>0.0385</td>
<td>7.9429</td>
<td>0.0083</td>
<td>0.0083</td>
<td>0.0051</td>
<td>0.2849</td>
<td>0.0009</td>
</tr>
<tr>
<td>gam.su</td>
<td>28.0218</td>
<td>172.3364</td>
<td>0.0256</td>
<td>0.0891</td>
<td>1.6021</td>
<td>3.5822</td>
<td>0.0011</td>
</tr>
<tr>
<td>gam.su.adp</td>
<td>51.9683</td>
<td>203.0960</td>
<td>0.0327</td>
<td>0.1020</td>
<td>3.0230</td>
<td>4.9254</td>
<td>0.0012</td>
</tr>
<tr>
<td>gam.su.adp2</td>
<td>25.5928</td>
<td>163.6145</td>
<td>0.0239</td>
<td>0.0828</td>
<td>1.5313</td>
<td>3.4610</td>
<td>0.0011</td>
</tr>
<tr>
<td>gam.sd</td>
<td>27.9388</td>
<td>170.6179</td>
<td>0.0253</td>
<td>0.0889</td>
<td>1.6011</td>
<td>3.5868</td>
<td>0.0011</td>
</tr>
<tr>
<td>gam.su.mod</td>
<td>27.9356</td>
<td>172.0356</td>
<td>0.0255</td>
<td>0.0912</td>
<td>1.5986</td>
<td>3.5780</td>
<td>0.0011</td>
</tr>
<tr>
<td>gam.sd.mod</td>
<td>27.8536</td>
<td>170.3390</td>
<td>0.0253</td>
<td>0.0910</td>
<td>1.5977</td>
<td>3.5825</td>
<td>0.0011</td>
</tr>
</tbody>
</table>

**Figure 3.25.** Results for \( m = 5000 \) tests, dependent \( p \)-values, \( \mu = 2 \), and \( A_0 = .8 \)

<table>
<thead>
<tr>
<th>E(V)</th>
<th>E(R)</th>
<th>FDR</th>
<th>FDP</th>
<th>s(V)</th>
<th>s(R)</th>
<th>s(FDR)</th>
<th>s(FDP)</th>
</tr>
</thead>
<tbody>
<tr>
<td>BenHoch</td>
<td>17.8850</td>
<td>33.8611</td>
<td>0.0276</td>
<td>0.0744</td>
<td>1.6459</td>
<td>1.8197</td>
<td>0.0013</td>
</tr>
<tr>
<td>BenHoch.adp</td>
<td>24.5374</td>
<td>40.7237</td>
<td>0.0279</td>
<td>0.0742</td>
<td>2.3176</td>
<td>2.4819</td>
<td>0.0013</td>
</tr>
<tr>
<td>BenLui</td>
<td>0.0451</td>
<td>2.0083</td>
<td>0.0092</td>
<td>0.0111</td>
<td>0.0060</td>
<td>0.0724</td>
<td>0.0009</td>
</tr>
<tr>
<td>gam.su</td>
<td>17.5673</td>
<td>32.8957</td>
<td>0.0257</td>
<td>0.0731</td>
<td>1.6587</td>
<td>1.8329</td>
<td>0.0012</td>
</tr>
<tr>
<td>gam.su.adp</td>
<td>24.6383</td>
<td>39.8823</td>
<td>0.0257</td>
<td>0.0713</td>
<td>2.5889</td>
<td>2.7758</td>
<td>0.0013</td>
</tr>
<tr>
<td>gam.su.adp2</td>
<td>16.6090</td>
<td>31.3797</td>
<td>0.0245</td>
<td>0.0697</td>
<td>1.6112</td>
<td>1.7811</td>
<td>0.0012</td>
</tr>
<tr>
<td>gam.sd</td>
<td>17.2190</td>
<td>32.4509</td>
<td>0.0252</td>
<td>0.0726</td>
<td>1.6520</td>
<td>1.8272</td>
<td>0.0012</td>
</tr>
<tr>
<td>gam.su.mod</td>
<td>17.5104</td>
<td>32.8172</td>
<td>0.0256</td>
<td>0.0748</td>
<td>1.6547</td>
<td>1.8289</td>
<td>0.0012</td>
</tr>
<tr>
<td>gam.sd.mod</td>
<td>17.1649</td>
<td>32.3722</td>
<td>0.0251</td>
<td>0.0742</td>
<td>1.6481</td>
<td>1.8233</td>
<td>0.0012</td>
</tr>
</tbody>
</table>

**Figure 3.26.** Results for \( m = 5000 \) tests, dependent \( p \)-values, \( \mu = 2 \), and \( A_0 = .95 \)

<table>
<thead>
<tr>
<th>E(V)</th>
<th>E(R)</th>
<th>FDR</th>
<th>FDP</th>
<th>s(V)</th>
<th>s(R)</th>
<th>s(FDR)</th>
<th>s(FDP)</th>
</tr>
</thead>
<tbody>
<tr>
<td>BenHoch</td>
<td>41.4640</td>
<td>951.7149</td>
<td>0.0378</td>
<td>0.1562</td>
<td>1.6380</td>
<td>2.0024</td>
<td>0.0010</td>
</tr>
<tr>
<td>BenHoch.adp</td>
<td>59.3765</td>
<td>980.6770</td>
<td>0.0457</td>
<td>0.1853</td>
<td>2.4143</td>
<td>2.5807</td>
<td>0.0011</td>
</tr>
<tr>
<td>BenLui</td>
<td>0.0409</td>
<td>349.3361</td>
<td>0.0058</td>
<td>0.0081</td>
<td>0.0053</td>
<td>2.7420</td>
<td>0.0007</td>
</tr>
<tr>
<td>gam.su</td>
<td>68.4469</td>
<td>994.3783</td>
<td>0.0624</td>
<td>0.3493</td>
<td>1.6689</td>
<td>2.1239</td>
<td>0.0010</td>
</tr>
<tr>
<td>gam.su.adp</td>
<td>249.6182</td>
<td>1195.3078</td>
<td>0.1881</td>
<td>0.9907</td>
<td>3.0661</td>
<td>3.5499</td>
<td>0.0010</td>
</tr>
<tr>
<td>gam.su.adp2</td>
<td>42.8616</td>
<td>954.0977</td>
<td>0.0389</td>
<td>0.1620</td>
<td>1.6159</td>
<td>2.1023</td>
<td>0.0009</td>
</tr>
<tr>
<td>gam.sd</td>
<td>67.3798</td>
<td>992.4951</td>
<td>0.0606</td>
<td>0.3462</td>
<td>1.6612</td>
<td>2.1879</td>
<td>0.0009</td>
</tr>
<tr>
<td>gam.su.mod</td>
<td>68.2372</td>
<td>994.0024</td>
<td>0.0622</td>
<td>0.3887</td>
<td>1.6651</td>
<td>2.1257</td>
<td>0.0010</td>
</tr>
<tr>
<td>gam.sd.mod</td>
<td>67.1857</td>
<td>992.1293</td>
<td>0.0605</td>
<td>0.3855</td>
<td>1.6576</td>
<td>2.1894</td>
<td>0.0009</td>
</tr>
</tbody>
</table>

**Figure 3.27.** Results for \( m = 5000 \) tests, dependent \( p \)-values, \( \mu = 4 \), and \( A_0 = .8 \)
<table>
<thead>
<tr>
<th></th>
<th>E(V)</th>
<th>E(R)</th>
<th>FDR</th>
<th>FDP</th>
<th>s(V)</th>
<th>s(R)</th>
<th>s(FDR)</th>
<th>s(FDP)</th>
</tr>
</thead>
<tbody>
<tr>
<td>BenHoch</td>
<td>20.9892</td>
<td>221.2076</td>
<td>0.0353</td>
<td>0.1144</td>
<td>1.6811</td>
<td>1.7595</td>
<td>0.0012</td>
<td>0.0032</td>
</tr>
<tr>
<td>BenHoch.adp</td>
<td>28.3471</td>
<td>229.5020</td>
<td>0.0364</td>
<td>0.1167</td>
<td>2.3798</td>
<td>2.4441</td>
<td>0.0013</td>
<td>0.0032</td>
</tr>
<tr>
<td>BenLui</td>
<td>0.0457</td>
<td>86.0626</td>
<td>0.0076</td>
<td>0.0089</td>
<td>0.0060</td>
<td>0.6788</td>
<td>0.0008</td>
<td>0.0009</td>
</tr>
<tr>
<td>gam.su</td>
<td>23.9622</td>
<td>228.9967</td>
<td>0.0444</td>
<td>0.1928</td>
<td>1.6922</td>
<td>1.7869</td>
<td>0.0012</td>
<td>0.0039</td>
</tr>
<tr>
<td>gam.su.adp</td>
<td>31.8494</td>
<td>236.1930</td>
<td>0.0453</td>
<td>0.1911</td>
<td>2.6468</td>
<td>2.7486</td>
<td>0.0012</td>
<td>0.0039</td>
</tr>
<tr>
<td>gam.su.adp2</td>
<td>19.3037</td>
<td>217.9352</td>
<td>0.0303</td>
<td>0.0975</td>
<td>1.6437</td>
<td>1.7397</td>
<td>0.0012</td>
<td>0.0030</td>
</tr>
<tr>
<td>gam.sd</td>
<td>23.3131</td>
<td>228.2389</td>
<td>0.0435</td>
<td>0.1911</td>
<td>1.6813</td>
<td>1.7837</td>
<td>0.0012</td>
<td>0.0039</td>
</tr>
<tr>
<td>gam.su.mod</td>
<td>23.8747</td>
<td>228.8447</td>
<td>0.0442</td>
<td>0.2172</td>
<td>1.6879</td>
<td>1.7833</td>
<td>0.0012</td>
<td>0.0041</td>
</tr>
<tr>
<td>gam.sd.mod</td>
<td>23.2356</td>
<td>228.0968</td>
<td>0.0434</td>
<td>0.2153</td>
<td>1.6771</td>
<td>1.7801</td>
<td>0.0012</td>
<td>0.0041</td>
</tr>
</tbody>
</table>

**Figure 3.28.** Results for $m = 5000$ tests, dependent $p$-values, $\mu = 4$, and $A_0 = .95$
Applications

4.1 Introduction

No doubt, the rapidly increasing literature on the FDR and other multiple testing error rates is due to the importance of the applications involving large-scale testing, where the FWER is inappropriately restrictive. Two prominent examples are gene microarrays, where potentially thousands of genes are simultaneously analyzed for expression, and astronomical data surveys, whose “datacubes” may involve millions of locations at each of thousands of radio frequencies—astronomical numbers indeed! Moreover, these numbers are ever growing with the increasing technology of the facilities involved in collecting the data. In this chapter, some of the methods discussed here are applied to each of these examples.

4.2 Astronomical source detection

The image in Figure 4.1 is of a single channel of a relatively small datacube of dimension $300 \times 301$ locations and with two visually apparent sources. This is an ideal image in that the two sources are easily identified at a glance. When the sources are not so obvious—and considering that surveys of this type may generate thousands of these images—a more systematic algorithm
is required. Essentially, each location, or pixel, in this image is treated with an individual hypothesis test of $H_0$: “source present” versus $H_a$: “source not present”. The resulting data consists of an extremely large set of $p$-values, and the threshold for source detection is subject to control of an appropriate error rate over the whole image. We discuss this process, beginning with a brief background of the situation. All images of datacubes in this chapter are credited to www.cv.nrao.edu.

![Figure 4.1. Visual detection of source signals in a single channel of a datacube](image)

### 4.2.1 The role of radio surveys in astronomy

Currently, astrophysicists believe that only about 4% of the total energy density in the universe is in the electromagnetic spectrum. The remaining 96% of this energy is divided into dark matter (22%) and dark energy (74%) [6]. Not only are these dark components invisible, but they also do not interact with electromagnetic forces, which makes them difficult to study. The Swiss astrophysicist Fritz Zwicky first found evidence of dark matter in 1933, but others have corroborated his results only in the last 40 years or so, beginning with Vera Rubin in the late 1960s [37]. Thus, there is still much to learn about dark matter, and determining its nature is considered one of the most important problems in modern cosmology.
According to theory, galactic structures collapse and merge together in a hierarchical fashion. Hot gases gravitate together and form stars, and the aggregation of stars make up galaxies. However, observations show that the visible matter of many galaxies collects together in a disc shape and exhibits temperature and pressure too high to be explained without a significant amount of dark matter acting to compress the visible part into the observable disc. Moreover, the amount and shape of dark matter within a galaxy is associated with the thickness of this disc: thinner galaxies tend to contain a greater ratio of dark matter to ordinary matter than that for thicker galaxies [21]. Moreover, many of the highly-flattened disc galaxies—the so-called superthins—have simple structures that allow for easier investigation of their morphology [35].

Unfortunately, these superthins are usually low-surface brightness galaxies (LSBGs), which are not represented as well as high surface brightness galaxies (HSBGs) in optical samples since HSBGs can be seen from longer distances [14]. To avoid this selection bias, astrophysicists commonly collect radio wave samples, rather than optical ones. The HI Parkes All Sky Survey (HIPASS), collected between 1977 and 2002, is one such sample. Radio wave detection achieves its selection impartiality by measuring proton counts of neutral hydrogen (HI), which are emitted from LSBGs in the same range as that for HSBGs [29]. Moreover, the spatial resolution for radio receivers is much better—up to a factor of 100—than that for visible wavelength telescopes such as the Hubble Space Telescope.

Radiotelescopes can detect HI emissions from LSBGs and HSBGs over multiple wavebands, much as a radio can pick up multiple audio stations. Additionally, two-dimensional coordinate location information is recorded, which provides the right ascension (longitude) and declination (latitude) of each response. These are plotted on the horizontal and vertical axes, respectively, of Figure 4.1. Thus, these radio surveys produce a “datacube” of information with two axes for location and a third axis for the wavebands. The illustration of the three-dimensional nature in Figure 4.2 is due to Meyer et al in [17].

Not surprisingly, these datacubes are quite large, five to six gigabytes of
data each with typically $512 \times 512$ locations and 128 channels, and they will continue to enlarge as new technology allows for higher spatial resolutions and wider frequencies. In particular, interferometry can combine signals from multiple antennae, such as the 27-antennae Expanded Very Large Array (EVLA) under construction in New Mexico, to often achieve the quantum limit. With recent upgrades to EVLA’s construction, it is expected to capture 16000 channels with sensitivity and resolution several times better than what we have previously seen. By 2012, radio telescopes will reach resolutions of $2048 \times 2048$ at each of 16000 channels. Moreover, the number of datacubes these telescopes will produce will be extremely large. After only a week of operation, these new telescopes will provide enough data to revolutionize our understanding of galaxy evolution and gas accretion [20].

4.2.2 Bump-hunting in radio datacubes

In order to keep up with the advancements in radio surveys and the extremely large numbers of datacubes they produce, our current analysis techniques must also advance. In addition to the neutral hydrogen signals emitted from the
distant galaxies of interest, a typical datacube includes various contaminants received by the radio telescopes that complicate matters. Quasars and radio galaxies emit detectable but uninteresting radio signals continuously throughout many channels. In contrast, the desirable HI emission is faint and restricted to a few channels. Additionally, radio frequency interference (RFI) from cell phones, satellites, and even the radio telescopes themselves can corrupt the data [19]. Subtracting these unwanted signals is difficult enough, but random noise (not necessarily white) complicates the situation further still.

When working with a small number of manageably sized datacubes, astronomers and astrophysicists can rely on a simple visual inspection of each channel and, at a glance, discern whether a certain location represents a true signal of interest or a contaminant of some kind. An example of this situation is illustrated in Figure 4.1.

Another approach involves establishing a cutoff flux level (how bright a certain location is) and claiming that any candidate signal above that level comes from a non-null source of interest. Two examples of this are the National Radio Astronomy Observatory’s (NRAO) SERCH algorithm and a similar method called PICASSO. These methods work well when the noise is Gaussian, but they are not inherently 3-dimensional [18]. Others, such as TopHat, cross-correlates the data with profiles and groups features in adjacent velocity planes [17]. There are still others with even more sophisticated mechanisms designed to detect non-null sources, but they still rely on visual inspection for their reliability, which is not practical for handling large numbers of datacubes of current size.

4.2.3 Statistical application

The problem at hand is to identify true faint source signals (non-null values) in extremely large datacubes with correlation within each 2-dimensional channel, possible correlation among channels. The goal is an algorithm that does not rely on human visual inspection to check reliability but does allow for
human intervention and modification if desired. Additionally, the algorithm should be supported by rigorous statistical techniques to validate the results. A central issue here, since multiple hypothesis tests are to be conducted, is an appropriate control for multiplicity. Thus, the methods discussed above lend themselves naturally to this problem.

Before applying the statistical procedures, the data is “cleaned” by suppressing the effects of the RFI and the continuum radio sources. This may be accomplished with Fourier filtering or polynomial fitting. Researchers in astronomy handle this stage, but the data values are still not in the form of $p$-values as described above. For example, a histogram of the null data (noise) in Figure 4.1 appears normally distributed—not uniform—reflecting the filtering applied to it. This is revealed in Figure 4.3a. Therefore, before application of the procedures, which expect $p$-values in $(0, 1)$, we apply the transformation $2\Phi(-|\frac{x}{\sigma}|)$ to each value $x$ in the datacube as motivated in the discussion in the previous chapter. The division by $\sigma \approx 0.0223$ is to standardize the variance. The resulting transformed values, both null and non-null, are the $p$-values we work with. Figure 4.3b illustrates these transformed values while preserving the spatial structure. The sources are not as obviously noticeable as in Figure 4.1.

After applying each procedure, we have a $p$-value threshold that we use to discretize the values in the datacube image, according to whether a source is deemed present at that location. Continuing with the image in Figure 4.1, the procedures provide the thresholds given in Figure 4.4. For all procedures, $\alpha = 0.05$. These results are consistent with the large $m$ simulations in Chapter 3. Here, the number of tests (pixel locations) is $m = 300(301) = 90300$. Despite the difference between the Benjamini and Hochberg threshold and those of $\gamma$SUP and $\gamma$SDP, the spatial arrangements of rejections for the procedures are indistinguishable, as can be seen in Figure 4.5.

For a particularly large datacube, it is more efficient to apply these procedures in two stages. The first step, which is represent above, essentially trims away most of the uninteresting noise from the parts of the data that could
Figure 4.3. a) Approximate normal density of noise in Figure 4.1 with mean $\approx 0$ and standard deviation $\approx 0.0223$ b) Datacube replaced with $p$-values according to the normal noise in a)

Figure 4.4. $p$-value thresholds for the datacube image in Figure 4.1 with $\alpha = .05$

potentially contain non-null sources. Since the vast majority of locations and channels will not contain non-null sources, investigating each individual location separately at this point is inefficient. However, if sub-cubes of manageable size are considered, say $50 \times 50$, whether a non-null source is contained somewhere inside can be satisfactorily determined. Then, in the next step, the exact location can be searched for more meticulously.
With the above application serving as the initial step identifying the source locations, consider the following sub-cube of size $51 \times 81$ in Figure 4.6, along with its transformed $p$-values. A similar transformation is applied to convert the location values into suitable $p$-values, and these are administered to the procedures. The thresholds are tabulated in Figure 4.7. With the larger proportion of non-null $p$-values present at this step, the more conservative
<table>
<thead>
<tr>
<th>Procedure</th>
<th>Cutoff</th>
<th>Rej</th>
</tr>
</thead>
<tbody>
<tr>
<td>BenHoch</td>
<td>.002951510</td>
<td>5931</td>
</tr>
<tr>
<td>BenHoch.adp</td>
<td>.002951510</td>
<td>5931</td>
</tr>
<tr>
<td>BenLui</td>
<td>.000011820</td>
<td>2200</td>
</tr>
<tr>
<td>gam.su</td>
<td>.002951510</td>
<td>5931</td>
</tr>
<tr>
<td>gam.su.adp</td>
<td>.002951510</td>
<td>5931</td>
</tr>
<tr>
<td>gam.su.adp2</td>
<td>.001752808</td>
<td>5032</td>
</tr>
<tr>
<td>gam.sd</td>
<td>.002951510</td>
<td>5931</td>
</tr>
<tr>
<td>gam.su.mod</td>
<td>.002951510</td>
<td>5931</td>
</tr>
<tr>
<td>gam.sd.mod</td>
<td>.002951510</td>
<td>5931</td>
</tr>
</tbody>
</table>

**Figure 4.7.** $p$-value thresholds for the data sub-cube image in Figure 4.6 with $\alpha = .05$

The **BenLui** procedure fares better. Figure 4.8 compares this procedure with the **gam.su**.

![Figure 4.8](image)

**Figure 4.8.** $p$-values in the data sub-cube, discretized according to rejection for the **gam.su** procedure (a) and the **BenLui** procedure (b)

### 4.2.4 Comments

Some of the most important questions in astrophysics depend on accurate interpretations and classification of HI radio surveys. The development of
modern radio telescopes in the near future and the quantity of data surveys they will produce demand fully automated algorithms for detecting faint signals. Not only is visual inspection impractical for the number of datacubes soon to be produced, but it also allows for ambiguity in conclusions. Multiple-testing methods, including the ones proposed in this dissertation, provide a systematic way to determine which locations contain true sources and which are merely noise.

One will note that the discussion of these so-called datacubes has focused on a single two-dimensional channel. This is for ease of explanation here. The entire cube may be addressed in similar fashion, with the exception being the treatment of dependency among channels relative to the dependency among locations within a channel. Although, all locations within a channel may be positively correlated—at least in a local sense—the channel-to-channel dependence is fundamentally different. Since sources of this nature tend to persist in multiple adjacent channels, one strategy is to combine multiple channels together into a single test statistic at a specific location. Then, a $p$-value may be constructed by consideration of the joint behavior of the multiple channels at that location. This is one of many interesting questions raised in the treatment of astronomical surveys.

### 4.3 Genetic expression

The data for this study consists of 1185 gene expressions for each of 84 samples of individuals arranged in a microarray. Each sample consists of three groups: normal, pancreatitis, and cancer. Among the 1185 expressions, 291 are removed due to missing values within the samples. This data is also considered by the authors in [36], although they use a nearest-neighbor approach to handling missing data.

In terms of the statistical application, this example is similar to the astronomical example above in that we work with a large collection of $p$-values, each of which is to be rejected or accepted based on a threshold determined
by the procedures. In the astronomical example, rejection means concluding that a source is present. Here, a p-value at a specific gene measures how much the three subject groups differ regarding their gene expression. Rejecting this a p-value is to conclude that the groups are different. To prepare such p-values from the raw microarray data, we use the techniques described below.

### 4.3.1 Normalization

Microarray data should be normalized for a variety of reasons, including unequal quantities in starting RNA, differences in detection efficiencies between the fluorescent dyes, and other systematic biases [23]. We follow the approach of Quackenbush [23], and further details of this normalization can be found there.

To detect various systematic biases we may plot the log ratio of red to green intensities, referred to as the “ratio-intensity” (R-I), against the log product. Specifically, if $R$ and $G$ are the red and green signals (foreground less background), then a plot of $M = \log_2(R/G)$ against $A = .5\log_2(RG)$ should exhibit a scatter about zero. The plots below for a normal group and for a cancer group illustrate significant departure from this. The pancreatitis groups behave similarly. Each of the plots in Figure 4.9 is of an individual sample, but the others behave similarly.

![Diagnostic plots before normalization](image)

**Figure 4.9.** Diagnostic plots before normalization
One way to correct for these issues is to replace R-I with the residuals obtained from regression on log product. The plots in Figure 4.10 show the results for representative normal and cancer groups. As expected, the scatter is about zero, but there is still not the desired random scatter. Quadratic regression is of little help as well. Instead, the next approach is loess regression.

Finally, the plots in Figure 4.11 below demonstrate the effect of the lowess fit. These show the random scatter we seek in order to perform various analyses. Therefore, the (normalized) lowess residuals replace the non-normalized log ratio as the gene expression response of interest.
4.3.2 Comparing gene expressions among groups

Since each of the 84 samples comes from exactly one of three groups, a question of interest is whether these groups exhibit different behavior. At each gene, we fit an analysis of variance (ANOVA) model comparing the three groups, where the response is the normalized log ratio of gene expression; residual plots for this model are in Figure 4.12 for a representative gene.

Figure 4.12. Residuals and histogram of studentized residuals of ANOVA model for a representative gene

The ANOVA $F$ statistic for this particular location is 1.061 with a $p$-value of 0.3542, so no further pairwise comparisons are necessary. After applying the testing procedures to this set of $p$-values we obtain the rejection thresholds in Figure 4.13. Plots of all $p$-values and their rejection indicators according to the $\gamma$SUP are given in Figure 4.14. Note that the $p$-values for this array are quite moderate, and only a few genes are identified as differentially expressed. Such a plot for the BHP is indistinguishable and is therefore omitted.

4.4 Conclusions

Both astronomical datacubes and gene expression microarrays provide large families of hypothesis tests to consider simultaneously. Because of the extreme number of tests, which can be much higher than that considered here, controlling the probability of any false positive result—referred to above as the
Procedure       Cutoff   Rej
BenHoch         .001745783 19
BenHoch.adp     .002006225 21
BenLui          .000099663  3
gam.su          .001349244 15
gam.su.adp      .001185699 14
gam.su.adp2     .001185699 14
gam.sd          .001349244 15
gam.su.mod      .001349244 15
gam.sd.mod      .001349244 15

Figure 4.13.  $p$-value thresholds for the microarray with $\alpha = .1$

Figure 4.14.  $p$-values for the ANOVA at each gene location

FWER—is prohibitively restrictive. Moreover, the examples here demonstrate situations where meaningful conclusions can be made, even in the presence of a small fraction of false positives.

In the first example, no doubt some of the specific locations are incorrectly identified as containing an astronomical source. Nevertheless, the general location of the sources is correctly determined because the correct discoveries sufficiently outweigh the incorrect ones. This is the defining property of the FDR and is precisely what makes it so appropriate for this example. Likewise,
in the microarray, we are essentially comparing only three groups, despite doing for over 1000 individual genes. If we wish to make a general conclusion about the similarities or disparities of these groups, we need not require that every one of our rejected gene tests be correct. A sufficiently high proportion of correct ones, as achieved by controlling the FDR, is adequate.
We have seen how the multiplicity problem arises in simultaneous testing and how the more recent false discovery rate lends itself well to these situations, specifically when the number of tests is too large for the traditional family-wise error rate to be applied. As long as meaningful conclusions are possible in the presence of a small fraction of incorrect discoveries, the false discovery rate is the error rate of choice.

Not only is the false discovery rate an appropriate quantity to consider, but the available procedures to control it are also easy and efficient to use. Ease and efficiency of use are necessary properties of a procedure in large-scale testing since the number of individual tests may number in the millions. In Chapter 1, we defined the step-up and step-down procedures, which are among the most popular FDR controlling methods available, and we saw that various modifications and generalizations allow them to apply to any situation. This dissertation provides a generalization to these step-wise procedures by introducing another quantity, $\gamma$, as the fraction of $p$-values required to be large or small in the relevant step-up or step-down procedure.

In Chapter 2, we focused on the limiting behavior of the new methods in relation to the well-known Benjamini and Hochberg step-up method. These asymptotic results are of particular interest in the large-scale scenarios because the number of tests is so high. Following the approach of Genovese and
Wasserman in [7], we showed that the limiting threshold of the proposed procedure is at least as large as that of Benjamini and Hochberg but also observed that FDR control was compromised without some sort of adjustment. We have recommended such adjustments separately and depending on whether the non-null distribution is moderate or extreme. The simulation results in Chapter 3 support the recommendations in the cases for which they are intended.

As a nature result of such an investigation, we have raised new questions to consider for future work. One of which is to re-define $\gamma$ as a function of $m$ that would allow the limiting threshold of the proposed procedure to converge to that of Benjamini and Hochberg, which has been established as optimal in certain ways [7]. For appropriately chosen $\gamma_m > 0$, increased power may be gained for finite $m$ without sacrificing FDR control.

Another possible avenue for investigation along the lines introduced here is to allow a more general function of the $p$-values to define the stopping rule. In the proposed procedures here, we have required that the fraction, or arithmetic mean, of $p$-values be sufficiently large or small. Perhaps an alternative function could incorporate a specific dependence structure for certain cases not covered here. The authors in [3] discuss examples where the positive regression dependence property does not hold.
A.1 Proof of the theorem

Recall that the $\gamma$SUP rejects the smallest $i^*$ $p$-values, where

$$i^* = \max \left\{ i : \frac{1}{i} \sum_{j=1}^{i} I \left( G_m^{-1} \left( \frac{j}{m} \right) \leq \frac{j}{m} \alpha \right) \geq 1 - \gamma \right\}.$$ 

To work with this explicitly as a sequence in $m$, we express this as $i^* = [v_m m]$, where

$$v_m = \max \left\{ v : \frac{1}{m} \sum_{j=1}^{[vm]} I \left( G_m^{-1} \left( \frac{j}{m} \right) \leq \frac{j}{m} \alpha \right) \geq 1 - \gamma \right\}.$$ 

We must show that $v_m \xrightarrow{P} v^*$ as $m \to \infty$, with

$$v^* = \sup \left\{ v : \frac{1}{v} \int_0^v I \left( G^{-1}(u) \leq u\alpha \right) \ du \geq 1 - \gamma \right\}.$$ 

We first show that $\frac{1}{m} \sum_{j=1}^{[vm]} I \left( G_m^{-1} \left( \frac{j}{m} \right) \leq \frac{j}{m} \alpha \right) \xrightarrow{P} \int_0^v I \left( G^{-1}(u) \leq u\alpha \right) \ du$ uniformly in $v \in (\delta, 1)$ for any small $\delta > 0$. In what follows, all limits consider $m \to \infty$.

Lemma 1.

$$\int_0^v I \left( G_m^{-1}(u) \leq u\alpha \right) \ du - \int_0^v I \left( G^{-1}(v) \leq u\alpha \right) \ du \xrightarrow{P} 0$$
uniformly in \((0,1)\).

**Proof.**

\[
\left| \int_0^v \left[ I \left( G_m^{-1}(u) \leq u\alpha \right) - I \left( G^{-1}(u) \leq u\alpha \right) \right] \, du \right| \\
\leq \int_0^1 I \left( G_m^{-1}(u) \leq u\alpha < G^{-1}(u) \right) \, du + \int_0^1 I \left( G^{-1}(u) \leq u\alpha < G_m^{-1}(u) \right) \, du.
\]

Fix \(u\). If \(u\alpha < G^{-1}(u)\), there is some \(\delta > 0\) such that \(u\alpha < G^{-1}(u) - \delta\). So,

\[
\mathcal{P} \left( G_m^{-1}(u) \leq u\alpha < G^{-1}(u) \right) \leq \mathcal{P} \left( G_m^{-1}(u) < G^{-1}(u) - \delta \right) \\
\leq \mathcal{P} \left( G^{-1}(u) - G_m^{-1}(u) > \delta \right) \to 0.
\]

Similarly, if \(u\alpha > G^{-1}(u)\), there is \(\delta > 0\) such that \(u\alpha > G^{-1}(u) + \delta\). So,

\[
\mathcal{P} \left( G^{-1}(u) < u\alpha < G_m^{-1}(u) \right) \leq \mathcal{P} \left( G^{-1}(u) + \delta < G_m^{-1}(u) \right) \\
\leq \mathcal{P} \left( G^{-1}(u) - G_m^{-1}(u) > \delta \right) \to 0.
\]

The proof is completed since there is at most one \(0 < u\) such that \(G^{-1}(u) = u\alpha\).

**Lemma 2.**

\[
\int_0^v I \left( G_m^{-1}(u) \leq (u + 1/m)\alpha \right) \, du - \int_0^v I \left( G_m^{-1}(v) \leq u\alpha \right) \, du \\ \xrightarrow{P} 0
\]

uniformly in \((0,1)\).

**Proof.** We can write this difference as \(d_m(v) = \int_0^v I(0 < G_m^{-1}(u) - u\alpha < \frac{\alpha}{m}) \, du\). Also, note that

\[
\sup_{0 < v < 1} d_m(v) = \int_0^1 I \left( 0 < G_m^{-1}(u) - u\alpha < \frac{\alpha}{m} \right) \, du.
\]
For small $\delta > 0$,
\[
P \left( \sup_{0 < v < 1} d_m(v) > \delta \right) \leq \frac{1}{\delta} \int_0^1 P \left( 0 < G_m^{-1}(u) - u\alpha < \frac{\alpha}{m} \right) du.
\]
By the Dominated Convergence Theorem, it is enough to show that
\[
P \left( 0 < G_m^{-1}(u) - u\alpha < \frac{\alpha}{m} \right) \to 0 \quad \text{(A.1)}
\]
for each $u \in (0, 1)$. We know that $G_m^{-1}(u) - u\alpha \to G^{-1}(u) - u\alpha$ for each such $u$. If $G^{-1}(u) - u\alpha > 0$, then $G^{-1}(u) - u\alpha \geq \delta > 0$. For all large $m$, we must have
\[
G_m^{-1}(u) - u\alpha \geq \frac{\delta}{2}. \quad \text{(A.2)}
\]
Similarly, if $G^{-1}(u) - u\alpha < 0$, then $G^{-1}(u) - u\alpha < -\delta$. For all large $m$, we must have $G_m^{-1}(u) - u\alpha < -\frac{\delta}{2}$. This, along with (A.2), gives (A.1).

Lemma 3.
\[
\frac{1}{m} \sum_{j=1}^{[m]} I \left( G_m^{-1} \left( \frac{j}{m} \right) \leq \frac{j}{m} \alpha \right) \to \int_0^v I(G^{-1}(u) \leq u\alpha) du
\]
uniformly in $(0, 1)$.

Proof. Denote the left and right sides above by $h'_m(v)$ and $h(v)$, respectively. By inspection at each interval $\frac{j-1}{m} < u \leq \frac{j}{m}$ for $j = 1, \ldots, m$, we have $G_m^{-1}(u) = G_m^{-1}(\frac{j}{m})$ and $u \leq \frac{j}{m} < u + \frac{1}{m}$. Therefore,
\[
\int_0^v I \left( G_m^{-1}(u) \leq u\alpha \right) du \leq h'_m(v) \leq \int_0^v I \left( G_m^{-1}(u) \leq (u + 1/m) \alpha \right) du.
\]
The result follows by Lemmas 1 and 2.

For the remainder of the proof, fix $\gamma > 0$. Since some unique $u_0 > 0$
satisfies $G^{-1}(u_0) = u_0 \alpha$ and $G^{-1}(u) < u \alpha$ for $0 < u < u_0$, we have
\[
v^* = \sup \left\{ v : \frac{1}{v} \int_0^v I \left( G^{-1}(u) \leq u \alpha \right) \, du \geq 1 - \gamma \right\}
= \sup \left\{ v : \frac{u_0}{v} \geq 1 - \gamma \right\}
= \min \left\{ 1, \frac{u_0}{1 - \gamma} \right\}.
\]

For $v^* < 1$, $h(v)/v$ is decreasing in $v$ for $v \geq u_0$. Fix small $\delta > 0$ such that $0 < v^* - \delta < v^* + \delta < 1$. By Lemma 3, $h'_m(v) \to h(v)$ uniformly in $(0, 1)$. Now, $h'_m(v)/v$ and
\[
\frac{h_m(v)}{v} = \frac{1}{[vm]} \sum_{j=1}^{[vm]} I \left( G^{-1}_m \left( \frac{j}{m} \right) \leq \frac{j}{m} \alpha \right)
\]
both converge to $h(v)/v$ uniformly in $[v^* + \delta, 1]$.

\[
\sup_{v \geq v^* + \delta} \frac{h_m(v)}{v} = \sup_{v \geq v^* + \delta} \left[ \frac{h(v)}{v} + \frac{h_m(v) - h(v)}{v} \right]
\leq \sup_{v \geq v^* + \delta} \frac{h(v)}{v} + \frac{1}{v^* + \delta} \sup_{v \geq v^* + \delta} \left[ h_m(v) - h(v) \right]
\]
and

\[
|h_m(v) - h(v)| \leq |h'_m(v) - h(v)| + \frac{1}{m} \frac{v}{[vm]} - |m|
\leq |h'_m(v) - h(v)| + \frac{vm - [vm]}{[vm]} \to 0
\]
uniformly in $v \geq v^* + \delta$. So,
\[
\limsup_{m \to \infty} \left[ \sup_{v \geq v^* + \delta} \frac{h_m(v)}{v} \right] \leq \sup_{v \geq v^* + \delta} \frac{h(v)}{v} \leq \frac{h(v^* + \delta)}{v^* + \delta} \leq 1 - \gamma - \epsilon
\]
for some $\epsilon > 0$. So,
\[
\sup_{v \geq v^* + \delta} \frac{h_m(v)}{v} < 1 - \gamma
\]
for all large \( m \). Then, \( v_m < v^* + \delta \) for all large \( m \). So, \( \limsup v_m \leq v^* \).

Now, \( h(v)/v \) is continuous at \( v^* \), and \( h(v)/v \) is strictly decreasing in a neighborhood of \( v^* \). For small fixed \( \delta > 0 \), there exists \( v_1 \in (v^* - \delta, v^*) \) such that \( h(v_1) > 1 - \gamma - \epsilon \) for some \( \epsilon > 0 \). Thus, \( h_m(v_1)/v_1 > 1 - \gamma \) for all large \( m \). So, \( v_m > v_1 \) for such \( m \). Hence \( \limsup v_m \geq v^* - \delta \), and finally, \( \lim v_m = v^* \).

As a final note, the assumption that \( v^* \) is strictly less than one can be justified as follows. \( v^* \) will be large when \( G \) is extreme. An upper bound for \( G(u) \) is \( A_0 u + (1 - A_0) \), which leads to

\[
v^* < \frac{1 - A_0}{(1 - A_0 \alpha)(1 - \gamma)}.
\]

Thus, \( v^* < 1 \) when \( A_0 > \frac{\gamma}{1 - (1 - \gamma) \alpha} \). This will be easily satisfied in practice. For example, with \( \gamma = \alpha = .1 \), we need \( A_0 > .1099 \). Smaller \( \alpha \) will impose a less restrictive bound on \( A_0 \). Typically, \( A_0 \) is close to one.

### A.2 Genovese and Wasserman FDP Control

Toward the end of Section 1.3.2, we argue that the procedure of Genovese and Wasserman in [8] controls the δFDP by modifying a step-down procedure with critical values \( \alpha_i = B^{-1}_{1,m+1-i}(\alpha) \), where \( B_{a,b} \) is the beta distribution with mean \( \frac{a}{a+b} \). If \( R \) is the number of rejections resulting from the SDP, then modifying this number to \( R' = \frac{R}{1-\delta} \) controls the δFDP. However, the authors describe their procedure with a different, and it is not immediately obvious why it reduces to the result we claim in Section 1.3.2. We reconcile these two approaches here.

Let \( S = \{1, 2, \ldots, m\} \) be the set of indices and \( S_0 \subset S \) be the set of indices corresponding to true nulls. Also, let \( \pi(i) \) be the index of the \( i \)th ordered \( p \)-value; that is, \( P_{\pi(i)} = P_{(i)} \). Let \( S(R) \subset S \) be a rejection region of the form \( \{\pi(1), \ldots, \pi(R)\} \). For given \( R \), the FDP is expressed as

\[
\text{FDP} = \Gamma(S(R)) = \frac{|S_0 \cap S(R)|}{R}.
\]
For given $\alpha$ and $\delta$, the goal is to find $R$ such that $\mathcal{P}(\Gamma(S(R)) \leq \delta) \geq 1 - \alpha$. Using this notation, the authors begin by finding $U$, a confidence superset for $S_0$ such that $\mathcal{P}(S_0 \in U) \geq 1 - \alpha$. Then, by the definition of $\Gamma(S(R))$,

$$
\mathcal{P} \left( \Gamma(S(R)) \leq \max_{B \in U} \frac{|B \cap S(R)|}{R} \forall S(R) \right) \\
\geq \mathcal{P} \left( \Gamma(S(R)) \leq \max_{B \in U} \frac{|B \cap S(R)|}{R} \forall S(R) \mid S_0 \in U \right) \mathcal{P}(S_0 \in U) \\
= \mathcal{P}(S_0 \in U).
$$

They then define a family of upper bounds $\Gamma(S(R))$ by

$$
\Gamma(S(R)) = \max_{B \in U} \frac{|B \cap S(R)|}{R}.
$$

(A.3)

Finally, they choose the largest $R$ such that $\Gamma(S(R)) \leq \delta$, which gives the desired result:

$$
\mathcal{P}(\Gamma(S(R)) \leq \delta) \geq \mathcal{P}(\Gamma(S(R)) \leq \Gamma(S(R))) \geq 1 - \alpha.
$$

The brute-force way to find $U$ is to test $H_0 : W \subset S_0$ versus $H_a : W \subset S$ at level-$\alpha$ for every (nonempty) $W \subset S$. In fact, the authors mention such a test involving comparing the $k$th ordered p-value with $B_{k,|W|-k+1}(\alpha)$, which is a reasonable approach since $P_{(k)}$ has such a distribution under $H_0$. However, with so many subsets $W \subset S$ to consider, this approach is computationally unacceptable in most multiple testing settings. Fortunately, as the authors prove, $\Gamma(S(R))$ can be computed directly:

$$
\Gamma(S(R)) = \max_{B \in U} \frac{|B \cap S(R)|}{R} = \frac{|\{\pi(k), \ldots, \pi(J_k - 1)\}^c \cap S(R)|}{R},
$$

where $J_k = \min\{j : P_{(j)} \geq B_{k,m-j+1}(\alpha)\}$.

For fixed $\delta$, the authors’ approach is then to define $R$ as the largest number satisfying $\Gamma(S(R)) \leq \delta$. To see why this reduces to our claim, note that $R$ is
expressed as

$$\max\{r : \Gamma(\pi(1), \ldots, \pi(r)) \leq \delta\}$$

$$= \max \left\{ r : \frac{|\{\pi(k), \ldots, \pi(J_k - 1)\}^c \cap \{\pi(1), \ldots, \pi(r)\}|}{r} \leq \delta \right\}$$

$$= \max \left\{ r : \frac{(k - 1) + |\{\pi(J_k), \ldots, \pi(r)\}|}{r} \leq \delta \right\}$$

$$= \max\{r : r - J_k + k \leq \delta r\}$$

$$= \max\left\{ r : r \leq \frac{J_k - k}{1 - \delta} \right\}$$

$$= \left\lfloor \frac{J_k - k}{1 - \delta} \right\rfloor.$$ 

For $k = 1$, $J_k - k = i_D$ as defined in Section 1.3.2, which satisfies the claim.
Appendix B

Computer code

B.1 R code for simulations

fdr = function(p,m0,gam,alpha)
{
  # returns the number of rejections
  # and number of false rejections
  # and fdp for the specified type of procedure
  # all cutoffs i*alpha/m unless specified
  # type of procedure:

  # Existing procedures for comparison
  # BenHoch = original Simes/BH stepup
  # BenLui = 1999 Benjamini Lui stepdown

  # Proposed procedures, use gamma = alpha
  # gam.su = max for which fraction of 1s is 1-gam
  # gam.sd = max for which fraction of 1s is 1-gam, and all below

  # Proposed procedures (modified), gamma = alpha, alpha’ = (1-gamma)alpha = (1-alpha)alpha
  # gam.su.mod = max for which fraction of 1s is 1-gam
  # gam.sd.mod = max for which fraction of 1s is 1-gam, and all below
# Perfunctory generalization of SUP
# \( \text{gam.min.su} = \text{min} \) for which fraction of 1s is \( \text{gam} \)
# \( \text{gam.min.sd} = \text{min} \) for which fraction of 1s is \( \text{gam} \), and all above

### adaptive ones using same procedure twice to get A0 estimate
# BenHoch.adp, BenLui.adp, etc

# \( p = \text{vector of pvalues}, m0 \text{ true ones last} \)
# \( m0 = \text{number of true pvalues} \)
# \( \text{gam} = \text{fraction parameter} \)
# \( \alpha = \text{fdr bound} \)

\[
m = \text{length}(p)
\]
\[
\text{ind} = c( \text{rep}(0,m-m0), \text{rep}(1,m0) ) \ # m0 1's last, corresponding to p
\]
\[
\text{ind} = \text{ind}[\text{order}(p)] \ # \text{ordered relative to p}
\]
\[
p = \text{sort}(p) \ # \text{so that (ind,p) consistent}
\]

################################
# existing for reference

\[
x = (1:m) \times \alpha / m \ # \text{critical values}
\]
\[
x = (p <= x) \ # \text{indicate whether } p <= \text{critical value}
\]

# BenHoch = original Simes/BH stepup
\[
R = \text{max}( 1:m ) \times x \ # \text{number rejections}
\]
\[
V = \text{sum}( c( \text{rep}(1,R), \text{rep}(0,m-R) ) \times \text{ind} ) \ # \text{number null rejections}
\]
\[
\text{BenHoch} = c(V, R, V / \text{max}(1,R), ((V / \text{max}(1,R)) > \alpha))
\]

# BenHoch adaptive
\[
q = \alpha / (1 + \alpha)
\]
\[
x = (1:m) \times q / m \ # \text{critical values}
\]
\[
x = (p <= x) \ # \text{indicate whether } p <= \text{critical value}
\]
\[
R = \text{max}( (1:m) \times x ) \ # \text{number rejections}
\]
\[
V = \text{sum}( c( \text{rep}(1,R), \text{rep}(0,m-R) ) \times \text{ind} ) \ # \text{number null rejections}
\]
\[
\text{if}( R > 0 \ && \ R < m ) \{
\]
\[
xx = (1:m) \times \alpha / (m-R)
\]
\( xx = (p <= xx) \)
\( R = \max( (1:m) * xx ) \)
\( V = \sum( c(\text{rep}(1,R), \text{rep}(0,m-R)), \text{ind} ) \)
BenHoch.adp = c(V, R, V/\max(1,R), ((V/\max(1,R)) > alpha))

# BenLui = original Simes/BH stepdown
\( bl = 1 - (1 - \min(1, m*alpha/(m-(1:m)+1)))^{(1/(m-(1:m)+1))} \)
\( R = m - \max( (m:1) * (1-b1) ) \) # number rejections
\( V = \sum( c(\text{rep}(1,R), \text{rep}(0,m-R)), \text{ind} ) \) # number null rejections
BenLui = c(V, R, V/\max(1,R), ((V/\max(1,R)) > alpha))

# proposed that start from bottom
\( y = \text{cumsum}(x) \)
\( y = (y >= (1:m) * (1-gam)) \)

# gam.su = max for which fraction of 1s is 1-gam
\( R = \max( (1:m) * y ) \)
\( V = \sum( c(\text{rep}(1,R), \text{rep}(0,m-R)), \text{ind} ) \)
gam.su = c(V, R, V/\max(1,R), ((V/\max(1,R)) > alpha))

\( V1 = V \)
\( R1 = R \)
if(R>0 && R<m) {
  gamm = R/m
  \( yy = \text{cumsum}(x) \)
  \( yy = (yy >= (1:m) * (1-gamm)) \)
  \( R1 = \max( (1:m) * yy ) \)
  \( V1 = \sum( c(\text{rep}(1,R1), \text{rep}(0,m-R1)), \text{ind} ) \)
gam.su.adp = c(V1, R1, V1/\max(1,R1), ((V1/\max(1,R1)) > alpha))
}

if(R>0 && R<m) {
  a1 = R/m
  #gamm = a1^2/(1-a1)/(1/alpha-1+a1)
gamm = a1/(1/alpha-1+a1)
yy = cumsum(x)
yy = (yy >= (1:m)*(1-gam))
R = max( (1:m)*yy )
V = sum(c(rep(1,R),rep(0,m-R))*ind) }
gam.su.adp2 = c(V, R, V/max(1,R), ((V/max(1,R))>alpha))

# gam.su.adp2 = max for which fraction of 1s is 1-gam, and all below
R = m - max( (m:1)*(1-y) )
V = sum(c(rep(1,R),rep(0,m-R))*ind)
gam.su = c(V, R, V/max(1,R), ((V/max(1,R))>alpha))

################################
# proposed that start from top

y = cumsum(x[m:1])
y = (y <= (1:m)*gam)

# gam.min.su = min for which fraction of 0s is gam
R = m - max( (1:m)*y )
V = sum(c(rep(1,R),rep(0,m-R))*ind)
gam.min.su = c(V, R, V/max(1,R), ((V/max(1,R))>alpha))

# gam.min.sd = min for which fraction of 0s is gam, and all above
R = max( (m:1)*(1-y) )
V = sum(c(rep(1,R),rep(0,m-R))*ind)
gam.min.sd = c(V, R, V/max(1,R), ((V/max(1,R))>alpha))

################################
# modifying with alpha' = (1-gamma)alpha

alpha = (1-gam)*alpha
x = (1:m)*alpha/m
x = (p <= x)
y = cumsum(x)
y = (y >= (1:m)*(1-gam))
# gam.su.mod = max for which fraction of 1s is 1-gam
R = max( (1:m)*y )
V = sum(c(rep(1,R),rep(0,m-R))*ind)
gam.su.mod = c(V, R, V/max(1,R), ((V/max(1,R))>alpha))

# gam.sd.mod = max for which fraction of 1s is 1-gam, and all below
R = m - max( (m:1)*(1-y) )
V = sum(c(rep(1,R),rep(0,m-R))*ind)
gam.sd.mod = c(V, R, V/max(1,R), ((V/max(1,R))>alpha))

rbind(BenHoch,
BenHoch.adp,
BenLui,
gam.su,
gam.su.adp,
gam.su.adp2,
gam.sd,
gam.su.mod,
gam.sd.mod,
gam.min.su,
gam.min.sd,
)

library(mnormt)
source("fdr.R")

a0 = .3
alt.val = 2
rho = 0
n = 500
n0 = as.integer(a0*n)
N = 10000
alpha = .05
gam = alpha

mu = c(rep(alt.val,n-n0),rep(0,n0))
\[ \text{sig} = \text{matrix}(\text{rep}(\rho, n^2), c(n, n)) + \text{diag}(1 - \rho, n, n) \]

\[
\text{fdp.obs} = \text{fdp.obs}2 = 0 \\
\text{for}(j \text{ in } 1:N) \{ \\
\text{p} = \text{as.vector}(2 \times (1 - \text{pnorm}(\text{abs}(\text{rmnorm}(1, \mu, \text{sig})))))) \\
\text{temp} = \text{fdr}(p, n0, \text{gam}, \alpha) \\
\text{fdp.obs} = \text{fdp.obs} + \text{temp} \\
\text{fdp.obs2} = \text{fdp.obs2} + \text{temp}^2 \\
\} \\
\text{fdp.obs2} = \text{as.table}(\text{sqrt}((\text{fdp.obs2} - \text{fdp.obs}^2 / N) / (N - 1))) \\
\text{fdp.obs} = \text{as.table}(\text{fdp.obs} / N) \\
\text{colnames}(\text{fdp.obs}) = \text{c}('E(V)', 'E(R)', 'FDR', 'FDP') \\
\text{colnames}(\text{fdp.obs2}) = \text{c}('s(V)', 's(R)', 's(FDR)', 's(FDP)') \\
\text{print}((\text{round}(\text{cbind}(\text{fdp.obs}, \text{fdp.obs2}), 4)))
\]

### B.2 R code for applications

\[
\text{X} = \text{readFITS}("a.fits") \\
\text{cube.w1} = (\text{X}[1])[0:300, 150:450, 16, 1] \\
\text{cube.s1} = (\text{X}[1])[0:300, 150:450, 19, 1] \\
\text{cube.s5} = (\text{X}[1])[150:200, 220:300, 19, 1] \\
\text{cube.w5} = (\text{X}[1])[150:200, 220:300, 16, 1] \\
\text{ncube.w} = ((\text{X}[1])[c((0:150), (200:300)), c((0:220), (300:450)), 16, 1]) \\
\text{ncube.s} = ((\text{X}[1])[c((0:150), (200:300)), c((0:220), (300:450)), 19, 1]) \\
\text{ms} = \text{mean}(\text{as.vector}(\text{ncube.s})) \\
\text{mw} = \text{mean}(\text{as.vector}(\text{ncube.w})) \\
\text{sds} = \text{sd}(\text{as.vector}(\text{ncube.s})) \\
\text{sdw} = \text{sd}(\text{as.vector}(\text{ncube.w})) \\
\text{x} = \text{ncube.s} \\
\text{hist}(x, \text{freq=F, main='\', xlab='Location noise'}) \\
\text{curve(\text{dnorm}(x, mean=ms, sd=sds), add=T, col='blue'}) \\
\text{x} = \text{ncube.w}
hist(x,freq=F,main='',xlab='Location noise')
curve(dnorm(x,mean=mw,sd=sdw),add=T,col='blue')

pvalue.s1 = 2*(1-pnorm(abs(cube.s1),sd=sds))
hist(as.vector(pvalue.s1),freq=F,main='',xlab='transformed p-values')
image(1-pvalue.s1,xlab='transformed p-values')

pvalue.s5 = 2*(1-pnorm(abs(cube.s5),sd=sds))
hist(as.vector(pvalue.s5),freq=F,main='',xlab='transformed p-values')
image(1-pvalue.s5,xlab='transformed p-values')

pvalue.w1 = 2*(1-pnorm(abs(cube.w1),sd=sdw))
hist(as.vector(pvalue.w1),freq=F,main='',xlab='transformed p-values')
image(1-pvalue.w1,xlab='transformed p-values')

pvalue.w5 = 2*(1-pnorm(abs(cube.w5),sd=sdw))
hist(as.vector(pvalue.w5),freq=F,main='',xlab='transformed p-values')
image(1-pvalue.w5,xlab='transformed p-values')

ptemp = pvalue.s1
ptemp = (cube.aov[,2])[is.finite(cube.aov[,2])]
ptemp = as.vector(ptemp)/5

alpha=.05
r=cutoff(ptemp,alpha,alpha)
cube.temp=cube.aov[,2]
cube.temp[cube.aov[,2]<r[2]]=1
cube.temp[cube.aov[,2]>r[2]]=0
image(cube.temp,xlab='pI',ylab='fraction')

temp=ptemp
temp[ptemp<=r[1]]=1
temp[ptemp>r[1]]=0
image(temp)

cutoff = function(p,gam,alpha)
{  
  # returns p-value threshold for each of the following procedures as vector  
  # 1) BenHoch = original Simes/BH stepup  
  # 2) BenLui = 1999 Benjamini Lui stepdown  
  # 3) gam.su = max for which fraction of 1s is 1-gam  
  # 4) gam.sd = max for which fraction of 1s is 1-gam, and all below  
  # 5) gam.su.mod = max for which fraction of 1s is 1-gam  
  # 6) gam.sd.mod = max for which fraction of 1s is 1-gam, and all below  
  # 7) gam.min.su = min for which fraction of 1s is gam  
  # 8) gam.min.sd = min for which fraction of 1s is gam, and all above  
  r = (fdr(p,1,gam,alpha))[1:10,2]  
  # gets the number of rejections for each procedure  
  p = c(0,sort(p))  
  p[r+1] # returns the threshold for each p-value  
}
Bibliography


nrao.edu. The expanded very large array project: a radio telescope to resolve cosmic evolution, 2008.


Vita
Scott Roths

Home Address:
411 Waupelani Dr Apt A346
State College, PA 16801

e-mail: sar320@gmail.com
phone: (814) 769-6563

Work Address:
333 Thomas Building
Pennsylvania State University
University Park, PA 16802

e-mail: sar320@psu.edu
phone: (814) 863-3374

Education

Ph.D. Statistics, Pennsylvania State University, (expected graduation date August 2011).

M.S. Statistics, Kansas State University, May, 2005.

B.S. Mathematics, Kansas State University, May, 2002.

Research


Employment

Graduate Instructor, Pennsylvania State University, July, 2006–present.

Graduate Research Assistant, Pennsylvania State University, January–August, 2010.