You have to be first, best, or different.
ACKNOWLEDGMENTS

I would like to thank my advisor, Malay Ghosh. It has been my honor to work with such a great mentor that has dedicated his entire life to research, teaching and enriching the lives of students. I have been grateful for the opportunity to learn under him and see my development in both research and personality grow from his endless knowledge not only in all areas of statistics but in everyday life. His enthusiasm for learning and working has shaped who I am as a person who will never stop chasing wonderful dreams. The life with Professor Ghosh has been the best memory for me in the past five years at the University of Florida.

I would like to thank Andrew Rosalsky for giving me the ability of understanding and applying probability theory. His great teaching in probability theory is benefiting me not only in completing my dissertation but also in my future research. I would also like to thank Nikolay Bliznyuk and Salvador Gezan for their insight that has lead to my work improving. Furthermore, I would like to thank all of the faculty, staff, students, and friends that made this dissertation possible.

Finally, I would like to thank my parents and my wife. Without their constant love and support, I would not have made it to where I am today. I would not be the person I am today without my parents’ constant encouragement of science, mathematics and statistics since I was very young.
# TABLE OF CONTENTS

<table>
<thead>
<tr>
<th>Section</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>ACKNOWLEDGMENTS</td>
<td>4</td>
</tr>
<tr>
<td>LIST OF TABLES</td>
<td>6</td>
</tr>
<tr>
<td>ABSTRACT</td>
<td>7</td>
</tr>
<tr>
<td>CHAPTER</td>
<td></td>
</tr>
<tr>
<td>1 INTRODUCTION</td>
<td>8</td>
</tr>
<tr>
<td>1.1 Multiple Testing Problems and Sparsity</td>
<td>8</td>
</tr>
<tr>
<td>1.2 Family Wise Error Rate and False Discovery Rate</td>
<td>9</td>
</tr>
<tr>
<td>1.3 Thresholding Equivalence of the Benjamini-Hochberg Procedure</td>
<td>10</td>
</tr>
<tr>
<td>1.4 Approximate Fixed Thresholding the Benjamini-Hochberg Procedure</td>
<td>11</td>
</tr>
<tr>
<td>1.5 Bayesian False Discovery Rate and Positive False Discovery Rate</td>
<td>12</td>
</tr>
<tr>
<td>1.6 Asymptotic Bayes Optimality under Sparsity</td>
<td>13</td>
</tr>
<tr>
<td>1.7 Multiple Testing for Exponential Distributions</td>
<td>15</td>
</tr>
<tr>
<td>1.8 Proposed Research</td>
<td>16</td>
</tr>
<tr>
<td>2 ASYMPTOTIC BAYES OPTIMALITY UNDER SPARSITY FOR EXPONENTIAL DISTRIBUTION</td>
<td>17</td>
</tr>
<tr>
<td>3 ASYMPTOTIC BAYES OPTIMALITY UNDER SPARSITY OF SOME PROCEDURES</td>
<td>27</td>
</tr>
<tr>
<td>3.1 Connecting the False Discovery Rate, the Bayesian False Discovery Rate and Asymptotic Bayes Optimality under Sparsity</td>
<td>27</td>
</tr>
<tr>
<td>3.2 Asymptotic Bayes Optimality under Sparsity of the Benjamini-Hochberg Procedure</td>
<td>31</td>
</tr>
<tr>
<td>3.3 Simulation Study</td>
<td>37</td>
</tr>
<tr>
<td>4 ALTERNATIVE CHOICE FOR THE DENSITY OF THE SCALE PARAMETER</td>
<td>39</td>
</tr>
<tr>
<td>5 CONCLUSIONS</td>
<td>48</td>
</tr>
<tr>
<td>REFERENCES</td>
<td>50</td>
</tr>
<tr>
<td>BIOGRAPHICAL SKETCH</td>
<td>51</td>
</tr>
</tbody>
</table>
## LIST OF TABLES

<table>
<thead>
<tr>
<th>Table</th>
<th>Description</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>1-1</td>
<td>Number of errors committed when testing $n$ null hypotheses</td>
<td>10</td>
</tr>
<tr>
<td>1-2</td>
<td>Matrix of Losses</td>
<td>13</td>
</tr>
<tr>
<td>2-1</td>
<td>Matrix of Losses</td>
<td>18</td>
</tr>
<tr>
<td>3-1</td>
<td>Sum Loss Comparison between Bayes Oracle and BH Procedure</td>
<td>36</td>
</tr>
<tr>
<td>4-1</td>
<td>Matrix of Losses</td>
<td>40</td>
</tr>
</tbody>
</table>
Multiple testing problems are gaining increasing prominence in statistical research. The prime reason behind this is that there is a growing need for statisticians to analyze large data sets involving many parameters. Several methods to solve multiple testing problems have been introduced since the breakthrough paper by Benjamini and Hochberg in 1995.

In this dissertation, we consider simultaneous testing of multiple exponential scale parameters under sparsity. Two general results provide thresholding conditions under which a given testing procedure achieves Bayesian optimality asymptotically. In particular, it is shown how multiple testing procedures of Benjamini and Hochberg's, Efron and Tibshirani's and Genovese and Wasserman's can achieve this Bayesian optimality asymptotically. An alternative density for the scale parameter has also been considered.
CHAPTER 1
INTRODUCTION

1.1 Multiple Testing Problems and Sparsity

Multiple testing problems are gaining increasing prominence in statistical research. The prime reason behind this is that there is a growing need for statisticians to analyze large data sets involving many parameters. To cite an example, one may refer to microarray data analysis where it is essential to test simultaneously expression levels of thousands of genes. Particularly in the field of genetic association studies, one may test simultaneously expression levels of thousands of genes to see the relationship between genes and, e.g., a disease.

Classical multiple comparison procedures focus on simultaneously considering moderate number of hypothesis tests, often in an analysis of variance. However, large-scale multiple testing may involve thousands or even greater numbers of tests, on which the traditional methods cannot work well. Thus, it is necessary to develop other methods to effectively solve problems of this type. In recent years, multiple testing procedures under sparsity have also become an important topic of research. In this case, sparsity means that the proportion of true alternatives among all hypothesis is very small.

Two of the most popular methods for solving problems of this type that are currently in vogue are the Bonferroni approach which controls the family wise error rate (FWER) and the Benjamini-Hochberg (1995) procedure which controls the false discovery rate (FDR). Some other optimality criteria have also been proposed in the context of multiple testing. For instance, in the classical Neyman-Pearson spirit, one may maximize the expected number of true discoveries, while keeping fixed one of the error measures such as FWER, FDR or expected number of false positives. Storey (2007) adopted such an approach. He referred to this as an oracle property. A second oracle property as defined in Sun and Cai (2007) assumed that the data are generated according to
a two-component mixture model. They maximized the marginal false nondiscovery rate, while controlling the marginal false discovery rate at a fixed level. The optimality was achieved asymptotically for any fixed (though unknown) proportion of alternatives.

In contrast to [10] Sun and Cai (2007), [2] Bogdan, Chakrabarti, Frommlet and Ghosh (henceforth, referred to as BCFG) (2011) took a decision theoretic point of view, where for each test, they assigned fixed losses for type I and type II errors. They also began with a two-component mixture model, and considered a fully asymptotic setup under which the proportion \( p \) of "true" alternatives among all tests converges to zero as the number of tests goes to infinity. With an additive loss structure, they considered an asymptotically optimal Bayesian procedure under sparsity. In the process, they could also demonstrate the asymptotic optimality of the BH procedure under sparsity from a purely Bayesian perspective. [6] Neuvial and Roquain (2012) extended BCFG's work to the exponential power family of distributions, including the double exponential distribution as a special case. This class of distributions arises as certain scale mixtures of the normal distribution.

1.2 Family Wise Error Rate and False Discovery Rate

FWER is quite adequate for detecting significant alternatives in the multiple testing context with small or moderate number of hypotheses. However, in high dimensional problems, for example in analyzing microarray data, FWER may fail to detect most of the significant alternatives, thus FWER cannot be used for exploratory study. On the other hand, for an exploratory study, or if significant results can easily be re-tested in an independent study, control of the FDR is often preferred. The FDR allows researchers to identify a set of "candidate positives," of which a high proportion are likely to be true. The false positives within the candidate set can then be identified in a follow-up study.

In a breakthrough paper, [1] Benjamini and Hochberg (1995) considered the problem of testing simultaneously \( n \) hypotheses, of which \( n_0 \) are true. Table 1-1 summarizes the situation. The specific \( n \) hypotheses are assumed to be known.
is the number of hypotheses rejected which is an observable random variable, while $U$, $V$, $S$ and $T$ are unobservable random variables.

Table 1-1. Number of errors committed when testing $n$ null hypotheses

<table>
<thead>
<tr>
<th></th>
<th>Declared non-significant</th>
<th>Declared significant</th>
</tr>
</thead>
<tbody>
<tr>
<td>True nulls</td>
<td>$U$</td>
<td>$V$</td>
</tr>
<tr>
<td>Non-true nulls</td>
<td>$T$</td>
<td>$S$</td>
</tr>
<tr>
<td></td>
<td>$n - R$</td>
<td>$R$</td>
</tr>
</tbody>
</table>

The False Discovery Rate (FDR) is defined as

$$FDR = E \left( \frac{V}{R} \right)$$

where $V$ is the number of "false" rejections and it is assumed that $\frac{V}{R} = 0$ when $R = 0$.

Two simple properties of FDR were shown. The first one is that if all null hypotheses are true, the FDR is equivalent to the FWER. The second one is that when $n_0 < n$, the FDR is smaller than or equal to the FWER. As a result, any procedure that controls the FWER also controls the FDR.

Benjamini and Hochberg created a procedure controlling false discovery rate. To control FDR at level $\gamma$, they consider sorting p-values in an ascending order $p_1 \leq p_2 \leq \ldots \leq p_n$ and rejecting the null hypotheses for which the corresponding p-values are smaller than or equal to $p(i)$ where

$$t = \arg\max_i \left\{ p(i) \leq \frac{i \gamma}{n} \right\}.$$ 

Benjamini-Hochberg procedure is shown to control the FDR at level $\gamma$. Benjamini and Hochberg also showed some properties of this procedure in the paper. Their paper gave an impetus to solve multiple testing problems, and it encouraged people to explore further the multiple testing world.

### 1.3 Thresholding Equivalence of the Benjamini-Hochberg Procedure

Consider the multiple testing problem of simultaneously testing $H_{0i}$ versus $H_{A\text{ni}}$ for each $i = 1, \ldots, n$. For each $i$, a random variable $X_{ni}$ with identical distribution is
observable. Many other multiple testing criteria use observable random variables $X_n$’s instead of using corresponding p-value’s.


$$
\hat{C}_n^{BH} = \inf \left\{ y : \frac{P_{H_0}(X_{n1} > y)}{1 - \hat{F}_n(y)} \leq \gamma \right\},
$$

where $\gamma$ is the FDR level and $\hat{F}_n(y)$ is the empirical cdf which is defined as $1 - \hat{F}_n(y) = \frac{\# \{X_n > y \}}{n}$. The equivalence theorem says that the BH procedure rejects the null hypothesis $H_{0ni}$ when $X_{ni} \geq \hat{C}_n^{BH}$.

Efron and Tibshirani’s equivalence theorem gives a tool for people to compare other thresholding multiple testing criterions with the BH procedure. We also need utilize this equivalence theorem to prove some results in Chapter 2.

### 1.4 Approximate Fixed Thresholding the Benjamini-Hochberg Procedure

For the same statistical model as given in Section 1.3, [5] Genovese and Wasserman (2002) investigated the asymptotic behavior of the random threshold for Benjamini-Hochberg procedure proposed in the equivalence theorem. They proved that while the number of tests tends to infinity and the fraction of true alternatives remains fixed, the random threshold of a Benjamini-Hochberg procedure controlling FDR at the level $\gamma$ can be approximated by a fixed threshold $C_n^{GW}$ with type I error $t_{1n}^{GW}$ and type II error $t_{2n}^{GW}$ which satisfies

$$
\gamma = \frac{t_{1n}^{GW}}{1 - p_n t_{1n}^{GW} + p_n (1 - t_{2n}^{GW})}.
$$

In Section 1.3, we have seen the relationship between the Benjamini-Hochberg procedure and the random thresholding procedure using $X_n$’s. Genovese and Wasserman’s result constructs another bridge which connects the Benjamini-Hochberg procedure, the random thresholding procedure and the fixed thresholding procedure.
1.5 Bayesian False Discovery Rate and Positive False Discovery Rate

As the False Discovery Rate plays an important role in multiple testing problem research, many other measures have also been proposed by people. One famous example is the Bayesian False Discovery Rate.

In addition to the statistical model in Section 1.3, assume that $H_{0ni}$ and $H_{A_ni}$ occur with probabilities $p_n$ and $1 - p_n$ respectively. Consider a fixed thresholding procedure (i.e. the same threshold for each individual test). Denote by $t_{1n}^*$ and $t_{2n}^*$ respectively the probabilities of type I error and type II error.

[4] Efron and Tibshirani (2002) defined the Bayesian False Discovery Rate (BFDR) as

$$BFDR = P(H_{0ni} \text{ is true} | H_{0ni} \text{ is rejected}) = \frac{(1 - p_n)t_{1n}^*}{(1 - p_n)t_{1n}^* + p_n(1 - t_{2n}^*)}.$$ 

To understand the relationship between the False Discovery Rate and the Bayesian False Discovery Rate, we may first have a look at another measure named the Positive False Discovery Rate which is defined by [8] Storey (2003) as

$$pFDR = E \left( \frac{V}{R} \left| R > 0 \right. \right).$$

For a two-group model, Storey proved a theorem which shows the equivalence between the Bayesian False Discovery Rate and the Positive False Discovery Rate, i.e.

$$BFDR = pFDR.$$ 

In addition, a corresponding corollary says that they are both equal to $\frac{E(V)}{E(R)}$ by noting that

$$E(V) = nP(H_{0ni} \text{ is true}) P(H_{0ni} \text{ is rejected} | H_{0ni} \text{ is true})$$

and

$$E(R) = nP(H_{0ni} \text{ is rejected}).$$
Table 1-2. Matrix of Losses

<table>
<thead>
<tr>
<th></th>
<th>choose $H_{0ni}$</th>
<th>choose $H_{Ani}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$H_{0ni}$ true</td>
<td>0</td>
<td>$\delta_{0n}$</td>
</tr>
<tr>
<td>$H_{Ani}$ true</td>
<td>$\delta_{An}$</td>
<td>0</td>
</tr>
</tbody>
</table>

Now it is easily seen that, comparing to $FDR = E \left( \frac{V}{R} \right)$, the BFDR (or pFDR) is a ratio of expectations $\frac{E(V)}{E(R)}$. More details about pFDR can be found in Storey’s paper. In Chapter 2, we discuss the relationship between BFDR and our proposed criterion by showing the equivalence under certain conditions.

1.6 Asymptotic Bayes Optimality under Sparsity

[2] Bogdan, Chakrabarti, Frommlet and Ghosh (2011) introduce another important measure named Asymptotic Bayes Optimality under Sparsity (ABOS). Suppose random variables $X_{n1}, \ldots, X_{nn}$ are independent identical distributed. For each $i = 1, \ldots, n$, $X_{ni}$ marginally has a mixture normal distribution

$$X_{ni} \sim (1 - p_n) N(0, \sigma_{n}^2) + p_n N(0, \sigma_{n}^2 + \tau_{n}^2)$$

where $p_n \in (0, 1)$, $\sigma_{n}^2 > 0$ and $\tau_{n}^2 > 0$. The word "sparsity" refers to the situation when $p_n \approx 0$. Consider the multiple testing problem of simultaneously testing $H_{0ni}$ versus $H_{Ani}$ for $i = 1, \ldots, n$. For each $i$, two states $H_{0ni}$ and $H_{Ani}$ happen with probabilities $(1 - p_n)$ and $p_n$ respectively. Under $H_{0ni}$, $X_{ni} \sim N(0, \sigma_{n}^2)$, and under $H_{Ani}$, $X_{ni} \sim N(0, \sigma_{n}^2 + \tau_{n}^2)$. The loss structure for making a decision in the $i^{th}$ test is indicated in Table 1-2.

Under an additive loss function, the Bayes decision problem leads to a procedure choosing the alternative hypothesis $H_{Ani}$ in the case such that

$$\frac{f_{An}(X_{ni})}{f_{0n}(X_{ni})} \geq \frac{(1 - p_n)\delta_{0n}}{p_n \delta_{An}} =: \frac{1 - p_n}{p_n} \frac{\delta_{An}}{\delta_{0n}}$$

where $f_{An}$ and $f_{0n}$ are the densities of $X_{ni}$ under alternative and null respectively.

Substituting the corresponding normal densities, one obtains a formal procedure:

reject $H_{0ni}$ if $\frac{X_{ni}^2}{\sigma_{n}^2} \geq c_{n}^2, i = 1, \ldots, n.$
where

\[ c^2_n = \frac{\sigma_n^2 + \tau_n^2}{\tau_n^2} \left[ \log \left( \frac{\sigma_n^2 + \tau_n^2}{\sigma_n^2} \right) + 2 \log \left( \frac{1 - \rho_n \delta_n}{\rho_n} \right) \right]. \]

This procedure is defined as *Bayes oracle* due to the use of unknown parameters.

BCFG introduce the following assumption under which both the type I and type II errors are nontrivial. They prove two basic results under this assumption and define the Asymptotically Bayes Optimal under Sparsity based on these results.

**Assumption** (Π) A sequence of vectors \( \{(\rho_n, \sigma_n^2, \tau_n^2, \delta_{0n}, \delta_{An})\}_{n=1}^{\infty} \) satisfies this assumption if the corresponding sequence of parameter vectors \((\rho_n, a_n, b_n)\) fulfills the following conditions: \( \rho_n \to 0, a_n \to \infty, b_n \to \infty \) and \( \frac{\log b_n}{a_n} \to C \in (0, \infty) \), as \( n \to \infty \), where \( a_n = \frac{\tau_n^2}{\sigma_n^2} \) and \( b_n = a_n \left(1 - \frac{\rho_n \delta_n}{\rho_n \delta_n}\right)^2 \).

**Result 1.1** Under Assumption (Π), the probabilities of type I error and type II error using the Bayes oracle are respectively

\[ t_{1n} = \exp \left( -\frac{C}{2} \right) \sqrt{\frac{2}{\pi a_n \log b_n}} (1 + o_n(1)) \]

and

\[ t_{2n} = (2 \Phi(\sqrt{C}) - 1)(1 + o_n(1)). \]

**Result 1.2** Under an additive loss function, the Bayes risk for a fixed threshold multiple testing procedure is given by

\[ R = \sum_{i=1}^{n} [(1 - \rho_n) t_{1n} \delta_{0n} + \rho_n t_{2n} \delta_{An}] = n [(1 - \rho_n) t_{1n} \delta_{0n} + \rho_n t_{2n} \delta_{An}]. \]

Therefore, under Assumption (Π), the Bayes risk of using Bayes oracle is

\[ R_{\text{optimal}} = n \rho_n \delta_{An} (2 \Phi(\sqrt{C}) - 1)(1 + o_n(1)). \]

**Definition 1.1** For a sequence of parameters satisfying Assumption (Π), a multiple testing rule is called *Asymptotically Bayes Optimal under Sparsity (ABOS)* if its risk \( R \) satisfies
where $R_{\text{optimal}}$ is the optimal risk given in Result 1.2.

BCFG investigated the relationship between the FDR, the BFDR and the ABOS, and they have also proved that the Benjamini-Hochberg procedure is ABOS with certain conditions.

[6] Neuvial and Roquain (2012) extended BCFG’s work to the exponential power family of distributions, including the double exponential distribution as a special case. This class of distributions arises as certain scale mixtures of the normal distribution.

1.7 Multiple Testing for Exponential Distributions


One possible application is related to multiple lifetime analysis. They consider the failure times of many comparable independent systems, where a small proportion of the systems may have significantly higher expected lifetimes than the typical systems. The failure times have exponential distributions, and we need an approach to detect the significant ones.

There is another application for multiple testing with Lehmann alternatives. Suppose that we conduct many independent statistical tests, each providing a $p$-value, say $p_i$, and that most of these tests are the cases with true nulls while a small proportion correspond to cases where a Lehmann alternative is true. Then $\log(1/p_i) \sim \text{Exp}(\mu_i)$, where most of the $\mu_i$ are 1, corresponding to true null hypotheses, while the rest are greater than 1, corresponding to Lehmann alternatives.

They proved asymptotic minimaxity of false discovery rate thresholding for exponential data under sparsity. Their work motivates us to consider the multiple testing problems in the context of exponential distributions.
1.8 Proposed Research

The genesis of my dissertation is based on the two papers by [2] Bogdan et al. (2011) and [3] Donoho and Jin (2006). While the former considered a two component mixture normal model, we consider instead mixture inverse-gamma distributions for exponential scale parameters. Unlike [3] Donoho and Jin (2006), our aim is not to prove asymptotic minimaxity, but establish the asymptotic Bayes optimality under sparsity.

CHAPTER 2
ASYMPTOTIC BAYES OPTIMALITY UNDER SPARSITY FOR EXPONENTIAL DISTRIBUTION

From this chapter on, the notations in the previous chapter are voided, and some of the notations are reassigned to other representations.

Suppose, for each fixed \( x \in \mathbb{R} \), \( x \neq 0 \), \( x \in \mathbb{R} \), \( x \in \mathbb{R} \), \( x \in \mathbb{R} \) are independent exponentials with respective pdfs

\[
f_{\sigma_i}(x) = \sigma_i^{-1} \exp(-x/\sigma_i), \quad x > 0, \quad \sigma_i > 0; \quad i = 1, \ldots, n.
\]

We will denote these pdfs by \( \text{Exp}(\sigma_i), i = 1, \ldots, n \).

Our objective is to test simultaneously \( H_{0i} \) versus \( H_{1i} \), where under \( H_{0i} \), \( \sigma_i \) has Inverse-Gamma pdf

\[
\pi_{\sigma_i}(\sigma) = [\exp(-\beta_j/\sigma_i)\beta_j^{\alpha_j}/\sigma_i^{\alpha_j+1}\Gamma(\alpha)], \quad j = 0, 1, \quad \beta_{1n} > 0, \quad \alpha > 1.
\]

Then, marginally under \( H_{1i} \), \( X_{ni} \) has pdf

\[
f_{\sigma_i}(x) = \alpha^{\alpha_j}(x + \beta_{1j})^{\alpha_j+1}, \quad j = 0, 1.
\]

where \( \beta_{1n} > 0, \) both unknown and \( \alpha > 1 \) known.

We assume that \( H_{0i} \) and \( H_{1i} \) occur with probabilities \( p_n \) and \( 1 - p_n \) respectively. Thus marginally, \( X_{ni} \) has the two-group mixture Pareto pdf

\[
(1 - p_n)\alpha^{\alpha_j}(x + \beta_{0n})^{\alpha_j+1} + p_n\alpha^{\alpha_j}(x + \beta_{1n})^{\alpha_j+1}.
\]

The word sparsity refers to the situation \( p_n \approx 0 \).

Following BCFG, consider the additive loss \( \sum_{i=1}^{n} L(X_{ni}) \), where \( L(X_{ni}) = \delta_{0n}\phi(X_{ni}) + \delta_{1n}(1 - \phi(X_{ni})) \). Here \( \phi(X_{ni}) \) equals 1 or 0 according to the rejection or acceptance of \( H_0 \).

Table 2-1 describes this loss structure.
Table 2-1. Matrix of Losses

<table>
<thead>
<tr>
<th></th>
<th>choose $H_{0ni}$</th>
<th>choose $H_{Ani}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$H_{0ni}$ true</td>
<td>0</td>
<td>$\delta_{0n}$</td>
</tr>
<tr>
<td>$H_{Ani}$ true</td>
<td>$\delta_{1n}$</td>
<td>0</td>
</tr>
</tbody>
</table>

Under this loss structure, the Bayes decision problem leads to a procedure choosing the alternative hypothesis $H_{Ani}$ in the case

$$\frac{f_{1n}(X_{ni})}{f_{0n}(X_{ni})} \geq \frac{(1 - p_n)\delta_{0n}}{p_n\delta_{1n}}$$

where $f_{1n}$ and $f_{0n}$ are the densities of $X_{ni}$ under alternative and null respectively. This simplifies to

reject $H_{0ni}$ if $X_{ni} \geq C_n$, $i = 1, \ldots, n.$

where

$$C_n = \frac{K_n\beta_{1n} - \beta_{0n}}{1 - K_n}$$

and

$$K_n = \left[ \frac{(1 - p_n)\delta_{0n}}{p_n\delta_{1n}} \left( \frac{\beta_{0n}}{\beta_{1n}} \right)^\alpha \right]^{-\frac{1}{\alpha+1}}.$$  

This procedure is defined as Bayes oracle due to the use of unknown parameters. [2]

Bogdan et al. (2011) defined a similar Bayes oracle property for the mixture normal model.

In a Bayesian framework, the Bayes risk is an important quantity that can be used to measure the performance of a rule. In order to derive the Bayes risk of the Bayes rule (the Bayes oracle), we firstly need the probabilities of type I and type II errors. We denote the probabilities of type I and type II errors respectively, for all $i$, as $t_{1n} = P_{H_{0ni}}(H_{0ni}$ is rejected$), t_{2n} = P_{H_{Ani}}(H_{0ni}$ is accepted$)$ and $t_{2n} = P_{H_{Ani}}(H_{0ni}$ is accepted$)$, since the marginal distribution of $X_{ni}$ or the threshold does not depend on $i$. Observing that under $H_{0ni}$,

$$Y_0 := -\log \left( \frac{\beta_{0n}}{X_{ni} + \beta_{0n}} \right)^\alpha \sim \text{Exp}(1).$$
and under $H_{A_n}$,\[ Y_1 := -\log \left( \frac{\beta_{1n}}{\lambda_n + \beta_{1n}} \right)^\alpha \sim \text{Exp}(1), \]

the probabilities of a type I and type II errors using the Bayes oracle are given by

\[
t_{1n} = P \left( Y_0 \geq -\log \left[ \frac{\beta_{0n}}{\beta_{1n} - \beta_{0n}} \left( 1 - K_n \right) \right]^\alpha \right) = \left[ \frac{\beta_{0n}}{\beta_{1n} - \beta_{0n}} \left( 1 - K_n \right) \right]^\alpha
\]

\[
= \left( \frac{\beta_{0n}}{C_n + \beta_{0n}} \right)^\alpha \tag{2-1}
\]

and

\[
t_{2n} = P \left( Y_1 < -\log \left[ \frac{\beta_{1n}}{\beta_{1n} - \beta_{0n}} \left( 1 - K_n \right) \right]^\alpha \right) = 1 - \left[ \frac{\beta_{1n}}{\beta_{1n} - \beta_{0n}} \left( 1 - K_n \right) \right]^\alpha
\]

\[
= 1 - \left( \frac{\beta_{1n}}{C_n + \beta_{1n}} \right)^\alpha . \tag{2-2}
\]

In order to have nontrivial probabilities of type II errors, i.e. $\lim_{n \to \infty} t_{2n} \in (0, 1)$, we need an assumption.

**Assumption (⋆)** A sequence of vectors $\{ (\rho_n, \beta_{0n}, \beta_{1n}, \delta_{0n}, \delta_{1n}) \}_{n=1}^\infty$ satisfies the following conditions, as $n \to \infty$,

\[ p_n \to 0, \beta_{1n} \to \infty, \beta_{0n} \to \beta_0 \in (0, \infty), \delta_{0n} \to \delta \in (0, \infty), \frac{\delta_{0n}}{\beta_{1n}} < K_n < 1, \text{ and } K_n \to A \in (0, 1). \]

Starting from this assumption, we can easily derive the following useful results which indicate the convergence rates of probabilities of type I and type II errors and the convergence relationship among different quantities.

**Theorem 2.1** Defining $M_n = \frac{\beta_{1n}}{\rho_n}$, Assumption (⋆) implies that

\[ \frac{\rho_n \delta_{1n}}{(1 - \rho_n) \delta_{0n}} \sim A^{-(\alpha + 1)} M_n^{-\alpha}. \]

**Theorem 2.2** Under Assumption (⋆), from (2–1) and (2–2), the probabilities of type I and type II errors using the Bayes oracle are given respectively by

\[ t_{1n} \sim \left( \frac{1 - A}{A} \right)^\alpha M_n^{-\alpha} \text{ and } t_{2n} \to 1 - (1 - A)^\alpha. \]
Corollary 2.1 Under Assumption (*),
\[
\frac{p_n \delta_{1n}}{(1 - p_n) \delta_{0n}} \sim \frac{1}{A(1 - A)^\alpha} t_{1n}.
\]

Proof: Follows immediately from Theorem 2.2.

Then we can define the Bayes risk of the Bayes rule (Bayes oracle), which has the minimal Bayes risk among all rules, in the following way.

Definition 2.1 Under an additive loss function, the Bayes risk for a fixed threshold multiple testing procedure with respective probabilities \( t'_{1n} \) and \( t'_{2n} \) of type I and type II errors is given by
\[
R' = \sum_{i=1}^{n} \left[ (1 - p_n) t'_{1n} \delta_{0n} + p_n t'_{2n} \delta_{1n} \right] = n \left[ (1 - p_n) t'_{1n} \delta_{0n} + p_n t'_{2n} \delta_{1n} \right].
\]

Thus, the Bayes risk corresponding to Bayes oracle which we will refer to as the optimal risk is
\[
R_{\text{optimal}} = n \left[ (1 - p_n) t_{1n} \delta_{0n} + p_n t_{2n} \delta_{1n} \right]. \tag{2–3}
\]

In addition, following BCFG, we have Definition 2.2 which forms the basis for judging whether an arbitrary multiple testing procedure is good enough or not by comparing the Bayes risk of the procedure with the optimal one.

Definition 2.2 For a sequence of parameters satisfying Assumption (*), a multiple testing rule is called asymptotically Bayes optimal under sparsity (ABOS) if its risk \( R \) satisfies
\[
\frac{R}{R_{\text{optimal}}} \to 1,
\]
as \( n \to \infty \), where \( R_{\text{optimal}} \) is the optimal risk given by (2–3).

In the rest of this paper, ABOS will be referred to asymptotically Bayes optimal (or optimality) under sparsity up to the context.

Based on Definition 2.2, we have the following three theorems which show the ABOS of two general classes of multiple testing procedures.

The first class consists of the fixed threshold multiple testing procedures.
Theorem 2.3 Define a fixed threshold multiple testing procedure \( FTM \) by reject \( H_{0ni} \) if \( X_{ni} \geq C_n^* \), \( i = 1, \ldots, n \), where \( C_n^* \geq 0 \) is a constant.

The corresponding probabilities of type I and type II errors are respectively given by

\[
t_1^* = \left( \frac{\beta_{1n}}{C_n^* + \beta_{1n}} \right)^\alpha \text{ and } t_2^* = 1 - \left( \frac{\beta_{1n}}{C_n^* + \beta_{1n}} \right)^\alpha.
\]

Theorem 2.4 indicates the condition under which a fixed threshold multiple testing procedure is ABOS.

Theorem 2.4 A fixed threshold multiple testing procedure \( FTM \) defined in Theorem 2.3 is ABOS if and only if \( C_n^* \sim C_n \).

Proof: The risk of the multiple testing procedure \( FTM \) with fixed threshold \( C_n^* \) is

\[
R^* := n[(1 - p_n) t_{1n}^* \delta_{0n} + p_n t_{2n}^* \delta_{1n}].
\]

Then, \( FTM \) is ABOS if and only if \( \frac{R^*}{R_{optimal}} \to 1 \)

\[
\iff \frac{R^* - R_{optimal}}{R_{optimal}} \to 0 \iff \frac{(1 - p_n) \delta_{0n}(t_{1n}^* - t_1n) + p_n \delta_{1n}(t_{2n}^* - t_2n)}{(1 - p_n) \delta_{0n} t_{1n} + p_n \delta_{1n} t_{2n}} \to 0
\]

\[
\iff \frac{(t_{1n}^* - t_{1n}) + \frac{p_n \delta_{1n}}{(1 - p_n) \delta_{1n}}(t_{2n}^* - t_{2n})}{t_{1n} + \frac{p_n \delta_{1n}}{(1 - p_n) \delta_{1n}} t_{2n}} \to 0 \iff \frac{(t_{1n}^* - t_{1n}) + \frac{t_{1n}}{A(1 - A)^\alpha}(t_{2n}^* - t_{2n})}{t_{1n} + \frac{1}{A(1 - A)^\alpha} t_{1n}} \to 0.
\]

Define \( L = \frac{A(1 - A)^\alpha}{1 - (1 - A)^\alpha + 1} \frac{t_{1n}^* - t_{1n}}{t_{1n}} + \frac{1}{1 - (1 - A)^\alpha + 1} (t_{2n}^* - t_{2n}) \). There are three possible cases.

1. If \( C_n^* = o(C_n) \), then \( \frac{t_{1n}^* - t_{1n}}{t_{1n}} \to \infty \) and \( |t_{2n}^* - t_{2n}| \leq 2 \).

Then \( L \to \infty \neq 0 \).

2. If \( C_n = o(C_n^*) \), then \( \frac{t_{1n}^* - t_{1n}}{t_{1n}} \to -1 \) and \( t_{2n}^* \to 1 \).

Then \( L \to \frac{(1 - A)^\alpha + 1}{1 - (1 - A)^\alpha + 1} \neq 0 \).
(3) If $C_n^* \sim \sigma C_n$ where $\sigma \in (0, \infty)$, then $t_{n1}^* \sim \sigma^{-\alpha} t_{n1}$ and $t_{n2}^* \rightarrow 1 - \left( \frac{1-A}{1-(\sigma-1)A} \right)^\alpha$. Thus 
\[ \frac{t_{n1}^*-t_{n1}}{t_{n1}} \rightarrow \sigma^{-\alpha}-1 \quad \text{and} \quad t_{n2}^*-t_{n2} \rightarrow (1-A)^\alpha - \left( \frac{1-A}{1+(\sigma-1)A} \right)^\alpha. \]
Therefore,
\[ L \rightarrow \frac{A(1-A)^\alpha}{1-(1-A)^{\alpha+1}}(\sigma^{-\alpha}-1) + \frac{1}{1-(1-A)^{\alpha+1}} \left[ (1-A)^\alpha - \left( \frac{1-A}{1+(\sigma-1)A} \right)^\alpha \right] \]
\[ = \frac{(1-A)^\alpha}{1-(1-A)^{\alpha+1}} \{ A\sigma^{-\alpha} + (1-A) - [A\sigma + (1-A)]^{-\alpha} \}. \]

Since $\alpha > 1$ and $A \in (0, 1)$, by Jensen’s inequality $A\sigma^{-\alpha} + (1-A) \geq [A\sigma + (1-A)]^{-\alpha}$ with equality holding if and only if $\sigma = 1$.

Thus $L \rightarrow 0$ if and only if $\sigma = 1$.

Hence, $\{FTM\}$ is ABOS if and only if $C_n^* \sim C_n$. \hfill \Box

**Remark:** Theorem 2.4 gives a necessary and sufficient condition for achieving ABOS of fixed threshold multiple testing procedure, in a simple form which involves the ratio of two fixed thresholds. Unlike the previous theorem, the following Theorem 2.5 gives a sufficient condition of ABOS for a given random threshold multiple testing procedure requiring a certain convergence rate for the difference of the two thresholds to zero.

**Theorem 2.5** Define a random threshold multiple testing procedure $\{RTM\}$ by reject $H_{0ni}$ if $X_{ni} \geq \hat{C}_n$, $i = 1, \ldots, n$, where $\hat{C}_n$ is a function of random variables.

If for all $\epsilon > 0$,
\[ P\left( |\hat{C}_n - C_n| \geq \epsilon \right) = o(M_n^{-\alpha}), \quad (2-4) \]
then $\{RTM\}$ is ABOS.

**Proof:** Note that by the model assumption,
\[ P\left( |\hat{C}_n - C_n| \geq \epsilon \right) = (1-p_n)P_{H_{0ni}} \left( |\hat{C}_n - C_n| \geq \epsilon \right) + p_nP_{H_{ni}} \left( |\hat{C}_n - C_n| \geq \epsilon \right). \]

From (2–4) and Theorem 2.1, this implies
\[ P_{H_{0ni}} \left( |\hat{C}_n - C_n| \geq \epsilon \right) = o(M_n^{-\alpha}) \quad (2–5) \]
and

\[ P_{H_{\text{av}}} \left( |\hat{\mathcal{C}}_n - C_n| \geq \epsilon \right) = o(1). \]  

(2–6)

Define the probabilities of type I and type II errors respectively by

\[ \hat{\mathcal{t}}_{1ni} = P_{H_{\text{av}}} \left( X_{ni} \geq \hat{\mathcal{C}}_n \right) \]

and \[ \hat{\mathcal{t}}_{2ni} = P_{H_{\text{av}}} \left( X_{ni} < \hat{\mathcal{C}}_n \right). \]

We have, for all \( \epsilon > 0 \),

\[
\hat{\mathcal{t}}_{1ni} = P_{H_{\text{av}}} \left( X_{ni} \geq \hat{\mathcal{C}}_n - C_n + C_n \right) \\
= P_{H_{\text{av}}} \left( [X_{ni} \geq \hat{\mathcal{C}}_n - C_n + C_n] \cap [|\hat{\mathcal{C}}_n - C_n| \geq \epsilon] \right) \\
+ P_{H_{\text{av}}} \left( [X_{ni} \geq \hat{\mathcal{C}}_n - C_n + C_n] \cap [|\hat{\mathcal{C}}_n - C_n| < \epsilon] \right) \\
\leq P_{H_{\text{av}}} \left( |\hat{\mathcal{C}}_n - C_n| \geq \epsilon \right) + P_{H_{\text{av}}} \left( |X_{ni} \geq C_n - \epsilon| \cap [|\hat{\mathcal{C}}_n - C_n| < \epsilon] \right) \\
= P_{H_{\text{av}}} \left( X_{n1} \geq C_n - \epsilon \right) + P_{H_{\text{av}}} \left( |\hat{\mathcal{C}}_n - C_n| \geq \epsilon \right)
\]

and

\[
\hat{\mathcal{t}}_{1ni} = P_{H_{\text{av}}} \left( X_{ni} \geq \hat{\mathcal{C}}_n - C_n + C_n \right) \\
= P_{H_{\text{av}}} \left( [X_{ni} \geq \hat{\mathcal{C}}_n - C_n + C_n] \cap [|\hat{\mathcal{C}}_n - C_n| \geq \epsilon] \right) \\
+ P_{H_{\text{av}}} \left( [X_{ni} \geq \hat{\mathcal{C}}_n - C_n + C_n] \cap [|\hat{\mathcal{C}}_n - C_n| < \epsilon] \right) \\
\leq P_{H_{\text{av}}} \left( |\hat{\mathcal{C}}_n - C_n| \geq \epsilon \right) + P_{H_{\text{av}}} \left( |X_{ni} \geq C_n - \epsilon| \cap [|\hat{\mathcal{C}}_n - C_n| < \epsilon] \right) \\
\leq P_{H_{\text{av}}} \left( |\hat{\mathcal{C}}_n - C_n| \geq \epsilon \right) + P_{H_{\text{av}}} \left( X_{ni} \geq C_n - \epsilon \right) \\
= P_{H_{\text{av}}} \left( X_{n1} \geq C_n - \epsilon \right) + P_{H_{\text{av}}} \left( |\hat{\mathcal{C}}_n - C_n| \geq \epsilon \right)
\]
Similarly,

\[ \hat{t}_{2ni} = P_{H_{an}} \left( X_{ni} < \hat{\zeta}_n - C_n + C_n \right) \]

\[ = P_{H_{an}} \left( [X_{ni} < \hat{\zeta}_n - C_n + C_n] \cap [|\hat{\zeta}_n - C_n| \geq \epsilon] \right) \]

\[ + P_{H_{an}} \left( [X_{ni} < \hat{\zeta}_n - C_n + C_n] \cap [|\hat{\zeta}_n - C_n| < \epsilon] \right) \]

\[ \geq P_{H_{an}} \left( [X_{ni} < C_n - \epsilon] \cap [|\hat{\zeta}_n - C_n| < \epsilon] \right) \]

\[ \geq P_{H_{an}} \left( X_{ni} < C_n - \epsilon \right) - P_{H_{an}} \left( \hat{\zeta}_n - C_n \geq \epsilon \right) \]

\[ = P_{H_{an1}} \left( X_{n1} < C_n - \epsilon \right) - P_{H_{an}} \left( \hat{\zeta}_n - C_n \geq \epsilon \right) \]

and

\[ \hat{t}_{2ni} = P_{H_{an}} \left( X_{ni} < \hat{\zeta}_n - C_n + C_n \right) \]

\[ = P_{H_{an}} \left( [X_{ni} < \hat{\zeta}_n - C_n + C_n] \cap [|\hat{\zeta}_n - C_n| \geq \epsilon] \right) \]

\[ + P_{H_{an}} \left( [X_{ni} < \hat{\zeta}_n - C_n + C_n] \cap [|\hat{\zeta}_n - C_n| < \epsilon] \right) \]

\[ \leq P_{H_{an}} \left( |\hat{\zeta}_n - C_n| \geq \epsilon \right) + P_{H_{an}} \left( [X_{ni} < C_n + \epsilon] \cap [|\hat{\zeta}_n - C_n| < \epsilon] \right) \]

\[ \leq P_{H_{an}} \left( |\hat{\zeta}_n - C_n| \geq \epsilon \right) + P_{H_{an}} \left( X_{ni} < C_n + \epsilon \right) \]

\[ = P_{H_{an1}} \left( X_{n1} < C_n + \epsilon \right) + P_{H_{an}} \left( |\hat{\zeta}_n - C_n| \geq \epsilon \right) \]
Therefore, as \( n \to \infty \), the ratio of the Bayes risk \( \hat{R}_n \) of \( \{ R TM \} \) to the Bayes optimal risk

\[
\frac{\hat{R}_n}{R_{\text{optimal}}} = \frac{\sum_{i=1}^{n} [(1 - p_n) \hat{t}_{1ni} \delta_{0n} + p_n \hat{t}_{2ni} \delta_{1n}]}{n[(1 - p_n) t_{1n} \delta_{0n} + p_n t_{2n} \delta_{1n}]} + \frac{1}{n} \sum \hat{t}_{2ni} (1 - p_n) t_{1n} \delta_{0n} + p_n t_{2n} \delta_{1n}
\]

\[
\leq \frac{P_{H_{\text{opt}}} (X_{n1} \geq C_n - \epsilon) - \frac{1}{n} \sum P_{H_{\text{av}}} (|\hat{C}_n - C_n| \geq \epsilon)}{P_{H_{\text{opt}}} (X_{n1} \geq C_n)} \times \frac{(1 - p_n) t_{1n} \delta_{0n} + p_n t_{2n} \delta_{1n}}{(1 - p_n) t_{1n} \delta_{0n} + p_n t_{2n} \delta_{1n}}
\]

\[
\to 1
\]

which follows from (2–5) and (2–6). Similarly, from (2–5) and (2–6),

\[
\frac{\hat{R}_n}{R_{\text{optimal}}} = \frac{\sum_{i=1}^{n} [(1 - p_n) \hat{t}_{1ni} \delta_{0n} + p_n \hat{t}_{2ni} \delta_{1n}]}{n[(1 - p_n) t_{1n} \delta_{0n} + p_n t_{2n} \delta_{1n}]} + \frac{1}{n} \sum \hat{t}_{2ni} (1 - p_n) t_{1n} \delta_{0n} + p_n t_{2n} \delta_{1n}
\]

\[
\leq \frac{P_{H_{\text{opt}}} (X_{n1} \geq C_n - \epsilon) + \frac{1}{n} \sum P_{H_{\text{av}}} (|\hat{C}_n - C_n| \geq \epsilon)}{P_{H_{\text{opt}}} (X_{n1} < C_n)} \times \frac{(1 - p_n) t_{1n} \delta_{0n} + p_n t_{2n} \delta_{1n}}{(1 - p_n) t_{1n} \delta_{0n} + p_n t_{2n} \delta_{1n}}
\]

\[
\to 1
\]

Thus, as \( n \to \infty \),

\[
\frac{\hat{R}_n}{R_{\text{optimal}}} \to 1.
\]
Hence, $\{RTM\}$ is ABOS.

Theorem 2.4 and 2.5 provide useful tools which can be applied to prove the ABOS of any specific multiple testing procedure within either of these two classes. In the following section, we investigate three examples given respectively by [4] Efron and Tibshirani (2002), [5] Genovese and Wasserman (2002) and [1] Benjamini and Hochberg (1995). We prove that they are all ABOS under certain conditions.
3.1 Connecting the False Discovery Rate, the Bayesian False Discovery Rate and Asymptotic Bayes Optimality under Sparsity

For a fixed thresholding procedure, Efron and Tibshirani (2002) defined another measure of the accuracy of multiple testing procedure, the Bayesian False Discovery Rate (BFDR) as

\[ BFDR = P(H_{0ni} \text{ is true} | H_{0ni} \text{ was rejected}) = \frac{(1 - \rho_n) t_{1n}^*}{(1 - \rho_n) t_{1n}^* + \rho_n (1 - t_{2n}^*)}. \]

The following theorem shows that the Bayes oracle procedure is asymptotically controlling BFDR at a certain level.

**Theorem 3.1.1** Under Assumption (*), the Bayes oracle procedure is controlling BFDR at the level \( \gamma_n^{BO} \) which satisfies, as \( n \to \infty \),

\[ \gamma_n^{BO} \to \frac{A}{A + \delta}. \]

**Proof:** By the definition of BFDR,

\[ \gamma_n^{BO} = \frac{(1 - \rho_n) t_{1n}}{(1 - \rho_n) t_{1n} + \rho_n (1 - t_{2n})}. \]

From Theorem 2.1 and Theorem 2.2, we have

\[ \gamma_n^{BO} \to \frac{A}{A + \delta}. \]

On the other hand, the following Theorem 3.1.2 indicates that, for a fixed thresholding procedure, under certain conditions, controlling the BFDR at a certain level is asymptotically equivalent to achieving the ABOS, and it specifies one necessary and sufficient condition. This theorem offers for this an opportunity to compare methods based on BFDR with those based on ABOS.
We first provide the specific form of the threshold of a fixed threshold procedure controlling BFDR at a given level in Lemma 3.1.1. Then we prove Theorem 3.1.2 based on the result of Lemma 3.1.1.

**Lemma 3.1.1** Under Assumption (*), for a fixed threshold rule controlling BFDR at the level \( \gamma / x_{42} \), the threshold \( C_n^B \) has the form

\[
C_n^B = (\beta_1 n - \beta_0 n) \left[ 1 - \frac{\beta_0 n}{\beta_1 n} \left( \frac{1 - p_n}{\gamma_n^B} \right)^{\frac{1}{\alpha}} \right]^{-1} - \beta_1 n.
\]

**Proof:** By the definition of BFDR, Theorem 2.1 and Theorem 2.2, we can obtain \( C_n^B \) by solving

\[
\hat{\gamma}_n^B = \frac{(1 - p_n) \gamma_n^B}{(1 - p_n) \gamma_n^B + p_n (1 - \gamma_n^B)} = \frac{(1 - p_n) \left( \frac{\beta_n}{\gamma_n^B + \gamma_n^B} \right)^{\alpha}}{(1 - p_n) \left( \frac{\beta_n}{\gamma_n^B + \gamma_n^B} \right)^{\alpha} + p_n \left( \frac{\beta_n}{\gamma_n^B + \gamma_n^B} \right)^{\alpha}}.
\]

Direct calculation gives

\[
C_n^B = (\beta_1 n - \beta_0 n) \left[ 1 - \frac{\beta_0 n}{\beta_1 n} \left( \frac{1 - p_n (1 - \gamma_n^B)}{p_n \gamma_n^B} \right)^{\frac{1}{\alpha}} \right]^{-1} - \beta_1 n.
\]

**Theorem 3.1.2** Under Assumption (*), a fixed thresholding rule with the threshold \( C_n^B \) controlling BFDR at the level \( \gamma_n^B \in (0, 1) \) is ABOS if and only if

\[
\gamma_n^B \overset{\text{A}}{\to} \frac{A}{A + \delta}.
\]

**Proof:** The proof follows from Theorem 2.4 and Lemma 3.1.1.

In an earlier work, [1] Benjamini and Hochberg (1995) introduced the False Discovery Rate (FDR), as

\[
FDR = E \left( \frac{V}{R} \right)
\]

where \( R \) is the total number of null hypotheses rejected, \( V \) is the number of "false" rejections and it is assumed that \( \frac{V}{R} = 0 \) when \( R = 0 \).

[5] Genovese and Wasserman (2002) proved that while the number of tests tends to infinity and the fraction of true alternatives remains fixed, a fixed threshold multiple
testing procedure with threshold $C_{n}^{GW}$ can approximately control FDR at level $\gamma_n$, if its corresponding type I error $t_{1n}^{GW}$ and type II error $t_{2n}^{GW}$ satisfy

$$\gamma_n = \frac{t_{1n}^{GW}}{(1 - p_n)t_{1n}^{GW} + p_n(1 - t_{2n}^{GW})}. $$

We may note that $\gamma_n$ is similar to $\gamma^B_n$ with the exception of the factor $1 - p_n$ appearing in the numerator of the latter. Thus similar to Lemma 3.1.1, Lemma 3.1.2 gives an expression for the fixed threshold of GW. Then we have the following Theorem 3.1.3 for the relationship between controlling FDR and achieving ABOS.

**Lemma 3.1.2** Under Assumption (*), for a fixed thresholding rule controlling FDR at the level $\gamma_n \in (0, 1)$, the Genovese-Wasserman approximation threshold $C_{n}^{GW}$ has the form

$$C_{n}^{GW} = (\beta_{1n} - \beta_{0n}) \left[ 1 - \frac{\beta_{0n}}{\beta_{1n}} \left( \frac{1 - (1 - p_n)\gamma_n}{p_n \gamma_n} \right)^{\frac{1}{\alpha}} \right]^{-1} - \beta_{1n}. $$

**Proof:** Invoking Theorem 2.3 and Theorem 2.2, one obtains $C_{n}^{GW}$ by solving

$$\gamma_n = \frac{t_{1n}^{GW}}{(1 - p_n)t_{1n}^{GW} + p_n(1 - t_{2n}^{GW})} = \frac{\left( \frac{\beta_{0n}}{C_{n}^{GW} + \beta_{0n}} \right)^{\alpha}}{(1 - p_n)\left( \frac{\beta_{0n}}{C_{n}^{GW} + \beta_{0n}} \right)^{\alpha} + p_n \left( \frac{\beta_{0n}}{C_{n}^{GW} + \beta_{0n}} \right)^{\alpha}}.$$

Direct calculation gives

$$C_{n}^{GW} = (\beta_{1n} - \beta_{0n}) \left[ 1 - \frac{\beta_{0n}}{\beta_{1n}} \left( \frac{1 - (1 - p_n)\gamma_n}{p_n \gamma_n} \right)^{\frac{1}{\alpha}} \right]^{-1} - \beta_{1n}. $$

**Theorem 3.1.3** Under Assumption (*), a fixed thresholding rule with the threshold $C_{n}^{GW}$ controlling FDR at the level $\gamma_n \in (0, 1)$ is ABOS if and only if

$$\gamma_n \rightarrow \frac{A}{A + \delta}. $$

**Proof:** The result follows from Theorem 2.4 and Lemma 3.1.2.

**Remark:** For a fixed thresholding procedure, Theorems 3.1.2 and 3.1.3 demonstrate that under controlling either BFDR or FDR, ABOS holds if and only if there is one-to-one asymptotic corresponding between the loss ratio $\delta$ and the level where BFDR or FDR is
held fixed. As a result there of, it is the loss ratio which asymptotically dictates the level at which we want to control FDR or BFDR. We may consider the loss ratio as a “tuning parameter” in this case.

In addition, here we prove Theorem 3.1.4 which gives a result of independent interest and will be needed in Section 3.2.

**Theorem 3.1.4** Under Assumption (*), as \( n \to \infty \), a necessary and sufficient condition such that \( C_n^{GW} = C_n + o(1) \) is

\[
\gamma_n = \frac{A}{A + \delta} + o(M_n^{-1}).
\]

**Proof:** By Lemma 3.1.2, we have

\[
C_n^{GW} - C_n = o(1),
\]

if and only if

\[
(\beta_{1n} - \beta_{0n}) \left[ 1 - \frac{\beta_{0n}}{\beta_{1n}} \left( \frac{1 - (1 - p_n) \gamma_n}{p_n \gamma_n} \right)^{\frac{1}{n}} \right]^{-1} - \beta_{1n} - \frac{K_n \beta_{1n} - \beta_{0n}}{1 - K_n} = o(1)
\]

\[
\Leftrightarrow (\beta_{1n} - \beta_{0n}) \left[ 1 - \frac{\beta_{0n}}{\beta_{1n}} \left( \frac{1 - (1 - p_n) \gamma_n}{p_n \gamma_n} \right)^{\frac{1}{n}} \right]^{-1} - (1 - K_n)^{-1} = o(1)
\]

\[
\Leftrightarrow (\beta_{1n} - \beta_{0n}) \left[ \frac{\beta_{0n}}{\beta_{1n}} \left( \frac{1 - (1 - p_n) \gamma_n}{p_n \gamma_n} \right)^{\frac{1}{n}} - K_n \right] = o(1)
\]

\[
\Leftrightarrow \left[ \frac{1 - (1 - p_n) \gamma_n}{p_n \gamma_n} \right]^{\frac{1}{n}} - \frac{\beta_{1n} K_n}{\beta_{0n}} = o(1)
\]

\[
\Leftrightarrow M_n^{-1} K_n^{-1} \left[ \frac{1 - (1 - p_n) \gamma_n}{p_n \gamma_n} \right]^{\frac{1}{n}} = 1 + o(M_n^{-1}).
\]

By Taylor’s expansion, this holds if and only if

\[
\frac{1 - (1 - p_n) \gamma_n}{M_n^\alpha K_n^\alpha p_n \gamma_n} = \left[ 1 + o(M_n^{-1}) \right]^\alpha = 1 + o(M_n^{-1}),
\]
By Theorem 2.1, this is equivalent to

\[ \frac{1}{\gamma_n} = \frac{A + \delta}{A} + o(M_n^{-1}), \]

or equivalently

\[ \gamma_n = \frac{A}{A + \delta} + o(M_n^{-1}). \quad \square \]

### 3.2 Asymptotic Bayes Optimality under Sparsity of the Benjamini-Hochberg Procedure

The Benjamini-Hochberg procedure controls FDR at level \( \gamma_n \), by sorting p-values in an ascending order \( \rho(1) \leq \rho(2) \leq \ldots \leq \rho(n) \) and rejecting the null hypotheses for which the corresponding p-values are smaller than or equal to \( \frac{i \gamma_n}{n} \) where

\[ t = \arg\max_i \left\{ \rho(i) \leq \frac{i \gamma_n}{n} \right\}. \]

Instead of using p-values, to consider an equivalent procedure with a random threshold for observations \( X_{ni}, i = 1, \ldots, n \), we denote \( 1 - \hat{F}_n(y) = \frac{\# \{ X_n > y \} }{n} \), and

\[ 1 - F(y) = P(X_{n1} > y) = (1 - \rho_n) P_{H_{0n1}}(X_{n1} > y) + \rho_n P_{H_{1n1}}(X_{n1} > y). \]

It is shown, e.g. [4] Efron and Tibshirani (2002), that the BH procedure rejects the null hypothesis \( H_{0ni} \) when \( X_{ni} \geq \hat{c}_n^{BH} \) where the random threshold is given by

\[ \hat{c}_n^{BH} = \inf \left\{ y : \frac{P_{H_{1n1}}(X_{n1} > y)}{1 - \hat{F}_n(y)} \leq \gamma_n \right\}. \]

To prove that the random thresholding BH procedure is ABOS, we need some restrictions on the parameter space in addition to Assumption (*). The following Theorem 3.2.1 provides a sufficient condition under which the random thresholding BH procedure is ABOS.

**Theorem 3.2.1** Under Assumption (*), the random thresholding BH procedure controlling FDR at level of \( \gamma_n \) is ABOS if

\[ \gamma_n = \frac{A}{A + \delta} + o(M_n^{-1}), \]

\[ 31 \]
and

\[ M_n^{\alpha + \rho^2 \log n + \alpha \log(M_n]} = o(n), \]

where \( \rho > 0 \) is arbitrary.

**Proof:** Let \( \tilde{C}_{n}^{BH} \) be the BH random threshold at level \( \gamma_n \). The proof of the theorem is based on several technical lemmas. Lemma 3.2.1 is a technical device to Lemma 3.2.2 which shows the connection between the random threshold of BH and a modified version of the fixed threshold of GW. Then, Lemma 3.2.3 and Lemma 3.2.4 show that

\[ P\left( |\tilde{C}_{n}^{BH} - C_n| \geq \epsilon \right) = o(M_n^{-\alpha}). \]

Hence the result of Theorem 3.2.1 follows from Theorem 2.5. \( \square \)

**Lemma 3.2.1** Suppose Assumption (\( \star \)) holds and \( \gamma_n = \frac{A}{A+\delta} + o(M_n^{-1}) \). Let \( C_n^{GW} \) be the GW threshold at level \( \gamma_n \). For any constant \( \xi \in (0, 1) \), we have

\[ P\left( \frac{1 - \hat{P}_n(C_n^{GW})}{1 - F(C_n^{GW})} > 1 + \xi \right) \leq \exp \left\{ -\frac{1}{4} \left[ \frac{A + \delta}{A} \left( \frac{1 - A}{A} \right)^{\alpha} \right] nM_n^{-\alpha} \xi^2 [1 + o(1)] \right\}, \]

and

\[ P\left( \frac{1 - \hat{P}_n(C_n^{GW})}{1 - F(C_n^{GW})} < 1 - \xi \right) \leq \exp \left\{ -\frac{1}{4} \left[ \frac{A + \delta}{A} \left( \frac{1 - A}{A} \right)^{\alpha} \right] nM_n^{-\alpha} \xi^2 [1 + o(1)] \right\}. \]

**Proof:** By Theorem 2.3 and Theorem 3.1.4,

\[ 1 - F(C_n^{GW}) = (1 - p_n)P_{H_{n1}}(X_{n1} > C_n^{GW}) + p_nP_{H_{n1}}(X_{n1} > C_n^{GW}) \]

\[ = \left[ \frac{A + \delta}{A} \left( \frac{1 - A}{A} \right)^{\alpha} \right] M_n^{-\alpha} [1 + o(1)] \]

Note that \( 1 - \hat{P}_n(C_n^{GW}) \) is the average of \( n \) Bernoulli random variables with success probability \( 1 - F(C_n^{GW}) \). Hence, by Bernstein’s inequality (cf. [7] Serfling, 1980, p. 96),
one gets

\[
P \left( \left[ 1 - \hat{F}_n(C_{n}^{GW})\right] - \left(1 - F(C_{n}^{GW})\right) \right] > (1 - F(C_{n}^{GW})) \xi \right)
\]

\[
\leq \exp \left[-\frac{n\xi^2(1 - F(C_{n}^{GW}))^2}{2(1 + \xi)(1 - F(C_{n}^{GW}))} \right]
\]

\[
\leq \exp \left[-(n\xi^2/4)(1 - F(C_{n}^{GW})) \right]
\]

\[
= \exp \left\{ -\frac{1}{4} \left[ \frac{A + \delta}{A} \left( \frac{1 - A}{A} \right)^{\alpha} \right] nM_n^{-\alpha} \xi^2 \left[ 1 + o(1) \right] \right\}
\]

and similarly

\[
P \left( \left[ 1 - \hat{F}_n(C_{n}^{GW})\right] - \left(1 - F(C_{n}^{GW})\right) \right] < -(1 - F(C_{n}^{GW})) \xi \right) \leq
\]

\[
\exp \left\{ -\frac{1}{4} \left[ \frac{A + \delta}{A} \left( \frac{1 - A}{A} \right)^{\alpha} \right] nM_n^{-\alpha} \xi^2 \left[ 1 + o(1) \right] \right\}.
\]

Hence

\[
P \left( \frac{1 - \hat{F}_n(C_{n}^{GW})}{1 - F(C_{n}^{GW})} > 1 + \xi \right) \leq \exp \left\{ -\frac{1}{4} \left[ \frac{A + \delta}{A} \left( \frac{1 - A}{A} \right)^{\alpha} \right] nM_n^{-\alpha} \xi^2 \left[ 1 + o(1) \right] \right\},
\]

and

\[
P \left( \frac{1 - \hat{F}_n(C_{n}^{GW})}{1 - F(C_{n}^{GW})} < 1 - \xi \right) \leq \exp \left\{ -\frac{1}{4} \left[ \frac{A + \delta}{A} \left( \frac{1 - A}{A} \right)^{\alpha} \right] nM_n^{-\alpha} \xi^2 \left[ 1 + o(1) \right] \right\}.
\]

**Lemma 3.2.2** Suppose Assumption (\(\ast\)) holds and \(\gamma_n = \frac{A}{A+\delta} + o(M_n^{-1})\). Let \(\hat{C}_n^{BH}\) be the BH random threshold at level \(\gamma_n\) and let \(C_{1n}^{GW}\) be the GW threshold at level \(\gamma_{1n} = \gamma_n(1 - \xi_n)\) where \(\xi_n = o(M_n^{-1})\) as \(n \to \infty\). It follows that

\[
C_{1n}^{GW} = C_n + o(1),
\]

and

\[
P \left( \hat{C}_n^{BH} > C_{1n}^{GW} \right) \leq \exp \left\{ -\frac{1}{4} \left[ \frac{A + \delta}{A} \left( \frac{1 - A}{A} \right)^{\alpha} \right] nM_n^{-\alpha} \xi_n^2 \left[ 1 + o(1) \right] \right\}.
\]

**Proof:** Note that

\[
\gamma_{1n} = \gamma_n(1 - \xi_n) = \frac{A}{A+\delta} + o(M_n^{-1}).
\]
By Theorem 3.1.4,
\[ C_{1n}^{GW} = C_n + o(1). \]

Observe that from the definition of \( \hat{C}_{n}^{BH} \),
\[ \frac{P_{H_0}(X_{n1} > C_{1n}^{GW})}{1 - F_n(C_{1n}^{GW})} \leq \gamma_n \text{ implies } \hat{C}_{n}^{BH} \leq C_{1n}^{GW}. \]

Note \( \frac{P_{H_0}(X_{n1} > C_{1n}^{GW})}{1 - F(C_{1n}^{GW})} = \gamma_1 n = \gamma_n(1 - \xi_n) \). Therefore,
\[
P\left( \hat{C}_{n}^{BH} \leq C_{1n}^{GW} \right) \geq P\left( \frac{P_{H_0}(X_{n1} > C_{1n}^{GW})}{1 - F_n(C_{1n}^{GW})} \leq \gamma_n \right) 
= P\left( \frac{P_{H_0}(X_{n1} > C_{1n}^{GW})}{1 - F(C_{1n}^{GW})} \leq \gamma_n(1 - \xi_n) \right) 
= P\left( 1 - \frac{F_n(C_{1n}^{GW})}{1 - F(C_{1n}^{GW})} \geq 1 - \xi_n \right) .
\]

Hence,
\[
P\left( \hat{C}_{n}^{BH} > C_{1n}^{GW} \right) \leq P\left( \frac{1 - F_n(C_{1n}^{GW})}{1 - F(C_{1n}^{GW})} < 1 - \xi_n \right) 
\leq \exp \left\{ -\frac{1}{4} \left[ \frac{A + \delta}{A} \left( \frac{1 - A}{A} \right)^{\alpha} \right] nM_{\alpha}^{\alpha-\alpha} \xi_n^2 [1 + o(1)] \right\} . \quad \square
\]

**Lemma 3.2.3** Suppose Assumption (\( \ast \)) holds, \( \gamma_n = \frac{A}{A+\delta} + o(M_n^{-1}) \) and \( M_{n}^{\alpha+\rho+2} \log(M_n^{\alpha}) = o(n) \). Let \( \hat{C}_{n}^{BH} \) be the BH random threshold at level \( \gamma_n \). It follows that
\[ P(\hat{C}_{n}^{BH} > C_n + \epsilon) = o(M_n^{-\alpha}). \]

**Proof:** Note that \( M_{n}^{\alpha+\rho+2} \log n + \alpha \log(M_n) = o(n) \), which implies that \( M_{n}^{\alpha+\rho+2} \log(M_n^{\alpha}) = o(n) \). Choosing \( \xi_n = \frac{2M_{n}^{\alpha+\rho+2}}{\sqrt{2\alpha}} \), by Lemma 3.2.2,
\[
P(\hat{C}_{n}^{BH} > C_n + \epsilon) \leq \exp \left\{ -\frac{1}{4} \left[ \frac{A + \delta}{A} \left( \frac{1 - A}{A} \right)^{\alpha} \right] nM_{\alpha}^{\alpha-\alpha} \xi_n^2 [1 + o(1)] \right\} 
= \exp \left\{ -\frac{1}{4} \left[ \frac{A + \delta}{A} \left( \frac{1 - A}{A} \right)^{\alpha} \right] nM_n^{\alpha+\alpha+\rho+2} [1 + o(1)] \right\} 
= o(M_n^{-\alpha}). \quad \square
Lemma 3.2.4 Suppose Assumption (*) holds, \( \gamma_n = \frac{A}{\lambda + \delta} + o(M_n^{-1}) \) and \( M_n^{\alpha + \rho + 2} \log(\lambda/n) = o(n) \). Let \( \hat{C}_{BH}^n \) be the BH random threshold at level \( \gamma_n \). It follows that
\[
P(\hat{C}_{BH}^n < C_n - \epsilon) = o(M_n^{-\alpha}).
\]

**Proof:** Let \( C_{GW}^n \) be GW threshold at level \( \gamma_{2n} = \gamma_n(1 + \xi_n) \) where \( \xi_n = \frac{2M_n^{-(1+\rho/2)}}{\sqrt{\log(\lambda/n)}}. \)
Then,
\[
C_{GW}^n = C_n + o(1).
\]
Since \( \frac{P_{H_{BH}}(X_{n1} > y)}{1 - F(y)} \) is monotonically decreasing in \( y \),
\[
\hat{C}_{BH}^n < C_{GW}^n \quad \text{if and only if} \quad \frac{P_{H_{BH}}(X_{n1} > \hat{C}_{BH}^n)}{1 - F(\hat{C}_{BH}^n)} > \frac{P_{H_{GW}}(X_{n1} > C_{GW}^n)}{1 - F(C_{GW}^n)} = \gamma_n(1 + \xi_n).
\]
By the definition of \( \hat{C}_{BH}^n \),
\[
\frac{P_{H_{BH}}(X_{n1} > \hat{C}_{BH}^n)}{1 - F(\hat{C}_{BH}^n)} \leq \gamma_n.
\]
Thus, the event \( \hat{C}_{BH}^n < C_{GW}^n \) implies that
\[
\frac{1 - F(\hat{C}_{BH}^n)}{1 - F(C_{GW}^n)} > 1 + \xi_n.
\]
Therefore,
\[
P(\hat{C}_{BH}^n < C_{GW}^n) \leq P\left( \frac{1 - F(\hat{C}_{BH}^n)}{1 - F(C_{BW}^n)} > 1 + \xi_n \right) \leq P\left( \sup_{c \in [0, C_{GW}^n]} \frac{1 - F_n(c)}{1 - F(c)} > 1 + \xi_n \right).
\]
Let \( t = F(c), z_n = F(C_{GW}^n) \) and \( \hat{G}_n(t) = \text{empirical cdf of } n \text{ iid } \text{Uniform}(0, 1). \)
\[
P(\hat{C}_{BH}^n < C_{GW}^n) \leq P\left( \sup_{t \in [0, z_n]} \frac{1 - \hat{G}_n(t)}{1 - t} > 1 + \xi_n \right).
\]
Note that \( z_n = 1 - \left[ \frac{A + \delta}{A} \left( \frac{1 - A}{A} \right)^{\alpha} \right] M_n^{-\alpha} [1 + o(1)] \). Let \( u_i = \frac{i}{n} \) and \( k_n = \left\lfloor n \left( 1 - \left[ \frac{A + \delta}{A} \left( \frac{1 - A}{A} \right)^{\alpha} \right] M_n^{-\alpha} \right) \right\rfloor \).
From the monotonicity of \( \hat{G}_n(t) \) and \( t \), it follows that for \( t \in [u_i, u_i + \frac{1}{n}] \), \( t - \hat{G}_n(t) < u_i + \frac{1}{n} - \hat{G}_n(u_i) \), and \( 1 - t > 1 - u_i - \frac{1}{n} \). This implies \( \frac{t - \hat{G}_n(t)}{1 - t} < \frac{u_i + \frac{1}{n} - \hat{G}_n(u_i)}{1 - u_i - \frac{1}{n}} \). Thus
\[
\frac{1 - \hat{G}_n(u_i)}{1 - u_i - \frac{1}{n}}.
\]
Table 3-1. Sum Loss Comparison between Bayes Oracle and BH Procedure

<table>
<thead>
<tr>
<th>$(\rho_n, \beta_{1n})$</th>
<th>(0.001,10)</th>
<th>(0.01,10)</th>
<th>(0.1,10)</th>
<th>(0.001,100)</th>
<th>(0.01,100)</th>
<th>(0.1,100)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Loss of BO</td>
<td>6</td>
<td>65</td>
<td>679</td>
<td>0</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>Loss of BH</td>
<td>6</td>
<td>67</td>
<td>684</td>
<td>0</td>
<td>1</td>
<td>65</td>
</tr>
</tbody>
</table>

Therefore for sufficiently large $n$

$$P\left(\sup_{t \in [0, z_n]} \frac{1 - \hat{G}_n(t)}{1 - t} > 1 + \xi_n\right) \leq \sum_{i=0}^{k_n} P\left(1 - \hat{G}_n(u_i) > (1 - u_i - \frac{1}{n}(1 + \xi_n)\right).$$

Now observe that for every $i = 0, \ldots, k_n$,

$$1 - u_i + \frac{k_n}{n} \geq \left[\frac{A + \delta}{A} \left(\frac{1 - A}{A}\right)^{\alpha}\right] M_n^{-\alpha}.$$

Letting $\tau_{ni} = \frac{1}{n(1 - u_i)}$, we have $1 - u_i - \frac{1}{n} = (1 - u_i)(1 - \tau_{ni})$, and $\tau_{ni} \leq \left[\frac{A + \delta}{A} \left(\frac{1 - A}{A}\right)^{\alpha}\right]^{-1} M_n^{\alpha} = o(\xi_n)$. Then by the Bernstein inequality again, for every $i = 0, \ldots, k_n$,

$$P\left(1 - \hat{G}_n(u_i) > (1 - u_i - \frac{1}{n}(1 + \xi_n)\right) = P\left(1 - \hat{G}_n(u_i) > (1 - u_i)(1 + \xi_n)(1 - \tau_{ni})\right)$$

$$\leq \exp\left(-\frac{1}{4} n(1 - u_i)\xi_n^2 [1 + o(1)]\right)$$

$$\leq \exp\left(-\frac{1}{4} \left[\frac{A + \delta}{A} \left(\frac{1 - A}{A}\right)^{\alpha}\right] n M_n^{-\alpha} \xi_n^2 [1 + o(1)]\right)$$

Therefore,

$$P(\hat{c}_n^{BH} < C_n^{GW}) \leq P\left(\sup_{t \in [0, z_n]} \frac{1 - \hat{G}_n(t)}{1 - t} > 1 + \xi_n\right)$$

$$\leq \sum_{i=0}^{k_n} P\left(1 - \hat{G}_n(u_i) > (1 - u_i - \frac{1}{n}(1 + \xi_n)\right)$$

$$\leq n \exp\left(-\frac{1}{4} \left[\frac{A + \delta}{A} \left(\frac{1 - A}{A}\right)^{\alpha}\right] n M_n^{-\alpha} \xi_n^2 [1 + o(1)]\right)$$

$$= o(M_n^{-\alpha}).$$

Hence, $P(\hat{c}_n^{BH} < C_n - \epsilon) = o(M_n^{-\alpha})$. \hfill \Box.
3.3 Simulation Study

In this section, we conduct a simulation study to evaluate the performance of the procedure using Bayes oracle and the performance of Benjamin-Hochberg procedure using random threshold of p-value's.

WLOG, we choose $n = 10000$, $\alpha = 100$, $\beta_0 = 1$, and $\delta_0 = \delta_1 = 1$, and consider $\{0.001, 0.01, 0.1\}$ as the set of candidate $p_n$'s and $\{10, 100\}$ as the set of candidate $\beta_1$'s. We get the fixed threshold $C_n$ by using the true values of the parameters. We calculate the FDR level $\gamma_n$ which the BH procedure is controlling at using the first equation from Theorem 3.2.1 by replacing $A$ with $K_n$ and ignoring the $o\left(\frac{1}{n}\right)$ term.

We generate data from six (three $p_n$'s by two $\beta_1$'s) different two-group mixture Pareto distribution. Then using the data from each different model, we compare Bayes oracle procedure and BH procedure in the sense of comparing the sum of loss.

Table 3-1 shows the sum loss comparison between the Bayes oracle and the BH procedure. While the ratio $M_n = \frac{\beta_1}{\beta_0}$ is 10, the Bayes oracle and the BH procedure have almost the same sum loss, and they both claim significance for about 65% of the true alternatives. While the ratio $M_n = \frac{\beta_1}{\beta_0}$ is 100 and $p_n$ is either 0.001 or 0.01, both the Bayes oracle and the BH procedure make the correct decisions for almost all the hypotheses. All the above cases show that the BH procedure well approximates the Bayes rule in the sense of minimizing the Bayes risk.

However, while the ratio $M_n = \frac{\beta_1}{\beta_0}$ is 100 and $p_n = 0.1$ which means the sparsity is moderate, the Bayes oracle's sum loss is 1, but the BH procedure has a sum loss of 65. The Bayes oracle performs much better than the BH procedure in this case, which also means that the BH procedure fails to approximate the Bayes rule in the sense of minimizing the Bayes risk.

Note that none of the six combinations for $p_n$ and $\beta_1$ satisfies the sufficient condition in Theorem 3.2.1. However, only the last combination is an example of
failure for approximation. It tells us that the sufficient conditions for the BH procedure to achieve ABOS can be relaxed in advance.
CHAPTER 4
ALTERNATIVE CHOICE FOR THE DENSITY OF THE SCALE PARAMETER

From this chapter on, the notations in the previous chapter are voided, and some of the notations are reassigned to other representations.

Suppose, for each fixed $x_{ni}$, $x_{58} \in E$, $x_{31} \in C$, are independent exponentials with respective pdfs

$$f_{\sigma_{ni}}(x) = \sigma_{ni}^{-1} \exp(-x/\sigma_{ni}), \quad x > 0, \sigma_{ni} > 0; \quad i = 1, \ldots, n.$$  

We will denote these pdfs by $\text{Exp}(\sigma_{ni})$, $i = 1, \ldots, n$.

Our objective is to test simultaneously $H_{0ni}$ versus $H_{1ni}$, where under $H_{jni}$, $\sigma_{ni}$ has Gamma pdf

$$\pi_{nj}(\sigma) = \left[\exp(-\frac{\sigma_{ni}}{\beta_{jn}})\sigma_{ni}^{\alpha-1}\right]/[\beta_{jn}^\alpha \Gamma(\alpha)], \quad j = 0, 1, \quad \beta_{2n} > \beta_{0n} > 0, \quad \alpha > 0.$$  

Then, marginally under $H_{jni}$, $X_{ni}$ has pdf

$$f_{jn}(x) = \int_0^\infty\frac{1}{\beta_{jn}^\alpha \Gamma(\alpha)}\sigma_{ni}^{\alpha-2}\exp\left(-\frac{\sigma_{ni}}{\beta_{jn}} - \frac{x}{\sigma_{ni}}\right) d\sigma_{ni}, \quad j = 0, 1,$$

where $\beta_{1n} > \beta_{0n}$ both unknown and $\alpha > 0$ known.

Let $\lambda = 2x$ and $\mu_{jn} = \sqrt{x/\beta_{jn}}$, then for $j = 0, 1$,

$$f_{jn}(x) = \frac{1}{\beta_{jn}^\alpha \Gamma(\alpha)}\left(\frac{2\pi}{\lambda}\right)^{1/2}\exp\left(-\frac{\lambda}{\mu_{jn}}\right)$$

$$\int_0^\infty \sigma_{ni}^{\alpha-1/2}\left(\frac{2\pi}{\lambda}\right)^{-1/2}\sigma_{ni}^{-3/2}\exp\left[-\frac{\lambda}{2\sigma_{ni}}\left(\frac{\sigma_{ni}}{\beta_{jn}^\alpha - 1}\right)\right] d\sigma_{ni}$$

$$= \frac{\pi^{1/2}}{\Gamma(\alpha)}\beta_{jn}^{-\alpha}x^{-1/2}\exp\left(-2\sqrt{\frac{x}{\beta_{jn}}}\right) E\left(Y_{jn}^{\alpha-1/2}\right),$$

where $Y_{jn} \sim \text{Inverse Gaussian}(\mu_{jn}, \lambda) = IG(\sqrt{x/\beta_{jn}}, 2x)$. For $\alpha = 1/2, 3/2, 5/2, \ldots$, the closed form expression of $E\left(Y_{jn}^{\alpha-1/2}\right)$ may be derived. For simplicity, we choose $\alpha = 1/2$. In this case, $E\left(Y_{jn}^{\alpha-1/2}\right) = 1$. 

39
Table 4-1. Matrix of Losses

<table>
<thead>
<tr>
<th></th>
<th>choose $H_{0ni}$</th>
<th>choose $H_{Ani}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$H_{0ni}$ true</td>
<td>0</td>
<td>$\delta_{0n}$</td>
</tr>
<tr>
<td>$H_{Ani}$ true</td>
<td>$\delta_{1n}$</td>
<td>0</td>
</tr>
</tbody>
</table>

We assume that $H_{0ni}$ and $H_{1ni}$ occur with probabilities $p_n$ and $1 - p_n$ respectively. Thus marginally, $X_{ni}$ has the two-group mixture pdf

$$(1 - p_n) f_{0n}(x) + p_n f_{1n}(x).$$

The word sparsity refers to the situation $p_n \approx 0$.

Consider the additive loss $\sum_{i=1}^{n} L(X_{ni})$, where $L(X_{ni}) = \delta_{0n} \phi(X_{ni}) + \delta_{1n} (1 - \phi(X_{ni}))$. Here $\phi(X_{ni})$ equals 1 or 0 according to the rejection or acceptance of $H_0$. Table 4-1 describes this loss structure.

Under this loss structure, the Bayes decision problem leads to a procedure which chooses the alternative hypothesis $H_{Ani}$ when

$$\frac{f_{1n}(X_{ni})}{f_{0n}(X_{ni})} \geq \frac{(1 - p_n) \delta_{0n}}{p_n \delta_{1n}},$$

where $f_{1n}$ and $f_{0n}$ are the densities of $X_{ni}$ under alternative and null respectively. This simplifies to reject $H_{0ni}$ if

$$\sqrt{\frac{\beta_{0n}}{\beta_{1n}}} \exp \left[ 2 \sqrt{X_{ni}} \left( \frac{1}{\sqrt{\beta_{0n}}} - \frac{1}{\sqrt{\beta_{1n}}} \right) \right] \geq \frac{(1 - p_n) \delta_{0n}}{p_n \delta_{1n}},$$

or equivalently, reject $H_{0ni}$ if

$$X_{ni} \geq C_n, i = 1, \ldots, n.$$  

where

$$C_n = \left[ \frac{1}{2} \sqrt{\frac{\beta_{0n} \beta_{1n}}{\beta_{1n} - \sqrt{\beta_{0n}}}} \log \left( \frac{(1 - p_n) \delta_{0n}}{p_n \delta_{1n}} \sqrt{\frac{\beta_{1n}}{\beta_{0n}}} \right) \right]^2.$$  

This procedure is defined as Bayes oracle due to the use of unknown paramters.

In a Bayesian framework, the Bayes risk is an important quantity that can be used to measure the performance of a rule. In order to derive the Bayes risk of the
Bayes rule (the Bayes oracle), we firstly need the probabilities of type I and type II errors. We denote the probabilities of type I and type II errors respectively, for all $i$, as $t_{1n} = P_{H_{0i}}(H_{0i} \text{ is rejected}) = P_{H_{1i}}(X_{ni} \geq C_n)$ and $t_{2n} = P_{H_{0i}}(H_{0i} \text{ is accepted}) = P_{H_{1i}}(X_{ni} < C_n)$, since the marginal distribution of $X_{ni}$ or the threshold does not depend on $i$.

Note that under $H_{ji}$, $Z_{ni} := \sqrt{X_{ni}} \sim \text{Exponential} \left( \frac{\beta_{ij}}{2} \right)$. The probabilities of a type I and type II errors using the Bayes oracle are given by

$$t_{1n} = P_{H_{1i}} \left( Z_{ni} \geq C_n \right) = \exp \left( -2 \sqrt{\frac{C_n}{\beta_{1n}}} \right)$$  \hspace{1cm} (4–1)

and

$$t_{2n} = P_{H_{1i}} \left( Z_{ni} < C_n \right) = 1 - \exp \left( -2 \sqrt{\frac{C_n}{\beta_{1n}}} \right)$$ \hspace{1cm} (4–2)

In order to have nontrivial probabilities of type II errors, i.e. $\lim_{n \to \infty} t_{2n} \in (0, 1)$, we need an assumption.

**Assumption (**)** Let $M_n = \frac{\beta_n}{\beta_{0n}}$. A sequence of vectors $\{ (p_n, \beta_{0n}, \beta_{1n}, \delta_{0n}, \delta_{1n}) \}_{n=1}^\infty$ satisfies the following conditions, as $n \to \infty$,

$$p_n \to 0, \beta_{1n} \to \infty, \beta_{0n} \to \beta_0 \in (0, \infty), \frac{\delta_{0n}}{\delta_{1n}} \to \delta \in (0, \infty), -\frac{\log \rho_n}{\sqrt{M_n}} \to A \in (0, \infty).$$

Starting from this assumption, we can easily derive the following useful result which indicate the convergence rates of probabilities of type I and type II errors.

**Theorem 4.1** Under Assumption (***), from (4–1) and (4–2), the probabilities of type I and type II errors using the Bayes oracle are given respectively by

$$t_{1n} \sim \frac{p_n \delta_{1n}}{(1 - p_n) \delta_{0n}} M_n^{-1/2} = O(p_n M_n^{-1/2})$$ and $t_{2n} \to 1 - \exp(-A)$.

Then we can define the Bayes risk of the Bayes rule (Bayes oracle), which has the minimal Bayes risk among all rules, in the following way.
**Definition 4.1** Under an additive loss function, the Bayes risk for a fixed threshold multiple testing procedure with respective probabilities $t'_{1n}$ and $t'_{2n}$ of type I and type II errors is given by

$$R' = \sum_{i=1}^{n} [(1 - p_n) t'_{1n} \delta_{0n} + p_n t'_{2n} \delta_{1n}] = n [(1 - p_n) t'_{1n} \delta_{0n} + p_n t'_{2n} \delta_{1n}].$$

Thus, the Bayes risk corresponding to Bayes oracle which we will refer to as the optimal risk is

$$R_{\text{optimal}} = n [(1 - p_n) t_{1n} \delta_{0n} + p_n t_{2n} \delta_{1n}].$$

(4–3)

In addition, we have Definition 4.2 to give a criterion to judge whether an arbitrary multiple testing procedure is good enough or not by comparing the Bayes risk of the procedure with the optimal one.

**Definition 4.2** For a sequence of parameters satisfying Assumption (**), a multiple testing rule is called asymptotically Bayes optimal under sparsity (ABOS) if its risk $R$ satisfies

$$\frac{R}{R_{\text{optimal}}} \to 1,$$

as $n \to \infty$, where $R_{\text{optimal}}$ is the optimal risk given by (4–3).

In the rest of this paper, ABOS will be referred to asymptotically Bayes optimal (or optimality) under sparsity up to the context.

Based on Definition 4.2, we have the following three theorems which show the ABOS of two general classes of multiple testing procedures.

The first class is of the fixed threshold multiple testing procedures.

**Theorem 4.2** For a fixed threshold multiple testing procedure $\{FTM\}$ defined by

reject $H_{0ni}$ if $X_{ni} \geq C^*_n$, $i = 1, \ldots, n$, where $C^*_n \geq 0$.

The corresponding probabilities of type I and type II errors are respectively given by

$$t^*_1 = \exp \left( -2 \sqrt{\frac{C^*_n}{\lambda_n}} \right) \quad \text{and} \quad t^*_2 = 1 - \exp \left( -2 \sqrt{\frac{C^*_n}{\lambda_n}} \right).$$

Theorem 4.3 provides necessary and sufficient conditions under which a fixed threshold multiple testing procedure is ABOS.
Theorem 4.3 A fixed threshold multiple testing procedure defined in Theorem 4.2 is asymptotically Bayes optimal under sparsity (ABOS) if and only if one of the following conditions is satisfied (a) \( C_n^* \sim C_n \); (b) \( C_n^* \sim \beta_{0,n} \left( \log \frac{\delta}{[1-\exp(-A)]\rho_n} \right)^2 \).

Proof: The risk of the multiple testing procedure \{FTM\} with fixed threshold \( C_n^* \) is

\[
R^* := n[(1 - \rho_n) t_{1n}^* \delta_{0n} + \rho_n t_{2n}^* \delta_{1n}].
\]

Then, by Theorem 4.1, \{FTM\} is ABOS if and only if

\[
\frac{R^*}{R_{\text{Joint}}} \to 1
\]

\[
\iff \frac{(1 - \rho_n) \delta_{0n} t_{1n}^* + \rho_n \delta_{1n} t_{2n}^*}{(1 - \rho_n) \delta_{0n} t_{1n} + \rho_n \delta_{1n} t_{2n}} \to 1 \iff \frac{t_{1n}^* + \rho_n \delta_{1n} t_{2n}^*}{t_{1n} + \rho_n \delta_{1n} t_{2n}} \to 1
\]

\[
\iff \frac{\delta}{1 - \exp(-A)} \frac{t_{1n}^*}{\rho_n} + \frac{t_{2n}^*}{1 - \exp(-A)} \to 1.
\]

Define \( L = \frac{\delta}{1 - \exp(-A)} \frac{t_{1n}^*}{\rho_n} + \frac{t_{2n}^*}{1 - \exp(-A)} \). There are five possible cases.

1. If \( C_n^* = o(C_n) \) and \( C_n^* \sim \beta_{0,n} \left( \log \frac{\delta}{[1-\exp(-A)]\rho_n} \right)^2 \), then

\[
\frac{\delta}{1 - \exp(-A)} \frac{t_{1n}^*}{\rho_n} = \exp \left( -\sqrt{\frac{C_n^*}{\rho_n}} \right) \frac{\delta}{1 - \exp(-A)} \to 1 \text{ and } \frac{t_{2n}^*}{1 - \exp(-A)} \to 0.
\]

Thus \( L \to 1 \).

2. If \( C_n^* = o(C_n) \) and \( C_n^* \sim \beta_{0,n} \left( \log \frac{\delta}{[1-\exp(-A)]\rho_n} \right)^2 \), then

\[
\frac{\delta}{1 - \exp(-A)} \frac{t_{1n}^*}{\rho_n} = \exp \left( -\sqrt{\frac{C_n^*}{\rho_n}} \right) \frac{\delta}{1 - \exp(-A)} \to 1 \text{ and } \frac{t_{2n}^*}{1 - \exp(-A)} \to 0.
\]

Thus \( L \to 1 \).

3. If \( C_n = o(C_n^*) \), then \( \frac{\delta}{1 - \exp(-A)} \frac{t_{1n}^*}{\rho_n} \to 0 \) and \( \frac{t_{2n}^*}{1 - \exp(-A)} \to \frac{1}{1 - \exp(-A)} \).

Thus \( L \to 1 \).

4. If \( C_n^* \sim C_n \), then \( \frac{\delta}{1 - \exp(-A)} \frac{t_{1n}^*}{\rho_n} \to 0 \) and \( \frac{t_{2n}^*}{1 - \exp(-A)} \to 1 \).

Thus \( L \to 1 \).

5. If \( C_n^* \sim rC_n \) where \( r \neq 1 \), then \( \frac{\delta}{1 - \exp(-A)} \frac{t_{1n}^*}{\rho_n} \to 0 \) and

\[
\frac{t_{2n}^*}{1 - \exp(-A)} \to \frac{1 - \exp(-\sqrt{r}A)}{1 - \exp(-A)}.
\]

Thus \( L \to \frac{1 - \exp(-\sqrt{r}A)}{1 - \exp(-A)} \neq 1 \).

Hence, \{FTM\} is ABOS if and only if either \( C_n^* \sim \beta_{0,n} \left( \log \frac{\delta}{[1-\exp(-A)]\rho_n} \right)^2 \) or \( C_n^* \sim C_n \). □
Remark: Theorem 4.3 gives a necessary and sufficient condition of achieving ABOS for a given fixed threshold multiple testing procedure in a simple form which only involves a ratio of two fixed thresholds. Unlike the above, the following Theorem 4.4 gives a sufficient condition of ABOS for a given random threshold procedure requiring a certain convergence rate for the difference of the two thresholds to zero.

**Theorem 4.4** Define a random threshold multiple testing procedure \( \{ RTM \} \) by

\[
\text{reject } H_{0ni} \text{ if } X_{ni} \geq \hat{C}_n, \ i = 1, \ldots, n,
\]

If for all \( \epsilon > 0 \),

\[
P \left( |\hat{C}_n - C_n| \geq \epsilon \right) = o(p_n), \quad (4-4)
\]

then \( \{ RTM \} \) is ABOS.

**Proof:** Note that by the model assumption,

\[
P \left( |\hat{C}_n - C_n| \geq \epsilon \right) = (1 - p_n)P_{H_{0av}} \left( |\hat{C}_n - C_n| \geq \epsilon \right) + p_nP_{H_{av}} \left( |\hat{C}_n - C_n| \geq \epsilon \right).
\]

By (4–4),

\[
P_{H_{0av}} \left( |\hat{C}_n - C_n| \geq \epsilon \right) = o(p_n) \quad (4–5)
\]

and

\[
P_{H_{av}} \left( |\hat{C}_n - C_n| \geq \epsilon \right) = o(1). \quad (4–6)
\]

Define the probabilities of type I and type II errors respectively by \( \hat{\imath}_{1ni} = P_{H_{0av}} \left( X_{ni} \geq \hat{C}_n \right) \) and \( \hat{\imath}_{2ni} = P_{H_{av}} \left( X_{ni} < \hat{C}_n \right) \).
We have, for all $\epsilon > 0$,

$$\hat{r}_{1n} = P_{\text{Hon}} \left( X_{ni} \geq \hat{c}_n - C_n + C_n \right)$$

$$= P_{\text{Hon}} \left( |X_{ni} \geq \hat{c}_n - C_n + C_n| \cap ||\hat{c}_n - C_n| \geq \epsilon \right)$$

$$+ P_{\text{Hon}} \left( |X_{ni} \geq \hat{c}_n - C_n + C_n| \cap ||\hat{c}_n - C_n| < \epsilon \right)$$

$$\geq P_{\text{Hon}} \left( |X_{ni} \geq C_n + \epsilon| \cap ||\hat{c}_n - C_n| < \epsilon \right)$$

$$\geq P_{\text{Hon}} \left( X_{ni} \geq C_n + \epsilon \right) - P_{\text{Hon}} \left( |\hat{c}_n - C_n| \geq \epsilon \right)$$

$$= P_{\text{Hon}} \left( X_{n1} \geq C_n + \epsilon \right) - P_{\text{Hon}} \left( |\hat{c}_n - C_n| \geq \epsilon \right)$$

and

$$\hat{r}_{1n} = P_{\text{Hon}} \left( X_{ni} \geq \hat{c}_n - C_n + C_n \right)$$

$$= P_{\text{Hon}} \left( |X_{ni} \geq \hat{c}_n - C_n + C_n| \cap ||\hat{c}_n - C_n| \geq \epsilon \right)$$

$$+ P_{\text{Hon}} \left( |X_{ni} \geq \hat{c}_n - C_n + C_n| \cap ||\hat{c}_n - C_n| < \epsilon \right)$$

$$\leq P_{\text{Hon}} \left( |\hat{c}_n - C_n| \geq \epsilon \right) + P_{\text{Hon}} \left( |X_{ni} \geq C_n - \epsilon| \cap ||\hat{c}_n - C_n| < \epsilon \right)$$

$$\leq P_{\text{Hon}} \left( |\hat{c}_n - C_n| \geq \epsilon \right) + P_{\text{Hon}} \left( X_{ni} \geq C_n - \epsilon \right)$$

$$= P_{\text{Hon}} \left( X_{n1} \geq C_n - \epsilon \right) + P_{\text{Hon}} \left( |\hat{c}_n - C_n| \geq \epsilon \right)$$

Similarly,

$$\hat{r}_{2n} = P_{\text{Hon}} \left( X_{ni} < \hat{c}_n - C_n + C_n \right)$$

$$= P_{\text{Hon}} \left( |X_{ni} < \hat{c}_n - C_n + C_n| \cap ||\hat{c}_n - C_n| \geq \epsilon \right)$$

$$+ P_{\text{Hon}} \left( |X_{ni} < \hat{c}_n - C_n + C_n| \cap ||\hat{c}_n - C_n| < \epsilon \right)$$

$$\geq P_{\text{Hon}} \left( |X_{ni} < C_n - \epsilon| \cap ||\hat{c}_n - C_n| < \epsilon \right)$$

$$\geq P_{\text{Hon}} \left( X_{ni} < C_n - \epsilon \right) - P_{\text{Hon}} \left( |\hat{c}_n - C_n| \geq \epsilon \right)$$

$$= P_{\text{Hon}} \left( X_{n1} < C_n - \epsilon \right) - P_{\text{Hon}} \left( |\hat{c}_n - C_n| \geq \epsilon \right)$$
Therefore, as $n \to \infty$, the ratio of the Bayes risk $\hat{R}_n$ to the optimal risk $R_{\text{optimal}}$ is

$$\frac{\hat{R}_n}{R_{\text{optimal}}} = \frac{\sum_{i=1}^{n} [(1 - p_n) \hat{t}_{1ni} \delta_{0n} + p_n \hat{t}_{2ni} \delta_{1n}]}{n[(1 - p_n) t_{1n} \delta_{0n} + p_n t_{2n} \delta_{1n}]}$$

$$\geq \frac{\frac{1}{n} \sum \hat{t}_{1ni} (1 - p_n) t_{1n} \delta_{0n} + \frac{1}{n} \sum \hat{t}_{2ni} (1 - p_n) t_{1n} \delta_{0n} + p_n t_{2n} \delta_{1n}}{P_{H_{\text{an}}} (X_{n1} \geq C_n) + \frac{1}{n} \sum P_{H_{\text{an}}} (|\hat{C}_n - C_n| \geq \epsilon)}$$

$$\times \frac{\frac{1}{n} \sum (1 - p_n) t_{1n} \delta_{0n}}{P_{H_{\text{an}}} (X_{n1} < C_n) + \frac{1}{n} \sum P_{H_{\text{an}}} (|\hat{C}_n - C_n| \geq \epsilon)}$$

$$\times \frac{p_n t_{2n} \delta_{1n}}{(1 - p_n) t_{1n} \delta_{0n} + p_n t_{2n} \delta_{1n}}$$

$$\to 1$$
which follows from (4–5), (4–6) and Theorem 4.1. Similarly,

\[
\hat{R}_n = \frac{\hat{R}_n}{R_{\text{optimal}}} = \frac{\sum_{i=1}^{n} \left[ (1 - p_n) \hat{t}_{1ni} \delta_{0n} + p_n \hat{t}_{2ni} \delta_{1n} \right]}{n \left[ (1 - p_n) t_{1n} \delta_{0n} + p_n t_{2n} \delta_{1n} \right]}
\]

\[
= \frac{1}{n} \sum_{i=1}^{n} \frac{(1 - p_n) t_{1n} \delta_{0n}}{t_{1n}} + \frac{1}{n} \sum_{i=1}^{n} \frac{p_n t_{2n} \delta_{1n}}{t_{2n}}
\]

\[
\leq \frac{P_{H_{0n1}} (X_{n1} \geq C_n - \epsilon) + \frac{1}{n} \sum P_{H_{0n}} (|\hat{C}_n - C_n| \geq \epsilon)}{P_{H_{0n1}} (X_{n1} \geq C_n)}
\]

\[
\times \frac{(1 - p_n) t_{1n} \delta_{0n}}{(1 - p_n) t_{1n} \delta_{0n} + p_n t_{2n} \delta_{1n}}
\]

\[
+ \frac{P_{H_{0n1}} (X_{n1} < C_n + \epsilon) + \frac{1}{n} \sum P_{H_{0n}} (|\hat{C}_n - C_n| \geq \epsilon)}{P_{H_{0n1}} (X_{n1} < C_n)}
\]

\[
\times \frac{p_n t_{2n} \delta_{1n}}{(1 - p_n) t_{1n} \delta_{0n} + p_n t_{2n} \delta_{1n}}
\]

\[
\rightarrow 1
\]

Thus, as \( n \to \infty \),

\[
\frac{\hat{R}_n}{R_{\text{optimal}}} \to 1.
\]

Hence, \( \{ RTM \} \) is ABOS. \( \square \)

Similar to Theorem 2.4 and 2.5, Theorem 4.3 and 4.4 provide useful tools which can be applied to prove the ABOS of any specific multiple testing procedure within either of these two classes.
CHAPTER 5
CONCLUSIONS

Multiple testing problems originated from multiple sources including genomics research and witnessed a dramatic development during the past eighteen years. We briefly described the importance and development of research on the multiple testing problem in Chapter 1. We reviewed the literature on multiple testing problems most relevant to our research, including papers of [1] Benjamini and Hochberg (1995), Bogdan, Chakrabarti, Frommlet and Ghosh (2011), [4] Efron and Tibshirani (2002), [5] Genovese and Wasserman (2002) and [8] Storey(2003).

In Chapter 2, for a statistical model of mixture exponential distributions, we constructed a multiple testing procedure based on Bayesian decision rule with an assumption of the parameter space. We derived the important quantities such as probabilities of type I and type II errors and the corresponding optimal Bayes risk. Then, we introduced the concept of the Asymptotic Bayes Optimality under Sparsity based on the optimal Bayes risk. Eventually we proved two general theorems about the Asymptotic Bayes Optimality under Sparsity for the fixed threshold multiple testing procedure and the random threshold multiple testing procedure. These two theorems provide useful tools to verify the optimality of other thresholding procedures.

Next, in Chapter 3, we extended the above work to investigate the relationship between the Asymptotic Bayes Optimality under Sparsity, the False Discovery Rate and the Bayesian False Discovery Rate. We proved a series of theorems to provide the conditions under which the above three measures can be equivalent. We proved a set of lemmas by applying previous results. These lemmas helped us to prove the Asymptotic Bayes Optimality under Sparsity of the Benjamini-Hochberg procedure under certain conditions, particularly for the exponential distribution. We also ran a simulation study.
Finally, in Chapter 4, we considered a Gamma prior as an alternative choice on the scale parameters. We defined the Asymptotic Bayes Optimality under Sparsity based on this model and derived a series of corresponding results.
REFERENCES


BIOGRAPHICAL SKETCH

Ke Li started his graduate study at Department of Statistics, University of Florida in the fall of 2008. He received his Ph.D. from the University of Florida in the summer of 2013.