# Adaptive step up tests for the false discovery rate (FDR) under independence and dependence

Inaugural-Dissertation

zur Erlangung des Doktorgrades
der Mathematisch-Naturwissenschaftlichen Fakultät
der Heinrich-Heine-Universität Düsseldorf

vorgelegt von


**Philipp Heesen**
aus Willich


Düsseldorf, Mai 2014

aus dem Institut für Mathematik
der Heinrich-Heine Universität Düsseldorf

Gedruckt mit der Genehmigung der
Mathematisch-Naturwissenschaftlichen Fakultät der
Heinrich-Heine-Universität Düsseldorf

Referent: Prof. Dr. Arnold Janssen
Korreferent: Prof. Dr. Helmut Finner

Tag der mündlichen Prüfung: 15. Oktober 2014

# Contents

**Abstract** **iii**

**Zusammenfassung** **iv**

**Acknowledgements** **v**

**Preface** **vi**

**1 Introduction** **1**

1.1 From hypotheses testing to multiple hypotheses testing . . . . . . 1

1.2 Model assumptions and statistical framework . . . . . . . . . . . 2

1.3 Error rates . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 7

1.4 Multiple testing procedures . . . . . . . . . . . . . . . . . . . . . 10

1.5 New ideas for adaptive step up tests . . . . . . . . . . . . . . . . 16

1.6 Outline . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 18

**2 Inequalities for the FDR and critical values** **20**

2.1 Dependence structures . . . . . . . . . . . . . . . . . . . . . . . 21

2.2 Central lemmas . . . . . . . . . . . . . . . . . . . . . . . . . . . 24

2.3 Applications of the central lemmas . . . . . . . . . . . . . . . . . 31

2.4 Asymptotic worst case FDR of step-up-down tests coming from a concave rejection curve . . . . . . . . . . . . . . . . . . . . . . . 42

2.5 Converse Benjamini Hochberg Theorem . . . . . . . . . . . . . . 49

2.6 Asymptotics of adaptive SU tests . . . . . . . . . . . . . . . . . . 51

**3 A new estimation concept for adaptive test procedures** **57**

3.1 The original Storey Estimator . . . . . . . . . . . . . . . . . . . 58

3.2 New estimation concept and new estimators . . . . . . . . . . . . 61

3.3 Integral type estimators . . . . . . . . . . . . . . . . . . . . . . . 69
3.4 Other estimators . . . . . . . . . . . . . . . . . . . . . . . . . . . . 73
3.4.1 Estimators of Zeisel, Zuk and Domany . . . . . . . . . . . 73
3.4.2 Estimator of Benjamini, Krieger and Yekutieli . . . . . . . 75
3.5 Estimation for a generalized error rate . . . . . . . . . . . . . . . . 76

**4 Generalized error rate control of adaptive SU tests 79**
4.1 Model assumptions . . . . . . . . . . . . . . . . . . . . . . . . . . . 80
4.2 Conditions for finite sample error rate control and the generalized Storey estimator . . . . . . . . . . . . . . . . . . . . . . . . . . . . 84
4.3 Dynamic adaptive SU tests . . . . . . . . . . . . . . . . . . . . . . . 93
4.4 Simulation . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 101

**5 FDR control of adaptive SU tests under dependence 103**

**6 Appendix 109**

**List of Abbreviations and Symbols 115**

**List of Figures 119**

**List of Tables 120**

**Bibliography 121**

**Erklärung 128**

## Abstract

The false discovery rate (FDR) is widely used and accepted as error criterion in multiple hypothesis testing and much effort has been done to develop multiple tests which control the FDR under different model assumptions. The FDR is often used when the number of false rejections of a multiple test is allowed to be a reasonable portion of all rejections. It is defined as the expected portion of false rejections among all rejections. The most famous multiple test with respect to the FDR is the linear step-up test of Benjamini and Hochberg [2] which controls the FDR under different assumptions, but still does not exhausts it and thus may have a lack of power. Therefore, the adaptive step-up test of Storey et al. [61] has been proposed which includes an estimation of the portion of true null hypotheses.

In the introductory first chapter we introduce the basic model assumptions for this thesis and give a brief summary of well known results. Chapter 2 provides three central lemmas from which we derive various results concerning the FDR of non adaptive step-up tests, the critical values of step-up tests with FDR control and the asymptotic FDR of adaptive step-up tests. We consider several dependence structures of the $p$-values, including independence, a reverse martingale structure, positive regression dependence and arbitrary dependence. In Chapter 3 we extend the results of Storey [59, 60] to $p$-values of null hypotheses whose marginal distributions are stochastically larger than the uniform distribution on $[0, 1]$ which may occur for one sided hypotheses. We motivate a new class of estimators for adaptive step-up tests and show that the common estimation of the portion of true null hypotheses is not appropriate in this case. In Chapter 4 we establish a new sufficient condition for finite sample FDR control of adaptive step-up tests under independence and we prove that a slightly modified estimator from Chapter 3 satisfies this condition. It turns out that the selection of the estimator for the adaptive step-up test may even be performed in a data dependent manner. A reasonable selection method is discussed in a practical guide. Chapter 5 is devoted to finite sample FDR control of adaptive step-up tests under a specific kind of block dependence.

## Zusammenfassung

Die "false discovery rate" (FDR) ist als Fehlerkriterium in der multiplen Hypothesentesttheorie weit verbreitet und akzeptiert. Es wurde viel Aufwand betrieben, um multiple Tests zu entwickeln, welche die FDR unter verschiedenen Modellannahmen kontrollieren. Die FDR wird oft verwendet wenn die Anzahl der falschen Ablehnungen eines multiplen Tests ein angemessener Anteil aller Ablehnungen sein darf. Sie ist definiert als der erwartete Anteil falscher Ablehnungen an allen Ablehnungen. Der bekannteste multiple Test in Bezug auf die FDR ist der lineare step-up Test von Benjamini und Hochberg [2], welcher die FDR unter verschiedenen Annahmen kontrolliert, sie aber nicht ausschöpft und somit einen Mangel an Güte haben kann. Deshalb wurde ein adaptiver step-up Test von Storey et al. [61] vorgeschlagen, der eine Schätzung des Anteils der wahren Nullhypothesen mit einbezieht.

Im einleitenden ersten Kapitel führen wir die grundlegenden Modellannahmen für diese Arbeit ein und geben eine kurze Zusammenfassung wohlbekannter Resultate. Kapitel 2 stellt drei zentrale Lemmata bereit, aus denen wir eine gewisse Anzahl von Resultaten herleiten, welche die FDR von nicht adaptiven step-up Tests, die kritischen Werte von step-up Tests mit FDR Kontrolle und die asymptotische FDR von adaptiven step-up Tests betreffen. Wir betrachten mehrere Abhängigkeitsstrukturen. Dazu gehören unabhängige $p$-Werte, eine Rückwärtsmartingalstruktur, positive Regressionsabhängigkeit der $p$-Werte und beliebig abhängige $p$-Werte. In Kapitel 3 erweitern wir die Resultate von Storey [59, 60] auf $p$-Werte von Nullhypothesen deren Randverteilungen stochastisch größer als die Gleichverteilung auf $[0, 1]$ sind. Dies kann bei einseitigen Hypothesen auftreten. Wir motivieren eine neue Klasse von Schätzern für adaptive step-up Tests und zeigen, dass die übliche Schätzung des Anteils der wahren Nullhypothesen in diesem Fall nicht angemessen ist. In Kapitel 4 erarbeiten wir eine neue hinreichende Bedingung für finite FDR Kontrolle von adaptiven step-up Tests unter Unabhängigkeit und wir beweisen, dass ein leicht modifizierter Schätzer aus Kapitel 3 dieser Bedingung genügt. Es stellt sich heraus, dass die Auswahl des Schätzers für den adaptiven step-up Test sogar datenabhängig erfolgen kann. Eine vernünftige Auswahlmethode wird in einem praktischen Leitfaden erörtert. Kapitel 5 ist der finiten FDR Kontrolle von adaptiven step-up Tests unter einer bestimmten Art der Blockabhängigkeit gewidmet.

# Acknowledgements

I am greatly indebted to my advisor Prof. Dr. Arnold Janssen and wish to express my thanks for his invaluable support, his guidance and supervision of this thesis and for many fruitful discussions during the time. Moreover, I wish to thank my colleagues at the chair for mathematical statistics and probability theory at the Heinrich-Heine-University Düsseldorf for their support, in particular Dr. habil. Markus Pauly.

I wish to express my thanks to Prof. Dr. Helmut Finner for introducing me to the field of multiple testing, many stimulating and fruitful conversations and for refereeing this thesis. Furthermore, I also wish to thank his team at the Deutsches Diabetes-Zentrum (DDZ).

The author wishes to thank the Deutsche Forschungsgemeinschaft (DFG) for financial support.

Finally, I wish to thank my family and Genia Korobschenko for their love and support.

# Preface

The Theory of multiple hypothesis tests has become a large area in mathematical statistics. We often notice the following fact: The larger the number of null hypotheses, the more difficult to make a reasonable conclusion about these null hypotheses. Therefore, the theory is based on the control of some type 1 error rate like the familywise error rate or the false discovery rate. The last one has been promoted by Benjamini and Hochberg [2] and is the focus of many recent publications. Currently, the publication of Benjamini and Hochberg [2] has been cited more than 23.000 times, see Google Scholar.

This thesis mainly splits into two parts. The first part (Chapter 2 and 5) provides new results for the false discovery rate of several multiple tests under different dependence structures. The second part (Chapter 3 and 4) particularly considers one sided hypotheses, where the distributions of $p$-values of true null hypotheses may be stochastically larger than the uniform distribution on the unit interval. Usually, distributions which lie deep inside the null hypotheses disturb the detection of false null hypotheses. It is particularly disturbed for adaptive multiple tests which incorporate an estimation of the number of true null hypotheses. This has already been mentioned by Dickhaus [11] and Pounds and Cheng [44]. Both try to improve the estimation of the number of true null hypotheses.

In this thesis, we will start with the basic theory of adaptive multiple tests and develop a new concept along the former considerations of Storey [59, 60] using the possible advantages of one sided null hypotheses. In some proofs we try to avoid some standard methods like so-called least favorable parameter configurations and non increasing testing procedures. Finally, we obtain adaptive multiple tests which work well under distributions which lie deep inside the null hypotheses. One could say, the deeper the better. Furthermore, in standard cases the new tests still behave like the often applied adaptive multiple test of Storey et al. [61]. As already mentioned, the larger the number of null hypotheses, the more difficult to make a reasonable conclusion. Therefore, the new multiple tests basically include a form of an adaptive reduction of the dimension. Obviously, a $p$-value of a true null hypothesis which tends to be large has a lower effect on the FDR than a $p$-value which is uniformly distributed on the unit interval. Thus, we try to include an estimation of the effective number of true null hypotheses which have an effect on the FDR.

Large parts of this thesis are based on the works of Heesen and Janssen [28, 29]. These works resulted from a joint work of the two authors. Each part contains essential ideas, aspects and work of both contributors. In the individual chapters, we go into detail and specify in which way these chapters rely on Heesen and Janssen [28, 29]. For some theorems, propositions and lemmas we explain where the corresponding statements may be found in Heesen and Janssen [28, 29]. Some statements presented in this thesis include no or minor changes and others include significant changes. This thesis and both works [28, 29] were supported by a project of the Deutsche Forschungsgemeinschaft (DFG). The topic of the project is "Signalerkennung in hochdimensionalen statistischen Modellen mit Anwendungen in den Lebenswissenschaften" (English: Signal Detection in high dimensional statistical models with applications in the Life Sciences).

A list of abbreviations and symbols is given at the end of this thesis.

# Chapter 1

# Introduction

## 1.1 From hypotheses testing to multiple hypotheses testing

Hypotheses testing is based on a statistical experiment $(\Omega, \mathcal{A}, \{P_\vartheta : \vartheta \in \Theta\})$, where the index set $\Theta$ of possible distributions may be parametric or non parametric. The parameter set $\Theta$ is divided into a null hypothesis $\mathcal{H} \subset \Theta$ and an alternative hypothesis $\mathcal{K} = \Theta \setminus \mathcal{H}$. Based on an observation $\omega \in \Omega$ coming from an underlying unknown distribution $P_{\vartheta^*}$, where $\vartheta^* \in \Theta$, we would like to decide whether $\vartheta^* \in \mathcal{H}$ or $\vartheta^* \in \mathcal{K}$ holds true for $\vartheta^*$. Since $\omega$ may be sampled from any considered distribution, we have to make a decision under uncertainty. $\mathcal{H}$ is called true if $\vartheta^* \in \mathcal{H}$ otherwise $\mathcal{K}$ is called true. Each decision may then lead to one of the following two errors. A decision for $\mathcal{K}$ when $\mathcal{H}$ is true is called type 1 error (or error of first kind). To overcome this problem, a specific level $\alpha \in (0,1)$ for the type 1 error is usually allowed and the decision rule is given by a so called hypothesis test $\phi : \Omega \to \{0,1\}$ which decides on $\mathcal{K}$ if $\phi(\omega) = 1$ and on $\Theta$ if $\phi(\omega) = 0$. The decision for $\Theta$ when $\mathcal{K}$ is true is called type 2 error (error of second kind). Then $\phi$ is called level $\alpha$ test when the probability of a type 1 error is bounded by $\alpha$ for all $\vartheta \in \mathcal{H}$, i.e. when

$$\sup_{\vartheta \in \mathcal{H}} P_\vartheta(\{\omega : \phi(\omega) = 1\}) \leq \alpha. \tag{1.1}$$

Thus, a decision for $\mathcal{K}$ when $\mathcal{H}$ is true could occur at most with probability $\alpha$ which is often chosen less than or equal to 0.1. Then one is often interested in tests which minimize the probability of a type 2 error for every $\vartheta \in \mathcal{K}$ over all level $\alpha$ tests.

Modern technical procedures enable the simultaneous measurement of high dimensional data. In the life sciences the Omics technologies like Genomics, Proteomics and Metabolomics are worth mentioning. For some diseases like diabetes it is assumed that a few genes may contribute to it. In order to identify these genes genome wide association studies are applied. Typically, a hypothesis is formulated for each gene which leads to a huge amount of null hypotheses. These null hypotheses have to be judged simultaneously. As we will see, this is a sensitive matter.

So let us switch to multiple hypotheses $\mathcal{H}_i \subset \Theta$ with corresponding alternatives $\mathcal{K}_i = \Theta \setminus \mathcal{H}_i$ and tests $\phi_i$, $i = 1, \ldots, n$. The following is based on the introductory story of Tukey [63] for the Higher Criticism concept. It can also be found in Donoho and Jin [13]. For now, let us assume that each test $\phi_i$ is an exact level $\alpha$ test for the single hypothesis test problem $\mathcal{H}_i$ versus $\mathcal{K}_i$. There often exists a $\vartheta \in \Theta$ such that every $\mathcal{H}_i$ is true and $P_\vartheta(\{\omega : \phi_i(\omega) = 1\}) = \alpha$ for $i = 1, \ldots, n$. Then the expected number of type 1 errors is simply given by

$$\mathbb{E}\left(\text{"number of type 1 errors"}\right) = n\alpha \tag{1.2}$$

which can be much larger than 1. Moreover, if the tests $\phi_i$ are also independent, then the probability of at least one type 1 error is given by

$$P(\text{"at least one type 1 error"}) = 1 - (1-\alpha)^n \underset{n\to\infty}{\longrightarrow} 1 \tag{1.3}$$

for fixed level $\alpha$. Thus, standard level $\alpha$ tests are often not appropriate to judge multiple hypotheses simultaneously and new concepts were developed. Before we get to that, we will first set up the statistical framework for this thesis.

For the sake of completeness, if the null hypotheses $\mathcal{H}_i$ are disjoint, then level $\alpha$ tests are known to be appropriate to judge the hypotheses simultaneously in the sense that the probability of at least one type 1 error is less than or equal to $\alpha$.

## 1.2 Model assumptions and statistical framework

Let us start with the following definition of the univariate stochastically larger property.

**Definition 1.1** (Stochastically larger)
*Let $X$ and $Y$ be real random variables with distributions $P_1$ and $P_2$ and distribution functions (df) $F_1$ and $F_2$, respectively. Then $X$ is called to be* **stochastically larger** *than $Y$ if and only if (iff) $F_1(t) \leq F_2(t)$ for all $t \in \mathbb{R}$. We also call $X$ to be stochastically larger than $P_2$ and $P_1$ to be stochastically larger than $P_2$ or write $X \overset{st}{\geq} Y$, $X \overset{st}{\geq} P_2$, $P_1 \overset{st}{\geq} P_2$ and $F_1 \overset{st}{\geq} P_2$ instead of saying that $X$ is stochastically larger than $Y$.*

Note that the above definition includes that a random variable $X$ or distribution $P$ is stochastically larger than itself, i.e. we have $X \overset{st}{\geq} X$ and $P \overset{st}{\geq} P$.

During this thesis we will consider different models which are all special cases of the following Basic Model. In particular we assume that every testing procedure is based on $p$-values, to be more precise, on the ordered $p$-values.

**Model 1.2** (Basic Model)
*Let $(\Omega, \mathcal{A}, \{P \in \mathcal{P}\})$ be a statistical experiment with a family of probability measures $\mathcal{P}$ on $\Omega$. Throughout, we assume that we have $n \in \mathbb{N}$ null hypotheses $\mathcal{H}_i$ which may be true or false and corresponding alternative hypotheses $\mathcal{K}_i, 1 \leq i \leq n$. The status of the null hypotheses, i.e. if they are true or false, is not fixed in advance but random. Moreover, we only observe the vector of p-values $p = (p_1, \ldots, p_n)$ of the corresponding null hypotheses $(\mathcal{H}_1, \ldots, \mathcal{H}_n)$. The p-values of this model are now constructed as follows. Consider the random vector*

$$(H_i, \zeta_i, \xi_i)_{i \leq n} : \Omega \longrightarrow \left(\{0,1\} \times [0,1]^2\right)^n. \tag{1.4}$$

*The random variable $H_i$ codes the occurrence of a true null hypothesis $\mathcal{H}_i$ when $H_i = 0$ and false null hypothesis when $H_i = 1$, respectively, for $1 \leq i \leq n$. For convenience, we will also talk about true and false p-values instead of true and false null hypotheses. $(\zeta_i)_{i \leq n}$ denotes the vector of possible true p-values and $(\xi_i)_{i \leq n}$ the possible false ones. Furthermore, let $H = (H_1, \ldots, H_n)$ be the vector which contains the status of the null hypotheses. At this point, the only distributional assumption of $(H_i, \zeta_i, \xi_i)_{i \leq n}$ is that the conditional marginal distributions $\mathcal{L}(\zeta_i | H = h)$, $h \in \{0,1\}^n$, of the possible true p-values are either stochastically larger than the uniform distribution on $[0,1]$ or are the uniform distribution on $[0,1]$ itself. The* **p-values** *are then given by*

$$p_i = (1 - H_i)\zeta_i + H_i \xi_i, \quad 1 \leq i \leq n, \tag{1.5}$$

*and the* **random number of true p-values** *is given by*

$$N_0 = \sum_{i=1}^{n} (1 - H_i). \tag{1.6}$$

*We denote by* $I_0 = \{i : H_i = 0\}$ *the random index set of true p-values and by* $I_1 = \{i : H_i = 1\}$ *the random index set of false ones. Moreover, let us define the vectors of true and false p-values*

$$p_{I_0} = (p_i : i \in I_0) \quad \textit{and} \quad p_{I_1} = (p_i : i \in I_1), \tag{1.7}$$

*respectively. Every subsequently considered testing procedure only relies on the* **empirical cumulative distribution function (ecdf)**

$$\hat{F}_n(t) = \frac{1}{n} \sum_{i=1}^{n} \mathbb{1}\{p_i \le t\}, \quad t \in [0,1], \tag{1.8}$$

*and thus only on the* **order statistics**

$$p_{1:n} \le p_{2:n} \le \ldots \le p_{n:n} \tag{1.9}$$

*of the p-values* $p_1, \ldots, p_n$*. Furthermore, we assume that* $\mathbb{E}(N_0) > 0$*.*

The random number of true null hypotheses $N_0$ goes back to Efron's two group model, cf. Efron et al. [14].

In the following, let us denote by $U(a, b)$ the **uniform distribution** on the interval $[a, b]$. We will simply say uniform distribution or uniformly distributed without mentioning the precise interval if we relate to the uniform distribution on $[0, 1]$.

**Remark 1.3**

(a) The Basic Model is a **two stage model**. In the first stage the true and false null hypotheses are determined by sampling $H$ from $P$. Then $(\zeta_i, \xi_i)_{i \le n}$ is sampled from the conditional distribution $\mathcal{L}(P|H)$ and for the $p$-value $p_i$ in the second stage, the corresponding true $p$-value is picked out of $(\zeta_i)_{i \le n}$ iff $H_i = 0$. Otherwise, the corresponding false $p$-value is picked out of $(\xi_i)_{i \le n}$. Note that given the first stage, the marginal distribution of the sampled $p$-value $p_i$ may also depend on all other $(H_j)_{j \ne i}$ and also on the other $p$-values.

(b) All error measures which will be considered in this thesis are based on the number of type 1 errors and would be controlled anyway if $\mathbb{E}(N_0) = 0$. The

considered error measures do not take into account the number of type 2 errors or directional errors.
(c) Basic Model 1.2 makes almost no assumption about the distribution of $(p_1, \ldots, p_n)$ and $(H_i, \zeta_i, \xi_i)_{i \le n}$. We will soon consider models with different dependence structures, marginal distributions of true and false $p$-values and distributions of the occurrence of true and false $p$-values. This will happen by specifying the set of probability measures $\mathcal{P}$. Although $(\Omega, \mathcal{A}, \mathcal{P})$ in Basic Model 1.2 is introduced as an arbitrary statistical experiment, let us think of $\mathcal{P}$ as a maximum possible model. Then all following models are submodels of Basic Model 1.2. As already said, for the true $p$-values marginal distributions $\mathcal{L}(\zeta_i | H = h) \overset{st}{\ge} U(0,1)$, $h \in \{0,1\}^n$, are considered. For the false $p$-values usually marginal distributions $\mathcal{L}(\xi_i | H = h) \overset{st}{\le} U(0,1)$ are considered.
(d) In the literature, the hypotheses $\mathcal{H}_i$ and $\mathcal{K}_i$, $i = 1, \ldots, n$, often may be regarded as subsets of some nonparametric space of probability measures $\mathcal{P}$. This does not work here in general since the **null hypotheses are randomly true or false**. But note that those models are included in the Basic Model. An example will be given below in Model 1.4 (b).

Basic Model 1.2 is a very general model which contains the most popular models of the FDR literature. These are the following:

**Model 1.4**

**(a) (i.i.d. mixture model with uniformly distributed true p-values)**
*Starting from Basic Model 1.2 assume that* $H$, $(\zeta_i)_{i \le n}$ *and* $(\xi_i)_{i \le n}$ *are jointly independent random vectors. Moreover, let* $H_1, \ldots, H_n$ *be i.i.d.* $\mathcal{B}(1, 1 - \pi_0)$ *Bernoulli distributed for some* $\pi_0 \in (0,1]$. *Let* $(\zeta_i)_{i \le n} = (U_i)_{i \le n}$, *with on* $(0,1)$ *i.i.d. uniformly distributed random variables* $U_1, \ldots, U_n$ *and let* $(\xi_i)_{i \le n}$ *be i.i.d. random variables which are distributed according to some distribution function* $F_1$ *on* $[0,1]$. *For every fixed distribution under the present model the p-values* $p_1, \ldots, p_n$ *are i.i.d. with* $p_i \sim F_{\pi_0, F_1}$ *and* $F_{\pi_0, F_1}(t) = \pi_0 t + (1 - \pi_0) F_1(t)$, $0 \le t \le 1$.
**(b) (Model, where the hypotheses are subsets of the parameter space)**
*Consider the multiple hypothesis testing problem* $(\Omega, \mathcal{A}, \{P_\vartheta : \vartheta \in \Theta\})$, $\mathcal{H}_i \subset \Theta$, $i = 1, \ldots, n$, *of Section 1.1. Let* $p_i$, $i = 1, \ldots, n$, *be p-values for the hypotheses* $\mathcal{H}_i$ *which fulfill the usual assumption* $P_\vartheta(p_i \le t) \le t$ *for all* $i$ *and all* $\vartheta \in \mathcal{H}_i$.

*Then for each fixed $\vartheta \in \Theta$ the p-values $p_1, \ldots, p_n$ can be represented as p-values in the Basic Model 1.2 by a specific probability measure $P \in \mathcal{P}$ for which $H_i = \mathbb{1}\{\vartheta \notin \mathcal{H}_i\}$, $i = 1, \ldots, n$, are deterministic.*
**(c) (Dirac-Uniform (DU) configuration)** *Starting from Basic Model 1.2 assume that $(H_i, \xi_i)_{i \leq n}$ is deterministic with $N_0 = n_0$ and $\xi_i = 0$, $i = 1, \ldots, n$. Moreover, let $\zeta_i$, $i = 1, \ldots, n$ be i.i.d. uniformly distributed. Then this model is called $DU(n, n_0)$ configuration.*

Model 1.4 (a) includes a Bayesian approach and is based on Efrons two group model, cf. Efron et al. [14], and has also been considered by Storey [59, 60], Storey et al. [61] and Genovese and Wasserman [22, 23]. The mixture model of Donoho and Jin [13] is also based on this model. Although Model 1.4 (a) and hence Basic Model 1.2 include a Bayesian model approach, we will exclusively work with frequentist methods during this thesis, i.e. without the knowledge of the prior distribution.

Model 1.4 (b) is the most popular model in the FDR literature and has, for instance, been considered by Finner and Roters [20], Finner et al. [17] and Dickhaus [11]. In the literature, there are usually various assumptions concerning the dependence structure of the $p$-values. In Model 1.4 (b), these are ignored for the moment. But later on we will introduce various dependence structures for the Basic Model 1.2. Dirac-Uniform configurations have been considered by Finner and Roters [20], Finner et al. [17], Finner and Gontscharuk [18] and Finner et al. [19] for instance.

The null hypotheses $\mathcal{H}_1, \ldots, \mathcal{H}_n$ of Basic Model 1.2 are judged by a multiple hypothesis test

$$\phi(p) = (\phi_1(p), \ldots, \phi_n(p)) : \Omega \to \{0, 1\}^n \tag{1.10}$$

based on $p$, where $\phi_i(p) : \Omega \to \{0, 1\}$ are single hypothesis tests for $\mathcal{H}_i$ versus $\mathcal{K}_i$. The multiple test has the common interpretation: reject $\mathcal{H}_i$ iff $\phi_i(p) = 1$. For simplicity, we will also talk about rejecting $p$-values if we mean rejecting the corresponding null hypotheses. Note that the single hypothesis tests may depend on the entire vector $p$ which will be the case for step-wise tests. Let us denote by **R the number of rejected hypotheses**. In the following, we consider multiple tests which reject the smallest $p$-values $p_{1:n}, \ldots, p_{R:n}$ since their null hypotheses are most significant. For these multiple tests it suffices to define $R$. Once $R$ is defined, the test is defined. For the sake of convenience we will only deal with $R$ in the following. As already mentioned, we will only work with $p$-values $p_1, \ldots, p_n$ instead of test statistics and all upcoming multiple tests are particularly based on the order statistics $p_{1:n} \leq \ldots \leq p_{n:n}$. Therefore, we will also write $R(p)$ to emphasize that $R$ is a function of $p$.

At this point, ties may cause a problem if $p_{R:n} = p_{R+1:n}$. Then it would not be clear which of the corresponding null hypotheses should be rejected and which ones should not be rejected. But this problem does not occur for the step-wise multiple tests considered below since they reject all null hypotheses corresponding to $p$-values $p_{i:n} \leq p_{R:n}$, $i = 1, \ldots, n$.

By applying the multiple test, the single hypothesis tests may cause several type 1 and type 2 errors simultaneously. As we mentioned, for a single hypothesis test it is appropriate to control the probability of a type 1 error. Otherwise, for a multiple test it seems to be appropriate to control properties of the unobservable number of type 1 errors. Throughout this thesis the number of type 1 errors, i.e. the number of false rejections, of a multiple test will be denoted by $V$. For the Basic Model 1.2 we then have

$$V = \sum_{i \in I_0} \mathbb{1}\{\text{“}\mathcal{H}_i \text{ is rejected”}\}. \tag{1.11}$$

Later we will give a nice and easy manageable representation of $V$ for step-wise testing procedures which will also be defined later. Moreover, let us define

$$R(t) = n\hat{F}_n(t) = \sum_{i=1}^{n} \mathbb{1}\{p_i \leq t\} \quad \text{and} \tag{1.12}$$

$$V(t) = \sum_{i \in I_0} \mathbb{1}\{p_i \leq t\} \tag{1.13}$$

for $0 \leq t \leq 1$. These are the quantities $R$ and $V$ for the multiple test which exactly rejects the $p$-values $p_i$ which are less than or equal to the fixed threshold $t$.

## 1.3 Error rates

As already mentioned, the control of the probability of a type 1 error for each single hypothesis test is not appropriate in multiple testing. Therefore, several other error concepts have been developed in the past. These error concepts are usually based on the random variables $V$ and $R$. Let us start with the best known error rates.

**Definition 1.5** (FWER, FDR, ENFR)
*(a) The* **familywise error rate (FWER)** *of a multiple test* $\phi$ *for a distribution* $P \in \mathcal{P}$ *is given by*

$$FWER_P(\phi) = P(V > 0) \tag{1.14}$$

*which is the probability of at least one type 1 error.*
*(b) The* **false discovery rate (FDR)** *of a multiple test* $\phi$ *for a distribution* $P \in \mathcal{P}$ *is given by*

$$FDR_P(\phi) = \mathbb{E}_P\left(\frac{V}{R \vee 1}\right), \tag{1.15}$$

*where* $a \vee b = \max(a, b)$*. Then the FDR is the expected portion of false rejections among all rejections of the multiple test.*
*(c) The* **expected number of false rejections (ENFR)** *of a multiple test* $\phi$ *for a distribution* $P \in \mathcal{P}$ *is given by*

$$ENFR_P(\phi) = \mathbb{E}_P(V). \tag{1.16}$$

*(d) To simplify notation, we write FDR (FWER, ENFR) instead of* $FDR_P(\phi)$ *(*$FWER_P(\phi)$*,* $ENFR_P(\phi)$*).*
*(e) The multiple test is said to control the FDR (FWER) by* $\alpha \in (0,1)$ *for the model* $\mathcal{P}$ *iff* $FDR_P(\phi) \leq \alpha$ *(*$FWER_P(\phi) \leq \alpha$*) holds for every distribution* $P \in \mathcal{P}$*. Moreover, the multiple test is said to control the ENFR by* $\tilde{\alpha} \in (0, n)$ *for the model* $\mathcal{P}$ *iff* $ENFR_P(\phi) \leq \tilde{\alpha}$ *holds for every distribution* $P \in \mathcal{P}$*. For abbreviation, let us say the multiple test controls the FDR (FWER, ENFR) by* $\alpha$ *if it is clear which model* $\mathcal{P}$ *is meant.*

The FWER is the oldest error rate for multiple hypothesis tests and is known to be very restrictive for large $n$. Multiple tests controlling the FWER often have a lack of power (i.e. the number of type 2 errors tends to be large).

To overcome this problem, Benjamini and Hochberg [2] promoted the FDR as error criterion in their pioneering work "Controlling the false discovery rate: a practical and powerful approach to multiple testing". See Benjamini [1] for the background of this paper. Nowadays, the FDR is widely used and accepted as error criterion and much effort has been done to develop multiple tests controlling it under different model assumptions. It is used when the number of false rejections $V$ is allowed to be a reasonable portion of the number of all rejections $R$. For instance, this is the case for exploratory data analysis. Genome wide association studies are often evaluated in an exploratory manner and the significant genes are judged again by a follow-up study. Benjamini and Hochberg also discussed the properties of the **positive FDR** which is defined as

$$pFDR_P = \mathbb{E}_P \left( \frac{V}{R \vee 1} \middle| R > 0 \right). \tag{1.17}$$

The Bayesian interpretation of this error rate has particularly been studied by Storey [60].

The ENFR is also called per family error rate, PFE for short. It has been particularly investigated by Scheer [54].

Note that the FDR is the expectation over the so called **false discovery proportion** $FDP = \frac{V}{R \vee 1}$. A main criticism of the FDR is that it says little about the actual distribution of the FDP. To be more precise, it is not guaranteed that a specific portion of the rejected $p$-values is correctly rejected with specific probability. Therefore, the $\gamma$-$FDP$ has been proposed which is defined by $\gamma$-$FDP = P(FDP > \gamma)$ for some $0 < \gamma < 1$. Clearly, $\gamma$-$FDP \leq \alpha$ ensures that at most $\gamma R$ hypotheses of the $R$ rejected ones are at least correctly rejected with probability $1 - \alpha$.

A less conservative alternative to the FWER is the $k$-$FWER$ for some integer $k \geq 1$. It is given by $k$-$FWER = P(V \geq k)$ and allows up to $k - 1$ false rejections. In comparison to the FDP based error rates it does not relate the number of false rejections to the number of all rejections.

The $\gamma$-$FDP$ and $k$-$FWER$ are only listed for the sake of completeness and are not addressed in this thesis. For a treatment of these error rates we refer the reader to Lehmann and Romano [34], Romano and Shaikh [46, 47], Romano and Wolf [48] and Döhler [12].

During this thesis we will mainly consider the FDR as error criterion and in parts the ENFR. Later we will also allow the control of a generalized error rate which has been introduced by Meskaldji et al. [39]. But the main examples for this generalized error rate will again be the FDR and ENFR.

## 1.4 Multiple testing procedures

This thesis mainly deals with step-up and adaptive step-up tests which are defined as follows.

**Definition 1.6** (SU and adaptive SU test)
*Let* $0 < \alpha_{1:n} \leq \alpha_{2:n} \leq \ldots \leq \alpha_{n:n} < 1$ *be a sequence of fixed critical values. Then the multiple test which results from*

$$R = R(p) = \max\{i : p_{i:n} \leq \alpha_{i:n}\} \tag{1.18}$$

*with* $\max \emptyset = 0$ *is called* **step-up test (SU test)**. *The test then rejects the null hypotheses corresponding to the p-values* $p_{1:n}, \ldots, p_{R:n}$*, for short, the test rejects the p-values* $p_{1:n}, \ldots, p_{R:n}$*. Moreover, if the critical values* $\alpha_{i:n}$*,* $i = 1, \ldots, n$*, are replaced by data dependent critical values* $0 < \hat{\alpha}_{1:n} \leq \ldots \leq \hat{\alpha}_{n:n} < 1$ *in (1.18) (i.e. if the critical values* $\hat{\alpha}_{i:n} = \hat{\alpha}_{i:n}(p)$ *are functions of p), then it is called* **adaptive step-up test (adaptive SU test)**.

**Remark 1.7**
(a) We will also use the representation $\hat{\alpha}_{i:n} = \hat{\alpha}_{i:n}((\hat{F}_n(t))_{t \geq \lambda})$ if we want to accentuate that the critical values may also be defined as function of $(\hat{F}_n(t))_{t \geq \lambda}$ for some tuning parameter $0 < \lambda < 1$. Observe that SU tests are permutation-invariant in the $p$-values. Thus, the data dependent critical values of the adaptive SU tests should also be permutation-invariant in the $p$-values and hence be based on the order statistics.
(b) The step-wise multiple tests exactly reject all null hypotheses corresponding to $p$-values $p_i \leq \alpha_{R:n}$ and $p_i \leq \hat{\alpha}_{R:n}$, respectively. This can alternatively serve as definition for the exact rejection procedure and the problems with reference to ties which are mentioned above do not occur anymore.
(c) A SU test with $\alpha_{n:n} \geq 1$ would always reject every null hypothesis. Thus, we exclude this case.

In their pioneering work, Benjamini and Hochberg [2] proposed a SU test and showed that it controls the FDR by $\alpha$ under an independence assumption. The so called **Benjamini Hochberg SU test (BH test)** is given by the critical values

$$\alpha_{i:n}^{BH} = \frac{i}{n}\alpha, \quad i = 1, \ldots, n. \tag{1.19}$$

It is also called linear step-up test and is based on the Simes test which has been introduced as test for the global null hypothesis and controls the FWER in the weak sense (i.e. $FWER \leq \alpha$ holds for fixed $N_0 = n$), see Simes [58]. For now, let us assume that $H$ is fixed and not random anymore, all $p$-values are independent and the true $p$-values are distributed according to the uniform distribution. Then Benjamini and Hochberg [2] showed

$$FDR \leq \frac{N_0}{n}\alpha. \tag{1.20}$$

Later, Benjamini and Yekutieli [5] and Finner and Roters [20] obtained that "$\leq$" can be replaced by "$=$" in (1.20). We will refer to this as **Benjamini Hochberg Theorem**. Moreover, Benjamini and Yekutieli [5] also showed that (1.20) still holds if $H$ is fixed, the true $p$-values are distributed according to the uniform distribution and the $p$-values are positively regression dependent on the subset of true $p$-values which will be defined later on.

By Benjamini and Yekutieli [5] it is well-known that (1.20) no longer applies for arbitrary dependent $p$-values. Under the assumption of fixed $H$, uniformly distributed true $p$-values and arbitrary dependence structure of $p_1, \ldots, p_n$, they showed that (1.20) applies for the SU test with critical values

$$\alpha_{i:n}^{BY} = \frac{i}{n\sum_{j=1}^{n}\frac{1}{j}}\alpha, \quad i = 1, \ldots, n. \tag{1.21}$$

Later, Blanchard and Roquain [7] obtained that these critical values may also be replaced by

$$\alpha_{i:n}^{BR} = \frac{\alpha}{n}\int_0^i x d\nu(x), \quad i = 1, \ldots, n, \tag{1.22}$$

where $\nu$ is an arbitrary probability measure on $(0, \infty)$ and (1.20) also holds for true $p$-values which are stochastically larger than the uniform distribution. They noted that the choice of $\nu(\{k\}) = \left(k\sum_{j=1}^{n}\frac{1}{j}\right)^{-1}$, $k = 1, \ldots, n$, just leads to the critical values (1.21). Further results for the FDR of specific multiple tests under dependence have been obtained by Blanchard and Roquain [7, 8], Farcomeni [15], Finner et al. [16, 17], Gontscharuk [24], Guo and Rao [25], Roquain and Villers [50, 49], and Sarkar [51, 53] for instance.

Moreover, Blanchard and Roquain [7] showed that two simple conditions are sufficient for FDR control, namely, a so called dependency condition and a self consistency condition. Basically, the dependency condition imposes a condition on a SU test for a fixed dependency structure. Then every multiple test which additionally fulfills the self consistency condition based on the critical values of the SU test exhibits FDR control. As rule of thumb, all $p$-values which are rejected by the multiple test have to be less than or equal to $\alpha_{R':n}$, where $R'$ is the number of rejections of the multiple test and $\alpha_{i:n}$, $i = 1, \ldots, n$, are the critical values of the SU test mentioned above. Each of these multiple tests reject not more null hypotheses than the SU test. The self consistency condition is very interesting for a reduction of the rejections of a SU test when FDR control is still desired. This may be the case if a follow up study can only be carried out for a certain small number of hypotheses or genes for instance. By the results of Blanchard and Roquain [7] it follows easily that we don't even need to reject the smallest $p$-values while still preserving FDR control as long as the self consistency is satisfied. In what follows, we pursue a different aim and try to increase the number of rejections.

Note that the consideration of fixed $H$ corresponds to a conditional expectation in the above definition of the FDR and is also a special case of random $H$.

In the literature, also other SU tests with deterministic critical values $0 < \alpha_{1:n} \leq \ldots \leq \alpha_{n:n} < 1$ have been considered. A common assumption for the critical values is that

$$i \mapsto \frac{\alpha_{i:n}}{i} \quad \text{is non-decreasing.} \tag{1.23}$$

Under (1.23) and the assumptions of the BH Theorem, Benjamini and Yekutieli [5, Theorem 5.3] showed that Dirac-Uniform configurations (without mentioning these explicitly) are least favorable for the FDR, i.e.

$$FDR_P \leq FDR_{DU(n,n_0)} \tag{1.24}$$

holds for all considered distributions $P$. The DU configuration provides an easily manageable upper FDR bound and proving $FDR_{DU(n,n_0)} \leq \alpha$, $1 \leq n_0 \leq n$ suffices to show FDR control under the setting of the BH Theorem.

It is easy to verify that (1.23) holds for critical values coming from a concave rejection curve which will be defined as follows.

**Definition 1.8**

*Let* $r : [0,1] \to [0,\infty)$ *be a continuous, concave and non decreasing function with*

$r(0) = 0$ *and* $r(x_0) = 1$ *for some* $x_0 < 1$. *We refer to* $r$ *as* **concave rejection curve**. *Moreover, let* $r^{-1}$ *be the left continuous inverse and* $n \in \mathbb{N}$ *be fixed, then*

$$\alpha_{i:n} = r^{-1}\left(\frac{i}{n}\right), \quad i = 1, \dots, n, \tag{1.25}$$

*defines a set of critical values for step-wise tests. The function* $r^{-1}$ *is also called* **critical value curve**.

**Remark 1.9**
(a) The assumption $r(x_0) = 1$ for some $x_0 < 1$ ensures $\alpha_{n:n} < 1$ which is reasonable for SU tests. This assumption is often replaced by the weaker assumption $r(1) \geq 1$. For this thesis, we mainly consider the more stringent definition of rejection curves.
(b) Furthermore, each of these rejection curves provides sequences of step-wise tests which can be analyzed asymptotically. In Section 2.4 we will establish the asymptotics of the worst case FDR of this sequence as fix point equation.

The critical values (1.19) of the BH test are given by the rejection curve $r^{BH}(t) = \frac{t}{\alpha}$, $t \in [0, 1]$, see Finner et al. [17] for instance. Moreover, Finner et al. [17] introduced the **Asymptotically Optimal Rejection Curve (AORC)** which is constructed to have $FDR = \alpha$ in an asymptotic Dirac-Uniform configuration. This addresses the fact that the predetermined FDR level $\alpha$ is not exhausted by the BH test if there is at least one false null hypothesis. For fixed $\alpha \in (0, 1)$ the AORC is given by

$$r_\alpha(t) = \frac{t}{t(1-\alpha) + \alpha}, \quad t \in [0, 1], \tag{1.26}$$

and leads to the critical values

$$\alpha_{i:n}^{AORC} = \frac{i\alpha}{n - i(1-\alpha)}, \quad i = 1, \dots, n. \tag{1.27}$$

Observe that the assumptions of Definition 1.8 are not fulfilled for the AORC since $\alpha_{n:n} = r_\alpha^{-1}(1) = 1$. There exist several adjustments of the AORC and critical values to overcome this. Finner et al. [17] consider the AORC on a specific interval, say $[0, t^*]$ with $t^* < 1$, and different continuations on $[0, 1]$ such that the concavity just holds or at least (1.23) is just fulfilled. Furthermore, Finner et al. [19] consider so called step-up-down tests with critical values

$$\alpha_{i:n}^{AORC} = \frac{i\alpha}{n + \beta_n - i(1-\alpha)}, \quad i = 1, \dots, n, \tag{1.28}$$

with $\beta_n > 0$. These are given by the rejection curve $r(t) = (1+\frac{\beta_n}{n})r_\alpha(t)$ which is a modification of the AORC for fixed $n$. Gavrilov et al. [21] previously discussed the **step-down test (SD test)** with critical values (1.28) given by $\beta_n = 1$. We will give a definition of SD tests and discuss some of these adjustments later in Section 2.3.

Scheer [54] uses the possibility to define step-wise multiple tests in terms of crossing points of the rejection curve and the ecdf $\hat{F}_n$. We will revert to this technique in some proofs for the asymptotics of the FDR.

As already seen in (1.20), the predetermined FDR level $\alpha$ is not exhausted by the BH test if there is at least one false $p$-value. In order to improve this, adaptive SU tests have been proposed. The best known of these is the **adaptive SU test of Storey et al. [61]** which is related to the results of Schweder and Spjøtvoll [55] in some way. Let $\lambda \in (0,1)$ be a tuning parameter which is often chosen close to 0.5 and let $a \wedge b = \min(a,b)$. The test is then based on the data dependent critical values

$$\hat{\alpha}_{i:n} = \left(\frac{i}{\hat{n}_0}\alpha\right) \wedge \lambda \tag{1.29}$$

with estimator

$$\hat{n}_0 = \hat{n}_0(\lambda) = n\frac{1-\hat{F}_n(\lambda)+\frac{1}{n}}{1-\lambda}. \tag{1.30}$$

For fixed $H$ the estimator $\hat{n}_0(\lambda)$ is an estimator for the number of true null hypotheses $N_0$. Storey et al. [61] showed that this test controls the FDR by $\alpha$ under the same assumptions as in Benjamini and Hochberg [2]. Adaptive SU tests of the form (1.29) with different estimators $\hat{n}_0$ for $N_0$ (or estimators for related terms) were often considered in the recent literature. Sometimes taking the minimum with $\lambda$ in (1.29) has been omitted for the definition of the critical values. We refer to adaptive SU tests with critical values of the form (1.29) as **adaptive SU tests of Storey type**, but for convenience, we will often just talk about adaptive SU tests.

A frequently used motivation for adaptive SU tests of Storey type with estimator $\hat{n}_0$ for $N_0$ under these assumptions is the following. If the predetermined level $\alpha$ in the BH test is replaced by the data dependent level $\frac{n}{\hat{n}_0}\alpha$, then

$$FDR \approx \frac{N_0}{n} \cdot \frac{n}{\hat{n}_0}\alpha \approx \alpha \tag{1.31}$$

would approximately hold. Another motivation based on Storey [61] is given in Chapter 3. This motivation will also hold for true $p$-values whose marginal distributions are stochastically larger than the uniform distribution.

In the recent literature, several sufficient conditions for finite sample FDR control of adaptive SU tests of Storey type have been developed, see Benjamini et al. [4], Sarkar [52] and Zeisel et al. [65] for instance. Moreover, much effort has been done to develop and analyze estimators of $N_0$ and related terms, see Benjamini and Hochberg [3], Benjamini et al. [4], Blanchard and Roquain [8], Celisse and Robin [9], Chen and Doerge [10], Dickhaus [11], Langaas et al. [32], Liang and Nettleton [36], Meinshausen and Rice [37], Pounds and Cheng [43, 44], Schweder and Spjøtvoll [55], Storey [59, 60], Storey et al. [61] and Zeisel et al. [65]. Some approaches are based on the choice of $\lambda$ for the Storey estimator (1.30), see Langaas et al. [32] for a short summary and also Liang and Nettleton [36].

When the critical values of a step-wise test coincide, i.e. when $\alpha_{1:n} = \ldots = \alpha_{n:n}$, then the test degenerates to a **single step test** which rejects a $p$-value $p_i$ iff $p_i \leq \alpha_{1:n}$. The most famous single step test is the **Bonferroni test** which is defined by $\alpha_{1:n} = \frac{\alpha}{n}$ and controls the FWER by $\alpha$. Furthermore, the **Šidàk test** is defined by $\alpha_{1:n} = 1 - (1-\alpha)^{\frac{1}{n}}$ and controls the FWER by $\alpha$ under an independence assumption. Finner and Gontscharuk [18] introduced adaptive versions of these tests, where $n$ is replaced by a Storey type estimator of the form

$$\hat{n}_0(\lambda, \kappa) = n \frac{1 - \hat{F}_n(\lambda) + \frac{\kappa}{n}}{1 - \lambda} \tag{1.32}$$

for some $0 < \lambda < 1$ and $\kappa > 0$. They were the first who showed that the FWER of these adaptive tests is controlled by $\alpha$ under an independence assumption for several choices of $\lambda$ and $\kappa$.

The adaptive and non adaptive SU tests yield a nice representation of the **number of false rejections** given by

$$V = \sum_{i \in I_0} \mathbb{1}\{p_i \leq \hat{\alpha}_{R:n}\}. \tag{1.33}$$

This equality easily follows by (1.11) and Remark 1.7. Many of the following proofs will use this representation.

## 1.5 New ideas for adaptive step up tests

Inference in multiple hypothesis testing is known to be more difficult than inference in single hypothesis testing. However, it also offers some possibilities of improvements which do not exist for single hypothesis testing. For instance, consider a one-sided single hypothesis testing problem, say $\{\vartheta \leq \vartheta_0\}$ versus $\{\vartheta > \vartheta_0\}$. If the real distribution lies deep inside the null hypothesis, then the actual probability of a type 1 error will often be much smaller than the allowed type 1 error level $\alpha$. In other words, the allowed type 1 error level $\alpha$ is often far from being exhausted.

The adaptive test of Storey et al. [61] already tries to improve the exhaustion of the predetermined allowed FDR level $\alpha$ since (1.20) holds for the BH test. However, if the marginal distributions of some true $p$-values are stochastically larger than the uniform distribution, then the adaptive test of Storey et al. [61] may perform badly in the sense, that it rejects less hypotheses than the BH test and that the allowed FDR level $\alpha$ is again far from being exhausted. In this case, even the actual FDR level of the BH test decreases. True $p$-values which are stochastically larger than the uniform distribution often occur in **one-sided hypotheses testing problems**. But now, in comparison to a single hypothesis test, there is more than one $p$-value available and the present information given by the $p$-values is beyond the information for the estimation of the number of true null hypotheses. In this thesis, it is shown how to use and incorporate this information into a new estimation concept to improve the exhaustion of the allowed FDR level $\alpha$ when the marginal distributions of some true $p$-values are stochastically larger than the uniform distribution. These improvements, which are the main results of Chapter 3 and 4, hold under a basic independence assumption and some slight regularity assumptions.

Pounds and Cheng [44] and Dickhaus [11] already mentioned the lack of exhaustion of the FDR in these cases and tried to improve the commonly conducted estimation (mostly for $N_0$ and related terms). Therefore, Dickhaus [11] considered randomized $p$-values which allow an improved estimation of $N_0$. But note that this randomization particularly loses the useful information on which our new estimation concept is based. Moreover, we will show that $N_0$ (or related terms like $\pi_0$) is not the parameter which should really be estimated for adaptive SU tests when stochastically larger true $p$-values are allowed. The estimation of the crucial parameter is quite more complex.

Note that for single hypothesis tests a better exhaustion of the allowed type 1 error level would only be desirable if the power of the tests got larger. Otherwise it would only produce more errors. Moreover, for the separability of the hypotheses, it is quite desirable that the type 1 error level is close to zero if the present distribution lies deeper inside the null hypothesis. In addition, it is desirable that the type 2 error level tends to zero if the present distribution goes deeper inside the alternative hypothesis. But observe that this behavior is not desirable for multiple hypothesis tests. One or more null $p$-values whose distributions lie deep inside the null hypotheses often disturb the detection of false $p$-values. Moreover, an increasing sample size $n$ already deteriorates this detection. Thus, one would like to have multiple tests with maximal detection power of false null hypotheses with simultaneous FDR control. Therefore, we will focus on the exhaustion of the predetermined FDR level. In some way, the new procedures include a reduction of the dimension $n$.

In the literature, it is often dealt with **least favorable parameter configurations (LFCs)** and non-increasing testing procedures. LFCs often provide an easily manageable upper FDR bound. However, this approach sometimes works against the idea of exhausting the allowed FDR level $\alpha$ which is the idea of adaptive multiple tests. By only considering LFCs one misses a chance of improvement of the exhaustion. For example, Blanchard and Roquain [8], Blanchard et al. [6], Finner and Roters [20], Finner et al. [17], Finner and Gotscharuk [18] Finner et al. [19], Etienne and Roquain [50] and Gontscharuk [24] worked with LFCs. In the literature, it turned out that LFCs are often given by Dirac-Uniform configurations.

Furthermore, **non-increasing testing procedures** sometimes miss possible improvements. A multiple testing procedure is called non-increasing if the mapping $p_i \mapsto R$ is non-increasing in each $p$-value $p_i$. Thus, for adaptive non-increasing multiple tests with critical values of the form (1.29), estimators $\hat{n}_0$ are considered which are non-decreasing, i.e. the mapping $p_i \mapsto \hat{n}_0$ is non-decreasing in each $p$-value. A common motivation for non-increasing multiple tests is the following. The larger a single $p$-value the less significant the corresponding hypothesis and the less $p$-values should be rejected. But this motivation is based on a one dimensional incomplete view of things. Observe that a multiple test usually rejects a number of small $p$-values. A consideration of all $p$-values thus yields that an excessive number of large $p$-values indicates that there are too few small true $p$-values in the area, where the $p$-values may be rejected. Hence we are able to reject more $p$-values by still controlling the FDR as the statements of Chapter 4 will show. The mapping $p_i \mapsto R$ should rather be non-increasing in each $p$-value only on the area, where a $p$-value really may be rejected. We will especially consider adaptive multiple tests which are not non-increasing in order to improve the exhaustion of the predetermined FDR level $\alpha$ over a wide range. Non-increasing testing procedures and non-decreasing estimators have, for instance, been considered by Benjamini et al. [4], Blanchard and Roquain [7, 8], Sarkar [52] and Zeisel et al. [65]. In particular, Benjamini et al. [4] and Blanchard and Roquain [7, 8] used that the uniform distribution is least favorable for the true $p$-values (among stochastically larger distributions) for the FDR of non-increasing (adaptive) SU tests. This is a relative easy way to show that these procedures work in these cases, but they do not work always well, as we will see.

## 1.6 Outline

Chapter 2 provides three central lemmas (Lemma 2.5, 2.9 and 2.10) from which we derive a certain amount of results for several dependence structures. Section 2.1 gives a definition of these dependence structures, including independence, a reverse martingale structure, positive regression dependence and arbitrary dependence. The central lemmas are stated in Section 2.2 and each of them gives attention to some of these dependence structures. Section 2.3 contains results concerning the FDR of non adaptive SU tests and the critical values of SU tests with FDR control. In Section 2.4 we establish the asymptotic worst case FDR of step-wise tests coming from a concave rejection curve as solution of a fix point equation and Section 2.5 provides a converse result of the BH Theorem. Moreover, Section 2.6 is devoted to the study of the asymptotic FDR of adaptive SU tests.

In Chapter 3 we extend the results of Storey [59, 60] to $p$-values whose marginal distributions are stochastically larger than the uniform distribution which may occur for one sided hypotheses. The often applied estimator of Storey (1.30) does not work well here. Therefore, we motivate a new class of estimators for adaptive SU tests and show that the common estimation of the number of true null hypotheses is not appropriate for the exhaustion of the FDR in this case. Moreover, we will discuss some other estimators.

In Chapter 4 we establish a new sufficient condition for finite sample FDR control of adaptive SU tests under independence which does not need the assumption of non decreasing estimators. Moreover, we prove that a slightly modified estimator from Chapter 3 satisfies this condition. It turns out that the selection of the estimator for the adaptive SU test may even be performed in a data dependent manner. A reasonable selection method with proven FDR control which yields a better exhaustion of the FDR and thus more power is discussed in a practical guide. In this chapter, an important assumption is the convexity of the marginal distribution functions of the true $p$-values. We will show that this is a rather weak assumption.

Chapter 5 is devoted to finite sample FDR control of adaptive SU tests under a specific kind of dependence. The $p$-values may line up into independent blocks, where the $p$-values within each block may form a reverse martingale.

Finally, the appendix contains some technical results which are used in the proofs of the previous chapters.

# Chapter 2

# Inequalities for the FDR and critical values

In this chapter, we provide a number of results for the FDR of several multiple tests under different dependence structures. These include independent true $p$-values, a reverse martingale dependency, $p$-values which are positively regression dependent on the subset of true $p$-values (PRDS), negatively regression dependent true $p$-values (NRDS) and an arbitrary dependency, see Section 2.1. In Section 2.2 we establish some powerful technical and central lemmas. The inequalities in these lemmas enable the development of inequalities for the FDR itself and for the critical values of SU tests which exhibit FDR control. This is part of Section 2.3. Furthermore, we represent the asymptotic worst case FDR of SU tests which come from a concave rejection curve as unique solution of a fix point equation, see Section 2.4. This fix point equation merely relies on the concave rejection curve. In Section 2.5 we derive a converse Benjamini Hochberg type Theorem and show that the BH test is the only SU test which is not influenced by the distribution of false $p$-values when uniformly distributed true $p$-values are present. In Section 2.6 we have a look at the asymptotics of adaptive SU tests. Adaptive tests of Storey type (1.29) are considered for reverse martingale dependent $p$-values (including independent ones) and another type of adaptive tests is proposed for arbitrary dependent $p$-values. The central lemmas easily establish a common sufficient condition for the estimators or distributions which ensures asymptotic FDR control.

The results of this chapter mainly rely on Heesen and Janssen [29]. In their work, they focused on uniformly distributed true $p$-values, whereas the present results often include true $p$-values whose marginal distributions are stochastically larger than the uniform distribution. However, these generalizations for non-adaptive multiple tests are usually clear by slight changes in the proofs. In particular, the notation has changed to treat all cases at once. For uniformly distributed true $p$-values it often suffices to condition under $N_0$, but for stochastically larger ones it is convenient to condition under $H$. Section 2.5 relies on Heesen and Janssen [28].

## 2.1 Dependence structures

The simplest dependence structure contains independent true $p$-values and the most results in the literature refer to this structure. In view of Basic Model 1.2, the status of a null hypothesis, i.e. if it is true or false, is random. Hence, our simplest dependence structure contains a conditional independence assumption.

**Model 2.1** (Basic Independence Model (BI) and generalized Basic Independence Model (gBI))
*Starting from Basic Model 1.2, assume that conditioned under $H = (H_1, \ldots, H_n)$ the true p-values $p_i$, $i \in I_0$, and the vector of false p-values $p_{I_1}$ are jointly independent. Note that $p_{I_1}$ is considered as one random variable, whereas $p_i$, $i \in I_0$, are considered as individual random variables in terms of independence.*
*(a) Conditioned under $H$ let*

$$p_i \sim U(0,1), \quad i \in I_0. \tag{2.1}$$

*Then we call this model* **Basic Independence Model (BI)***.*
*(b) Conditioned under $H$ let*

$$p_i \sim F_{0,H,i}, \quad i \in I_0, \tag{2.2}$$

*with a df satisfying $F_{0,H,i}(t) \leq t$, $t \in [0,1]$. Then we call this model* **generalized Basic Independence Model (gBI)***. Here each true p-value is allowed to have a different marginal distribution.*

For notational convenience the definition of the model above distinguishes between uniformly distributed true $p$-values and stochastically larger ones. As already seen in the introduction, the largest part of the FDR literature focuses on uniformly distributed true $p$-values, whereas only a small part also considers stochastically larger ones.

For two vectors $c = (c_1, \ldots, c_n)$ and $c' = (c'_1, \ldots, c'_n)$ with $c, c' \in [0,1]^n$ let $c \leq c'$ denote the component-by-component property $c_i \leq c'_i$, $i = 1, \ldots, n$. Moreover, let $\mathcal{B}([0,1]^n)$ denote the Borel sets of $[0,1]^n$, then a set $C \in \mathcal{B}([0,1]^n)$ is said to be decreasing iff $c' \in C$ and $c \leq c'$ imply $c \in C$. The next definition allows for more dependence between all $p$-values.

**Model 2.2** (Reverse martingale dependence, PRDS, NRDS)
*Starting from Basic Model 1.2 consider the following submodels.*
*(a) Conditioned under $H$ let*

$$\frac{\mathbb{1}\{p_i \leq t\}}{t}, \quad 0 < t \leq 1, \tag{2.3}$$

*be a reverse martingale for every $i \in I_0$ with respect to the reverse filtration $\mathcal{F}_t = \sigma\big(H, \mathbb{1}\{p_j \leq s\} : s \geq t, 1 \leq j \leq n\big)$. Then we call this model* **Reverse Martingale Model (RM Model)**.
*(b) For every decreasing set $C \in \mathcal{B}([0,1]^n)$, $h \in \{0,1\}^n$ and $i \in I_0$ let*

$$t \mapsto P(p \in C | p_i \leq t, H = h) \tag{2.4}$$

*be non-increasing. Then we call this model* **PRDS Model** *(positive regression dependent on the subset of true null hypotheses).*
*(c) Under the assumptions in (b), let (2.4) be non-decreasing and let $\mathcal{L}(\zeta_i | H = h) = U(0,1)$, $i = 1, \ldots, n$, for all $h \in \{0,1\}^n$, then we call this model* **NRDS Model** *(negative regression dependent on the subset of true null hypotheses).*

**Remark 2.3**
(a) Since $P(p_i \leq 1 | H = h) = 1$ obviously holds for every $1 \leq i \leq n$ and $h \in \{0,1\}^n$, the martingale property directly implies that $\mathcal{L}(p_i | H = h) = U(0,1)$ for every $i \in I_0$ under this condition.
(b) An interesting generalization of the RM Model would be a reverse super martingale model, i.e.

$$\mathbb{E}\left(\frac{\mathbb{1}\{p_i \leq t\}}{t} \Big| \mathcal{F}_s, H = h\right) \leq \frac{\mathbb{1}\{p_i \leq s\}}{s} \tag{2.5}$$

holds for $t < s$ and $i \in I_0$ with $H = h \in \{0,1\}^n$. This leads to convex marginal distribution functions for the true $p$-values which will be a crucial condition in

Chapter 4 for the gBI Model.

(c) The BI Model is a submodel of the gBI, RM, PRDS and NRDS Model and the gBI Model is a submodel of the PRDS Model. In this sense, the PRDS and NRDS Model are not comparable with the RM Model. But the intersection of the PRDS and RM Model are at least greater than the BI Model. It is easy to check that the maximal dependent uniform case, i.e. when $\zeta_1 = \ldots = \zeta_n \sim U(0,1)$ and $H_1 = \ldots = H_n = 0$, is included in both models.

(d) Assumption (2.4) in the definition of the PRDS Model is slightly weaker than the original assumption of Benjamini and Yekutieli [5]. Instead of (2.4) they demand

$$t \mapsto P(p \in C | p_i = t, H = h) \tag{2.6}$$

to be non-increasing in our notation and in terms of $p$-values. Originally they only considered deterministic $H$ and formulated (2.6) in terms of test statistics. It is well known that (2.6) implies (2.4), cf. Finner et al. [17] for instance. Nevertheless, we will call it PRDS Model.

In addition to identical true $p$-values the RM Model yields a rich class of other underlying distributions for the $p$-values. The next example occurs in multivariate extreme value theory and risk analysis when the $p$-values have a joint risk component.

**Example 2.4** (RM Model, cf. Heesen and Janssen [29] Example 2.1)
Let $X_1, \ldots, X_n, Y$ be independent, continuous and real-valued random variables, where $X_1, \ldots, X_n$ are i.i.d.. Moreover, let $Z_i = \max(X_i, Y)$, $i = 1, \ldots, n$ and $H(t) = P(Z_1 \leq t)$, $t \in \mathbb{R}$. The transformed $p$-values $p_i = H(Z_i)$ then fulfill (2.3) and the distribution of $p_1, \ldots, p_n$ is included in the RM Model.

**Proof**. Define

$$M_t^{(i)} = \frac{\mathbb{1}\{Z_i \leq t\}}{H(t)} \quad \text{for } t \in \{H > 0\},\ i = 1, \ldots, n.$$

For $n = 1$ it is well known that $M_t^{(1)}$ is a reverse martingale w.r.t. $\mathcal{G}_t = \sigma(M_s^{(1)} : s \geq t)$. We now prove that $M_t^{(1)}$ is a reverse martingale w.r.t. $\mathcal{F}_t = \sigma((M_s^{(j)})_{s \geq t} : j = 1, \ldots, n) = \sigma((\mathbb{1}\{Z_i \leq s\})_{s \geq t} : j = 1, \ldots, n)$ in the case of $n > 1$. Therefore let $t < s$. Obviously, $\mathbb{E}(M_t^{(1)} | \mathcal{F}_s) = 0 = M_s^{(1)}$ if $Z_1 > s$. Otherwise, $Z_1 \leq s$ implies $X_1 \leq s$ and $Y \leq s$. Thus, $\mathbb{1}\{Z_i \leq \tau\} = \mathbb{1}\{X_i \leq \tau\}$

holds for all $i \geq 2$ and $\tau \geq s$. By the independence of $Z_1$ and $X_2, \ldots, X_n$ it follows that

$$\begin{aligned}
\mathbb{E}\left(M_t^{(1)}|\mathcal{F}_s\right) &= \mathbb{E}\left(M_t^{(1)}|\mathcal{F}_s\right) \mathbb{1}\{Z_1 \leq s\} \\
&= \mathbb{E}\left(M_t^{(1)}|\mathbb{1}\{Z_i \leq \tau\} : \tau \geq s,\, i = 1, \ldots, n\right) \mathbb{1}\{Z_1 \leq s\} \\
&= \mathbb{E}\left(M_t^{(1)}|\mathbb{1}\{Z_1 \leq s\},\, \mathbb{1}\{X_i \leq \tau\} : \tau \geq s,\, i = 2, \ldots, n\right) \mathbb{1}\{Z_1 \leq s\} \\
&= \mathbb{E}\left(M_t^{(1)}|\mathbb{1}\{Z_1 \leq s\}\right) \mathbb{1}\{Z_1 \leq s\} \\
&= \mathbb{E}\left(M_t^{(1)}|\mathcal{G}_s\right) \mathbb{1}\{Z_1 \leq s\} \\
&= \mathbb{E}\left(M_t^{(1)}|\mathcal{G}_s\right) = M_s^{(1)}.
\end{aligned}$$

Finally, observe that the time change

$$u \mapsto M_{H^{-1}(u)}^{(i)} = \frac{\mathbb{1}\{H(Z_i) \leq u\}}{u}, \quad i = 1, \ldots, n,$$

preserves the reverse martingale property for the pseudo-inverse distribution function $H^{-1}$ and yields the desired distribution for the $p$-values. □

In the literature, conditional versions of the BI, gBI and PRDS Models have often been considered, see Benjamini and Hochberg [2], Benjamini and Yekutieli [5], Blanchard and Roquain [7], Finner and Roters [20] and Finner et al. [17] for instance. The conditional versions also correspond to deterministic $H$ which is a special case of the models presented above.

Note that a model which allows arbitrary dependent $p$-values is just given by Basic Model 1.2 and obviously includes all other models.

## 2.2 Central lemmas

Almost all results of this Chapter are based on the following central lemmas which yield equalities and inequalities depending on the present model which enable conclusions for interesting aspects of the FDR. The theory of this section for the BI Model, the RM Model, the PRDS Model with uniformly distributed true $p$-values and the NRDS Model is treated in Heesen and Janssen [29] Lemma 6.1 (a) and (b).

Let us start with the basic independence type models and the RM Model. The technique of Lemma 2.5 also provides a short proof of the Benjamini Hochberg Theorem, see (1.20).

**Lemma 2.5** (cf. Heesen and Janssen [29] Lemma 6.1 (a))
*Let* $0 < \hat{\alpha}_{1:n} \leq \ldots \leq \hat{\alpha}_{n:n} \leq \lambda < 1$ *be data dependent critical values given by measurable functions*

$$\hat{\alpha}_{i:n} = \hat{\alpha}_{i:n}((\hat{F}_n(t))_{t \geq \lambda}), \quad i = 1, \ldots, n. \tag{2.7}$$

*Moreover, introduce* $\hat{\alpha}_{0:n} = \hat{\alpha}_{1:n}$ *and* $\gamma(i) = n\hat{\alpha}_{i:n}$*. Then*

$$\mathbb{E}\left(\frac{V}{\gamma(R)}\right) = \frac{\mathbb{E}(N_0)}{n} \tag{2.8}$$

*holds for the SU test under the RM Model (including the BI Model) and "$\leq$" holds in (2.8) under the gBI Model.*

**Proof**. Let us first prove the assumption for the RM Model. Therefore, observe that the SU test can be represented by the reverse stopping time

$$\tau = \sup\{\hat{\alpha}_{i:n}, i = 1, \ldots, n \,:\, p_{i:n} \leq \hat{\alpha}_{i:n}\} \vee \hat{\alpha}_{1:n},$$

which is adapted to the reverse filtration $(\mathcal{F}_t)_{0<t\leq 1}$ and where $\sup \emptyset = 0$. Then every $p$-value $p_i \leq \tau$ is rejected and we have $\frac{V}{\gamma(R)} = \frac{V(\tau)}{\gamma(R(\tau))}$ by the definitions (1.12) and (1.13). Due to the measurability, conditioned under $\mathcal{F}_\lambda$ the critical values $\hat{\alpha}_{1:n}, \ldots, \hat{\alpha}_{n:n}$ may be considered as fixed. Thus, $\tau$ is a discrete stopping time w.r.t. the reverse martingale (2.3) for the period $\hat{\alpha}_{1:n} \leq t \leq \lambda$ under this condition. Furthermore, due to the construction of $\tau$ and $\hat{\alpha}_{0:n} = \hat{\alpha}_{1:n}$ we observe that $\hat{\alpha}_{R(\tau):n} = \tau$. Thus, by (2.3) and the discrete version of the optional stopping theorem we obtain

$$\begin{aligned}
\mathbb{E}\left(\frac{V}{\gamma(R)}\Big|\mathcal{F}_\lambda\right) &= \mathbb{E}\left(\frac{V(\tau)}{\gamma(R(\tau))}\Big|\mathcal{F}_\lambda\right) = \mathbb{E}\left(\frac{V(\tau)}{n\hat{\alpha}_{R(\tau):n}}\Big|\mathcal{F}_\lambda\right) \\
&= \frac{1}{n}\mathbb{E}\left(\frac{V(\tau)}{\tau}\Big|\mathcal{F}_\lambda\right) = \frac{1}{n}\mathbb{E}\left(\frac{V(\lambda)}{\lambda}\Big|\mathcal{F}_\lambda\right)
\end{aligned}$$

and again by (2.3), integration yields

$$\mathbb{E}\left(\frac{V}{\gamma(R)}\right) = \frac{1}{n}\mathbb{E}\left(\frac{V(\lambda)}{\lambda}\right) = \frac{1}{n}\mathbb{E}\left(\frac{V(1)}{1}\right) = \frac{\mathbb{E}(N_0)}{n}.$$

Let us now consider the gBI Model. Observe that conditioned under $H$ the index set of true $p$-values $I_0$ is fixed. Let $p^{(i)} = (p_1, \dots, p_{i-1}, 0, p_{i+1}, \dots, p_n)$ be the vector of $p$-values, where the $i$-th $p$-value is decreased to zero, $R^{(i)} = R(p^{(i)})$ and $\hat{\alpha}^{(i)}_{i:n} = \hat{\alpha}_{i:n}(p^{(i)})$. Then by Lemma 6.1 (a) of the appendix we have

$$\begin{aligned}\mathbb{E}\left(\frac{V}{\gamma(R)}\Big| H\right) &= \sum_{i\in I_0}\mathbb{E}\left(\frac{\mathbb{1}\{p_i \le \hat{\alpha}_{R:n}\}}{n\hat{\alpha}_{R:n}}\Big| H\right)\\ &= \sum_{i\in I_0}\mathbb{E}\left(\frac{\mathbb{1}\{p_i \le \hat{\alpha}^{(i)}_{R^{(i)}:n}\}}{n\hat{\alpha}_{R^{(i)}:n}}\Big| H\right) = (\star).\end{aligned}$$

Due to the conditional independence under $H$ Fubini's Theorem yields

$$(\star) = \sum_{i\in I_0}\mathbb{E}\left(\frac{F_{0,H,i}(\hat{\alpha}^{(i)}_{R^{(i)}:n})}{n\hat{\alpha}^{(i)}_{R^{(i)}:n}}\Big| H\right) \le \frac{N_0}{n}$$

with equality in the BI Model. The statement again follows by integration. □

**Remark 2.6**

(a) The data dependent critical values (2.7) only use the information of the ecdf $\hat{F}_n$ on $[\lambda, 1]$ and the adaptive test only rejects $p$-values smaller or equal to $\lambda$, where $\lambda$ is an arbitrary but fixed tuning parameter. In Chapter 4 we will consider adaptive SU tests which also have this estimation area and rejection area, say.

(b) The critical value $\hat{\alpha}_{0:n}$ may be set to zero without changing the adaptive SU test. Nonetheless, we set $\hat{\alpha}_{0:n} = \hat{\alpha}_{1:n}$ in order to avoid notational problems in the proofs when $R = 0$ occurs with positive probability.

(c) Storey et al. [61] already incorporated martingale arguments which have been outlined by Scheer [54].

**Remark 2.7** (BH theorem)

By Remark 6.2 (b) of the appendix and Fubini's Theorem the proof of the BH theorem now reduces to

$$\begin{aligned}\mathbb{E}\left(\frac{V}{R\vee 1}\Big| H\right) &= N_0\mathbb{E}\left(\frac{\mathbb{1}\{p_1 \le \alpha^{BH}_{R:n}\}}{R}\Big| H\right) = N_0\mathbb{E}\left(\frac{\mathbb{1}\{p_1 \le \alpha^{BH}_{R^{(1)}:n}\}}{R^{(1)}}\Big| H\right)\\ &= N_0\mathbb{E}\left(\frac{\alpha^{BH}_{R^{(1)}:n}}{R^{(1)}}\Big| H\right) = \frac{N_0}{n}\alpha\end{aligned}$$

in the BI Model, where $p_1$ is assumed to be a true $p$-value without restrictions and $R^{(1)} = R(0, p_2, \dots, p_n)$.

The technique of Remark 2.7 is also applicable to shorten the proof of the statement of Benjamini and Yekutieli [5, Theorem 5.3] that the Dirac-Uniform configurations are least favorable for the FDR of SU tests with critical values (1.23) under the model of the BH theorem. This result can actually be extended to the gBI Model with convex dfs of true $p$-values and is stated in the next lemma.

**Lemma 2.8**

*Let $0 < \alpha_{0:n} = \alpha_{1:n} \leq \ldots \leq \alpha_{n:n} < 1$ be deterministic critical values which fulfill (1.23) and consider the corresponding SU test. Furthermore, consider a distribution $P$ of the gBI Model with fixed $H = h \in \{0,1\}^n$, $N_0 = n_0$ and assume that the dfs of true p-values $F_{0,H,i}$, $i \in I_0$, in (2.2) are convex. Then*

$$FDR_P \leq FDR_{DU(n,n_0)} \tag{2.9}$$

*holds for the $DU(n, n_0)$ configuration.*

**Proof**. Define $R^{(i)} = R(p_1, \ldots, p_{i-1}, 0, p_{i+1}, \ldots, p_n)$ as in the proof of Lemma 2.5. Along the lines of the technique in Remark 2.7 observe that

$$\begin{aligned} \mathbb{E}_Q\left(\frac{V}{R \vee 1}\right) &= \sum_{i \in I_0} \mathbb{E}_Q\left(\frac{\mathbb{1}\{p_i \leq \alpha_{R:n}\}}{R}\right) = \sum_{i \in I_0} \mathbb{E}_Q\left(\frac{\mathbb{1}\{p_i \leq \alpha_{R^{(i)}:n}\}}{R^{(i)}}\right) \\ &= \sum_{i \in I_0} \mathbb{E}_Q\left(\frac{F_{0,H,i}\left(\alpha_{R^{(i)}:n}\right)}{R^{(i)}}\right) \end{aligned} \tag{2.10}$$

holds for the distributions $Q \in \{P, DU(n, n_0)\}$. Thus, it suffices to show

$$\mathbb{E}_P\left(\frac{F_{0,H,i}\left(\alpha_{R^{(i)}:n}\right)}{R^{(i)}}\right) \leq \mathbb{E}_{DU(n,n_0)}\left(\frac{F_{0,H,i}\left(\alpha_{R^{(i)}:n}\right)}{R^{(i)}}\right) \tag{2.11}$$

for $i \in I_0$. Let $1 \leq j < k \leq n$. By (1.23) and the convexity of $F_{0,H,i}$, it follows that

$$F_{0,H,i}(\alpha_{j:n}) \leq F_{0,H,i}\left(j\frac{\alpha_{k:n}}{k}\right) \leq \left(1 - \frac{j}{k}\right) F_{0,H,i}(0) + \frac{j}{k} F_{0,H,i}(\alpha_{k:n}), \tag{2.12}$$

where $F_{0,H,i}(0) = 0$. Hence, $j \mapsto \frac{F_{0,H,i}(\alpha_{j:n})}{j}$ is non decreasing. It is easy to check that

$$\mathcal{L}(R^{(i)}|P) \overset{st}{\leq} \mathcal{L}(R^{(i)}|DU(n, n_0)) \tag{2.13}$$

which now implies (2.11) by an alternative definition of the stochastically larger property, see Theorem 1.2.8 of Müller and Stoyan [41]. Without loss of generality let $I_0 = \{1, \ldots, n_0\}$, $i = 1$ and $U_2, \ldots, U_{n_0}$ be i.i.d. uniformly distributed random variables which fulfill $U_i \le p_i$ *a.s.* for $i = 2, \ldots, n_0$. From this and since $R^{(i)}$ is non increasing in each $p$-value it follows that $R(0, p_2, \ldots, p_n) \le R(0, U_2, \ldots, U_{n_0}, 0, \ldots, 0)$ holds almost surely. This finally implies (2.13) since $\mathcal{L}(R(0, U_2, \ldots, U_{n_0}, 0, \ldots, 0)|P) = \mathcal{L}(R^{(i)}|DU(n, n_0))$. □

The next lemma for the PRDS and NRDS Model only applies to deterministic critical values and we only obtain inequalities instead of equalities in the conditional uniform case.

**Lemma 2.9** (cf. Heesen and Janssen [29] Lemma 6.1 (b) and Meskaldji et al. [39] Corollary 3.9)
*Let* $0 < \alpha_{0:n} = \alpha_{1:n} \le \ldots \le \alpha_{n:n} < 1$ *be deterministic critical values and* $\gamma(i) = n\alpha_{i:n}$. *Then*

$$\mathbb{E}\left(\frac{V}{\gamma(R)}\right) \le \frac{\mathbb{E}(N_0)}{n} \tag{2.14}$$

*holds for the SU tests under the PRDS Model and "≥" holds in (2.14) for the NRDS model.*

**Proof**. First, let us consider the PRDS Model conditioned under $H = h = (h_1, \ldots, h_n) \in \{0, 1\}^n$. By (2.4) we have

$$\mathbb{E}\left(\frac{V}{\gamma(R)}\Big|H = h\right) = \sum_{i\,:\,h_i=0} \mathbb{E}\left(\frac{1\{p_i \le \alpha_{R:n}\}}{\gamma(R)}\Big|H = h\right)$$

$$= \sum_{i\,:\,h_i=0} \sum_{j=1}^{n} \frac{1}{\gamma(j)} P(R = j|p_i \le \alpha_{j:n}, H = h) P(p_i \le \alpha_{j:n}|H = h)$$

$$\le \sum_{i\,:\,h_i=0} \sum_{j=1}^{n} \frac{1}{n} P(R = j|p_i \le \alpha_{j:n}, H = h)$$

$$= \frac{1}{n} \sum_{i\,:\,h_i=0} \sum_{j=1}^{n} \left[P(R \ge j|p_i \le \alpha_{j:n}, H = h) - P(R \ge j+1|p_i \le \alpha_{j:n}, H = h)\right]$$

$$= \frac{1}{n} \sum_{i\,:\,h_i=0} P(R \ge 1|p_i \le \alpha_{1:n}, H = h)$$

$$+ \frac{1}{n} \sum_{i\,:\,h_i=0} \sum_{j=2}^{n} \left[P(R \ge j|p_i \le \alpha_{j:n}, H = h) - P(R \ge j|p_i \le \alpha_{j-1:n}, H = h)\right]$$

$$\le \frac{1}{n} \sum_{i\,:\,h_i=0} P(R \ge 1|p_i \le \alpha_{1:n}, H = h) = \frac{N_0}{n}$$

since $\{R \geq j\}$ is a decreasing set for every $j = 2, \ldots, n$. Now, integration yields the statement for the PRDS Model and the statement for the NRDS Model follows in the same way. Note that the true $p$-values in the NRDS Model follow a uniform distribution conditioned under $H$. Thus, the first inequality in the above formula becomes an equality. □

Except for a rescaling of $\gamma$, a deterministic view of $N_0$ and a weighting of the hypotheses (which is due to Blanchard and Roquain [7]), Lemma 2.9 coincides with Corollary 3.9 of Meskaldji et al. [39]. Their proof follows from Proposition 3.6 of Blanchard and Roquain [7] and can easily be extended to random $N_0$. In contrast, the proof of Lemma 2.9 is a more direct one which is based on the technique of Theorem 4.1 of Finner et al. [17]. For a further discussion see Remark 2.12 below.

The following lemma obtains an inequality for the Basic Model 1.2, where the $p$-values may have an arbitrary dependence structure. Therefore, we consider adaptive SU tests with data dependent critical values which are based on the deterministic critical values (1.22) introduced by Blanchard and Roquain [7]. To the best of the author's knowledge, such kind of data dependent critical values have not been considered before.

**Lemma 2.10** (cf. Heesen and Janssen [29] Lemma 6.1 (c))
*Let* $0 < \widetilde{\gamma}(0) = \widetilde{\gamma}(1) \leq \ldots \leq \widetilde{\gamma}(n)$ *be data dependent values given by measurable functions*

$$\widetilde{\gamma}(i) = \widetilde{\gamma}(i, (\hat{F}_n(t))_{0 \leq t \leq 1}), \quad i = 0, \ldots, n. \tag{2.15}$$

*Moreover, let* $\nu$ *be a probability measure on* $(0, \infty)$ *and define data dependent critical values via*

$$\hat{\alpha}_{i:n} = \left( \frac{\alpha}{n} \int_0^{\widetilde{\gamma}(i)} x d\nu(x) \right) \wedge \lambda, \quad i = 0, \ldots, n, \tag{2.16}$$

*where* $0 < \lambda < 1$*. Then*

$$\mathbb{E}\left( \frac{V}{\alpha \widetilde{\gamma}(R)} \right) \leq \frac{\mathbb{E}(N_0)}{n} \tag{2.17}$$

*holds for the SU test under the Basic Model 1.2 with arbitrary dependence structure.*

**Proof**. Similar as in the proof of Lemma 3.2 of Blanchard and Roquain [7] on page 988, rewriting $1/z = \int_z^\infty v^{-2}dv$, Fubini's Theorem and (2.16) yield

$$
\begin{aligned}
\mathbb{E}\left(\frac{V}{\alpha\widetilde{\gamma}(R)}\Big|H\right) &= \sum_{i\,:\,H_i=0}\mathbb{E}\left(\frac{\mathbb{1}\{p_i \le \hat{\alpha}_{R:n}\}}{\alpha\widetilde{\gamma}(R)}\Big|H\right)\\
&= \sum_{i\,:\,H_i=0}\mathbb{E}\left(\mathbb{1}\{p_i \le \hat{\alpha}_{R:n}\}\int_0^\infty v^{-2}\mathbb{1}\{v > \alpha\widetilde{\gamma}(R)\}dv\Big|H\right)\\
&= \sum_{i\,:\,H_i=0}\int_0^\infty v^{-2}\mathbb{E}\left(\mathbb{1}\{p_i \le \hat{\alpha}_{R:n}\}\mathbb{1}\{v > \alpha\widetilde{\gamma}(R)\}\Big|H\right)dv\\
&\le \sum_{i\,:\,H_i=0}\int_0^\infty v^{-2}\mathbb{E}\left(\mathbb{1}\left\{p_i \le \frac{\alpha}{n}\int_0^{\widetilde{\gamma}(R)} x d\nu(x)\right\}\mathbb{1}\{v > \alpha\widetilde{\gamma}(R)\}\Big|H\right)dv\\
&\le \sum_{i\,:\,H_i=0}\int_0^\infty v^{-2}P\left(p_i \le \frac{\alpha}{n}\int_0^{\frac{v}{\alpha}} x d\nu(x)\Big|H\right)dv\\
&\le \frac{\alpha N_0}{n}\int_0^\infty v^{-2}\int_0^{\frac{v}{\alpha}} x d\nu(x)dv\\
&= \frac{\alpha N_0}{n}\int_0^\infty x\int_{\alpha x}^\infty v^{-2}dv d\nu(x) = \frac{N_0}{n}.
\end{aligned}
$$

The last inequality follows since conditioned under $H$ the marginal distribution of each true $p$-value is stochastically larger than the uniform distribution. □

**Remark 2.11**

(a) The Benjamini Yekutieli [5] choice of $\nu(\{k\}) = \left(k\sum_{j=1}^n \frac{1}{j}\right)^{-1}$ in (2.16) leads to the adaptive critical values

$$\hat{\alpha}_{i:n} = \frac{\lfloor\widetilde{\gamma}(i)\rfloor}{n\sum_{j=1}^n \frac{1}{j}} \cdot \alpha, \quad i = 1, \ldots, n. \tag{2.18}$$

(b) In contrast to Lemma 2.5, Lemma 2.10 also applies to data dependent critical values which are functions of the entire ecdf $(\hat{F}_n(t))_{0\le t\le 1}$. For some dependence structures, observing one $p$-value allows to draw conclusions on other $p$-values. Assume that $n = N_0 = 2$ and $p_1 = 1 - p_2 = U \sim U(0,1)$ holds. Then observing $(\hat{F}_n(t))_{0.5\le t\le 1}$ suffices at least to reconstruct the complete ecdf $(\hat{F}_n(t))_{0\le t\le 1}$ if the distribution is known. Thus, the division into an estimation and rejection area for arbitrary dependent $p$-values would make no sense.

**Remark 2.12** (Family of generalized error rates)

Meskaldji et al. [39] and Meskaldji [38] introduced an entire family of generalized error rates. Each non decreasing function $\rho : \mathbb{N} \cup \{0\} \to (0, \infty)$ defines an error rate which is given by

$$\mathbb{E}\left(\frac{V}{\rho(R)}\right). \tag{2.19}$$

The main examples are the FDR, given by $\rho(i) = i \vee 1$, $i \in \mathbb{N} \cup \{0\}$, and the ENFR, given by $\rho \equiv 1$. Weighted versions of the FDR and ENFR are also listed by Meskaldji et al. [39]. For each error rate, they proposed a corresponding SU test with critical values

$$\alpha_{i:n} = \frac{\rho(i)}{n}\alpha, \quad i = 1, \ldots, n, \tag{2.20}$$

where $\rho$ is the same function as in the error rate (2.19). These SU tests control (2.19) by $\alpha$ under independent true $p$-values or the common PRDS assumption, see Remark 2.3 (d). Furthermore, they showed that a SU test with

$$\alpha_{i:n} = \frac{\alpha}{n}\int_0^{\rho(i)} x d\nu(x), \quad i = 1, \ldots, n, \tag{2.21}$$

based on (1.22), controls (2.19) by $\alpha$ under arbitrary dependence (basically under the assumptions of Lemma 2.10 with deterministic $N_0$). Thus, Lemma 2.9 and non data dependent versions of Lemma 2.5 for the gBI Model and Lemma 2.10 may also be proved by their results which are straightforward applications of the results of Blanchard and Roquain [7]. The representation of the results only differs in the rescaling $\gamma = \alpha\rho$. Note that $\rho(n) < \frac{n}{\alpha}$ is needed in our setting to ensure $\alpha_{n:n} < 1$ for (2.20) and (2.21). Furthermore, a simple rescaling of $\rho$ does not really change the properties of the error rate.

We will now use the central lemmas to derive finite and asymptotic results for the FDR of adaptive and non-adaptive multiple tests. In Chapter 4, we will come back to this generalized error concept and develop adaptive SU tests. In the following, we suggest the reader to mainly focus on the FDR and ENFR when we consider the error rate (2.19).

In a further work, Meskaldji et al. [40] treated the question of an optimal choice of the error rate, see also Meskaldji [38].

## 2.3 Applications of the central lemmas

We will now derive inequalities for the FDR under different models and inequalities for critical values which have overall FDR control under the BI Model. In this section we focus on SU tests with deterministic critical values

$$0 < \alpha_{0:n} = \alpha_{1:n} < \ldots < \alpha_{n:n} < 1. \tag{2.22}$$

Let us start with the following question.

**What can be said about the FDR of a SU test with critical values (2.22) under different models?**

The results of the following lemma are intuitive for the gBI Model but not clear in advance since the distribution of false $p$-values is arbitrary.

**Lemma 2.13** (cf. Heesen and Janssen [29] Lemma 3.1)
*Let $1 \le k \le n$, $0 < c < 1$ and consider the SU test with critical values (2.22).*
*(a) If $\alpha_{i:n} \le \frac{ic}{n}$ holds for all $1 \le i \le k$, then*

$$FDR \le \frac{\mathbb{E}(N_0)}{n} c + P(R > k) \tag{2.23}$$

*follows under the gBI, RM and PRDS Model (which all include the BI Model).*
*(b) Moreover, if $\alpha_{i:n} \le \frac{ic}{n}$, $i = 1, \ldots, n$, holds with strict inequality for a fixed $i = j$ and $P(R = j, V > 0) > 0$, then we obtain the strict inequality $FDR < \frac{\mathbb{E}(N_0)}{n} c$ at least under the BI Model.*
*(c) If $\alpha_{i:n} \ge \frac{ic}{n}$ holds for all $1 \le i \le n$, then*

$$FDR \ge \frac{\mathbb{E}(N_0)}{n} c \tag{2.24}$$

*follows under the RM and NRDS Model (which also include the BI Model).*
*(d) Moreover, under the assumption of (c), if in addition $\alpha_{j:n} > \frac{jc}{n}$ holds for some fixed $j \in \{1, \ldots, n\}$ with $P(R = j, V > 0) > 0$, then $FDR > \frac{\mathbb{E}(N_0)}{n} c$ holds at least under the BI Model.*

**Proof**. (a) Let $\gamma(i) = n\alpha_{i:n}$, $i = 0, \ldots, n$, and observe that $\frac{\gamma(j)}{j} \le c$ holds for all $1 \le j \le k$. Since $R = 0$ implies $V = 0$ we obtain

$$\begin{aligned} FDR &\le \mathbb{E}\left( \frac{V}{\gamma(R)} \frac{\gamma(R)}{R \vee 1} \mathbb{1}\{1 \le R \le k\} \right) + P(R > k) \\ &\le \mathbb{E}\left( \frac{V}{\gamma(R)} \right) c + P(R > k) \le \frac{\mathbb{E}(N_0)}{n} c + P(R > k) \end{aligned}$$

by Lemma 2.5 and 2.9 under the gBI, RM and PRDS Model.
(b) By (a) we already know $FDR \le \frac{c\mathbb{E}(N_0)}{n}$. Suppose "=" would hold in the

previous formula. We would then obtain $0 = \mathbb{E}\left(\frac{cV}{\gamma(R)} - \frac{V}{R\vee 1}\right)$ by Lemma 2.5. But notice that $\frac{c}{\gamma(i)} - \frac{1}{i\vee 1} \geq 0$ holds for $i = 1, \dots, n$. Thus,

$$\mathbb{E}\left(\frac{cV}{\gamma(R)} - \frac{V}{R\vee 1}\right) \geq \left(\frac{c}{\gamma(j)} - \frac{1}{j\vee 1}\right) \cdot P(R = j, V > 0) > 0$$

follows by our assumptions, a contradiction.
(c) Observe that $\frac{\gamma(i)}{i} \geq c$ holds for $1 \leq i \leq n$. Thus, Lemma 2.5 and 2.9 imply

$$FDR = \mathbb{E}\left(\frac{V}{\gamma(R)}\frac{\gamma(R)}{R\vee 1}\mathbb{1}\{R > 0\}\right) \geq \mathbb{E}\left(\frac{V}{\gamma(R)}\right) c \geq \frac{\mathbb{E}(N_0)}{n} c$$

under the RM and NRDS Model.
(d) Observe that $\frac{c}{\gamma(i)} - \frac{1}{i\vee 1} \geq (>) 0$ holds for $i = 1, \dots, n$, $(j)$ and the assertion follows in the same way as (b). □

**Remark 2.14**
Note that $P(R = j, V > 0) > 0$ and even $P(R = j) > 0$ can not be guaranteed in general. These assumptions depend strongly on the distribution of false $p$-values. Assume that $\xi_i \in (\alpha_{j:n}, \alpha_{j+1:n}]$, $i = 1, \dots, n$, and $N_0 < n$ almost surely, where $(\alpha_{j:n}, \alpha_{j+1:n}]$ is a non empty interval. It follows easily that $R = j$ can not occur.

The next proposition establishes some lower and upper bounds for the FDR of SU tests which are based on critical values (2.22).

**Proposition 2.15** (cf. Heesen and Janssen [29] Proposition 3.1)
*Consider the SU test with critical values (2.22). Then*

$$FDR \leq \frac{\mathbb{E}(N_0)}{n} \max_{1\leq i\leq n} \frac{n\alpha_{i:n}}{i} \tag{2.25}$$

*holds under the gBI, RM and PRDS Model (including the BI Model) and*

$$FDR \geq \frac{\mathbb{E}(N_0)}{n} \min_{1\leq i\leq n} \frac{n\alpha_{i:n}}{i} \tag{2.26}$$

*under the RM and NRDS Model (including the BI Model).*

**Proof**. Let $\gamma(i) = n\alpha_{i:n}$, $i = 0, \dots, n$, again and note that $R = 0$ implies $V = 0$. A direct application of Lemma 2.5 and 2.9 yields

$$\begin{aligned} FDR &= \mathbb{E}\left(\frac{V}{\gamma(R)}\frac{\gamma(R)}{R\vee 1}\mathbb{1}\{R > 0\}\right) \\ &\leq \mathbb{E}\left(\frac{V}{\gamma(R)}\right) \max_{1\leq i\leq n} \frac{n\alpha_{i:n}}{i} \leq \frac{\mathbb{E}(N_0)}{n} \max_{1\leq i\leq n} \frac{n\alpha_{i:n}}{i} \end{aligned}$$

under the gBI, RM and PRDS Model. (2.26) follows in the same way. □

With different methods Guo and Rao [26] already showed that (2.25) holds under the PRDS property of Remark 2.3 (d).

Under regularity assumptions it can be shown that the inequalities of Proposition 2.15 are asymptotically sharp, i.e. if the critical values are generated from a concave rejection curve which has to fulfill some regularity assumptions, then there exist sequences of distributions such that the asymptotic FDR coincides with the asymptotic bound. The lower bound (2.26) can actually be improved for a specific class of SU tests under a Dirac-Uniform model with martingale structure.

**Proposition 2.16** (cf. Heesen and Janssen [29] Remark 3.1)
*Consider the RM Model with deterministic* $H = h \in \{0,1\}^n$, $\xi_i = 0$, $i = 1,\dots,n$, *and* $N_0 = n_0$ *(including the* $DU(n, n_0)$ *configuration). Furthermore, assume a SU test with critical values (2.22) which fulfills the common requirement (1.23), then*

$$FDR \geq n_0 \frac{\alpha_{n+1-n_0:n}}{n+1-n_0}. \tag{2.27}$$

**Proof**. The result directly follows by a reinspection of the proof of Proposition 2.15 since $R \geq n - n_0 + 1$ holds on $\{V > 0\}$ for the present models. Thus, by (1.23) we obtain $\frac{\gamma(R)}{R \vee 1} \mathbb{1}\{R > 0\} \geq \frac{n\alpha_{n+1-n_0:n}}{n-n_0+1}$. □

We now move to the next question.

**What can be said about the critical values $\alpha_{i:n}$ when the FDR is controlled by $\alpha$ ($FDR \leq \alpha$) under different models for fixed sample size $n$?**

For this question, let us again consider SU tests with deterministic critical values (2.22) which fulfill the common assumption (1.23), i.e. $i \mapsto \frac{\alpha_{i:n}}{i}$ is non-decreasing. The next lemma gives necessary conditions for FDR control in the BI model.

**Lemma 2.17** (cf. Heesen and Janssen [29] Lemma 3.2)
*Consider the SU test given by critical values (2.22) with (1.23) and assume that* $FDR \leq \alpha$ *holds for all distributions in the BI Model.*
*(a) Then* $\alpha_{i:n} \leq \frac{i\alpha}{n+1-i}$ *follows for all* $1 \leq i \leq n$.

*(b) Furthermore, if $\frac{\alpha_{k:n}}{k} < \frac{\alpha_{k+1:n}}{k+1}$ holds for some $1 \leq k \leq n-1$, then $\alpha_{i:n} < \frac{i\alpha}{n+1-i}$ follows for all $i \leq k$.*

**Proof**. (a) Observe that $FDR_{DU(n,n_0)} \leq \alpha$ holds particularly for the $DU(n, n_0)$ configurations, $1 \leq n_0 \leq n$. Thus, by Proposition 2.16 we directly observe the result by setting $i = n - n_0 + 1$.

The statement may also be proved in a direct way. Therefore, by Lemma 2.5, (1.23) and by the notation of the proof of Proposition 2.15 we have

$$\alpha \geq FDR_{DU(n,n+1-i)} = \mathbb{E}_{DU(n,n+1-i)}\left(\frac{V}{\gamma(R)}\frac{\gamma(R)}{R \vee 1}\mathbb{1}\{R > 0\}\right) \tag{2.28}$$

$$\geq \mathbb{E}_{DU(n,n+1-i)}\left(\frac{V}{\gamma(R)}\right)\frac{\gamma(i)}{i} = \frac{n+1-i}{i}\alpha_{i:n} \tag{2.29}$$

for $1 \leq i \leq n$ since $R \geq i$ holds on $\{V > 0\}$ under the present model.
(b) The inequality in (2.29) becomes a strict inequality for $i \leq k$ since

$$P_{DU(n,n+1-i)}(R > k) \geq P_{DU(n,n+1-i)}(p_{n:n} \leq \alpha_{n:n}) > 0.$$

□

**Remark 2.18**
The necessary conditions of Lemma 2.17 naturally hold under the gBI, RM and PRDS Model since they include the BI Model. By the results of Section 2.2, we get no further necessary conditions for these models.

**Corollary 2.19** (cf. Heesen and Janssen [29] Corollary 3.1)
*Let the assumptions of Lemma 2.17 hold.*
*(a) Then $\alpha_{1:n} \leq \frac{\alpha}{n}$ follows.*
*(b) If $\alpha_{1:n} = \frac{\alpha}{n}$ holds, then the SU test is already a BH test, i.e. $\alpha_{i:n} = \frac{i}{n}\alpha$, $i = 1, \ldots, n$.*
*(c) If $\alpha_{1:n} = \frac{\beta}{n}$ holds for some $\beta \leq \alpha$, then $FDR \geq \frac{\mathbb{E}(N_0)}{n}\beta$ follows under the BI Model.*

**Proof**. (a) The statement is a special case of Lemma 2.17 (a).
(b) Let us assume that the SU test is no Benjamini Hochberg test. Thus, by (1.23) there exists some $1 < k < n$ with $\frac{\alpha_{k:n}}{k} < \frac{\alpha_{k+1:n}}{k+1}$. But Lemma 2.17 (b) then implies $\alpha_{1:n} < \frac{\alpha}{n}$ which contradicts our assumption.
(c) Observe that $\alpha_{1:n} = \frac{\beta}{n}$ and (1.23) imply $\alpha_{i:n} \geq \frac{i}{n}\beta$, $i = 1, \ldots, n$. The statement then follows by Lemma 2.13 (c). □

Let us now have a look at some explicit SU tests and discuss the necessary conditions which are introduced above.

**Example 2.20**
(a) Stepwise tests with critical values of the form

$$\alpha_{i:n} = \frac{i\alpha}{n + b - ia}, \quad i = 1, \dots, n, \tag{2.30}$$

for non-negative parameters $a, b$ with $1 - \alpha + \frac{b}{n} > a$ are frequently discussed in the literature, see Finner et al. [17, 19] and Gavrilov et al. [21] for instance. The requirement $\alpha_{n:n} < 1$ for all $n \in \mathbb{N}$ for SU tests initially implies $0 \le a \le 1 - \alpha$. By Lemma 2.17 a necessary condition for $FDR \le \alpha$ is given by $a \le b$. Moreover, if $a > 0$, then the stricter condition $a < b$ is necessary.
(b) Consider the adjusted critical values

$$\alpha_{i:n} = \frac{i\alpha}{n - i(1 - \alpha)}, \quad i = 1, \dots, k_n < n, \tag{2.31}$$

of the AORC, see (1.27). There are several possibilities for the choice of $\alpha_{i:n}$, $i = k_n + 1, \dots, n$, such that (1.23) stays true or that a corresponding rejection curve stays concave, cf. Example 3.2 in Finner et al. [17]. By them it is well known that a SU test with adjusted critical values (2.31) does not exhibit finite sample FDR control but asymptotic FDR control for the BI Model. By (a) we directly observe that finite sample FDR control can not hold since $a = 1 - \alpha$ and $b = 0$ do not fulfill the necessary conditions. Actually, the first critical value $\alpha_{1:n} = \frac{\alpha}{n-(1-\alpha)} > \frac{\alpha}{n}$ is too large to allow FDR control.

Our inequalities include a device for the choice of adequate parameters $a, b$ for the critical values (2.30). Proposition 2.21 offers an approach for the adjustment of $a$ for fixed parameter $b$. Below, we mainly restrict ourselves to adjustments under the BI Model. Several adjustments have been introduced by Finner et al. [17, 19] and Gontscharuk [24].

**Proposition 2.21** (cf. Heesen and Janssen [29] Proposition 3.2)
*Consider the SU test with critical values (2.30) with $a, b \ge 0$ and $a < 1 - \alpha + \frac{b}{n}$.*
*(a) Under the RM Model (including the BI Model) we have*

$$FDR = \frac{\mathbb{E}(N_0)}{n + b}\alpha + \frac{\mathbb{E}(V)}{n + b}a \tag{2.32}$$

*with "$\leq$" under the gBI and PRDS Model and "$\geq$" under the NRDS Model.*
*(b) Let $P$ be a distribution with $\xi_1 = \ldots = \xi_n = 0$, $\zeta_1, \ldots, \zeta_n \sim U(0,1)$ and deterministic $H$ with $N_0 = n_0$, then*

$$\mathbb{E}_P(V) \geq n_0 \alpha_{n+1-n_0:n} \tag{2.33}$$

*holds for the ENFR. In particular, $P$ may be a Dirac-Uniform configuration.*
*(c) Now let $a > 0$ since otherwise, the present SU test would be a BH test and let $b$ be fixed. Moreover, let $h(n_0, \alpha) = \mathbb{E}_{DU(n,n_0)}\left(V^{BH,\alpha}\right)$ denote the ENFR of the level $\alpha$ BH test and introduce $a_0$ as unique positive solution of*

$$\alpha = \max_{1 \leq n_0 \leq n} \left( \frac{\alpha n_0}{n+b} + a \cdot \frac{h(n_0, \alpha')}{n+b} \right), \tag{2.34}$$

*where $\alpha' = \frac{\alpha n}{n+b}$. Let $\mathcal{P}_{BI}$ be the set of all possible distributions of the BI Model for fixed $n$ and let $FDR_P(a,b)$ denote the FDR of the SU test with critical values (2.30). Then there exists a unique parameter $a_1 \in [0, \min(a_0, 1-\alpha+\frac{b}{n}))$ with*

$$\sup_{P \in \mathcal{P}_{BI}} FDR_P(a_1, b) = \alpha. \tag{2.35}$$

*The worst case FDR is strictly smaller (larger) than $\alpha$ for $a < a_1$, ($a > a_1$).*

**Proof**. (a) By Lemma 2.5 we obtain

$$\frac{\mathbb{E}(N_0)}{n} = \frac{1}{n\alpha}\mathbb{E}\left(\frac{V(n+b) - RVa}{R \vee 1}\right) = \frac{n+b}{n\alpha}FDR - \frac{a}{n\alpha}\mathbb{E}(V)$$

for the RM Model since $R = 0$ implies $V = 0$ and the assertion for the other models follows in the same way by Lemma 2.5 and 2.9 with "$\leq$" and "$\geq$".
(b) We observe

$$\mathbb{E}_P(V) = \sum_{i \in I_0} \mathbb{E}_P\left(\mathbb{1}\{p_i \leq \alpha_{R:n}\}\right) \geq \sum_{i \in I_0} \mathbb{E}_P\left(\mathbb{1}\{p_i \leq \alpha_{n+1-n_0:n}\}\right) = n_0 \alpha_{n+1-n_0:n}$$

since $I_0$ is deterministic under $P$ and since $R \geq n+1-n_0$ holds on $\{p_i \leq \alpha_{R:n}\}$ for all true $p_i$.
(c) Observe that the critical values fulfill (1.23). Hence, we may restrict ourselves to Dirac-Uniform configurations since the worst case FDR is given by one for some $1 \leq n_0 \leq n$, see (2.45) for instance. It follows that

$$a \mapsto \sup_{P \in \mathcal{P}_{BI}} FDR_P(a,b) = \max_{1 \leq n_0 \leq n} \left( \frac{n_0}{n+b}\alpha + \frac{\mathbb{E}_{DU(n,n_0)}(V)}{n+b} a \right) \tag{2.36}$$

is continuous and strictly monotonic increasing on $(0, 1-\alpha+\frac{b}{n})$ because the critical values do and hence $\mathbb{E}_{DU(n,n_0)}(V)$ does. Furthermore, we have

$$\sup_{P\in\mathcal{P}_{BI}} FDR_P(0,b) = \frac{n}{n+b}\alpha \le \alpha \quad \text{and} \quad \lim_{a\nearrow 1-\alpha+\frac{b}{n}} FDR_{DU(n,n)}(a,b) = 1$$

since $a_{n:n} \to 1$ for $a \nearrow 1-\alpha+\frac{b}{n}$. Finally, if $a_0 < 1-\alpha+\frac{b}{n}$ holds, then $\sup_{P\in\mathcal{P}_{BI}} FDR_P(a_0,b) > \alpha$ holds true because of (2.34), (2.36) and since $\alpha_{i:n} \ge \frac{\alpha' i}{n}$, $i = 1,\dots,n$, implies $h(n_0,\alpha') < \mathbb{E}_{DU(n,n_0)}(V)$, $n_0 = 1,\dots,n$. □

**Remark 2.22**

(a) The value $a_0$ in Proposition 2.21 (c) should be regarded as initial value for the search for $a_1$ which yields FDR control.

(b) The ENFR of the BH test under a Dirac-Uniform configuration is easy to compute and given by

$$h(n_0,\alpha) = \mathbb{E}_{DU(n,n_0)}\left(V^{BH}\right) = \sum_{i=1}^{n_0}(n-n_0+1)\binom{n_0}{i} i! \left(\frac{\alpha}{n}\right)^i, \tag{2.37}$$

cf. Finner and Roters [20] page 991. This ENFR can also be computed by the simple recursion $\mathbb{E}_{DU(n,1)}\left(V^{BH}\right) = \alpha$ and

$$\mathbb{E}_{DU(n,n_0)}\left(V^{BH}\right) = \frac{n_0\alpha}{n}\cdot\left(\mathbb{E}_{DU(n,n_0-1)}\left(V^{BH}\right) + n - n_0 + 1\right), \tag{2.38}$$

$n_0 = 2,\dots,n$, which just gives (2.37) by induction. The proof of the recursion is simple and short: Similar to Remark 2.7 we observe

$$\begin{aligned}\mathbb{E}_{DU(n,n_0)}\left(V^{BH}\right) &= n_0\mathbb{E}_{DU(n,n_0)}\left(\alpha_{R^{(1),BH}:n}\right) = \frac{n_0\alpha}{n}\mathbb{E}_{DU(n,n_0)}\left(R^{(1),BH}\right)\\ &= \frac{n_0\alpha}{n}\mathbb{E}_{DU(n,n_0)}\left(V^{(1),BH}+n-n_0\right)\\ &= \frac{n_0\alpha}{n}\mathbb{E}_{DU(n,n_0-1)}\left(V^{BH}+n-n_0+1\right)\end{aligned}$$

for the BH test, where $V^{(1),BH}$ is analogously defined as $R^{(1),BH}$.

(c) Note that Proposition 2.21 (b) also holds for critical values of the form (2.22) and there is no assumption about the dependence structure of the true $p$-values.

By similar methods, Scheer [54] basically obtained (2.32) for $a = 1-\alpha$ for the BI Model, cf. the proof of Theorem 3.2 in [54]. Therefore, he used a Lemma which utilizes rejection curves and martingale arguments and which is similar to the statement of Lemma 2.5 restricted to the BI model and non adaptive SU tests, cf. Lemma 3.6 in [54].

In the next proposition, we establish another adjustment of the critical values which may have some advantage in practice. The proposal relies on the following observation. Typically, the largest critical values in (2.30) are responsible for situations with $FDR > \alpha$, cf. Finner et al. [17]. For these reasons, we propose to bound the largest critical values as follows.

**Proposition 2.23** (cf. Heesen and Janssen [29] Proposition 3.3)
*Let $\epsilon > 0$ be a small constant and consider a SU test with critical values satisfying (1.23) and $\alpha_{1:n} < \frac{\alpha}{n}$. For fixed $1 \le k \le n$ we introduce new critical values*

$$\alpha_{i:n}^{(k)} = \min\left(\alpha_{i:n}, \frac{i}{k}\alpha_{k:n}\right), \quad i = 1, \ldots, n, \tag{2.39}$$

*and denote the FDR of the corresponding SU test by $FDR(\alpha_{\cdot:n}^{(k)})$. If there is a distribution $P \in \mathcal{P}_{BI}$ with $FDR_P(\alpha_{\cdot:n}) > \alpha + \epsilon$, then there exists some $1 \le k_0 < n$ with*

$$\sup_{P\in\mathcal{P}_{BI}} FDR(\alpha_{\cdot:n}^{(k)}) \le \sup_{P\in\mathcal{P}_{BI}} FDR(\alpha_{\cdot:n}^{(k_0)}) \le \alpha + \epsilon \tag{2.40}$$

*for all $k \le k_0$. Moreover, if $k_0$ is chosen to be maximal, then ">" holds in (2.40) for $k > k_0$.*

**Proof**. The critical values $\alpha_{i:n}^{(k)}$, $i = 1, \ldots, n$, are constructed so that (1.23) is just fulfilled. Hence, as in the proof of Proposition 2.21 (c) we may restrict ourselves to Dirac-Uniform configurations for worst case considerations. Introduce $V\left(\alpha_{\cdot:n}^{(k)}\right)$ as the number of false rejections of the specified SU test. Under the Dirac-Uniform configurations, it is easy to see that $V\left(\alpha_{\cdot:n}^{(k)}\right)$ is increasing in $k$ since the critical values do. It follows that

$$FDR_{DU(n,n_0)}\left(\alpha_{\cdot:n}^{(k)}\right) = \mathbb{E}_{DU(n,n_0)}\left(\frac{V(\alpha_{\cdot:n}^{(k)})}{n - n_0 - V(\alpha_{\cdot:n}^{(k)})}\right)$$

is increasing in $k$ since $x \mapsto \frac{x}{n-n_0-x}$ is increasing. Finally, observe that $k = 1$ yields a BH test with $FDR < \alpha$ which proves the assertion. $\square$

The modification (2.39) of the critical values has also been considered by Finner et al. [17, Example 3.2] for the special case of critical values coming from the AORC to construct a feasible SU test. Moreover, for this type of modification with some preselected $k$, Finner et al. [19] propose to increase the parameter $b$ of (2.30) with $a = 1 - \alpha$ in a further step to reduce the FDR below $\alpha$. In contrast to earlier work, Proposition 2.23 works for general critical values which fulfill (1.23) and $\epsilon$ can even be replaced by 0. In the worst case we have $k_0 = 1$ and the resulting SU test is simply a BH test.

**Example 2.24**

Let us consider the BI Model and the critical values

$$\alpha_{i:n} = \frac{i\alpha}{n + 1 - i(1 - \alpha)}, \quad i = 1, \ldots, n, \tag{2.41}$$

of Gavrilov et al. [21] which have been introduced for SD tests, see Definition 2.28. It is well-known that this SD test has finite sample FDR control, whereas this does not apply for the corresponding SU test. However, the necessary conditions for finite sample FDR control, shown in Lemma 2.17, are fulfilled. The conditions are necessary but not sufficient. In this case, our results give at least a meaningful lower bound for $FDR_{DU(n,n_0)}$ based on (2.27). Figure 2.1

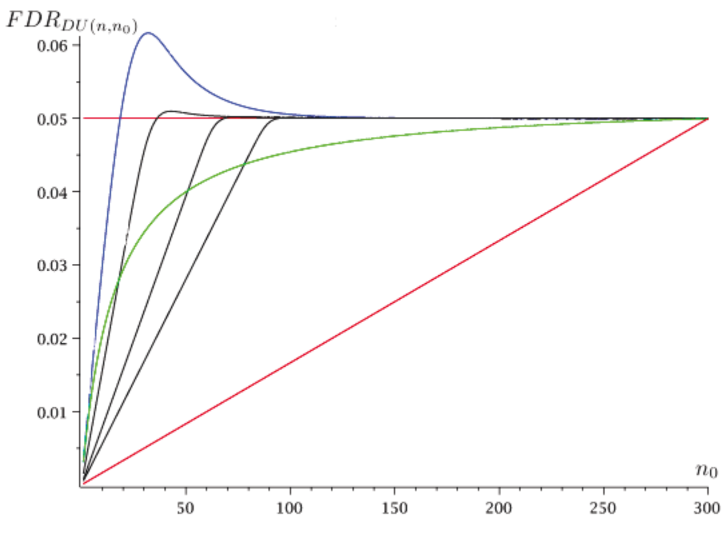



Figure 2.1: $FDR_{DU(n,n_0)}$ plot with $n = 300$ for (2.41) (blue curve), (2.39) with $k = 283,\ 250, 223$ based on (2.41) (black curves in decreasing order), BH test (increasing red straight line) and lower bound (2.26) (green curve).

shows $FDR_{DU(n,n_0)}$ of several SU tests for fixed $n = 300$ and different values of $n_0$. Note that the worst case FDR of these SU tests is just attained by one of those Dirac-Uniform configurations. The upper (blue) curve belongs to (2.41).

Observe that the lower bound (2.26) is based on

$$\min_{1\leq i\leq n} \frac{n\alpha_{i:n}}{i} = \frac{\alpha}{1+\frac{\alpha}{n}} \underset{n\to\infty}{\longrightarrow} \alpha. \tag{2.42}$$

Thus, the bound is close to $FDR_{DU(n,n_0)}$ for fixed $N \approx n$, as one can see in the picture. Moreover, the black curves are based on (2.39) for $k = 283, 250, 223$ and are plotted with graphs decreasing in $k$. A choice of $k = 300$ would just lead to the blue curve. The increasing straight line represents the FDR of the BH test and the green curve is the lower bound (2.26). The calculation of the blue and black curves have been done by using a program of the workgroup of Helmut Finner which is based on Bolshev's recursion, see Shorack and Wellner [56, page 366] for instance. Numerical results yield the value $k_0 = 283$ for $\epsilon = 10^{-3}$ and $k_0 = 223$ for $\epsilon = 10^{-4}$ given by (2.40), see also Table 2.1.

| $k$ | 300 | 283 | 250 | 223 |
|---|---|---|---|---|
| $\sup_{1\leq n_0\leq n} FDR_{DU(n,n_0)}$ | 0.06165 | 0.05098 | 0.05020 | 0.05009 |
| $\operatorname{argmax}_{1\leq n_0\leq n} FDR_{DU(n,n_0)}$ | 32 | 43 | 74 | 100 |

Table 2.1: Worst case FDR for different choices of $k$ in (2.39) based on (2.41).

The results given in Figure 2.1 are quite promising. A minor modification of the critical values (2.41) yields a FDR which is pretty much the predetermined FDR level $\alpha$ over a wide range. The $FDR_{DU(n,n_0)}$ is quite good for large and medium $n_0$, where the power of the multiple test is really needed. In a small area, it has a slightly increased FDR level $\alpha + \epsilon$ which is often accepted in practice. Basically, a further adjustment is not needed. Otherwise, a small reduction of the critical values (2.41) yields finite sample FDR control. This could be done by a reduction of $\alpha$ in (2.41) or by an increase of $b$ in (2.30) with $a = 1 - \alpha$. As already mentioned, the latter has been proposed by Finner et al. [19] for another preselection of $k$ in (2.39). As another approach, they considered a direct adjustment of $b$.

Note that these considerations have been done for Dirac-Uniform configurations and the actual FDR level for distributions of the BI Model can be smaller, cf. Finner et al. [17]. Moreover, if the true $p$-values get stochastically larger, then the exhaustion of the predetermined FDR level is arbitrarily bad. In Chapter 4, we will show how adaptive SU tests with critical values of the form (1.29) can be used to tackle this problem.

## 2.4 Asymptotic worst case FDR of step-up-down tests coming from a concave rejection curve

In this section we consider the worst case FDR asymptotics of a sequence of step-wise multiple tests which is based on a concave rejection curve $r$, see Definition 1.8. Recall, that the critical values of the step-wise tests are determined by (1.25). We will first consider sequences of SU tests under all possible distributions of the BI Models. The results also apply for a submodel of the gBI Model which is larger than the BI model and for step-down and step-up-down tests which are defined later on.

For our asymptotic considerations, let $V_n$ and $R_n$ denote the quantities $V$ and $R$ of the corresponding tests for fixed $n \in \mathbb{N}$. Note that $H = (H_1, \dots, H_n)$ and $N_0$ also depend on $n$ even if we do not spend an extra index.

The next theorem establishes the asymptotic worst case FDR of SU tests as solution of a fix point equation. This fix point equation only depends on the concave rejection curve which determines the critical values for the SU tests.

**Theorem 2.25** (cf. Heesen and Janssen [29] Theorem 3.1)
*Let $\mathcal{P}_{n,BI}$ be the set of all possible distributions of the BI Model for fixed $n \in \mathbb{N}$. Consider the sequence of SU tests given by a concave rejection curve $r$ via the critical values (1.25) and introduce the asymptotic worst case FDR as*

$$\beta = \limsup_{n\to\infty} \sup_{P_n \in \mathcal{P}_{n,BI}} FDR_{P_n}. \tag{2.43}$$

*Then $0 < \beta < 1$ and $\beta$ is the unique solution of the fix point equation $H(1-\beta) = \beta$, where*

$$H(s) = \sup\left\{ \frac{x}{1-x} \cdot \frac{1-r(x)}{r(x)} \ : \ 0 < x \text{ and } \frac{r(x)-x}{1-x} \le s \right\} \quad s \in (0,1). \tag{2.44}$$

**Proof**. We begin by proving the inequalities $0 < \beta < 1$. For now choose $n_0 = \lfloor \frac{n}{2} \rfloor$ and consider the $DU(n, n_0)$ configurations for each $n$. Observe that $\frac{R_n}{n} \ge \frac{n-n_0}{n} \ge \frac{1}{2}$ holds for the present model. Hence, at least every $p$-value $p_i \le r^{-1}(\frac{1}{2})$ will be rejected since $r^{-1}(\frac{1}{2}) \le r^{-1}(\frac{R_n}{n}) = \alpha_{R_n:n}$ holds. It follows that

$$\beta \ge \liminf_{n\to\infty} FDR_{DU(n.n_0)} \ge \liminf_{n\to\infty} \mathbb{E}_{DU(n.n_0)}\left(\frac{V_n}{n}\right) \ge \liminf_{n\to\infty} \frac{n_0}{n} r^{-1}\left(\frac{1}{2}\right) > 0.$$

Furthermore, observe that the concavity of $r$ implies (1.23) for the critical values. Hence,

$$\alpha_{i:n} \leq \frac{i}{n}\alpha_{n:n} = \frac{i}{n}r^{-1}(1) \quad \text{with } r^{-1}(1) < 1$$

holds and by Lemma 2.13 (a) it follows that $FDR_{P_n}$ is simultaneously bounded by $r^{-1}(1) < 1$ for all $P_n \in \mathcal{P}_{n,BI}$. Otherwise, this fact may also be proved by observing that $P_n(V_n = 0) \geq P_n(p_i > r^{-1}(1) : i \in I_0) > 0$ holds for every $P_n \in \mathcal{P}_{n,BI}$.

We proceed by proving that $\beta$ is a solution of $H(1-s) = s$. It is easy to check that this is the only solution since $s \mapsto H(1-s)$ is non-increasing. Let $P_n^H$ be the distribution of $H$ under $P_n$ and $P_n^{\cdot\,|\,H=h} = P_n(\,\cdot\,|H = h)$ be the conditional distribution given $H = h \in \{0,1\}^n$. Then by Lemma 2.8 (or Benjamini and Yekutieli [5, Theorem 5.3])

$$\begin{aligned} FDR_{P_n} &= \int FDR_{\left(P_n^{\cdot\,|\,H=h}\right)} dP_n^H(dh) \\ &\leq \sup_{h\in\{0,1\}^n} FDR_{\left(P_n^{\cdot\,|\,H=h}\right)} \leq \sup_{1\leq n_0\leq n} FDR_{DU(n,n_0)} \end{aligned} \tag{2.45}$$

holds for all $P_n \in \mathcal{P}_{n,BI}$ since the Dirac-Uniform configurations are least favorable for the SU test under the present conditional distributions. As $DU(n,n_0) \in \mathcal{P}_{n,BI}$ we obtain the representation

$$\beta = \limsup_{n\to\infty} \sup_{1\leq n_0\leq n} FDR_{DU(n,n_0)}$$

and thus, there exists a subsequence of $(n,n_0)_n = (n, n_{0,n})_n$, again denoted by $(n,n_0)_n$, which satisfies

$$FDR_{DU(n,n_0)} \underset{n\to\infty}{\longrightarrow} \beta \quad \text{and} \quad \frac{n_0}{n} \underset{n\to\infty}{\longrightarrow} 1-y \tag{2.46}$$

for some $1-y \in [0,1]$. Furthermore, since $V_n \leq n_0$ we have

$$FDR_{DU(n,n_0)} = \mathbb{E}\left(\frac{V_n}{n-n_0+V_n}\right) \leq \frac{n_0}{n}$$

and hence $1-y \in [\beta,1]$ by (2.46).

Let us first consider an arbitrary sequence of $(n,n_0)$ which fulfills $\frac{n_0}{n} \to 1-y \in [\beta,1)$ with positive $y$ and compute the limit FDR of the corresponding sequence of Dirac-Uniform configurations. In addition to (1.25) let us define $\alpha_{0:n} = r^{-1}(\frac{0}{n}) = 0$. Then

$$\hat{F}_n(\alpha_{R_n:n}) = \frac{R_n}{n} = r(\alpha_{R_n:n}) \tag{2.47}$$

follows by the subsequent considerations. The first equality directly follows by the definition of SU tests since the critical values are non decreasing. The second equality follows by (1.25) and the fact, that the concave rejection curve $r$ is invertible on $[0, \inf\{x : r(x) \geq 1\}]$. Let us introduce the straight line $g(t) = y + (1-y)t$, $t \in [0,1]$, which crosses the points $(0,y)$, $(1,1)$ and has the unique crossing point $(x, K)$, $0 < x, K < 1$, with $r$. Moreover, let $Z$ be a weak accumulation point of $\alpha_{R_n:n}$, i.e. there is a subsequence so that $\mathbb{E}(f(\alpha_{R_n:n})) \to \mathbb{E}(f(Z))$ for all bounded and continuous functions $f$. Then we have

$$r(Z) = g(Z) = y + (1-y)Z,$$

because $r$ is continuous and $\hat{F}_n$ converges uniformly to $g$ with probability 1. There is only one crossing point and thus $Z = x$ is constant for each weak accumulation point $Z$. By the above we now deduce

$$\frac{R_n}{n} = r\left(\alpha_{R_n:n}\right) \underset{n\to\infty}{\overset{a.s.}{\longrightarrow}} r(x) = K \tag{2.48}$$

at least along further subsequences. A simple geometric argument for the gradient of $g$ yields

$$\frac{1-y}{1} = \frac{r(x)-y}{x}$$

which easily gives

$$y = \frac{r(x)-x}{1-x} \quad \text{and} \quad 1-y = \frac{1-r(x)}{1-x}. \tag{2.49}$$

By $\frac{n_0}{n} \to 1-y \in [\beta, 1)$, (2.48), (2.49), subsequence arguments and dominated convergence we now obtain

$$\begin{aligned} \lim_{n\to\infty} FDR_{DU(n,n_0)} &= \lim_{n\to\infty} \mathbb{E}_{DU(n,n_0)} \left( \frac{\frac{R_n}{n} - \frac{n-n_0}{n}}{\frac{R_n}{n}} \right) \\ &= \frac{r(x)-y}{r(x)} = \frac{x}{1-x} \cdot \frac{1-r(x)}{r(x)}. \end{aligned} \tag{2.50}$$

In particular, for the subsequence in (2.47) we have $\beta = \frac{x}{1-x} \cdot \frac{1-r(x)}{r(x)}$.

We now turn to the case $\frac{n_0}{n} \to 1$. By the same arguments as above, but with unique crossing point $(x, K) = (0,0)$, it follows that $\frac{R_n}{n} \to 0$ a.s. at least along subsequences. For every $x' > 0$ let us define $z' = r(x') > 0$ and observe that

$$\frac{n\alpha_{j:n}}{j} \leq \frac{n\alpha_{\lfloor nz' \rfloor:n}}{\lfloor nz' \rfloor} = \frac{nr^{-1}\left(\frac{\lfloor nz' \rfloor}{n}\right)}{\lfloor nz' \rfloor} < 1$$

holds for all $j \leq \lfloor nz' \rfloor$ by (1.23) which is true for critical values from a concave rejection curve. Thus, by Lemma 2.13 (a) we obtain

$$\begin{aligned} FDR_{DU(n,n_0)} &\leq \frac{r^{-1}\left(\frac{\lfloor nz' \rfloor}{n}\right)}{\frac{\lfloor nz' \rfloor}{n}} \frac{n_0}{n} + P_{DU(n,n_0)}(R_n \geq \lfloor nz' \rfloor) \\ &\underset{n\to\infty}{\longrightarrow} \frac{r^{-1}(z')}{z'} = \frac{x}{r(x)} \end{aligned}$$

for all subsequences $\frac{n_0}{n} \to 1$ and hence

$$\limsup_{n\to\infty} FDR_{DU(n,n_0)} \leq \liminf_{x\searrow 0} \frac{x}{r(x)} = \liminf_{x\searrow 0} \frac{x}{r(x)} \cdot \frac{1-r(x)}{1-x} \tag{2.51}$$

since $\lim_{x\searrow 0} \frac{1-r(x)}{1-x} = 1$.

Finally, we have

$$\begin{aligned} H(1-\beta) &= \sup\left\{\lim_{n\to\infty} FDR_{DU(n,n_0)} \,:\, \frac{n_0}{n} \to 1-y \text{ with } \beta \leq 1-y < 1\right\} \\ &\leq \beta \leq H(1-\beta), \end{aligned}$$

where the equality follows by (2.50) if we set $y = \frac{r(x)-x}{1-x}$. Note that $\frac{r(x)-x}{1-x} = 0$ can only occur for $x = 0$ which is excluded in the formula above. The first inequality is obvious and the second inequality follows by considering the appropriate subsequence $(n, n_0)_n$ for which $\lim_{n\to\infty} FDR_{DU(n,n_0)} = \beta$ holds and by (2.50) and (2.51). □

**Remark 2.26**

Note that

$$\frac{x}{1-x} \cdot \frac{1-r_\alpha(x)}{r_\alpha(x)} = \alpha, \quad x \in (0,1), \tag{2.52}$$

holds for the AORC $r_\alpha$, defined in (1.26). Although the AORC is not feasible for Theorem 2.25, (2.50) supports its optimality for asymptotic Dirac-Uniform configurations.

In a series of theorems, Finner et al. [17, Theorem 5.1-5.4] computed asymptotic upper FDR bounds and limit FDRs for multiple tests based on the AORC under various assumptions on the limiting behavior of $\frac{R_n}{n}$. Some elements of the proof of Theorem 2.25 are similar to techniques which are used in this series of theorems. For finite $n$ and deterministic $N_0$ they also used that the asymptotic worst case FDR is attained by a sequence of Dirac-Uniform configurations. Moreover, from (2.48), (2.49) and Theorem 5.2 of Finner et al. [17] we may also deduce (2.50). But as we have seen in (2.50) itself, this result is easy to prove, at least in the special case of Dirac-Uniform configurations. In view of the worst case FDR, Theorem 2.25 pursues a consistent approach which pools all possible situations for SU tests. In fact, as we have seen, the limit of $\frac{n_0}{n}$ already determines the unique limit of $\frac{R_n}{n}$ for sequences of Dirac-Uniform configurations. Moreover, (2.48) and (2.50) may also be proved by Lemma 3.21 of Gontscharuk [24] and (2.49). This argument also applies to the case $\frac{n_0}{n} \to 1$. Similar arguments, as in the proof above, were also used by Scheer [54, Lemma 2.9] in his set up in order to prove that $\alpha_{R_n:n}$ converges to the crossing point $x$.

As the following corollary will show, Theorem 2.25 also holds under the submodel of the gBI Model, where the df's of true $p$-values are convex.

**Corollary 2.27**
*Let $\mathcal{P}^C_{n,gBI}$ be the set of all possible distributions of the gBI Model for fixed $n \in \mathbb{N}$ with convex dfs $F_{0,H,i}$ in (2.2). Under the assumptions of Theorem 2.25 we have*

$$\limsup_{n\to\infty} \sup_{P_n \in \mathcal{P}^C_{n,gBI}} FDR_{P_n} = \limsup_{n\to\infty} \sup_{P_n \in \mathcal{P}_{n,BI}} FDR_{P_n}. \tag{2.53}$$

**Proof**. Observe that (2.45) holds by Lemma 2.8 for every $P_n \in \mathcal{P}^C_{n,gBI}$ and it follows that "$\leq$" holds in (2.53). Furthermore, "$\geq$" obviously holds since $\mathcal{P}_{n,BI} \subset \mathcal{P}^C_{n,gBI}$. □

The results of Theorem 2.25 and Corollary 2.27 can actually be extended to step-up-down tests (SUD tests). We will point out that the asymptotic bound $\beta$ is the same for sequences of SUD tests with critical values which are generated from the same rejection curve. But first, let us give a definition.

**Definition 2.28** (Step-up-down (SUD) test)
*Let $0 < \alpha_{1:n} \leq \alpha_{2:n} \leq \ldots \leq \alpha_{n:n} < 1$ be a sequence of fixed critical values and $\lambda \in \{1, \ldots, n\}$. Then the multiple test defined by*

$$R_{SUD(\lambda),n} = R_{SUD(\lambda),n}(p)$$

$$= \begin{cases} \max\{i \in \{\lambda, \ldots, n\} : p_{j:n} \leq \alpha_{j:n} \text{ for all } \lambda \leq j \leq i\}, & \text{if } p_{\lambda:n} \leq \alpha_{\lambda:n}, \\ \max\{0, i \in \{1, \ldots, \lambda - 1\} : p_{i:n} \leq \alpha_{i:n}\}, & \text{if } p_{\lambda:n} > \alpha_{\lambda:n} \end{cases}$$

*is called* **step-up-down test** ($SUD(\lambda)$ **test**). *As already mentioned, the test then rejects the p-values* $p_{1:n}, \ldots, p_{R_{SUD(\lambda)}:n}$. *The* $SUD(1)$ *test is called* **step-down test (SD test)** *and the* $SUD(n)$ *test is the already defined SU test.*

The following lemma extends the results of Finner et al. [19, Theorem 2] if one of the SUD tests is a SU test and is needed in the announced theorem.

**Lemma 2.29**

*Let* $0 < \alpha_{0:n} = \alpha_{1:n} \leq \ldots \leq \alpha_{n:n} < 1$ *be deterministic critical values which fulfill (1.23) and consider the corresponding* $SUD(\lambda)$ *and SU tests for some* $\lambda \in \{1, \ldots, n\}$. *Furthermore, let* $P$ *be a distribution of the gBI Model for fixed* $n$ *and assume that the dfs of true p-values* $F_{0,H,i}$, $i \in I_0$, *in (2.2) are convex. Then the FDRs of the present tests satisfy*

$$FDR_{P,SUD(\lambda)} \leq FDR_{P,SU}. \tag{2.54}$$

**Proof**. Let $V_{SUD(\lambda),n}$ be the number of false rejections of the $SUD(\lambda)$ test and $V_{SU,n}$ of the SU test. The technique of the present proof is close to the one of Lemma 2.8. Along the lines of that proof we obtain

$$FDR_{P,SU} = \mathbb{E}_P\left(\sum_{i\in I_0}\mathbb{E}_P\left(\frac{F_{0,H,i}\left(\alpha_{R^{(i)}_{SU,n}:n}\right)}{R^{(i)}_{SU,n}}\Big|H\right)\right)$$

by Remark 6.2 (b) of the appendix and Fubini's Theorem. Similarly, from Remark 6.2 (c) and Fubini's Theorem we conclude that

$$\begin{aligned} FDR_{P,SUD(\lambda)} &= \mathbb{E}_P\left(\sum_{i\in I_0}\mathbb{E}_P\left(\frac{\mathbb{1}\{p_i \leq \alpha_{R_{SUD(\lambda),n}:n}\}}{R_{SUD(\lambda),n}}\Big|H\right)\right) \\ &\leq \mathbb{E}_P\left(\sum_{i\in I_0}\mathbb{E}_P\left(\frac{\mathbb{1}\left\{p_i \leq \alpha_{R^{(i)}_{SUD(\lambda),n}:n}\right\}}{R^{(i)}_{SUD(\lambda),n}}\Big|H\right)\right) \\ &= \mathbb{E}_P\left(\sum_{i\in I_0}\mathbb{E}_P\left(\frac{F_{0,H,i}\left(\alpha_{R^{(i)}_{SUD(\lambda),n}:n}\right)}{R^{(i)}_{SUD(\lambda),n}}\Big|H\right)\right) \end{aligned}$$

holds. The statement now follows since $j \mapsto \frac{F_{0,H,i}(\alpha_{j:n})}{j}$ is non decreasing (see (2.12)) and since $R^{(i)}_{SUD(\lambda),n} \leq R^{(i)}_{SU,n}$ holds for the step-wise tests with same critical values. □

In short, under the assumption (1.23) Lemma 2.8 says that the FDR of SU tests is largest under Dirac-Uniform configurations and Lemma 2.29 says that it is largest for SU tests in the class of SUD tests. Blanchard et al. [6] give a nice and broad overview of previous results of this kind, see also Figure 2 in their work. In the strict sense, they just focus on uniformly distributed true $p$-values.

We are now able to formulate the theorem.

**Theorem 2.30** (cf. Heesen and Janssen [29] Theorem 4.1)
*Consider a sequence of $SUD(\lambda_n)$ tests given by a concave rejection curve $r$ and an arbitrary sequence $(\lambda_n)_{n\in\mathbb{N}}$ with $1 \le \lambda_n \le n$, $n \in \mathbb{N}$. Then*

$$\begin{aligned}\beta &= \limsup_{n\to\infty} \sup_{P_n \in \mathcal{P}^C_{n,gBI}} FDR_{P_n, SUD(\lambda_n)} \\ &= \limsup_{n\to\infty} \sup_{P_n \in \mathcal{P}_{n,BI}} FDR_{P_n, SUD(\lambda_n)},\end{aligned} \tag{2.55}$$

*where $\beta$ is defined in (2.43) for the sequence of SU tests which is based on $r$.*

**Proof**. We only show the first equality in (2.55) since the other one follows analogously. By Lemma 2.29 we directly observe that "$\geq$" holds in (2.55). Let us now consider sequences of $DU(n, n_0)$ configurations with $\frac{n_0}{n} \to 1-y \in [\beta, 1)$. Along the lines of the proof of Theorem 2.25 we have

$$\frac{R_{SUD(\lambda_n),n}}{n} \xrightarrow[n\to\infty]{a.s.} K \tag{2.56}$$

at least along subsequences since (2.47) also holds for $SUD(\lambda_n)$ tests and where $0 < K < 1$ is the same value as for the sequence of SU tests. Again,

$$FDR_{DU(n,n_0),SUD(\lambda_n)} \xrightarrow[n\to\infty]{} \frac{x}{1-x} \cdot \frac{1-r(x)}{r(x)} \tag{2.57}$$

follows with $x$ defined as in that proof. A reinspection of the proof then shows that the worst case limit $\beta$ is attained by some sequence

$$FDR_{DU(n,n_0),SU} \xrightarrow[n\to\infty]{} \frac{x}{1-x} \cdot \frac{1-r(x)}{x} \tag{2.58}$$

with $\frac{n_0}{n} \to 1 - y \in [\beta, 1)$, see (2.50), or by $\liminf_{x\searrow 0} \frac{x}{1-x} \cdot \frac{1-r(x)}{r(x)}$, see (2.51). The statement now follows since the limits in (2.57) and (2.58) coincide. Note that $1-y$ close to 1 corresponds to $x$ close to 0. □

**Remark 2.31**
Consider a sequence of $SUD(\lambda_n)$ or SD tests as in Theorem 2.30 and assume that every $SUD(\lambda_n)$ test has finite sample FDR control by $\alpha$ under the gBI Model with convex dfs $F_{0,H,i}$ in (2.2). Then the above Theorem says that the corresponding sequence of SU tests, based on the same rejection curve, automatically has asymptotic FDR control by $\alpha$ under the present model. Note that $\alpha_{n:n} = r^{-1}(1) < 1$ is crucial for this statement.

## 2.5 Converse Benjamini Hochberg Theorem

As already mentioned, the FDR of the BH test under the BI Model is given by $\frac{\mathbb{E}(N_0)}{n}\alpha$. In this section we give a converse result and show that the BH test is the only SU test which exactly exhibits this FDR level. It is obviously clear that there are multiple tests which have this FDR level for a single or maybe a few distributions. As we will see, already a small submodel of the BI Model implies the result. Unlike the other results of this chapter, this section relies on Heesen and Janssen [28].

**Theorem 2.32** (cf. Theorem 5.1 in Heesen and Janssen [28])
*Consider a SU test with arbitrary deterministic critical values* $0 < \alpha_{1:n} \leq \alpha_{2:n} \leq \ldots \leq \alpha_{n:n} = \lambda < 1$ *or an adaptive SU test with data dependent critical values* $0 < \hat{\alpha}_{1:n} \leq \ldots \leq \hat{\alpha}_{n:n} \leq \lambda < 1$ *given by measurable functions*

$$\hat{\alpha}_{i:n} = \hat{\alpha}_{i:n}((\hat{F}_n(t))_{t\geq\lambda}), \quad i = 1, \ldots, n. \tag{2.59}$$

*(a) Suppose that we have* $FDR = \mathbb{E}(N_0)c$ *for a constant* $0 < c < \frac{1}{n}$ *for all distributions of Model 1.4 (a), where* $1 - \pi_0$ *varies over a non-trivial interval* $I \subset (0, 1)$ *and the df* $F_1$ *varies over all dfs with* $F_1(t) \geq t$ *for all* $t \in [0, 1]$ *and which have a Lebesgue density. Then the critical values already satisfy*

$$\hat{\alpha}_{i:n} = \frac{i}{n}\alpha \quad a.s. \tag{2.60}$$

*for* $i = 1, \ldots, n$*, where* $\alpha = nc$*. In the deterministic case the "a.s." can be dropped.*
*(b) Suppose that we have* $\mathbb{E}\left(\frac{V}{R\vee 1} \middle| N_0 = 1\right) = c$ *for all distributions in (a), then (2.60) holds again.*

**Proof**. This proof is based on results for complete statistical models which can be found in Lehmann and Romano [35, p. 116] and Pfanzagl [42, p. 17-22]. Our assumptions imply that

$$\int \left( \mathbb{E}\left( \frac{V}{R \vee 1} \Big| N_0 = n_0 \right) - n_0 c \right) \mathcal{L}(N_0)(dn_0) = 0 \tag{2.61}$$

holds for all distributions which are described in (a). Observe that $\mathcal{L}(N_0)$ is a binomial distribution. Thus, we have

$$\mathbb{E}\left( \frac{V}{R \vee 1} \Big| N_0 = n_0 \right) - n_0 c = 0 \tag{2.62}$$

for all $n_0 \in \{0, \dots, n\}$ since $N_0$ is a complete statistic for the exponential family of binomials.

From here on it suffices to prove (b) and we focus on $n_0 = 1$. Conditioned under $N_0 = 1$ the exact position of the true $p$-value does not matter since the SU tests are permutation-invariant in the $p$-values. In this way let $U$ be a uniformly distributed true $p$-value and independent of $U$ let $\xi_1, \dots, \xi_{n-1}$ be false $p$-values i.i.d. according to the alternative df $F_1$. Moreover, denote $\xi_{\cdot:n-1} = (\xi_{1:n-1}, \dots, \xi_{n-1:n-1})$. From (2.62) we conclude that

$$\begin{aligned} 0 &= \mathbb{E}\left( \frac{V(U, \xi_1, \dots, \xi_{n-1})}{R(U, \xi_1, \dots, \xi_{n-1}) \vee 1} - c \right) \\ &= \int \int_0^1 \left( \frac{V(u, \xi_{\cdot:n-1})}{R(u, \xi_{\cdot:n-1}) \vee 1} - c \right) du \, \mathcal{L}(\xi_{\cdot:n-1} | F_1) d\xi_{\cdot:n-1} \end{aligned} \tag{2.63}$$

holds. Observe that the family of distributions $\mathcal{L}(\xi_1 | F_1)$ is convex and complete in the sense of Pfanzagl [42, Theorem 1.5.10] and hence $\xi_{\cdot:n-1}$ is a complete statistic for this model. It follows that

$$\int_0^1 \left( \frac{V(u, \xi_{\cdot:n-1})}{R(u, \xi_{\cdot:n-1}) \vee 1} \right) du = c \tag{2.64}$$

holds $\mathcal{L}(\xi_{\cdot:n-1} | F_1)$ almost surely. By the same arguments as in the proof of Lemma 2.5 we now obtain that the left hand side of (2.64) is equal to $\frac{\hat{\alpha}_{R^{(1)}:n}}{R^{(1)}}$, where $R^{(1)} = R(0, \xi_1, \dots, \xi_{n-1})$. For the right hand side of (2.64) introduce $\alpha = nc$ which gives

$$\hat{\alpha}_{R^{(1)}:n} = \frac{R^{(1)}}{n} \alpha \qquad \mathcal{L}(\xi_{\cdot:n-1} | F_1) - a.s.. \tag{2.65}$$

It is easily seen that $\alpha < 1$ holds since $F_1$ may be the distribution function of the uniform distribution. In this case we have $\alpha = n\mathbb{E}\left( \frac{V}{R \vee 1} | N_0 = 1 \right) =$

$\mathbb{E}\left(\frac{R}{R}\mathbb{1}\{R > 0\}\right) = P(R > 0) < 1$. Moreover, it is easy to check that the sets $\{R^{(1)} = j\}$, $j = 1, \ldots, n$, have positive probability for at least one alternative distribution function $F_1$ which finally proves the assertion by (2.65). We may therefore again consider uniformly distributed false $p$-values and observe that

$$\{\xi_{j-1:n} \leq \hat{\alpha}_{j:n}, \xi_{j:n} > \lambda\} \subset \{R^{(1)} = j\}, \quad 1 \leq j \leq n,$$

holds for data driven critical values. Clearly, deterministic critical values are a special case of data dependent critical values. □

**Remark 2.33**
(a) The conditional distributions which are considered in (b) correspond to distributions of the BI Model with deterministic $H$ and $N_0 = 1$, where the false $p$-values are i.i.d. according to an alternative df $F_1$ as described in (a).
(b) The above theorem says that the BH test is the only adaptive SU test with "distribution free" FDR. Note that "distribution free" here shall mean that the df of false $p$-values $F_1$ has no influence on the FDR. Of course, the distribution of $N_0$ still has one.
(c) Theorem 2.32 also clarifies that there is no chance to obtain exact $FDR = \alpha$ for all cases for the adaptive SU tests given by (2.59).

Under the assumption (1.23) Benjamini and Yekutieli [5, Theorem 5.3] showed that the FDR of a SU test is non-decreasing when the distribution of false $p$-values becomes stochastically smaller. Otherwise, if $i \mapsto \frac{\alpha_{i:n}}{i}$ is non-increasing, they proved that the FDR is also non-increasing. In these cases Theorem 2.32 is not surprising, even though this statement actually only implies that the FDR of the BH test is distribution free. However, the present result holds under much more general assumptions for SU and adaptive SU tests without any monotonicity assumption of the form (1.23).

## 2.6 Asymptotics of adaptive SU tests

In the last sections, we mainly considered non adaptive SU tests and results based on some central lemmas. Observe that Lemma 2.5 and 2.10 are also valid for adaptive SU tests and we have not yet used the full potential of these central lemmas. Now we show that these two lemmas easily establish a sufficient condition which ensures asymptotic FDR control for particular sequences of adaptive SU tests. The next theorem applies to the RM and gBI Model and to adaptive SU tests of Storey type with arbitrary estimators.

**Theorem 2.34** (cf. Heesen and Janssen [29] Theorem 5.1)
*Let $\mathcal{P}_{n,RM}$ and $\mathcal{P}_{n,gBI}$ be the sets of all possible distributions of the RM and gBI Model for fixed $n \in \mathbb{N}$, respectively. Moreover, let $(P_n)_{n\in\mathbb{N}}$ be a sequence of distributions with $P_n \in \mathcal{P}_{n,RM}$ or $P_n \in \mathcal{P}_{n,gBI}$ for all $n$ and let $\lambda \in (0,1)$ be a tuning parameter. Consider the sequence of adaptive SU tests given by the critical values*

$$\hat{\alpha}_{i:n} = \left(\frac{i}{\hat{n}_{0,n}}\alpha\right) \wedge \lambda, \quad 1 \le i \le n, \tag{2.66}$$

*where*

$$\hat{n}_{0,n} = \hat{n}_{0,n}((\hat{F}_n(t))_{t\ge\lambda}) > 0, \quad n \in \mathbb{N}, \tag{2.67}$$

*is a sequence of estimators for $N_0$ given by measurable functions. If*

$$P_n\left(\frac{\hat{n}_{0,n}}{N_0} \le 1-\delta\right) \underset{n\to\infty}{\longrightarrow} 0 \tag{2.68}$$

*for all $\delta > 0$, then we have*

$$\limsup_{n\to\infty} FDR_{P_n} \le \alpha, \tag{2.69}$$

*where $\frac{x}{0} = \infty$ for $x > 0$.*

**Proof**. Obviously, $FDR = 0$ if $N_0 = 0$ almost surely. Thus, without loss of generality let $N_0 > 0$ hold almost surely for all $n$. By Lemma 2.5 and (2.66) we obtain

$$\frac{N_0}{n} \ge \mathbb{E}_{P_n}\left(\frac{V_n}{n\hat{\alpha}_{R_n:n}}\Big|H\right) \ge \mathbb{E}_{P_n}\left(\frac{\hat{n}_{0,n}}{n}\cdot\frac{V_n}{R_n\alpha}\Big|H\right) \tag{2.70}$$

under both models since the conditional case is also included as special case. Thus, by reordering and integration

$$\begin{aligned}
\alpha &\ge \mathbb{E}_{P_n}\left(\frac{\hat{n}_{0,n}}{N_0}\cdot\frac{V_n}{R_n}\right) \ge \mathbb{E}_{P_n}\left((1-\delta)\frac{V_n}{R_n}\mathbb{1}\left\{\frac{\hat{n}_{0,n}}{N_0} > 1-\delta\right\}\right) \\
&= (1-\delta)\mathbb{E}_{P_n}\left(\frac{V_n}{R_n}\right) - (1-\delta)\mathbb{E}_{P_n}\left(\frac{V_n}{R_n}1\left\{\frac{\hat{n}_{0,n}}{N_0} \le 1-\delta\right\}\right) \\
&\ge (1-\delta)\mathbb{E}_{P_n}\left(\frac{V_n}{R_n}\right) - (1-\delta)P_n\left(\frac{\hat{n}_{0,n}}{N_0} \le 1-\delta\right)
\end{aligned}$$

holds for every $\delta > 0$ and the assertion follows by (2.68). □

In a special case (i.e. if $P_n\left(\frac{\hat{n}_{0,n}}{N_0} < 1\right) \to 0$ and if $\hat{n}_{0,n}$ are non decreasing estimators), the above theorem can be easily proved by application of Lemma 4.3 of Blanchard and Roquain [7]. Due to a little bug, non decreasing has to be replaced by non increasing in [7, Lemma 4.3].

Finner and Gontscharuk [18] and Gontscharuk [24] already used (2.68) to show asymptotic FWER control of a specific sequence of adaptive Bonferroni tests and adaptive SD tests, respectively. These results hold under no additional dependence assumption, whereas Theorem 2.34 needs at least a reverse martingale structure. This is not surprising since the BH test itself does not exhibit FDR control under arbitrarily dependent $p$-values. Therefore, the SU tests with critical values (1.21) of Benjamini and Yekutieli [5] and (1.22) of Blanchard and Roquain [7] have been proposed. It turns out that a particular adaptive version of these SU tests has asymptotic FDR control under the same assumption, as the next theorem will show.

For the sake of completeness, note that Theorem 2.34 has a small intersection with Theorem 4.5 of Gontscharuk [24], where the asymptotic FDR of the adaptive SU test of Storey is computed. This theorem works under a weak dependence assumption which basically corresponds to (2.79) below and just gives (2.68) for the Storey estimator. In contrast to Theorem 2.34, it does not need any further dependence assumption like the RM Model, but it requires some additional assumption concerning the asymptotics of $\frac{R_n}{n}$.

Further asymptotic results for the FDR of adaptive test procedures under specific dependence structures have been obtained by Farcomeni [15]. His results rely on the consistency of an estimator for the portion of true null hypotheses.

**Theorem 2.35** (cf. Heesen and Janssen [29] Theorem 5.2)
*Let $\mathcal{P}_n$ be the set of all possible distributions of the Basic Model 1.2 and let $(P_n)_{n\in\mathbb{N}}$ be a sequence with $P_n \in \mathcal{P}_n$. Moreover, consider the sequence of adaptive SU tests given by the critical values (2.16) with*

$$\widetilde{\gamma}(i) = \widetilde{\gamma}_n(i) = i\frac{n}{\hat{n}_{0,n}}, \quad 1 \le i \le n, \tag{2.71}$$

*and*

$$\hat{n}_{0,n} = \hat{n}_{0,n}((\hat{F}_n(t))_{0\le t\le 1}) > 0, \quad n \in \mathbb{N}, \tag{2.72}$$

*is a sequence of estimators for $N_0$ given by measurable functions. If (2.68) holds for all $\delta > 0$, then (2.69) follows again.*

**Proof**. Again, let $N_0 > 0$ hold almost surely. By Lemma 2.10 and (2.71) we directly obtain that

$$\frac{N_0}{n} \geq \mathbb{E}_{P_n}\left(\frac{\hat{n}_{0,n}}{n} \cdot \frac{V_n}{R_n \alpha}\Big| H\right).$$

The statement now follows by the same arguments as in the proof of Theorem 2.34. □

**Remark 2.36**

(a) As in Lemma 2.10, the data dependent critical values (2.16) with (2.71) may again be functions of the entire ecdf $(\hat{F}_n(t))_{0\leq t\leq 1}$.

(b) Theorem 2.34 and 2.35 should not be used to look for estimators that yield asymptotic FDR control in the entire RM and Basic Model, respectively. Instead, they give a sufficient condition for asymptotic FDR control for a fixed sequence of distributions. Moreover, (2.68) describes a subclass of models with asymptotic FDR control for each fixed sequence of estimators. In the maximal dependent uniform case, i.e. when $\zeta_1 = \ldots = \zeta_n \sim U(0,1)$ and $H_1 = \ldots = H_n = 0$ hold, almost no sequence of estimators will satisfy (2.68) except for the very conservative estimators $\hat{n}_{0,n} = n$ which always work. However, whether (2.68) holds or not under the complete gBI Model is often easy to verify for non decreasing estimators. In the other models, it gets more complicated.

(c) The choice of $\nu(\{k\}) = \left(k\sum_{j=1}^{n}\frac{1}{j}\right)^{-1}$, $k = 1,\ldots,n$, which led to the SU test of Benjamini and Yekutieli [5], see (1.21), now yields an adaptive version with critical values

$$\hat{\alpha}_{i:n} = \frac{\lfloor \frac{n}{\hat{n}_{0,n}} \cdot i \rfloor}{n\sum_{j=1}^{n}\frac{1}{j}}\,\alpha, \quad i = 1,\ldots,n. \tag{2.73}$$

Lemma 4.3 of Blanchard and Roquain [7] also applies to the PRDS Model and Basic Model in the special case which is mentioned above. It is an open question whether Lemma 2.9 can be extended to data dependent critical values and thus whether Theorem 2.34 also holds under the PRDS Model. Note that the estimator in Theorem 2.35 scales the range of the integral which defines the adaptive critical values, whereas the adaptive critical values in Lemma 4.3 of Blanchard and Roquain [7] are directly scaled by the estimator.

We now show a converse result of Theorem 2.34. Under mild regularity assumptions, (2.69) already implies a stricter version of (2.68) for Storey type estimators.

**Proposition 2.37** (cf. Heesen and Janssen [29] Proposition 5.1)
*As in Theorem 2.34, let $(P_n)_{n\in\mathbb{N}}$ be a sequence of RM Models and assume that either*

*(i)* $\frac{N_0}{n} \to 1$ *or*

*(ii)* $\xi_i \leq \lambda$ *for all* $i \in \mathbb{N}$ *and* $0 < \eta \leq \frac{N_0}{n}$ *for all* $n \in \mathbb{N}$

*holds. Consider the sequence of adaptive SU tests (2.66) with Storey type estimators*

$$\hat{n}_{0,n} = n\frac{1 - \hat{F}_n(\lambda) + \kappa_n}{1-\lambda}, \quad n \in \mathbb{N}, \tag{2.74}$$

*for some positive sequence* $\kappa_n \to 0$. *Moreover, assume that* $P_n(\{\hat{\alpha}_{R:n} = \lambda\}) \to 0$ *for* $n \to \infty$ *and (2.69) hold. Then we have*

$$\frac{\hat{n}_{0,n}}{N_0} \underset{n\to\infty}{\longrightarrow} 1 \quad \textit{in } P_n\textit{-probability.} \tag{2.75}$$

**Proof**. Introduce the set $A_n = \{\hat{\alpha}_{R_n:n} < \lambda\}$ and its complement $A_n^c$. Observe that in the proof of Lemma 5.1 $\alpha' = \frac{n\hat{F}_n(\lambda)}{\lambda \hat{n}_{0,n}}\alpha \leq 1$ holds on $A_n$. Otherwise, if we assume $\alpha' > 1$, then in the notation of the proof $R = R_q = n\hat{F}_n(\lambda)$ follows and $A_n = \{\frac{n\hat{F}_n(\lambda)}{\hat{n}_{0,n}}\alpha < \lambda\}$ contradicts our assumption. Hence, by the arguments of the proof of Lemma 5.1 we obtain

$$\mathbb{E}_{P_n}\left(\frac{V_n}{R_n \vee 1}\mathbb{1}_{A_n}\right) = \frac{\alpha}{\lambda}\mathbb{E}_{P_n}\left(\frac{V_n(\lambda)}{\hat{n}_{0,n}}\mathbb{1}_{A_n}\right).$$

Define $S_n(\lambda) = n\hat{F}_n(\lambda) - V_n(\lambda)$, then by (2.69) and (2.74) it follows that

$$\alpha \geq \limsup_{n\to\infty} \frac{\alpha(1-\lambda)}{\lambda} \cdot \mathbb{E}_{P_n}\left(\frac{V_n(\lambda)}{n - V_n(\lambda) - S_n(\lambda) + \kappa_n}\mathbb{1}_{A_n}\right)$$

holds. Let us first consider case (ii). For each $\delta > 0$ we obtain

$$1 \geq \frac{1-\lambda}{\lambda}\limsup_{n\to\infty}\mathbb{E}_{P_n}\left(\frac{V_n(\lambda)/N_0}{1 - V_n(\lambda)/N_0 + \kappa_n/N_0 + \delta}\mathbb{1}_{A_n}\right). \tag{2.76}$$

Observe next that $0 \leq \frac{V_n(\lambda)}{N_0} \leq 1$ is tight and we consider an arbitrary distributional cluster point $Z$ of $\frac{V_n(\lambda)}{N_0}$. Let us denote the appertaining subsequence again by $n$, i.e. $\frac{V_n(\lambda)}{N_0} \to Z$ in distribution. Note that $\mathbb{E}(Z) = \lambda$ and

$$1 \geq \frac{1-\lambda}{\lambda}\mathbb{E}\left(\frac{Z}{1-Z+\delta}\right)$$

hold. This and Jensen's inequality implies

$$\frac{\lambda}{1-\lambda} \geq \mathbb{E}\left(\frac{Z}{1-Z}\right) \geq \frac{\lambda}{1-\lambda}$$

when $\delta \searrow 0$. Since $x \mapsto \frac{x}{1-x}$ is strictly convex we have $Z = \lambda$ a.e.. Since $Z$ was an arbitrary cluster point we conclude $\frac{V_n(\lambda)}{N_0} \to \lambda$ in $P_n$-probability which finally implies

$$\frac{\hat{n}_{0,n}}{N_0} = \frac{N_0 - V_n(\lambda) + n\kappa_n}{(1-\lambda)\cdot N_0} \underset{n\to\infty}{\longrightarrow} 1 \quad \text{in } P_n\text{-probability.}$$

The proof of case (i) is similar. Note that the assumption implies $\frac{(n-N_0)+S_n(\lambda)}{N_0} \to 1$ and we may proceed as in (2.76). □

**Remark 2.38**

Consider the RM Models and let us exclude the cases with $N_0 = 0$. As long as sufficient variability of the variables $((1-H_i)\mathbb{1}\{p_i \leq \lambda\})_{i\leq n}$ is present, condition (2.68) can easily be verified for the Storey type estimators (2.74) where $\kappa_n$ is some arbitrary positive sequence. Let

$$\hat{F}_{0,n}(\lambda) = \frac{1}{N_0}\sum_{i=1}^{n}(1-H_i)\mathbb{1}\{p_i \leq \lambda\} \tag{2.77}$$

be the ecdf of the true $p$-values. Then a sufficient condition for (2.68) is

$$Var_{P_n}\left(\hat{F}_{0,n}(\lambda)\middle|(H_i)_{i\leq n}\right) = \mathbb{E}_{P_n}\left(\left(\hat{F}_{0,n}(\lambda)-\lambda\right)^2\middle|(H_i)_{i\leq n}\right) \underset{n\to\infty}{\longrightarrow} 0 \tag{2.78}$$

in $P_n$-probability. For the gBI and Basic Model

$$P_n\left(\hat{F}_{0,n}(\lambda) \geq \lambda + \epsilon\middle|(H_i)_{i\leq n}\right) \underset{n\to\infty}{\longrightarrow} 0 \tag{2.79}$$

in $P_n$-probability for all $\epsilon > 0$ is sufficient. These cases correspond to so called weak dependency assumptions which were discussed in Gontscharuk [24], Section 4.1.

**Proof**. Observe that

$$\frac{\hat{n}_{0,n}}{N_0} \geq \frac{N_0 - V_n(\lambda) + \kappa_n}{(1-\lambda)\cdot N_0} \geq \frac{1-\hat{F}_{0,n}(\lambda)}{1-\lambda} \underset{n\to\infty}{\longrightarrow} 1 \quad \text{in } P_n\text{-probability}$$

holds for the first case, where the right hand side converges by (2.78) w.r.t. the conditional convergence. The second case follows similarly. □

In Chapter 5, we will give a counterexample for FDR control under the RM Model when adaptive SU tests of Storey type with estimator (2.74) and $\lambda < 1$ are applied.

# Chapter 3

# A new estimation concept for adaptive test procedures

Adaptive SU tests often incorporate an estimation of the proportion of true null hypotheses. In a model with random true and false null hypotheses (like Model 1.4 (a)), the estimator for the proportion of true null hypotheses can also be seen as an estimator for the expected proportion $\pi_0$ and vice versa. In this chapter we develop a new estimation concept in order to improve the estimators for adaptive SU tests. The possible improvements of estimation strongly depend on the model assumptions, i.e., on the possible distributions of true and false $p$-values. Furthermore, we will show that the possible improvements for adaptive SU tests are much broader than only an improved estimation of the portion and expected portion of true null hypotheses, respectively. In the next chapter we will study finite sample FDR control of adaptive SU tests.

In the first section we will give a brief summary of the well known and often applied Storey estimator. For this estimator, we provide an optimality result in a specific submodel of Model 1.4 (a). In Section 3.2 we will relax these model assumptions and consider a more general model. The important question will be: What does really account for an improvement of an estimator for adaptive SU tests? Therefore, we reconsider the studies of Storey [59] in this more general model and show that the expected proportion of true null hypotheses $\pi_0$ is not the parameter which should be estimated. The estimation of the correct parameter is quite more complex. In this sense, "new estimation concept" shall mean that we show how and why to estimate a different parameter. For this purpose, we introduce new estimators and give a further optimality result in a specific submodel. In Section 3.3 we will introduce integral versions of estimators, considered in the previous sections. In the subsequent Section 3.4 we will adjust estimators, which appeared in the recent literature, in order to make them work well in our more general model. Finally, Section 3.5 is devoted to the family of generalized error rates of Meskaldji et al. [39], see Remark 2.12.

## 3.1 The original Storey Estimator

We begin with a brief review of well known properties of the original Storey estimator which has been introduced by Schweder and Spjøtvoll [55] and later by Storey [59]. Throughout we call it original because it is the first of two versions of the estimator which appeared in the literature. The other slightly modified version appeared in Storey et al. [61] and has been introduced in order to provide finite sample FDR control for the adaptive SU test of Storey, see (1.29) and (1.30). We will have a detailed look at the approach of Storey [59] in the next section. For this section let us recall Model 1.4 (a) which has also been considered by Storey [59, 60]:

Starting from Basic Model 1.2, the $p$-values $p_1, \ldots, p_n$ are true or false according to the *i.i.d.* Bernoulli random variables $H_1, \ldots, H_n \sim B(1, 1-\pi_0)$, where $\pi_0 \in (0,1]$ and $H_i = 0$ codes the occurrence of a true $p$-value. The true $p$-values are independent and uniformly distributed. Moreover, the false $p$-values are independent and distributed according to some alternative df $F_1$ on $[0,1]$ and independent of the vector of true $p$-values. For every fixed distribution under the present model $p_1, \ldots, p_n$ are *i.i.d.* with $p_i \sim F_{\pi_0, F_1}$, where $F_{\pi_0, F_1}(t) = \pi_0 t + (1-\pi_0)F_1(t)$, $0 \leq t \leq 1$. In this model, at least the possible joint distributions of $(p_1, \ldots, p_n)$ can be represented by

$$\{P^n_{\pi_0, F_1} : \pi_0 \in (0,1],\, F_1 \text{ alternative df}\}, \tag{3.1}$$

where $P_{\pi_0, F_1} \sim F_{\pi_0, F_1}$. We will use a similar notation for further submodels without defining it every time.

**Remark 3.1**

A parametric statistical model $(\Omega, \mathcal{A}, \{P_\vartheta : \vartheta \in \Theta\})$ is called identifiable iff the mapping $\vartheta \mapsto P_\vartheta$ is bijective. Note that (3.1) is non-identifiable and the distributions $P^n_{\pi_0,F_1}$ of the $p$-values are not able to model the occurrence of true and false $p$-values. If $F_1$ is the df of $U(0,1)$, then every choice of $\pi_0$ leads to the same probability measure $P^n_{1,F_1}$. Therefore, the alternative df $F_1$ may be restricted to the subset of dfs which are stochastically smaller than $U(0,1)$ without the uniform distribution itself. Unfortunately, this does not work for generalizations of this model when the marginal distributions of true $p$-values are allowed to be stochastically larger than $U(0,1)$. For instance, $U(0,\frac{1}{2})$ distributed false $p$-values, $U(\frac{1}{2},1)$ distributed true $p$-values and $\pi_0 = \frac{1}{2}$ lead to the probability measure $P^n_{1,F_1}$. Nevertheless, this is an interesting and possible underlying model which should definitely be analyzed with respect to the FDR. This is the main focus of this chapter. Working without identifiability is a very challenging task. Therefore, we will begin by considering specific submodels which exhibit identifiability and where the estimation of the essential parameters for the FDR works well. The new estimators are developed in these submodels. In a further step, we will show that the introduced estimators still exhibit a conservative behavior in the complete non-identifiable model.

The original Storey estimator is given by

$$\hat{\pi}_0^{St,o.}(\lambda) = \frac{1-\hat{F}_n(\lambda)}{1-\lambda} \tag{3.2}$$

with some tuning parameter $\lambda \in (0,1)$. In the present model, $\hat{\pi}_0^{St,o.}(\lambda)$ is an estimator for the expected proportion of true null hypotheses $\pi_0$. Storey [59] incorporates $\hat{\pi}_0^{St,o.}(\lambda)$ into the estimator

$$\widehat{FDR}(t) = \frac{\hat{\pi}_0^{St,o.}(\lambda)\cdot t}{\hat{F}_n(t)\vee\frac{1}{n}}, \quad t\in[0,1], \tag{3.3}$$

for $FDR(t)$ which is defined as the FDR of the multiple test procedure which rejects every $p$-value $p_i$ iff $p_i \le t$. The bias of $\hat{\pi}_0^{St,o.}(\lambda)$ in the present model is given by

$$\begin{aligned}\mathbb{E}\left(\hat{\pi}_0^{St,o.}(\lambda)-\pi_0\right) &= \tfrac{1-\pi_0\lambda-(1-\pi_0)F_1(\lambda)}{1-\lambda}-\pi_0\\ &= (1-\pi_0)\tfrac{1-F_1(\lambda)}{1-\lambda}\ge 0,\end{aligned} \tag{3.4}$$

cf. Langaas et al. [32]. Furthermore, Chen and Doerge [10] calculated the bias under the gBI Model conditioned under $H$. Usually, the false $p$-values are assumed to be stochastically smaller than the uniform distribution. From (3.4) we can observe that $\hat{\pi}_0^{St,o.}$ has a small bias when $F_1(\lambda)$ is close to 1 and is actually unbiased for $F_1(\lambda) = 1$. Thus, with respect to the bias, one would like to choose a large $\lambda$ close to 1. However, one would also like to have an estimator with small variance, but this would require the choice of a small $\lambda$. This leads to a bias variance trade-off and $\lambda$ is often chosen close to 0.5, cf. Storey and Tibshirani [62].

The original Storey estimator only uses the ecdf $\hat{F}_n$ at a single position $\lambda$. The question which now arises is the following: Is $\hat{\pi}_0^{St,o.}$ improvable by incorporating more information of the ecdf $\hat{F}_n$? The next theorem shows that $\hat{\pi}_0^{St,o.}(\lambda)$ has minimum variance among all unbiased estimators which only use the information of $(\hat{F}_n(t))_{t\geq\lambda}$ when $F_1(\lambda) = 1$.

**Theorem 3.2**

*Let $\lambda \in (0,1)$ and consider the following submodel of Model 1.4 (a). Assume that $F_1(\lambda) = 1$ holds for the df of false p-values, i.e. all false p-values are almost surely located in $[0, \lambda]$. Then the original Storey estimator $\hat{\pi}_0^{St,o.}(\lambda)$ is a uniformly minimum variance unbiased (UMVU) estimator for the statistical functional $P_{\pi_0,F_1} \mapsto \pi_0$ in the class of estimators $\{\hat{\pi}_0 = f((\hat{F}_n(t))_{t\geq\lambda}) :$ $f$ is measurable$\}$.*

**Proof**. The statement follows by application of the well known Theorem of Lehmann and Scheffé (see Witting [64, Satz 3.35] or Lehmann and Casella [33, Theorem 1.11] for instance). Therefore, consider the transformed model

$$\left\{ \left(P^n_{\pi_0,F_1}\right)^{(\hat{F}_n(t))_{t\geq\lambda}} : \pi_0 \in (0,1],\, F_1 \text{ with } F_1(\lambda) = 1 \right\}. \tag{3.5}$$

Observe that $F_{\pi_0,F_1}(t) = (1-\pi_0) + \pi_0 t$ for $t \geq \lambda$ and all considered dfs $F_1$. Thus, we obtain $P_{\pi_0,F_1}(\,\cdot\,|(\lambda,1)) \sim U(\lambda,1)$ and

$$\left(P^n_{\pi_0,F_1}\right)^{(\hat{F}_n(t))_{t\geq\lambda}|\hat{F}_n(\lambda)=k/n} = \mathcal{L}\left( \left( \frac{k}{n} + \frac{n-k}{n}\hat{F}_{n-k}(t) \right)_{t\geq\lambda} \middle| U(\lambda,1)^{n-k} \right)$$

is independent of $\pi_0$ and $F_1$ for $0 \leq k \leq n$. Hence, $\hat{F}_n(\lambda)$ is a sufficient statistic for (3.5). Moreover, observe that

$$\mathcal{L}\left( n\hat{F}_n(\lambda) \middle| \left(P^n_{\pi_0,F_1}\right)^{(\hat{F}_n(t))_{t\geq\lambda}} \right) = B(n, F_{\pi_0,F_1}(\lambda))$$

is a binomial distribution and the image of $\{(\pi_0, F_1) : \pi_0 \in (0,1], F_1$ with $F_1(\lambda) = 1\}$ under the mapping $(\pi_0, F_1) \mapsto F_{\pi_0,F_1}(\lambda)$ is given by $[\lambda, 1)$. Hence,

$\hat{F}_n(\lambda)$ is also a complete statistic for (3.5) (see Lehmann and Romano [35, Example 4.3.1] for instance). Finally, $\hat{\pi}_0^{St,o.}(\lambda)$ is unbiased and factorizes after the sufficient and complete statistic $\hat{F}_n(\lambda)$ which proves the statement. □

## 3.2 New estimation concept and new estimators

Let us now switch to the following more general model which allows for significant improvements as we will see.

**Model 3.3** (*i.i.d.* mixture model)
*Assume a relaxed form of Model 1.4 (a), where the true p-values $(\xi_i)_{i\leq n}$ are still i.i.d. but distributed according to some null df $F_0$ whose distribution is stochastically larger than the uniform distribution. For every fixed distribution under the present model the p-values $p_1,\ldots,p_n$ are i.i.d. with $p_i \sim F$ and*

$$F(t) = F_{\pi_0,F_0,F_1}(t) = \pi_0 F_0(t) + (1-\pi_0)F_1(t), \quad 0 \leq t \leq 1. \tag{3.6}$$

*Hence, the possible joint distributions of $(p_1,\ldots,p_n)$ can be represented by*

$$\{P^n_{\pi_0,F_0,F_1} : \pi_0 \in (0,1],\ F_0 \overset{st}{\geq} U(0,1),\ F_1 \text{ alternative df}\},$$

*where $P_{\pi_0,F_0,F_1} \sim F_{\pi_0,F_0,F_1}$. We will use a similar notation for further submodels without defining it again.*

Note that this model is no artificial construct to allow the announced improvements. Null dfs $F_0$ with $F_0 \overset{st}{\geq} U(0,1)$ widely occur in one sided hypotheses testing problems. For more information we refer to Section 4.1.

We will soon show that the parameter $\pi_0$ is not really the parameter one should estimate for FDR controlling procedures in this model. But even the estimation of $\pi_0$ becomes more difficult when the null $p$-values get stochastically larger as the following example shows. The last fact has already been mentioned by Chen and Doerge [10] and Dickhaus [11]. In contrast to our approach, their approach is based on an improved estimation of $\pi_0$ and related parameters.

**Example 3.4**
Let us consider Model 3.3 with $F_1(t) = \mathbb{1}_{[0,1]}(t)$ and $F_0(t) = (1-\pi_2)t + \pi_2 \mathbb{1}_{\{1\}}(t)$, $0 \leq t \leq 1$, with $\pi_2 \in [0,1]$. The df $F_0$ is the df of a **mixed Dirac-Uniform (mDU) configuration** with Dirac part 1. Furthermore, the df

$$F(t) = (1-\pi_0)\mathbb{1}_{[0,1]}(t) + \pi_0(1-\pi_2)t + \pi_0\pi_2\mathbb{1}_{\{1\}}(t), \quad 0 \leq t \leq 1,$$

of the $p$-values $p_i$ is the df of a **mixed twofold Dirac-Uniform (m2DU) configuration** with Dirac parts 0 and 1. Observe, that $F$ is a linear function on the open interval $(0,1)$ and has jumps at 0 and 1. According to $\pi_2$, the estimator $\hat{\pi}_0^{St,o.}(\lambda)$ can have a very bad behavior even if $F_1(\lambda) = 1$. Table 3.1 shows $\mathbb{E}(\hat{\pi}_0^{St,o.}(0.5))$ for $\pi_0 = 0.9$ and different values of $\pi_2$ in order to indicate this. The larger the Dirac part $\pi_0\pi_2$ at 1 in $F$, the larger the upwards bias. The

| $\pi_2$ | 0 | 0.05 | 0.1 | 0.15 |
|---|---|---|---|---|
| $\mathbb{E}(\hat{\pi}_0^{St,o.}(0.5))$ | 0.9 | 0.945 | 0.99 | 1.035 |

Table 3.1: Expectation of the original Storey estimator.

same behavior shows up for different values of $\pi_0$ and $\lambda$ and also for different dfs $F_0$ and $F_1$. Actually, $\mathbb{E}(\hat{\pi}_0^{St,o.}(\lambda)) > 1$ is possible, cf. Table 3.1 for $\pi_2 = 0.15$.

Our next considerations are based on Storey [59, Section 3]. We will generalize his results by replacing Model 1.4 (a) by Model 3.3 and by reconsidering his approach step by step.

His purpose is to estimate $FDR(\Gamma)$ which is defined as the FDR of the multiple test procedure which is based on real test statistics $T_1, \ldots, T_n$ and rejects every $T_i$ iff $T_i \in \Gamma$ for a fixed rejection region $\Gamma \subseteq \mathbb{R}$. As in Model 3.3, $T_1, \ldots, T_n$ are *i.i.d.* randomly chosen to be true or false but distributed according to some fixed arbitrary null df $\widetilde{F}_0$ or alternative df $\widetilde{F}_1$. However, in terms of $p$-values $p_1, \ldots, p_n$, Storey only considers uniformly distributed true $p$-values and rejection areas $\Gamma = [0,t]$. The FDR of the rejection area $[0,t]$ is denoted by $FDR(t)$. He precisely calculates $pFDR(\Gamma)$ for $T_1, \ldots, T_n$ which is defined as the **positive FDR (pFDR)** of the test described before. Since our $p$-value Model 3.3 is contained in his model of $T_1, \ldots, T_n$ we immediately obtain the next theorem.

**Theorem 3.5** (cf. Theorem 1 in Storey [59] and [60])
*Consider Model 3.3. Then we have*

$$pFDR(t) = \frac{\pi_0 F_0(t)}{F(t)}, \quad t \in [0,1]. \tag{3.7}$$

**Proof**. The statement follows directly from Theorem 1 of Storey [59] as special case. □

Storey proposes to estimate $\pi_0$ by the conservative (upwards biased) estimate $\hat{\pi}_0^{St,o.}(\lambda)$ and $F(t)$ by the ecdf $\hat{F}_n(t)$. Furthermore, $F_0(t) = t$, $0 \leq t \leq 1$, holds for uniformly distributed true $p$-values, but Storey suggests to use the correction $t/(1-(1-t)^n)$, because $1-(1-t)^n$ is a lower bound for $P(R(t) > 0)$ and pFDR is defined as conditional expectation given $R(t) > 0$. This leads to the estimator

$$\widehat{pFDR}^{St}(t) = \frac{\hat{\pi}_0^{St,o.}(\lambda) \cdot t}{(\hat{F}_n(t) \vee \frac{1}{n})(1-(1-t)^n)}, \quad t \in [0,1]. \tag{3.8}$$

Retrospectively, he argues that

$$\widehat{FDR}^{St}(t) = \frac{\hat{\pi}_0^{St,o.}(\lambda) \cdot t}{\hat{F}_n(t) \vee \frac{1}{n}}, \quad t \in [0,1], \tag{3.9}$$

is an estimator for $FDR(t)$ since the FDR is not a conditioned quantity. For further understanding of these corrections for the estimation of $pFDR(t)$ and $FDR(t)$, we refer to Storey [59, Section 3 and 8].

For our next considerations we will only focus on $FDR(t)$ and estimators for this quantity. For uniformly distributed $p$-values one can clearly replace $F_0(t)$ by $t$ itself, but if we switch to Model 3.3, $F_0(t) = t$ no longer applies. In Model 3.3, $F_0(t) \leq t$ is valid. Depending on $F_0$, the replacement of $F_0(t)$ by $t$ may be far too conservative. Of course, $\pi_0$ itself is usually estimated by a conservatively biased estimate (in the sense that one would tend to overestimate $\pi_0$ and hence tend to overestimate $FDR(t)$), but one would truly like to have a low bias in order to have an accurate estimation of $FDR(t)$. As we have seen in Example 3.4, $\hat{\pi}_0^{St,o.}$ may already have an increased bias if we switch to Model 3.3. The next example will give a joint consideration of this problem.

**Example 3.6** (Example 3.4 continued)
Let us consider a conditional version of the m2DU configuration. Assume that about $(1-\pi_0)n$ $p$-values are false and have Dirac distribution $\epsilon_0$, about $\pi_0(1-\pi_2)n$ $p$-values are true, independently and uniformly distributed and about $\pi_0\pi_2 n$ $p$-values are true and have Dirac distribution $\epsilon_1$. The exact order of the $p$-values inside $p$ does not matter. If we now consider $FDR(t)$ with $t < 1$, about $\pi_0\pi_2 n$ true $p$-values (i.e. the true $p$-values with Dirac distribution $\epsilon_1$) are not in danger of getting rejected and becoming a so-called false positive by the corresponding testing procedure. Hence, these $p$-values do not have any influence on $FDR(t)$. This clearly remains true if the Dirac distribution $\epsilon_1$ is replaced by some $p$-value df $G$ with $G(t) = 0$. Even if $0 \leq G(t) < t$, the associated

$p$-values would have a lower effect on $FDR(t)$ than the uniformly distributed true $p$-values. This observation also carries over to SU tests, see Chapter 4 for more details.

Therefore, it is desirable to jointly estimate $\pi_0 F_0(t)$ in the present generalized model and to use

$$\widehat{FDR}(t) = \frac{\widehat{\pi_0 F_0}(t)}{\hat{F}_n(t) \vee \frac{1}{n}}, \quad t \in [0,1], \tag{3.10}$$

as estimator for $FDR(t)$. In Model 3.3, the term $1-(1-t)^n$ is no longer a lower bound for $P(R(t) > 0)$ and hence we cannot define an analogue estimator for $pFDR(t)$. But we still regard $\widehat{FDR}(t)$ as estimator for $FDR(t)$, see also (3.43) for a direct motivation.

We now derive new estimators for $\pi_0 F_0(t)$ which will be denoted by $\hat{\varpi}(t,\cdot)$. Note that $\varpi$ is the calligraphic form of $\pi$. This shall indicate that $\hat{\varpi}(t,\cdot)$ is the estimator for the crucial parameter $\pi_0 F_0(t)$ which is related to $\pi_0$, in some sense. Previously in Model 1.4 (a), $\pi_0$ has been the crucial parameter. It will turn out that the estimation of $\pi_0 F_0(t)$ is more difficult than merely estimating $\pi_0$. Therefore, we will develop the new estimators in a specific submodel in which $\pi_0 F_0(t)$ shows a nice behavior and proceed as described in Remark 3.1 due to the problem of identifiability. Let us now introduce the family of new estimators given by

$$\hat{\varpi}(t,\lambda,\gamma) = \frac{\hat{F}_n(\gamma) - \hat{F}_n(\lambda)}{\gamma - \lambda}\, t \tag{3.11}$$

with $0 \le t \le 1$ and $0 < \lambda < \gamma \le 1$. Similar estimators have already been mentioned in passing by Liang and Nettleton [36] in another context and without focus on finite sample FDR control. Further estimators will be described and discussed later.

**Proposition 3.7**
*Let $0 < \lambda^* < \gamma^* \le 1$ be fixed in advance and consider the following submodel of Model 3.3. Assume that $F_0(t) = (1-\pi_2)t + \pi_2 G(t)$ for some p-value df $G$ with $G(\gamma^*) = 0$ and for some $\pi_2 \in [0,1]$. Furthermore, assume that we have $F_1(\lambda^*) = 1$. If $t \le \gamma*$ and $\lambda^* \le \lambda < \gamma \le \gamma^*$, then $\hat{\varpi}(t,\lambda,\gamma)$ defined in (3.11) is an unbiased estimator for $\pi_0 F_0(t) = \pi_0(1-\pi_2)t$.*

**Proof**. Since $F(t) = (1-\pi_0) + \pi_0(1-\pi_2)t$ for all $t \in [\lambda^*, \gamma^*]$, it follows immediately that

$$\mathbb{E}\left(\hat{\varpi}(t,\lambda,\gamma)\right) = \frac{F(\gamma)-F(\lambda)}{\gamma-\lambda}\cdot t = \pi_0(1-\pi_2)t.$$

□

The next theorem gives an optimality result for the new estimators which is similar to the one of Section 3.1.

**Theorem 3.8**

*Consider the model given in Proposition 3.7. Then the estimator $\hat{\varpi}(t,\lambda^*,\gamma^*)$ is an UMVU estimator for the statistical functional $P_{\pi_0,\pi_2,F_0,G} \mapsto \pi_0 F_0(t) = \pi_0(1-\pi_2)t$ in the class of estimators $\{\hat{\varpi} = f((\hat{F}_n(t))_{\lambda^*\le t\le\gamma^*} \,:\, f$ is measurable$\}$ for every $t \le \gamma^*$.*

**Proof**. Similar to the proof of Theorem 3.2 we show that $(\hat{F}_n(\lambda^*), \hat{F}_n(\gamma^*))$ is a complete and sufficient statistic for the transformed model

$$\left\{ \left(P^n_{\pi_0,\pi_2,F_1,G}\right)^{(\hat{F}_n(t))_{\lambda^*\le t\le\gamma^*}} \;:\; \begin{array}{ll} \pi_0\in(0,1], & F_1 \text{ with } F_1(\lambda^*)=1, \\ \pi_2\in[0,1], & G \text{ with } G(\gamma^*)=0 \end{array} \right\}. \tag{3.12}$$

The statement then follows by the same arguments. Observe that $F(t) = F_{\pi_0,\pi_2,F_1,G}(t) = (1-\pi_0)+\pi_0(1-\pi_2)t$ for all $\lambda^*\le t\le\gamma^*$. Thus,
$P_{\pi_0,\pi_2,F_1,G}(\,\cdot\,|[\lambda^*,\gamma^*]) \sim U(\lambda^*,\gamma^*)$ follows and

$$\begin{aligned}&\left(P^n_{\pi_0,\pi_2,F_1,G}\right)^{(\hat{F}_n(t))_{\lambda^*\le t\le\gamma^*}|\hat{F}_n(\lambda^*)=k_1/n,\hat{F}_n(\gamma^*)=k_2/n} \\ &\qquad = \mathcal{L}\left(\left(\frac{k_1}{n}+\frac{k_2-k_1}{n}\hat{F}_{k_2-k_1}(t)\right)_{\lambda^*\le t\le\gamma^*}\Big| U(\lambda^*,\gamma^*)^{k_2-k_1}\right)\end{aligned}$$

is independent of $\pi_0,\pi_2,F_1$ and $G$ for $0\le k_1\le k_2\le n$. Hence, $(\hat{F}_n(\lambda^*),\hat{F}_n(\gamma^*))$ is a sufficient statistic for (3.12). Furthermore, it suffices to show that the multinomial distributed statistic

$$\left(n\hat{F}_n(\lambda^*), n(\hat{F}_n(\gamma^*)-\hat{F}_n(\lambda^*))\right) \sim \mathcal{M}\left(n, F(\lambda^*), F(\gamma^*)-F(\lambda^*), 1-F(\gamma^*)\right)$$

is complete with respect to (3.12). This follows directly by application of Lemma 6.3 of the appendix since

$$\{(F(\lambda^*), F(\gamma^*)-F(\lambda^*)) \,:\, \pi_0,\pi_2\in(0,1)\} \tag{3.13}$$

has a nonempty interior. Therefore, observe that the mapping

$$(0,1)^2 \ni (\pi_0, \pi_2) \mapsto (F(\lambda^*), F(\gamma^*) - F(\lambda^*))$$
$$= \big(1 - \pi_0 + \pi_0(1-\pi_2)\lambda^*, \pi_0(1-\pi_2)(\gamma^* - \lambda^*)\big)$$

has the continuous inverse

$$(x, y) \mapsto \left(1 - x + y\frac{\lambda^*}{\gamma^* - \lambda^*}, \frac{1 - x - y\frac{1-\lambda^*}{\gamma^*-\lambda^*}}{1 - x + y\frac{\lambda^*}{\gamma^*-\lambda^*}}\right) \tag{3.14}$$

on the set (3.13). The inverse is well-defined since $F(\lambda^*) < 1$ for all $\pi_0, \pi_2 \in (0,1)$ and thus, $1 - x + y\frac{\lambda^*}{\gamma^*-\lambda^*} > 0$ follows. Hence, (3.13) is a nonempty and open set as inverse image (of the continuous inverse) of $(0,1)^2$. □

**Remark 3.9**

The estimator

$$\hat{\varpi}(\lambda, \gamma) = \frac{\hat{F}_n(\gamma) - \hat{F}_n(\lambda)}{\gamma - \lambda} \tag{3.15}$$

(which results from (3.11) by removing $t$) can be seen as generalization of the original Storey estimator $\hat{\pi}_0^{St,o.}(\lambda)$. But in the model of Proposition 3.7, $\hat{\varpi}(\lambda, \gamma)$ estimates $\pi_0(1-\pi_2)$, whereas $\hat{\pi}_0^{St,o.}(\lambda)$ estimates $\pi_0$. In the more general Model 3.3 it is actually not clear what $\hat{\varpi}(\lambda, \gamma)$ is estimating exactly. Hence, one should rather compare $\hat{\varpi}(t, \lambda, \gamma)$ with $\hat{\pi}_0^{St,o.}(\lambda) \cdot t$ which both estimate $\pi_0 F_0(t)$. We will also refer to $\hat{\pi}_0^{St,o.}(\lambda) \cdot t$ as original Storey estimator. However, every estimator $\hat{\varpi}(t, \cdot)$ for $\pi_0 F_0(t)$ considered in this chapter is linear in $t$ and may be written as $\hat{\varpi}(t, \cdot) = t \cdot \hat{\varpi}(\cdot)$ for some estimator $\hat{\varpi}(\cdot)$ which does not depend on $t$. For convenience we will hence sometimes deal with $\hat{\varpi}(\cdot)$ instead of $\hat{\varpi}(t, \cdot)$.

A slightly modified version of (3.15) has been introduced in Heesen and Janssen [28]. In their model, it served as estimator for the random number of true null hypotheses $N_0$.

In Model 1.4 (a) the original Storey estimator $\hat{\pi}_0^{St,o.}(\lambda)$ exhibits a very nice behavior if all false $p$-values are almost surely located in $[0, \lambda]$. The new estimator $\hat{\varpi}(t, \lambda, \gamma)$ shows a very nice behavior in the model of Proposition 3.7. But the assumption of false $p$-values, which are less than or equal to $\lambda$, seems to be far more realistic than the assumptions of Proposition 3.7. We will now weaken the assumptions of Proposition 3.7 and investigate the behavior of the new estimators (3.11).

**Theorem 3.10**
*(a) Consider Model 3.3 and let* $0 < \lambda < \gamma \leq 1$. *Furthermore, assume that the df* $F_0$ *of true p-values is convex and differentiable. Then* $\hat{\varpi}(t,\lambda,\gamma)$ *is a conservatively biased estimator for* $\pi_0 F_0(t)$ *for all* $0 < t \leq \lambda$, *i.e. we have*

$$\mathbb{E}\left(\hat{\varpi}(t,\lambda,\gamma)\right) \geq \pi_0 F_0(t). \tag{3.16}$$

*(b) Under the assumptions of (a), the estimator* $\hat{\varpi}(t,\lambda,\gamma)$ *has a lower bias than* $\hat{\pi}_0^{St,o.}(\lambda) \cdot t$, *iff*

$$\frac{\gamma-\lambda}{1-\lambda} + \frac{1-\gamma}{1-\lambda}F(\lambda) \geq F(\gamma). \tag{3.17}$$

**Proof**. (a) By our assumptions on $F_0$, observe that

$$\frac{F_0(\gamma)-F_0(\lambda)}{\gamma-\lambda} \geq F_0'(\lambda) \quad \text{and} \quad \frac{F_0(t)}{t} = \frac{F_0(t)-F_0(0)}{t} \leq F_0'(t).$$

Thus, we obtain

$$\begin{aligned}
&\mathbb{E}\left(\hat{\varpi}(t,\lambda,\gamma) - \pi_0 F_0(t)\right) \\
&= \frac{\pi_0 F_0(\gamma) + (1-\pi_0)F_1(\gamma) - \pi_0 F_0(\lambda) - (1-\pi_0)F_1(\lambda)}{\gamma-\lambda} t - \pi_0 F_0(t) \\
&\geq \pi_0 t \left(\frac{F_0(\gamma)-F_0(\lambda)}{\gamma-\lambda} - \frac{F_0(t)}{t}\right) \geq \pi_0 t \left(F_0'(\lambda) - F_0'(t)\right) \geq 0
\end{aligned}$$

since $F_0'$ is non decreasing for convex differentiable functions.
(b) By definition, we have

$$\begin{aligned}
0 \;&\leq\; \mathbb{E}\left(\hat{\pi}_0^{St,o.}(\lambda)\cdot t - \hat{\varpi}(t,\lambda,\gamma)\right) \\
&=\; \frac{t}{(1-\lambda)(\gamma-\lambda)}\left(\gamma-\lambda+(1-\gamma)F(\lambda)-(1-\lambda)F(\gamma)\right)
\end{aligned}$$

iff (3.17). □

**Remark 3.11**
(a) Convex dfs of true $p$-values occur in a broad and intuitive setting of testing problems. We will go into details in Section 4.1.
(b) The smaller $F(\gamma)$ and hence the more $p$-values located in the upper tail, the better $\hat{\varpi}(t,\lambda,\gamma)$ becomes in terms of bias.

Some more general conditions for a conservative bias of several estimators are treated in Proposition 3.17.

The possible choice of $\lambda$ and $\gamma$ leads to a more complex bias variance trade-off as in the case for $\hat{\pi}_0^{St,o.}$ since the bias is particularly influenced by the choice of $\gamma$. At this point, we propose to use $\lambda = 0.5$ and $\gamma = 0.95$ (or $\gamma = 0.9$) for the new estimator $\hat{\varpi}(t, \lambda, \gamma)$. Based on the original Storey estimator $\hat{\pi}_0^{St,o.}(\lambda) \cdot t$, it has only a slightly increased variance but excludes a substantial part of the upper tail of the ecdf $\hat{F}_n$ which may include too many true $p$-values, in the sense that their part of the distribution does not affect $FDR(t)$ for small $t$.

As estimator for $FDR(t)$ we propose

$$\widehat{FDR}(t) = \begin{cases} \frac{\hat{\varpi}(t,\lambda,\gamma)}{\hat{F}_n(t) \vee \frac{1}{n}}, & 0 \leq t \leq \lambda, \\ 1, & \lambda < t \leq 1. \end{cases} \tag{3.18}$$

As seen in Proposition 3.7, the new estimator $\hat{\varpi}(t, \lambda, \gamma)$ is unbiased in a specific submodel for $t \leq \gamma^*$. Since we try to choose $\gamma$ close to the unknown parameter $\gamma^*$ we expect $\widehat{FDR}(t)$ not to be accurate for $t > \gamma$. Hence, one should rather estimate $FDR(t)$ for $t > \gamma$ by the conservative estimate 1. The estimator in (3.18) is actually a little more conservative and estimates $FDR(t)$ for $t > \lambda$ by 1. A similar approach is described in Storey et al. [61] who redefined $\widehat{FDR}^{St}(t) = 1$ for $t \geq \lambda$ when $\hat{\pi}_0^{St,o.}(\lambda)$ is used in $\widehat{FDR}^{St}(t)$. Note that $\gamma^*$ can only be defined in a specific submodel of Model 3.3.

$FDR(t)$ is an interesting value, but the FDR of the corresponding test, which rejects every $p$-value $p_i \leq t$, is not bounded by any predetermined level $\alpha$. Therefore, one would like to have a multiple testing procedure which exhausts a predetermined FDR level $\alpha$ as good as possible. This leads to considerations for finite sample FDR control which is the topic of the next chapter. Here, we merely show how to get from $\widehat{FDR}(t)$ to adaptive SU tests. Storey et al. [61] propose to reject every $p$-value with

$$p_i \leq \sup\left\{0 \leq t \leq 1 \, : \, \widehat{FDR}^{St}(t) \leq \alpha\right\} \tag{3.19}$$

for their estimator $\widehat{FDR}^{St}(t)$ and show that this procedure is the adaptive SU test with critical values

$$\alpha_{i:n} = \frac{i}{n\hat{\pi}_0^{St,o.}(\lambda)}\alpha, \quad 1 \leq i \leq n, \tag{3.20}$$

see Storey et al. [61, Lemma 2]. An analogue statement applies to (3.18).

**Theorem 3.12** (cf. Lemma 2 in Storey et al. [61])
*Consider the estimator (3.18). The multiple test procedure which rejects every p-value with*

$$p_i \leq \sup\left\{0 \leq t \leq 1 \,:\, \widehat{FDR}(t) \leq \alpha\right\} \tag{3.21}$$

*is the adaptive SU test with critical values*

$$\alpha_{i:n} = \left(\frac{i}{n\hat{\varpi}(\lambda,\gamma)}\alpha\right) \wedge \lambda, \quad 1 \leq i \leq n. \tag{3.22}$$

**Proof**. The statement follows directly from the proof of Lemma 2 of Storey et al. [61] since $\hat{\varpi}(t,\lambda,\gamma) = \hat{\varpi}(\lambda,\gamma) \cdot t$ is linear in $t$. One merely has to replace the estimator $\hat{\pi}_0^{St,o.}(\lambda)$ by $\hat{\varpi}(\lambda,\gamma)$ and take into account that $\widehat{FDR}(t) = 1 > \alpha$ for $t \geq \lambda$. □

**Remark 3.13**
Again, note that every estimator $\hat{\varpi}$ considered in this chapter is of the form $\hat{\varpi}(t,\cdot) = t \cdot \hat{\varpi}(\cdot)$. Thus, Theorem 3.12 also applies if we replace $\hat{\varpi}(t,\lambda,\gamma)$ in (3.18) by any estimator $\hat{\varpi}(t,\cdot) = t \cdot \hat{\varpi}(\cdot)$ which will be studied in the next Sections 3.3 and 3.4.

Moreover, to obtain finite sample FDR control, Storey et al. [61] modified $\hat{\pi}_0^{St,o.}$ and $\widehat{FDR}^{St}(t)$, respectively. We will also modify our estimators, cf. Chapter 4. In the next section we first introduce different estimators of integral type.

## 3.3 Integral type estimators

The tuning parameter $\lambda$ of the estimator $\hat{\pi}_0^{St,o.}(\lambda)$ is often chosen by a bias variance trade-off. The smaller $\lambda$ the larger the bias and the smaller $\lambda$ the smaller the variance of $\hat{\pi}_0^{St,o.}(\lambda)$. Moreover, under certain assumptions we have shown that $\hat{\pi}_0^{St,o.}(\lambda)$ is the best estimator in a specific class of estimators. If one believes that there are no false $p$-values above $\lambda$ and one would like to consider only the information of the ecdf $\hat{F}_n$ above $\lambda$, then Theorem 3.2 advises to take $\hat{\pi}_0^{St,o.}(\lambda)$. However, $\lambda$ has to be chosen heuristically in practice.

Instead of choosing a potential optimal $\lambda$ in this fashion and then applying $\hat{\pi}_0^{St,o.}(\lambda)$, we consider the following approach. Let $Q$ be a fixed probability measure on $[c_0, c_1]$ with $0 < c_0 \leq c_1 < 1$ and define the estimator

$$\hat{\pi}_0^{Int}(Q) = \int \hat{\pi}_0^{St,o.}(\lambda) Q(d\lambda). \tag{3.23}$$

The advantage of this estimator is that we are able to incorporate the original Storey estimator also for $\lambda$ that is little smaller than usually used. For this approach, we recommend to incorporate $\hat{\pi}_0^{St,o.}(\lambda)$ with small $\lambda$ only with small weights. If such a $\lambda$ is too small, in the sense that $\hat{\pi}_0^{St,o.}(\lambda)$ has a too large bias in comparison to standard $\lambda$, then it does not account so much for the complete estimation. Often $\lambda = 0.5$ is chosen as parameter for $\hat{\pi}_0^{St,o.}$. Based on this, we may consider

$$\hat{\pi}_0^{Int}\left(\frac{1}{10}\varepsilon_{0.4} + \frac{8}{10}\varepsilon_{0.5} + \frac{1}{10}\varepsilon_{0.6}\right) = \frac{1}{10}\hat{\pi}_0^{St,o.}(0.4) + \frac{8}{10}\hat{\pi}_0^{St,o.}(0.5) + \frac{1}{10}\hat{\pi}_0^{St,o.}(0.6)$$

which also takes $\lambda = 0.4$ and $\lambda = 0.6$ into account.

The same considerations can be done for $\hat{\varpi}(\lambda, \gamma)$ particularly with regard to $F_0$ and $\gamma$. Let $Q$ be a fixed probability measure on $[c_0, 1]^2$ with $0 < c_0 < 1$ and $Q(\{(\lambda, \gamma) : \gamma - \lambda > \epsilon\}) = 1$ for some $\epsilon > 0$. The tuning parameters $\lambda$ and $\gamma$ should never be too close together since this would dramatically increase the variance of $\hat{\varpi}(\lambda, \gamma)$. Let us then define the estimator

$$\hat{\varpi}^{Int}(Q) = \int \hat{\varpi}(\lambda, \gamma) Q(d(\lambda, \gamma)) \tag{3.24}$$

which is a natural generalization of $\pi_0^{Int}$. If $Q([c_0, 1] \times \{1\}) = 1$, then $\hat{\varpi}^{Int}(Q) = \hat{\pi}_0^{Int}(Q^{\mathrm{Pr}_1})$, where $\mathrm{Pr}_1$ is the projection $(\lambda, \gamma) \mapsto \lambda$.

**Theorem 3.14**

*Consider Model 3.3 and let the above assumptions hold for (3.24). Then*

$$\mathbb{E}\left(\hat{\varpi}^{Int}(Q)\right) = \int \frac{F(\gamma) - F(\lambda)}{\gamma - \lambda} Q(d(\lambda, \gamma)) \tag{3.25}$$

*and*

$$\begin{aligned} &Var\left(\hat{\varpi}^{Int}(Q)\right) \\ &= \frac{1}{n}\int\int \frac{P\big((\lambda, \gamma] \cap (\lambda', \gamma']\big)}{(\gamma - \lambda)(\gamma' - \lambda')} Q(d(\lambda, \gamma))Q(d(\lambda', \gamma')) - \frac{1}{n}\mathbb{E}\left(\hat{\varpi}^{Int}(Q)\right)^2, \end{aligned} \tag{3.26}$$

*where $P$ is the corresponding probability measure of the df $F$.*

**Proof**. Formula (3.25) follows directly by definition of $\hat{\varpi}(\lambda,\gamma)$ and Fubini's Theorem. Here, $(\lambda,\gamma)$ can be regarded as a $Q$ random variable. In this sense, let $(\lambda',\gamma')$ be an independent copy of $(\lambda,\gamma)$. Hence, by Fubini's Theorem, we obtain

$$\mathbb{E}\left(\hat{\varpi}^{Int}(Q)^2\right) = \mathbb{E}\left(\int \hat{\varpi}(\lambda,\gamma)Q(d(\lambda,\gamma)) \cdot \int \hat{\varpi}(\lambda',\gamma')Q(d(\lambda',\gamma'))\right) \quad (3.27)$$

$$= \mathbb{E}\left(\int\int \hat{\varpi}(\lambda,\gamma)\hat{\varpi}(\lambda',\gamma')\, Q(d(\lambda,\gamma))Q(d(\lambda',\gamma'))\right) \quad (3.28)$$

$$= \int\int \mathbb{E}\big[\hat{\varpi}(\lambda,\gamma)\hat{\varpi}(\lambda',\gamma')\big] Q(d(\lambda,\gamma))Q(d(\lambda',\gamma')). \quad (3.29)$$

Let $0 \le x \le y \le 1$. Since

$$\big(n\hat{F}_n(x), n(\hat{F}_n(y) - \hat{F}_n(x)), n - n\hat{F}_n(y)\big) \sim \mathcal{M}\big(n, F(x), F(y) - F(x), 1 - F(y)\big)$$

is distributed according to the multinomial distribution observe that

$$\mathrm{Cov}\big(n\hat{F}_n(x), n(\hat{F}_n(y) - \hat{F}_n(x))\big) = -nF(x)(F(y) - F(x)).$$

Thus, we have

$$\mathbb{E}\left(\hat{F}_n(x)\hat{F}_n(y)\right) = \frac{1}{n^2}\mathbb{E}\left(n\hat{F}_n(x)\big(n\hat{F}_n(y) - n\hat{F}_n(x)\big)\right) + \mathbb{E}\left(\hat{F}_n(x)^2\right)$$

$$= -\frac{1}{n}F(x)(F(y) - F(x)) + F(x)(F(y) - F(x)) + \frac{1}{n}F(x)(1 - F(x)) + F(x)^2$$

$$= (1 - \frac{1}{n})F(x)F(y) + \frac{1}{n}F(x)$$

and hence

$$\mathbb{E}\big[\hat{\varpi}(\lambda,\gamma)\hat{\varpi}(\lambda',\gamma')\big] = \mathbb{E}\left(\frac{\hat{F}_n(\gamma) - \hat{F}_n(\lambda)}{\gamma - \lambda} \cdot \frac{\hat{F}_n(\gamma') - \hat{F}_n(\lambda')}{\gamma' - \lambda'}\right)$$

$$= \left(1 - \frac{1}{n}\right) \cdot \frac{F(\gamma) - F(\lambda)}{\gamma - \lambda} \cdot \frac{F(\gamma') - F(\lambda')}{\gamma' - \lambda'}$$

$$+ \frac{F(\min\{\gamma,\gamma'\}) - F(\min\{\gamma,\lambda'\}) - F(\min\{\lambda,\gamma'\}) + F(\min\{\lambda,\lambda'\})}{n(\gamma - \lambda)(\gamma' - \lambda')}.$$

Moreover, observe that

$$
\begin{aligned}
&F(\min\{\gamma,\gamma'\}) - F(\min\{\gamma,\lambda'\}) - F(\min\{\lambda,\gamma'\}) + F(\min\{\lambda,\lambda'\}) \\
&= \begin{cases} F(\gamma) - F(\gamma) - F(\lambda) + F(\lambda) = 0 & \text{for } \lambda < \gamma \le \lambda' < \gamma', \\ F(\gamma') - F(\lambda') - F(\lambda) + F(\lambda) = F(\gamma') - F(\lambda') & \text{for } \lambda \le \lambda' < \gamma' \le \gamma, \\ F(\gamma) - F(\lambda') - F(\lambda) + F(\lambda) = F(\gamma) - F(\lambda') & \text{for } \lambda \le \lambda' \le \gamma \le \gamma' \end{cases} \\
&= P\big((\lambda,\gamma] \cap (\lambda',\gamma']\big).
\end{aligned}
$$

Note that in the strict sense there exist 3 more cases, but the other cases follow directly by interchanging $(\lambda,\gamma)$ and $(\lambda',\gamma')$. Altogether, we obtain

$$
(3.29) = (1-\frac{1}{n})\mathbb{E}\left(\hat{\varpi}^{Int}(Q)\right)^2 + \int\int \frac{P\big((\lambda,\gamma]\cap(\lambda',\gamma']\big)}{n(\gamma-\lambda)(\gamma'-\lambda')} Q(d(\lambda,\gamma))Q(d(\lambda',\gamma'))
$$

which proves the assertion. □

**Remark 3.15**
(a) The integral type estimator $\hat{\varpi}^{Int}(Q)$ is unbiased for $\pi_0(1-\pi_2)$ under the assumptions of Proposition 3.7 if in addition $Q([\lambda^*,\gamma^*]^2) = 1$.
(b) For $\hat{\pi}_0^{Int}(Q)$ the statements from Theorem 3.14 reduce to

$$
\mathbb{E}\left(\hat{\pi}_0^{Int}(Q)\right) = \int \frac{1-F(\lambda)}{1-\lambda} Q(d\lambda)
$$

and

$$
\mathrm{Var}\left(\hat{\pi}_0^{Int}(Q)\right) = \frac{1}{n}\int\int \frac{1-F(\max\{\lambda,\lambda'\})}{(1-\lambda)(1-\lambda')} Q(d\lambda)Q(d\lambda') - \frac{1}{n}\mathbb{E}\left(\hat{\pi}_0^{Int}(Q)\right)^2.
$$

The integral type estimator $\hat{\pi}_0^{Int}(Q)$ is unbiased for $\pi_0$ under Model 1.4 (a) if in addition $F_1(\lambda^*) = 1$ and $Q([\lambda^*,1)) = 1$ hold for some $\lambda^*$.

**Corollary 3.16** (of Theorem 3.10)
*Consider Model 3.3 and assume that the df $F_0$ of true p-values is convex and differentiable. Moreover, let $Q([\lambda^*,1]^2) = 1$ for some $0 < \lambda^* < 1$. Then $\hat{\varpi}^{Int}(t,Q) = \hat{\varpi}^{Int}(Q)\cdot t$ is a conservatively biased estimator for $\pi_0 F_0(t)$ in the sense of Theorem 3.10 for all $0 \le t \le \lambda^*$.*

**Proof**. The statement follows directly by Theorem 3.10 and Fubini's Theorem. □

## 3.4 Other estimators

Several estimators for $\pi_0$ and related parameters can be found in the literature. In particular, a model similar to Model 1.4 (b) with deterministic $N_0 = n_0$ has often been considered and estimators $\hat{n}_0$ for $n_0$ were developed. As already mentioned, $\pi_0$ often denotes the portion $\frac{n_0}{n}$, whereas in this chapter, it denotes a parameter of Model 1.4 (a) and 3.3. Nonetheless, the estimators for $n_0$ and $\pi_0$ can be rewritten by $\hat{\pi}_0 = \frac{1}{n}\hat{n}_0$ regardless of whether $\pi_0$ denotes the parameter or the portion. In other words, the estimator $\hat{\pi}_0$ yields a reasonable estimator $\hat{n}_0$ and vice versa. This also applies to the estimation of $N_0$ in the BI and gBI Model which were defined in Section 2.1. In the following, we give a brief discussion of those estimators and introduce some generalized versions.

### 3.4.1 Estimators of Zeisel, Zuk and Domany

Zeisel et al. [65] developed two estimators for the number of true null hypotheses $n_0$ based on

$$\hat{n}_0^{ZZD,1} = 2\sum_{i=1}^{n} p_i \tag{3.30}$$

and

$$\hat{n}_0^{ZZD,2} = -\sum_{i=1}^{n} \log(1 - p_i). \tag{3.31}$$

The first estimator $\hat{n}_0^{ZZD,1}$ has already been used by Pounds and Cheng [44] who also tried to improve the estimation of $\pi_0$ and related terms when the true $p$-values are allowed to be stochastically larger than the uniform distribution. In Model 1.4 (a) and 3.3, the corresponding estimators for $\pi_0$ would be written as

$$\hat{\pi}_0^{ZZD,i} = \frac{1}{n}\hat{n}_0^{ZZD,i}, \quad i = 1, 2. \tag{3.32}$$

Zeisel et al. [65] argue that false $p$-values which are usually considered to be stochastically smaller than the uniform distribution thus have only a weak influence on both estimators and they hence have only a small conservative bias. Let us now switch to Model 3.3 again. Since the $p$-values close to 1 have a strong influence on both estimators (especially for the logarithmic one), a change to null $p$-values which are stochastically larger than the uniform distribution would increase the bias and the estimators get worse. Moreover, we already showed that an estimation of $\pi_0 F_0(t)$ instead of $\pi_0$ is preferable. Since $F_0(t) \leq t$ holds,

$\hat{\pi}_0^{ZZD,i} \cdot t$, $i = 1, 2$, often have a bias which is increased once more. It follows that these estimators are not suitable for such generalized models.

Zeisel et al. [65] place special emphasis on conservatively biased estimators. Such estimators are very easy to obtain. The next proposition introduces a general class of estimators for $\pi_0 F_0(t)$ and gives sufficient conditions for achieving a conservative bias.

**Proposition 3.17**

*Consider Model 3.3. Let* $g : [0,1] \to \mathbb{R}_{\geq 0}$ *be a non negative function with* $\int g d\lambda\!\!\lambda_{|(0,1)} = 1$ *and introduce the following estimator*

$$\hat{\varpi}(t,g) = \frac{t}{n}\sum_{i=1}^{n} g(p_i). \tag{3.33}$$

*Then* $\mathbb{E}\left(\hat{\varpi}(t,g)\right) \geq \pi_0 F_0(t)$ *holds*

*(a) for all* $t \leq 1$ *if* $g$ *is non-decreasing,*

*(b) for all* $t \leq \lambda$ *if* $F_0$ *is a convex df with density* $f_0$ *and* $g_{|[0,\lambda]} = 0$*,*

*(c) for all* $t \leq \lambda$ *if* $F_0$ *is a convex df which posses a density* $f_0$ *on the interval* $[0,1)$ *and* $g_{|[0,\lambda]\cup(1-\epsilon,1]} = 0$ *for some* $\epsilon > 0$*.*

**Proof**. (a) This assertion follows from

$$\mathbb{E}\left(\hat{\varpi}(t,g)\right) = t(1-\pi_0)\mathbb{E}_{F_1}\left(g\right) + t\pi_0\mathbb{E}_{F_0}\left(g\right) \geq t\pi_0 \int_0^1 g(s)ds \geq F_0(t)\pi_0$$

since $F_0$ is stochastically larger than the uniform distribution and $g$ is non-decreasing.

(b) Without restrictions, we may assume that $f_0$ is non decreasing on $[0,1]$ since $F_0$ is convex. Hence, we obtain

$$\begin{aligned}\mathbb{E}\left(\hat{\varpi}(t,g)\right) &\geq t\pi_0 \int_\lambda^1 g(s)F_0(ds) = t\pi_0 \int_\lambda^1 g(s)f_0(s)ds \\ &\geq t\pi_0 f_0(t)\int_\lambda^1 g(s)ds = t\pi_0 f_0(t) \\ &\geq \pi_0 \int_0^t f_0(s)ds = \pi_0 F_0(t).\end{aligned} \tag{3.34}$$

(c) Here, we may assume that $f_0$ is non decreasing on $[0,1)$ and the assertion follows by the same arguments as in (b) by replacing the upper integral bound 1 in (3.34) by $1-\epsilon$. □

It is very easy to obtain conservatively biased estimators, but FDR control for the corresponding adaptive SU tests is still far away. However, we would like to have estimators with small conservative bias over a wide range of possible distributions in Model 3.3. For this purpose we use heuristics. At least in a specific submodel of Model 3.3, we have already shown that the estimators $\hat{\varpi}(t,\lambda,\gamma)$ are unbiased and have minimum variance among all unbiased estimators, see Theorem 3.8.

It is difficult to adapt $\hat{\pi}_0^{ZZD,i}$, $i=1,2$, to Model 3.3 since large $p$-values greatly increase the estimate. However, as we have seen, large $p$-values should rather be excluded or handled carefully. An estimator that tries to avoid these problems is given by

$$\hat{\varpi}^{log}(\lambda,\epsilon) = -c(\lambda,\epsilon)\frac{1}{n}\sum_{i=1}^{n}\log\left(\frac{p_i-\lambda}{1-\lambda}\right)\mathbb{1}_{[\lambda+\epsilon,1]}(p_i)$$

with $0<\lambda<1$, $\epsilon>0$, $\lambda+\epsilon<1$, where the convention $\infty\cdot 0=0$ is utilized. The constant $c(\lambda,\epsilon)$ is chosen such that the assumption of Proposition 3.17 is fulfilled. The most influential $p$-values lie in the neighborhood of $\lambda+\epsilon$ and large $p$-values have less influence on the estimate.

Note that the choice of $g(p_i)=\frac{\mathbb{1}\{\lambda<p_i\leq\gamma\}}{\gamma-\lambda}$ in (3.33) just yields $\hat{\varpi}(t,\lambda,\gamma)$ and each $p$-value $p_i\in(\lambda,\gamma]$ has the same influence on the estimate.

### 3.4.2 Estimator of Benjamini, Krieger and Yekutieli

Motivated by Schweder and Spjøtvoll [55], Benjamini and Hochberg [3] and Storey [59]

$$\hat{n}_0^{BKY} = \frac{n-k}{1-p_{k:n}} \tag{3.35}$$

with $k=\lfloor\frac{n}{2}\rfloor$ (and other choices of $k$) has been introduced by Benjamini et al. [4] as estimator for the number of true null hypotheses $n_0$. As model they considered a relaxed form of Model 1.4 (b), where the true $p$-values are allowed to be discrete and hence stochastically larger than the uniform distribution. It is well known that under certain regularity assumptions we basically have

$$\hat{\pi}_0^{St,o.}(p_{k:n}) \mathrel{\widehat{=}} \frac{1}{n}\hat{n}_0^{BKY}. \tag{3.36}$$

In this manner we get a new estimator by adjusting $\hat{\varpi}(\lambda,\gamma)$. Therefore, introduce

$$\hat{\varpi}^{BKY}(t,k_1,k_2) = \frac{t}{n}\cdot\frac{k_2-k_1}{(p_{k_2:n}-p_{k_1:n})\vee\frac{1}{n}} \tag{3.37}$$

with $k_1 = \lfloor \frac{n}{2} \rfloor$ and $k_2 = \lfloor 0.95n \rfloor$ as estimator for $\pi_0 F_0(t)$. Langaas et al. [32] already introduced

$$\hat{\pi}_0 = \min_{l \leq n} \frac{\hat{F}_n(p_{n:n}) - \hat{F}_n(p_{l:n})}{p_{n:n} - p_{l:n}} = \min_{l \leq n} \frac{1 - \hat{F}_n(p_{l:n})}{p_{n:n} - p_{l:n}}$$

as estimator for the parameter $\pi_0$ which has a slight resemblance to (3.37). This estimator is based on a decreasing density estimation by a non parametric maximum likelihood estimator, where the false null $p$-values are assumed to have a density which is decreasing.

Note that $\hat{n}_0^{BKY}$ may be very conservative if $n - n_0 > k$. In a $DU(n, n_0)$ configuration with $n - n_0 > k$ we obtain $\hat{n}_0^{BKY} = n - k > n_0$. Moreover, if $n - n_0 < k$ and the true $p$-values stochastically increase, then $p_{k:n}$ and hence $\hat{n}_0^{BKY}$ tend to be larger and the estimation may be very conservative again. Similar to (3.36) we basically have

$$\hat{\varpi}(t, p_{k_1:n}, p_{k_2:n}) \mathrel{\widehat{=}} \hat{\varpi}^{BKY}(t, k_1, k_2) \tag{3.38}$$

under certain regularity assumptions and some of the properties of $\hat{\varpi}(t, \lambda, \gamma)$ carry over to $\hat{\varpi}^{BKY}(t, k_1, k_2)$. Hence, the choice of $\hat{\varpi}^{BKY}(t, k_1, k_2)$ and related estimators as estimator for adaptive SU tests may also have some advantages.

## 3.5 Estimation for a generalized error rate

Until now the present chapter only considered the FDR as error rate. We derived (3.10) as estimator for the FDR of the multiple test which rejects every $p$-value which is less than or equal to a fixed threshold $t$. Theorem 3.12 then gave us the relation between the estimator (3.10) and adaptive SU tests. In particular, this theorem motivates the use of $\hat{\varpi}(\lambda, \gamma)$ instead of $\hat{\pi}_0^{St,o.}(\lambda)$ as estimator for the adaptive SU tests. Finally, we will briefly study the generalized error rates of Meskaldji et al. [39] which have been introduced in Remark 2.12.

**Theorem 3.18**

*Consider Model 3.3 and let $\rho : \{0, \ldots, n\} \to (0, \frac{n}{\alpha})$ be a non decreasing function. If $F(t) > 0$, then we have*

$$\mathbb{E}\left(\frac{V(t)}{\rho(R(t))}\right) = \frac{\pi_0 F_0(t)}{F(t)} g(F(t), \rho), \tag{3.39}$$

*where $g(F(t), \rho) = nF(t)E(\frac{1}{\rho(B+1)})$ and $B \sim \mathcal{B}(n-1, F(t))$.*

**Proof**. The first part of the proof of (3.39) is similar to the proof of Theorem 1 of Storey [60]. Observe that

$$\begin{aligned}
\mathbb{E}\left(\frac{V(t)}{\rho(R(t))}\right) &= \sum_{k=1}^{n} \mathbb{E}\left(\frac{V(t)}{\rho(k)}\Big| R(t)=k\right) P(R(t)=k) & (3.40)\\
&= \sum_{k=1}^{n} \frac{k}{\rho(k)} P(H_1=0|p_1\le t) P(R(t)=k) & (3.41)\\
&= \frac{\pi_0 F_0(t)}{F(t)} \sum_{k=1}^{n} \frac{k}{\rho(k)} P(R(t)=k), & (3.42)
\end{aligned}$$

where (3.41) follows by the same arguments as in the proof of Storey [60, Theorem 1] which we omit at this point. Furthermore, we obtain the remaining statement

$$\begin{aligned}
\sum_{k=1}^{n} \frac{k}{\rho(k)} P(R(t)=k) &= \sum_{k=1}^{n} \frac{k}{\rho(k)} \binom{n}{k} F(t)^k (1-F(t))^{n-k}\\
&= nF(t) \sum_{j=0}^{n-1} \frac{1}{\rho(j+1)} \binom{n-1}{j} F(t)^j (1-F(t))^{n-1-j}\\
&= g(F(t),\rho).
\end{aligned}$$

□

Observe that by Theorem 3.18

$$FDR(t) = \frac{\pi_0 F_0(t)}{F(t)} \cdot (1-(1-F(t))^n) \quad \text{and} \quad ENFR(t) = n\pi_0 F_0(t), \qquad (3.43)$$

where $ENFR(t)$ denotes the ENFR of the multiple test which rejects every $p$-value that is less than or equal to the fixed threshold $t$. By similar considerations as in Section 3.2, it follows that a reasonable estimator for $ENFR(t)$ is given by

$$\widehat{ENFR}(t) = n\hat{\varpi}(t,\lambda,\gamma). \qquad (3.44)$$

Based on the estimator (3.44), the adaptive test which rejects every $p$-value

$$p_i \le \sup\left\{0 \le t \le 1 \,:\, \widehat{ENFR}(t) \le \widetilde{\alpha}\right\} \qquad (3.45)$$

with $\widetilde{\alpha} \in (0,n)$ is an adaptive single step test with adaptive threshold $\frac{\widetilde{\alpha}}{n} \cdot \hat{\varpi}(\lambda,\gamma)$.

**Remark 3.19**
For the other error rates, an additional non trivial factor comes into play. At this point, it is not clear which estimator should be used for the estimation of $\mathbb{E}(\frac{V(t)}{\rho(R(t))})$ and hence for adaptive SU tests. However, in terms of finite sample control of the error rate $\mathbb{E}(\frac{V}{\rho(R)})$, we will show in Section 4.2 that the estimators for the FDR and ENFR work.

# Chapter 4

# Generalized error rate control of adaptive SU tests

In Chapter 3 we derived $\widehat{FDR}(t)$, see (3.10), as estimator for the FDR of the multiple test which rejects every $p$-value which is less than or equal to a fixed threshold $t$. Moreover, we introduced and motivated $\hat{\varpi}(t, \lambda, \gamma)$, see (3.11), as estimator for the crucial parameter $\pi_0 F_0(t)$ of Model 3.3 and used this estimator within the estimator $\widehat{FDR}(t)$. Theorem 3.12 then gave us the relation between $\widehat{FDR}(t)$ and adaptive SU tests. In particular, this theorem motivated the use of $n\hat{\varpi}(\lambda, \gamma)$, see in (3.15), as estimator for adaptive SU tests.

This section is devoted to finite sample FDR control and generalized error rate control of adaptive SU test under the gBI Model. We establish a new sufficient condition for generalized error rate control of adaptive SU tests which does not need the assumption of non decreasing estimators. In comparison to previous conditions, it is shown that this condition is more powerful in the situation of the gBI Model. Furthermore, we prove that a slightly modified version of $n\hat{\varpi}(\lambda, \gamma)$ satisfies this condition. It turns out that the selection of the estimator for the adaptive SU test may even be performed in a data dependent manner which leads to dynamic adaptive SU tests. Finally, a reasonable selection method with generalized error rate control is developed in a practical guide and we give a small simulation study.

The results of this chapter are based on Heesen and Janssen [28]. In their work, they focus on the BI Model, whereas the present results are based on the gBI Model, where the marginal distributions of the true $p$-values may be stochastically larger than the uniform distribution. In comparison to the BI Model, new interesting and intuitive opportunities arise for the gBI Model, including Model 3.3. These opportunities particularly affect the data dependent selection of the estimator of the adaptive SU test.

## 4.1 Model assumptions

Let us briefly recall the **gBI Model**: Conditioned under $H$ let $p_i$, $i \in I_0$, and $p_{I_1}$ be **jointly independent**. Moreover, let

$$p_i \sim F_{0,H,i}, \quad i \in I_0, \tag{4.1}$$

with dfs $F_{0,H,i}(t) \leq t$, $t \in [0,1]$.

Most of the adaptive SU tests which are considered below will not have error rate control at level $\alpha$ in the entire gBI Model and we have to restrict ourselves to specific submodels of the gBI Model. Therefore, we consider some weak forms of convexity for the dfs $F_{0,H,i}$, see Remark 4.5 for a one sided convexity condition. The next theorem shows that even convex dfs $F_{0,H,i}$ occur in a natural and wide range of one sided testing problems. In Chapter 2 and 3 we often used the assumption of convex dfs of true $p$-values in connection with the gBI Model.

**Theorem 4.1**
*Consider $(\Omega, \mathcal{F}, \{P_\vartheta : \vartheta \in \Theta\})$ with $\Theta \subseteq \mathbb{R}$ and monotone likelihood ratio in $T : \Omega \to \mathbb{R}$. Moreover, assume that the distributions $P_\vartheta^T$ have a Lebesgue density and the corresponding dfs $F_\vartheta^T$ are strictly increasing on the interval $S(\vartheta) = (\underline{S}(\vartheta), \bar{S}(\vartheta)) \subset \bar{\mathbb{R}}$, where $F_\vartheta^T(\underline{S}(\vartheta)) = 0$ and $F_\vartheta^T(\bar{S}(\vartheta)) = 1$ for all $\vartheta$. The quantities $\underline{S}(\vartheta)$ and $\bar{S}(\vartheta)$ may be infinite. Let us consider the one-sided testing problem*

$$\mathcal{H} : \{\vartheta \leq \vartheta_0\} \quad \text{versus} \quad \mathcal{K} : \{\vartheta > \vartheta_0\} \tag{4.2}$$

*with p-value*

$$p = 1 - F_{\vartheta_0}^T(T) \tag{4.3}$$

*for some $\vartheta_0 \in \Theta$. Then we have:*

(a) *The distribution $P_{\vartheta_0}^p$ is the uniform distribution on $(0,1)$.*

(b) *The distribution $P_\vartheta^p$ is stochastically larger than $U(0,1)$ for all $\vartheta \leq \vartheta_0$.*

*(c) The df $F^p_\vartheta$ of the p-value p is convex on $[0,1]$ for all $\vartheta \leq \vartheta_0$.*

**Proof**. (a) The statement follows directly since $F^T_{\vartheta_0}$ is continuous and thus we have $F^T_{\vartheta_0}(T) \sim U(0,1)$.
(c) Let us only consider $\vartheta < \vartheta_0$ since $F^p_{\vartheta_0}(x) = x$, $x \in [0,1]$, is obviously convex. We first show the convexity on the open interval $(0,1)$. According to the Lebesgue decomposition and (a) we obtain

$$F^p_\vartheta(x) = \int_0^x \frac{dP^p_\vartheta}{dP^p_{\vartheta_0}}(y)dy + P^p_\vartheta\left([0,x] \cap \left\{y : \frac{dP^p_\vartheta}{dP^p_{\vartheta_0}}(y) = \infty\right\}\right) \tag{4.4}$$

for $x \in [0,1]$. We continue by calculating

$$\frac{dP^p_\vartheta}{dP^p_{\vartheta_0}}(y) = \mathbb{E}_{P^T_{\vartheta_0}}\left(\frac{dP^T_\vartheta}{dP^T_{\vartheta_0}}(T)\Big|p = y\right), \quad y \in [0,1). \tag{4.5}$$

Since the experiment has a monotone likelihood ratio in $T$, observe that there exists a non-decreasing function $H_{\vartheta,\vartheta_0}$ with

$$\frac{dP_{\vartheta_0}}{dP_\vartheta}(\omega) = H_{\vartheta,\vartheta_0}(T(\omega)) \quad P_\vartheta + P_{\vartheta_0} \text{ a.e..}$$

For $y = 0$ observe that $\{p = y\} = \{F^T_{\vartheta_0}(T) = 1\} = \{T \geq \bar{S}(\vartheta_0)\}$. $\bar{S}(\vartheta)$ is non-decreasing since $H_{\vartheta,\vartheta_0}$ is non-decreasing and hence $\{T > \bar{S}(\vartheta_0)\}$ is a null set according to $P_\vartheta + P_{\vartheta_0}$. Moreover, $\{T = \bar{S}(\vartheta_0)\}$ is also a null set according to $P_\vartheta + P_{\vartheta_0}$. Thus, without restrictions, we may set $\frac{dP^T_\vartheta}{dP^T_{\vartheta_0}}(t) = 1/H_{\vartheta,\vartheta_0}(t) = 0$ for all $t \geq \bar{S}(\vartheta)$ and hence $\frac{dP^p_\vartheta}{dP^p_{\vartheta_0}}(0) = 0$ $P^p_{\vartheta_0}$-*a.s.*. For $y \in (0,1)$ observe that we have $\{p = y\} = \{(F^T_{\vartheta_0})^{-1}(1-y) = T\}$ since $F^T_{\vartheta_0}$ is strictly increasing on $S(\vartheta_0)$. Thus, by (4.5) we obtain

$$\frac{dP^p_\vartheta}{dP^p_{\vartheta_0}}(y) = \frac{dP^T_\vartheta}{dP^T_{\vartheta_0}}((F^T_{\vartheta_0})^{-1}(1-y)) = \frac{1}{H_{\vartheta,\vartheta_0}((F^T_{\vartheta_0})^{-1}(1-y))} \quad P^p_{\vartheta_0}\text{-a.s..} \tag{4.6}$$

The right hand side is a non-decreasing function in $y$ since $H_{\vartheta,\vartheta_0}$ and $(F^T_{\vartheta_0})^{-1}$ are non-decreasing. Moreover, since $(F^T_{\vartheta_0})^{-1}(1-y)) \in \overset{o}{S}(\vartheta_0)$ holds and since $P^T_{\vartheta|\overset{o}{S}(\vartheta_0)}$ is absolutely continuous with respect to $P^T_{\vartheta_0}$ we have

$$\frac{dP^p_\vartheta}{dP^p_{\vartheta_0}}(y) < \infty \quad P^p_{\vartheta_0} \text{ a.s.}$$

by (4.6). Thus, without restrictions we can assume $\frac{dP^p_\vartheta}{dP^p_{\vartheta_0}}$ to be non-decreasing and smaller than $\infty$ on $[0,1)$. Hence, by (4.4), we obtain

$$F^p_\vartheta(x) = \int_0^x \frac{dP^p_\vartheta}{dP^p_{\vartheta_0}}(y)dy \tag{4.7}$$

for $x \in (0,1)$, where $F_\vartheta^p$ is convex on $(0,1)$ (see Roberts and Varberg [45], Chapter 1, Section 12, Theorem A for instance).

We finally show convexity on the closed interval $[0,1]$. By (4.7) and since $\frac{dP_\vartheta^p}{dP_{\vartheta_0}^p}(y)$ is bounded for every $0 \leq y \leq \epsilon$ with fixed $0 < \epsilon < 1$, it follows that $F_\vartheta^p$ is continuous in 0. Furthermore, consider the continuous continuation

$$\widetilde{F}_\vartheta^p(x) = \begin{cases} F_\vartheta^p(x), & 0 \leq x < 1, \\ \lim_{y \nearrow 1} F_\vartheta^p(1), & x = 1, \end{cases}$$

of $F_\vartheta^p$ at 1. Then it is easily seen that $\widetilde{F}_\vartheta^p$ fulfills the convexity condition on the closed interval $[0,1]$ by turning to limits since $\widetilde{F}_\vartheta^p$ is continuous and already known to be convex on $(0,1)$. Finally, it is easy to show that $F_\vartheta^p$ is also convex on $[0,1]$ by using $\widetilde{F}_\vartheta^p(1) \leq F_\vartheta^p(1)$.
(b) By the convexity of $F_\vartheta^p$ in combination with $F_\vartheta^p(0) = 0$ and $F_\vartheta^p(1) = 1$ it follows that $F_\vartheta^p(t) \leq t$ holds for all $t \in [0,1]$. Note that the assertion also follows by standard arguments from the one sided testing theory. □

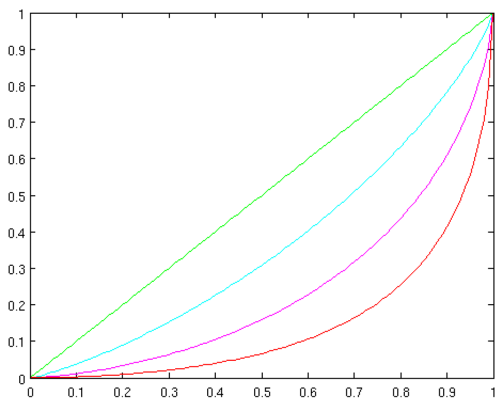



Figure 4.1: Distribution functions of true $p$-values in a one sided normal mean testing problem.

**Example 4.2**
(a) The assumptions of Theorem 4.1 are particularly satisfied for families of normal distributions. Consider the statistical experiment $(\Omega, \mathcal{B}, \{N(\vartheta, 1) : \vartheta \in \mathbb{R}\})$ and the one sided testing problem $\mathcal{H} : \{\vartheta \leq 0\}$ versus $\mathcal{K} : \{\vartheta > 0\}$. It is well known that the experiment has a monotone likelihood ratio in the identity

function. Thus, the $p$-value $p = 1 - \Phi(id)$ has a convex df for every $\vartheta \leq 0$, where $\Phi$ is the df of $N(0,1)$. Figure 4.1 shows the dfs of the $p$-value $p$ for $\vartheta = 0, -0.5, -1$ and $-1.5$ in decreasing order.

(b) Let us combine the situation of (a) with the gBI Model. Assume that we have $n = 1000$ independent $p$-values, where each $p$-value tests a one sided testing problem as described in (a). Furthermore, assume that $n_1 = 150$ of these $p$-values are false and distributed according to $\vartheta = 2$ and $n_0 = 850$ $p$-values are true. Let 250 of the true $p$-values be distributed according to $\vartheta = -1$ and 600 according to $\vartheta = 0$. The true $p$-values with $\vartheta = 0$ may also come from two sided testing problems of the form $\mathcal{H} : \{\vartheta = 0\}$ versus $\mathcal{K} : \{\vartheta \neq 0\}$. Figure 4.2 shows a typical realization of the ecdf of these $p$-values. Notice that this realization

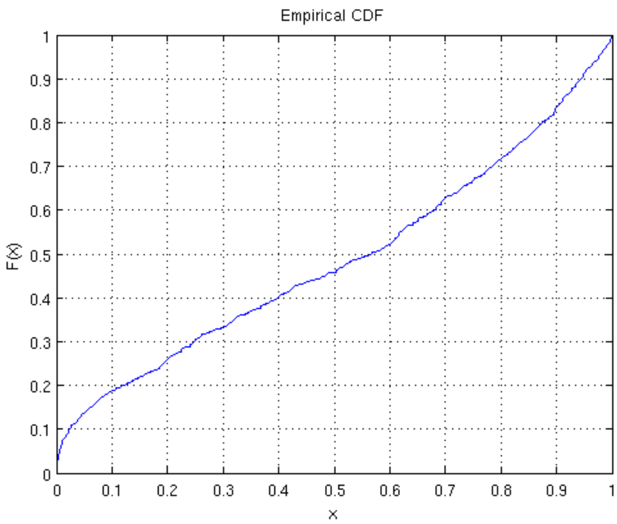



Figure 4.2: Realization of an ecdf of $p$-values in a one sided normal mean testing problem.

of the ecdf is sagging on the second half of the interval and that $\hat{F}_n(0.5) < 0.5$ holds. Further observe that $\hat{F}_n(0.5) < 0.5$ implies $\hat{n}_0(0.5) > n$ for the Storey estimator $\hat{n}_0(\lambda)$, see (1.30). In this situation, the adaptive SU test of Storey with critical values (1.29) and estimator $\hat{n}_0(0.5)$ tends to be more conservative than the BH test. Observe that the ecdf tends to be very steep near 1 due to the convexity of the marginal dfs which are based on $\vartheta = -1$. This fact enables the development of our dynamic adaptive SU tests in Section 4.3 and the effectiveness of the estimator $n\hat{\varpi}(\lambda, \gamma)$.

## 4.2 Conditions for finite sample error rate control and the generalized Storey estimator

We will now develop two conditions. Each of them ensures finite sample error rate control of adaptive SU tests of Storey type. Therefore, let us again consider the generalized error rate $\mathbb{E}\left(\frac{V}{\rho(R)}\right)$ of Meskaldji et al. [39] and let us assume that

$$\rho : \{0, \dots, n\} \to \left(0, \frac{n}{\alpha}\right) \quad \text{is a non decreasing function.} \tag{4.8}$$

As already mentioned, the generalized error rate, based on $\rho$, easily leads to SU tests with critical values $\alpha_{i:n} = \frac{\rho(i)}{n}\alpha$, $i = 1, \dots, n$, which fulfill $\mathbb{E}\left(\frac{V}{\rho(R)}\right) = \frac{\mathbb{E}(N_0)}{n}\alpha$ under the BI Model, cf. Meskaldij et al. [39] and also Remark 2.12. In the following, let us primarily consider the FDR, given by $\rho(i) = i \vee 1$, $i \in \mathbb{N} \cup \{0\}$, and the ENFR, given by $\rho \equiv 1$. At least the theoretical results hold for the entire family of generalized error rates based on (4.8). Again, a multiple test is said to control the generalized error rate $\mathbb{E}\left(\frac{V}{\rho(R)}\right)$ by $\alpha$ if $\mathbb{E}\left(\frac{V}{\rho(R)}\right) \leq \alpha$ holds under the present model. For abbreviation, we only talk about error rate control.

Let us divide the interval $[0, 1]$ into two areas: the rejection area $[0, \lambda]$ and the estimation area $[\lambda, 1]$, where $0 < \lambda < 1$ is a tuning parameter as before. Regardless, $\lambda$ is contained in both areas. The estimation part of the adaptive SU test is then carried out with the $p$-values which lie in the estimation area. Then the rejection part of the adaptive SU test is carried out with the $p$-values which lie in the rejection area. Storey et al. [61] implicitly used this concept.

The estimator $\hat{n}_0$ is based on the ecdf $\hat{F}_n$ on the **estimation area** $[\lambda, 1]$. Therefore, let

$$\hat{n}_0 = \hat{n}_0\left((\hat{F}_n(t))_{t \geq \lambda}\right) > 0 \tag{4.9}$$

be given by a measurable function. Again, note that for the FDR and ENFR the estimator $\hat{n}_0$ should not be regarded as estimator for $N_0$ but in the sense of Section 3.2. Nevertheless, we denote the estimator of the adaptive SU tests by $\hat{n}_0$ since it is not described in detail, at this point, and since it is used for all generalized error rates. For the other generalized error rates than the FDR and ENFR, based on (4.8), it is not clear which estimators may have a nice behavior, see Section 3.5. So far, $\hat{n}_0$ may at least be regarded as estimator for $N_0$.

The adaptive SU test is then based on the data dependent critical values

$$\hat{\alpha}_{i:n} = \left( \frac{\rho(i)}{\hat{n}_0} \alpha \right) \wedge \lambda, \quad 1 \le i \le n, \tag{4.10}$$

and only rejects a certain amount of $p$-values which lie in the **rejection area** $[0, \lambda]$. Each error rate based on $\rho$ leads to its own adaptive SU test with critical values (4.10). In this section we mainly consider a generalized Storey estimator which will be defined later and which is a slightly modified version of $n\hat{\varpi}(\lambda, \gamma)$. Moreover, in Section 4.3 we develop adaptive SU tests with adaptive estimators based on this generalized Storey estimator. We will refer to these tests as dynamic adaptive SU tests.

Observe that the random variables $R$ and $V$ are functions of the $p$-values, i.e. they may be written as $R = R(p)$ and $V = V(p)$. In the following, let $R = R(p)$ and $V = V(p)$ always refer to the present adaptive SU test when it is clear which test is meant. Some techniques and proofs are based on setting one true $p$-value to zero. Therefore, introduce

$$p^{(i)} = (p_1, \ldots, p_{i-1}, 0, p_{i+1}, \ldots, p_n), \tag{4.11}$$

$$R^{(i)} = R(p^{(i)}), \tag{4.12}$$

$$\hat{n}_0^{(i)} = \hat{n}_0(p^{(i)}) \tag{4.13}$$

$$\text{and} \quad \hat{\alpha}_{j:n}^{(i)} = \left( \frac{\rho(j)}{\hat{n}_0^{(i)}} \alpha \right) \wedge \lambda, \quad j = 1, \ldots, n, \tag{4.14}$$

for $i \in I_0$. Furthermore, we will sometimes condition under

$$\mathcal{F}_t = \sigma(H, \mathbb{1}\{p_i \le s\} : s \ge t, \, 1 \le i \le n). \tag{4.15}$$

for some $t \in (0,1)$. Conditioned under $\mathcal{F}_t$ the random variables $H$, $\mathbb{1}\{p_i \le s\}$, $s \ge t$, $1 \le i \le n$ and in particular $\hat{F}_n(t)$ can be treated as fixed values, due to measurability arguments.

Now we are able to formulate both conditions. The first one already showed up in Sarkar [52, Theorem 3.3] for a slightly different setting and only for the FDR. The inequalities of the next two theorems directly yield the desired conditions for the error rate control, see Remark 4.5 (a).

**Theorem 4.3**
*Assume (4.8). Consider the gBI Model and the adaptive SU test with critical values (4.10) and estimator (4.9). Then we have*

$$\mathbb{E}\left(\frac{V}{\rho(R)}\right) \leq \alpha \cdot \mathbb{E}\left(\sum_{i\in I_0} \frac{1}{\hat{n}_0^{(i)}}\right). \tag{4.16}$$

**Proof**. The proof is based on advanced techniques of Sarkar [52] and Benjamini et al. [4]. Conditioned under $H$, observe that the random index set of true $p$-values $I_0$ is fixed. Hence, by Lemma 6.1 (a) of the appendix, we obtain

$$\mathbb{E}\left(\frac{V}{\rho(R)}\Big|H\right) = \sum_{i\in I_0} \mathbb{E}\left(\frac{\mathbb{1}\{p_i \leq \hat{\alpha}_{R:n}\}}{\rho(R)}\Big|H\right) \tag{4.17}$$

$$= \sum_{i\in I_0} \mathbb{E}\left(\frac{\mathbb{1}\{p_i \leq \hat{\alpha}^{(i)}_{R^{(i)}:n}\}}{\rho(R^{(i)})}\Big|H\right). \tag{4.18}$$

Moreover, given $H$, observe that each true $p$-value $p_i$ is independent of all other $p$-values in the gBI Model. From this, Fubini's Theorem and (4.10) it follows that

$$(4.18) = \sum_{i\in I_0} \mathbb{E}\left(\frac{F_{0,H,i}\left(\hat{\alpha}^{(i)}_{R^{(i)}:n}\right)}{\rho(R^{(i)})}\Big|H\right) \leq \sum_{i\in I_0} \mathbb{E}\left(\frac{\hat{\alpha}^{(i)}_{R^{(i)}:n}}{\rho(R^{(i)})}\Big|H\right)$$

$$\leq \sum_{i\in I_0} \mathbb{E}\left(\frac{\alpha}{\hat{n}_0^{(i)}}\Big|H\right).$$

Finally, taking the expectation over $H$ yields the assertion. □

Basically, Sarkar [52, Theorem 3.3] considered the BI Model conditioned under $H$ for the FDR, where he does not need the estimation, rejection area concept. The estimator $\hat{n}_0$ may depend on the entire ecdf $\hat{F}_n$ and the adaptive SU test may reject $p$-values on the entire interval $[0, 1]$. In comparison, the restriction to the rejection area $[0, \lambda]$ in our case is not substantial, since it is often realistic in practice. Even in the FDR case, rejections of $p$-values larger than $\lambda$ are sometimes disliked, since the evidence of a large $p$-value against the corresponding null hypothesis is small. Moreover, the theorem of Sarkar [52] requires estimators $\hat{n}_0$ which are non-decreasing in each $p$-value $p_i$. As already mentioned in Section 1.5, this assumption is very popular in the FDR literature. However, as it will turn out later, some estimators which do not fulfill this assumption are exactly the ones which yield substantial improvements in the gBI Model when the dfs of true $p$-values are convex.

For the next condition and inequality, respectively, we have to make the additional assumption (4.19) below. Although the theorem contains this additional assumption, it has some advantages since the inequality (4.20) turns out to be sharper.

**Theorem 4.4**

*Assume (4.8). Consider the gBI Model and the adaptive SU test with critical values (4.10), estimator (4.9) and* $\lambda \in (0,1)$. *Moreover, assume that the dfs* $F_{0,H,i}$ *fulfill*

$$F_{0,H,i}(t\lambda) \leq t \cdot F_{0,H,i}(\lambda), \quad t \in [0,1],\ i \in I_0. \tag{4.19}$$

*Then we have*

$$\mathbb{E}\left(\frac{V}{\rho(R)}\right) \leq \frac{\alpha}{\lambda} \cdot \mathbb{E}\left(\frac{V(\lambda)}{\hat{n}_0}\right). \tag{4.20}$$

**Proof**. Conditioned under $\mathcal{F}_\lambda$, see (4.15), we have exactly $V(\lambda)$ true $p$-values smaller or equal to $\lambda$ (where $V(\lambda)$ is a fixed number) among $n(\lambda) = n\hat{F}_n(\lambda)$ $p$-values which are smaller or equal to $\lambda$. Without restrictions, we assume $N_0 \geq V(\lambda) > 0$ since everything is obviously true for the excluded cases. Let us now consider new rescaled $p$-values $q_i$, $i = 1, \ldots, n(\lambda)$, defined by

$$q_i = \frac{p_{j_i}}{\lambda}, \quad i = 1, \ldots, n(\lambda),$$

where $\{j_1 < \ldots < j_{n(\lambda)}\} = \{i : p_i \leq \lambda\}$ is the index set of $p$-values which are smaller or equal to $\lambda$. Moreover, define $\widetilde{I}_0 = \{i : j_i \in I_0\}$ as the index set of new $p$-values corresponding to true null hypotheses. Then $q_i$, $i \in \widetilde{I}_0$, and the vector $(q_i : i \notin \widetilde{I}_0)$ of new false $p$-values are jointly independent under the above condition. Furthermore, observe that

$$\begin{aligned} P\left(q_i \leq t \middle| \mathcal{F}_\lambda\right) &= P\left(q_i \leq t \middle| p_{j_i} \leq \lambda, H\right) \\ &= \frac{F_{0,H,j_i}(t\lambda)}{F_{0,H,j_i}(\lambda)} \leq t, \quad t \in [0,1],\ i \in \widetilde{I}_0, \end{aligned}$$

holds by independence and (4.19) under the above assumptions. Without restrictions, we can assume that $F_{0,H,j_i}(\lambda) > 0$ holds since the other case only occurs with probability zero by dropping out all conditions except for $H$.

We now apply Theorem 4.3 for non adaptive SU tests with critical values

$$\alpha^{(q)}_{i:n(\lambda)} = \frac{\rho(i)}{n(\lambda)}\alpha' \quad \text{with} \quad \alpha' = \frac{n(\lambda)}{\lambda \hat{n}_0}\alpha$$

on the $q$'s. By (4.9), the data dependent level $\alpha'$ only depends on the information given by $(n\hat{F}_n(t))_{t\geq\lambda}$. Conditioned under $\mathcal{F}_\lambda$, we thus have a usual non adaptive SU test on the $q$'s. Let $R_q$ and $V_q$ denote the number of rejections and false rejections, respectively, by the above SU test with critical values $\alpha^{(q)}_{i:n(\lambda)}$ for the $q$'s. We now have to consider different cases.

1. If $\alpha^{(q)}_{n(\lambda):n(\lambda)} < 1$ and $\alpha' < 1$, we directly obtain

$$E\left(\frac{V_q}{\rho(R_q)}\Big|\mathcal{F}_\lambda\right) \leq \frac{V(\lambda)}{n(\lambda)}\alpha' = \frac{V(\lambda)}{\lambda\hat{n}_0}\alpha \tag{4.21}$$

by Theorem 4.3.

2. If $\alpha^{(q)}_{n(\lambda):n(\lambda)} < 1$ holds, but $\alpha' \geq 1$, then the SU test on the $q$'s has a too big preselected level $\alpha'$, but may still reject not all hypotheses depending on the choice of $\rho$. For this case, let us introduce rescaled versions $\tilde{\alpha}' = \frac{1}{2}$ and $\tilde{\rho}(\cdot) = 2\alpha'\rho(\cdot)$ and consider the corresponding SU test with critical values $\tilde{\alpha}^{(q)}_{i:n(\lambda)} = \frac{\tilde{\rho}(i)}{n(\lambda)}\tilde{\alpha}'$. Now Theorem 4.3 is applicable to this SU test defined by $\tilde{\alpha}'$ and $\tilde{\rho}$. Since both tests coincide it follows that (4.21) holds again.

3. As last case suppose that $\alpha^{(q)}_{n(\lambda):n(\lambda)} \geq 1$ holds. Then every $p$-value $q_1, \ldots, q_{n(\lambda)}$ is rejected by the SU test. Hence, we have

$$\frac{V_q}{\rho(R_q)} = \frac{V(\lambda)}{\rho(n(\lambda))} \leq \frac{V(\lambda)}{n(\lambda)}\alpha' = \frac{V(\lambda)}{\lambda\hat{n}_0}\alpha. \tag{4.22}$$

Now observe that

$$\begin{aligned} R_q &= \max\{i \leq n(\lambda) : q_{i:n(\lambda)} \leq \alpha^{(q)}_{i:n(\lambda)}\} \\ &= \max\left\{i \leq n(\lambda) : \frac{p_{i:n}}{\lambda} \leq \frac{\rho(i)}{n(\lambda)}\frac{n(\lambda)}{\lambda\hat{n}_0}\alpha\right\} \\ &= \max\left\{i \leq n : p_{i:n} \leq \left(\frac{\rho(i)}{\hat{n}_0}\alpha\right) \wedge \lambda\right\} = R \end{aligned}$$

holds and hence $V_q = V$ follows since both tests, belonging to $R$ and $R_q$, are rejecting the same hypotheses. Thus, (4.21) and (4.22) yield

$$E\left(\frac{V}{\rho(R)}\right) = E\left(E\left(\frac{V}{\rho(R)}\Big|\mathcal{F}_\lambda\right)\right) = E\left(E\left(\frac{V_q}{\rho(R_q)}\Big|\mathcal{F}_\lambda\right)\right) \leq E\left(\frac{V(\lambda)}{\lambda\hat{n}_0}\alpha\right).$$

□

**Remark 4.5**

(a) For finite sample error rate control of $\mathbb{E}\left(\frac{V}{\rho(R)}\right)$ by $\alpha$ one merely has to show that the estimator $\hat{n}_0$ fulfills

$$\mathbb{E}\left(\sum_{i\in I_0}\frac{1}{\hat{n}_0^{(i)}}\right)\leq 1 \tag{4.23}$$

by Theorem 4.3 or

$$\mathbb{E}\left(\frac{V(\lambda)}{\hat{n}_0}\right)\leq \lambda \tag{4.24}$$

by Theorem 4.4. In fact, the control holds under the gBI Model and the gBI Model restricted by (4.19), respectively.

(b) The restriction (4.19) is quite natural. In other words, it says that a true $p$-value $p_i$ conditioned under $\{p_i \leq \lambda\}$ again has to be stochastically larger than the uniform distribution on the remaining interval $[0,\lambda]$. For our following considerations we also need (4.19) for other values than $\lambda$ in order to proceed with the further contemplation of the error rate control condition of Theorem 4.4 (except for the Storey estimator $\hat{n}_0(\lambda)$).

(c) Observe that the stronger restriction

$$F_{0,H,i}(tx)\leq t\cdot F_{0,H,i}(x),\quad x,t\in[0,1],\ i\in I_0, \tag{4.25}$$

corresponds to a one sided convexity condition of $F_{0,H,i}$. To be more precise, the usual convexity condition ($F(tx+(1-t)y)\leq tF(x)+(1-t)F(y)$, $0\leq t\leq 1$, $0\leq x,y\leq 1$) just has to be fulfilled for fixed $y=0$. As already seen in Theorem 4.1 and Example 4.2, the convexity of $F_{0,H,i}$ is given in a wide range of one sided testing problems.

The technique of Theorem 4.4 also gives an exact formula for the FDR under the BI Model, see Lemma 5.1 in the following chapter. The next theorem gives a comparison of both conditions and inequalities, respectively.

**Theorem 4.6**

*Consider the gBI Model. Then we have*

$$\frac{\alpha}{\lambda}\cdot\mathbb{E}\left(\frac{V(\lambda)}{\hat{n}_0}\right)\leq \alpha\cdot\mathbb{E}\left(\sum_{i\in I_0}\frac{1}{\hat{n}_0^{(i)}}\right). \tag{4.26}$$

*Moreover, under the BI Model "=" holds in (4.26).*

**Proof**. Let us again condition under $H$. Observe that $\hat{n}_0^{(i)} = \hat{n}_0$ holds on $\{p_i \leq \lambda\}$ for all $i \in I_0$ by (4.9). Thus, we obtain

$$\begin{aligned} E\left(\frac{V(\lambda)}{\hat{n}_0}\Big| H\right) &= \sum_{i\in I_0} E\left(\frac{1\{p_i \leq \lambda\}}{\hat{n}_0}\Big| H\right) = \sum_{i\in I_0} E\left(\frac{1\{p_i \leq \lambda\}}{\hat{n}_0^{(i)}}\Big| H\right) \\ &= \sum_{i\in I_0} F_{0,H,i}(\lambda)\cdot E\left(\frac{1}{\hat{n}_0^{(i)}}\Big| H\right) \leq \sum_{i\in I_0} \lambda \cdot E\left(\frac{1}{\hat{n}_0^{(i)}}\Big| H\right). \end{aligned}$$

The second to last equality holds due to the independence of $p_i$ and $\hat{n}_0^{(i)}$ and Fubini's Theorem. Moreover, under the BI Model we have $F_{0,H,i}(\lambda) = \lambda$ and it follows that "=" holds in the above formula. Finally, taking expectation yields the statement. □

Observe that the inequalities (4.16) and (4.20) of Theorem 4.3 and Theorem 4.4, respectively, coincide under the BI Model and (4.19) is then redundant. However, under the assumptions of Theorem 4.4, it turns out that condition (4.24) is more liberal than (4.23) since (4.23) implies (4.24) by Theorem 4.6. Note that $V(\lambda)$ factors the distributions of the true $p$-values into $\mathbb{E}(\frac{V(\lambda)}{\hat{n}_0})$. If the true $p$-values are stochastically larger than the uniform distribution, then $V(\lambda)$ tends to be small. Hence, $\hat{n}_0$ may also be small. To be more precise, it may be significantly smaller than $N_0$. Therefore, we will mainly consider the inequality of Theorem 4.4 and show that the new estimators fulfill (4.24).

In Chapter 3 we already introduced the estimator $\hat{\varpi}(\lambda, \gamma) = \frac{\hat{F}_n(\gamma)-\hat{F}_n(\lambda)}{\gamma-\lambda}$. As we have seen, it has some nice properties when included in the estimation of $FDR(t)$ and $ENFR(t)$ in Model 3.3 for convex dfs of true $p$-values. Furthermore, we related $\hat{\varpi}(\lambda, \gamma)$ and the estimation of $FDR(t)$ and $ENFR(t)$ to adaptive SU tests, see Theorem 3.12 and Section 3.5. Like Storey et al. [61] we have to consider a slightly modified version of $\hat{\varpi}(\lambda, \gamma)$ in order to achieve generalized error rate control. Therefore, let

$$\hat{m}(\lambda_1, \gamma_1) = n\frac{\hat{F}_n(\gamma_1) - \hat{F}_n(\lambda_1) + \frac{1}{n}}{\gamma_1 - \lambda_1} \tag{4.27}$$

for $\lambda \leq \lambda_1 < \gamma_1 \leq 1$. We call $\hat{m}(\lambda_1, \gamma_1)$ a **generalized Storey estimator**. In contrast to Section 3.2, it is not clear which exact parameter is estimated by $\hat{m}(\lambda_1, \gamma_1)$. Nevertheless, roughly speaking, $\hat{m}(\lambda_1, \gamma_1)$ estimates the effective number of true null hypotheses whose actual marginal distributions really concern the FDR via the SU test. Obviously, in comparison to uniformly distributed true $p$-values, true $p$-values whose marginal distributions are really stochastically larger than the uniform distribution often have a weaker effect on the FDR of many multiple tests since their probability of getting falsely rejected is lower. Although $\hat{m}(\lambda_1, \gamma_1)$ is particularly suitable for the FDR and ENFR (cf. Section 3.2 and 3.5) it will also work for the generalized error rates based on $\rho$ since the right hand side of (4.20) does not depend on the error rate itself.

For further considerations we need

$$F_{0,H,i}(tx) \leq t \cdot F_{0,H,i}(x), \quad t \in [0,1], \ i \in I_0, \tag{4.28}$$

for different values of $x \in [\lambda, 1]$.

**Lemma 4.7**
*Assume (4.8) and let $\hat{m}(\lambda_1, \gamma_1)$ be the generalized Storey estimator (4.27) for some $0 < \lambda \leq \lambda_1 < \gamma_1 \leq 1$. Consider the gBI Model and assume (4.28) for all $x \in \{\lambda, \gamma_1\}$. Then we have*

$$\mathbb{E}\left( \frac{V(\lambda)}{\hat{m}(\lambda_1, \gamma_1)} \Big| (n\hat{F}_n(t))_{t \geq \gamma_1} \right) \leq \lambda \tag{4.29}$$

*and the adaptive SU test with critical values (4.10) and estimator (4.27) has finite sample error rate control of $\mathbb{E}\left(\frac{V}{\rho(R)}\right)$ by $\alpha$.*

**Proof**. Observe that the mappings $p_i \mapsto \hat{m}(\lambda_1, \gamma_1)$, $i = 1, \ldots, n$, are non decreasing on $[0, \gamma_1]$. Thus, by Lemma 6.4 of the appendix, we only have to show that (4.29) holds for the BI Model. To be more precise, Lemma 6.4 shows that the FDR increases when, conditioned under $H$, the true $p$-values are replaced by independent on $[0, 1]$ uniformly distributed $p$-values. From this point, the proof is basically the same as in Heesen and Janssen [28, Lemma 3.2.].

Let us introduce the following simplifying notation

$$\begin{aligned} V(t) &= n_0(t) = \#\{p_i \leq t : H_i = 0\}, \\ S(t) &= n_1(t) = \#\{p_i \leq t : H_i = 1\}, \quad 0 \leq t \leq 1, \end{aligned}$$

and let us condition under $\mathcal{F}_{\gamma_1}$, see (4.15). In this case the quantities $n_0(\gamma_1)$ and $n_1(\gamma_1)$ can be treated as fixed values, whereas the random variables $V(\lambda)$,

$V(\lambda_1)$ and $S(\lambda)$ are still random. Since $n_1(\gamma_1) - S(\lambda_1) \geq 0$ holds, we obtain

$$\mathbb{E}\left(\frac{V(\lambda)}{n(\hat{F}_n(\gamma_1) - \hat{F}_n(\lambda_1) + \frac{1}{n})}\Big|\mathcal{F}_{\gamma_1}\right) \tag{4.30}$$

$$= \mathbb{E}\left(\frac{V(\lambda)}{n_0(\gamma_1) + 1 - V(\lambda) - (V(\lambda_1) - V(\lambda)) + n_1(\gamma_1) - S(\lambda_1)}\Big|\mathcal{F}_{\gamma_1}\right) \tag{4.31}$$

$$\leq \mathbb{E}\left(\frac{V(\lambda)}{n_0(\gamma_1) + 1 - V(\lambda) - (V(\lambda_1) - V(\lambda))}\Big|\mathcal{F}_{\gamma_1}\right). \tag{4.32}$$

The random vector $(V(\lambda), V(\lambda_1) - V(\lambda), V(\gamma_1) - V(\lambda_1))$ is distributed according to the multinomial distribution $\mathcal{M}\left(n_0(\gamma_1), \frac{\lambda}{\gamma_1}, \frac{\lambda_1-\lambda}{\gamma_1}, \frac{\gamma_1-\lambda_1}{\gamma_1}\right)$ under our conditions. Thus, Lemma 6.5 of the appendix yields

$$(4.32) = \frac{\lambda}{\gamma_1 - \lambda_1}\left(1 - \left(\frac{\lambda_1}{\gamma_1}\right)^{n_0(\gamma_1)}\right) \leq \frac{\lambda}{\gamma_1 - \lambda_1} \tag{4.33}$$

and integration gives

$$\mathbb{E}\left(\frac{V(\lambda)}{n(\hat{F}_n(\gamma_1) - \hat{F}_n(\lambda_1) + \frac{1}{n})}\Big|(n\hat{F}_n(t))_{t\geq\gamma_1}\right) \leq \frac{\lambda}{\gamma_1 - \lambda_1}$$

which proves (4.29). Finally, the error rate control follows by a further integration and Theorem 4.4. □

Due to conditioning under $(n\hat{F}_n(t))_{t\geq\gamma_1}$, inequality (4.29) is a little stronger than the required inequality (4.24). We will need this for dynamic adaptive SU tests presented in the next section. Furthermore, if the adaptive SU test with estimator $\hat{m}(\lambda_1, \gamma_1)$ is applied, then it is reasonable to set $\lambda = \lambda_1$.

**Remark 4.8**

The generalized Storey estimator (4.27) in Lemma 4.7 may be replaced by

$$\left(1 - \left(\frac{\lambda_1}{\gamma_1}\right)^{n\hat{F}_n(\gamma_1)\vee 1}\right) \cdot \hat{m}(\lambda_1, \gamma_1). \tag{4.34}$$

This simply follows by replacing the right hand side of (4.33) in the proof by

$$\frac{\lambda}{\gamma_1 - \lambda_1} \cdot \left(1 - \left(\frac{\lambda_1}{\gamma_1}\right)^{n\hat{F}_n(\gamma_1)\vee 1}\right).$$

This estimator is smaller and the adaptive SU test thus has more power. All in all this is only a very slight improvement which vanishes for increasing $n$ and hence we mainly focus on $\hat{m}(\lambda_1, \gamma_1)$.

## 4.3 Dynamic adaptive SU tests

Let us now develop the previously announced dynamic adaptive SU tests. We begin with a stationary approach which includes a non data dependent weighting of a family of estimators. This approach creates new estimators based on this family.

The following corollary of Theorem 4.4 is a generalization of Corollary 3.1 of Heesen and Janssen [28]. The statement there is only formulated for finitely many Storey estimators $\hat{n}_0(\lambda_i)$, $i = 1, \ldots, k$, with $\lambda \le \lambda_1 < \ldots < \lambda_k < 1$.

**Corollary 4.9** (Deterministic integral type combination of estimators)
*Assume (4.8) and consider the gBI Model. Furthermore, assume that (4.28) holds for all $x \in [\lambda, 1]$, where $0 < \lambda < 1$ is a tuning parameter. Let*

$$(\hat{m}(s))_{s\in S} = \left(\hat{m}\left(s, (\hat{F}_n(t))_{t\ge\lambda}\right)\right)_{s\in S} \tag{4.35}$$

*be a family of estimators indexed by some index set $S$ and let each estimator $\hat{m}(s)$ be given by a measurable function of $(\hat{F}_n(t))_{t\ge\lambda}$. Moreover, assume that $\hat{m}(s) > 0$ and*

$$\mathbb{E}\left(\frac{V(\lambda)}{\hat{m}(s)}\right) \le \lambda \quad \text{for all} \quad s \in S. \tag{4.36}$$

*If $\nu$ is a fixed probability measure on the index set $S$, then the adaptive SU test with critical values (4.10) and estimator*

$$\hat{m}_\nu = \int \hat{m}(s)\nu(ds), \tag{4.37}$$

*has error rate control of $\mathbb{E}\left(\frac{V}{\rho(R)}\right)$ by $\alpha$.*

**Proof**. By Jensen's inequality and Fubini's Theorem, we obtain

$$\mathbb{E}\left(\frac{V(\lambda)}{\hat{m}_\nu}\right) \le \mathbb{E}\left(\int \frac{V(\lambda)}{\hat{m}(s)}\nu(ds)\right) = \int \mathbb{E}\left(\frac{V(\lambda)}{\hat{m}(s)}\right)\nu(ds) \le \lambda$$

and the error rate control follows by Theorem 4.4. □

**Remark 4.10**
(a) In actual fact, the assumption that (4.28) holds for all $x \in [\lambda, 1]$ is not really needed since (4.36) shall hold. However, (4.36) often requires (4.28), see Lemma 4.7 for instance. Therefore, we additionally assume (4.28).
(b) In particular, the assumptions of Theorem 4.9 are fulfilled if $(\hat{m}(s))_{s\in S}$ is

the family of generalized Storey estimators introduced in (4.27).
(c) Theorem 4.9 corresponds to the ideas of Section 3.3.

The next theorem gives a dynamic approach and includes a data dependent weighting of estimators.

**Theorem 4.11** (Dynamic combination of estimators)
*Assume (4.8) and consider the gBI Model. Moreover, assume that (4.28) holds for all* $x \in [\lambda, 1]$*, where* $0 < \lambda < 1$ *is a tuning parameter. Let* $\lambda < \gamma_1 \leq \ldots \leq \gamma_k \leq 1$ *be inspection points and let* $\hat{m}(i)$*,* $i = 1, \ldots, k$*, be estimators of the form (4.9) which fulfill*

$$\mathbb{E}\left(\frac{V(\lambda)}{\hat{m}(i)}\Big|(n\hat{F}_n(t))_{t\geq\gamma_i}\right) \leq \lambda, \quad i = 1, \ldots, k. \tag{4.38}$$

*Furthermore, let* $\hat{\beta}_1, \ldots, \hat{\beta}_k$ *be non negative data dependent weights with* $\sum_{i=1}^k \hat{\beta}_i = 1$ *and let* $\hat{\beta}_i$ *be measurable with respect to* $\sigma((n\hat{F}_n(t))_{t\geq\gamma_i})$*,* $i = 1, \ldots, k$*. Then the dynamic adaptive SU test with critical values (4.10) and estimator*

$$\hat{m} = \sum_{i=1}^k \hat{\beta}_i \hat{m}(i) \tag{4.39}$$

*has error rate control of* $\mathbb{E}\left(\frac{V}{\rho(R)}\right)$ *by* $\alpha$*.*

**Proof**. The proof is basically the same as the proof of Theorem 6.1 of Heesen and Janssen [28]. Since $\hat{m}$ is a convex combination (or similar to the proof of Theorem 4.9 by Jensen's inequality), we observe that

$$\frac{V(\lambda)}{\sum_{i=1}^k \hat{\beta}_i \hat{m}(i)} \leq \sum_{i=1}^k \hat{\beta}_i \frac{V(\lambda)}{\hat{m}(i)}$$

holds for fixed $p$-values. Moreover, since the $\hat{\beta}_i$'s are $\sigma((n\hat{F}_n(t))_{t\geq\gamma_i})$-measurable, we obtain by (4.38)

$$\begin{aligned}
\mathbb{E}\left(\frac{V(\lambda)}{\hat{m}}\right) &\leq \mathbb{E}\left(\sum_{i=1}^k \hat{\beta}_i \frac{V(\lambda)}{\hat{m}(i)}\right) = \sum_{i=1}^k \mathbb{E}\left(\hat{\beta}_i \mathbb{E}\left(\frac{V(\lambda)}{\hat{m}(i)}\Big|\left(n\hat{F}_n(t)\right)_{t\geq\gamma_i}\right)\right) \\
&\leq \mathbb{E}\left(\sum_{i=1}^k \hat{\beta}_i \lambda\right) = \lambda
\end{aligned}$$

and the error rate control follows by Theorem 4.4. □

**Remark 4.12**
(a) Remark 4.10 (a) also applies for Theorem 4.11.
(b) In particular, (4.38) is fulfilled for the generalized Storey estimators $\hat{m}(\lambda_i, \gamma_i)$ with $\lambda \leq \lambda_i < \gamma_i \leq 1$, $i = 1, \dots, k$, see Lemma 4.7.

We now give a guideline how to use Theorem 4.11 for large $n$ in connection with the generalized Storey estimators $\hat{m}(\lambda_1, \gamma_1)$.

**Practical Guide 4.13**
For the Practical Guide let us again consider the gBI Model and assume that every possible df $F_{0,H,i}$ is convex which is slightly stricter than demanding for (4.28) for all $x \geq \lambda$. Furthermore, assume that conditioned under $H$ the false $p$-values $p_i$, $i \in I_1$, have marginal dfs $F_{1,H,i}$ which are concave or fulfill $F_{1,H,i}(\lambda) \approx 1$. Although the theorems of this chapter work without this assumption it is crucial for a significantly improved estimation part of the adaptive SU tests. Even in the BI Model the adaptive SU test of Storey et al. [61] with estimator $\hat{n}_0(\lambda)$ can perform very bad when arbitrary distributions of false $p$-values are allowed. Like the assumption of convex dfs of true $p$-values, this assumption is also very natural for false $p$-values and Theorem 4.1 may be extended in an analog way.

First, let us recapitulate some features of the estimator $n\hat{\varpi}(\lambda_1, \gamma_1)$ obtained in Section 3.2. These features basically transfer to the slightly modified generalized Storey estimators $\hat{m}(\lambda_1, \gamma_1)$ which are mainly considered in this chapter. As we have seen, the estimator $\hat{\varpi}(\lambda_1, \gamma_1)$ and hence $\hat{m}(\lambda_1, \gamma_1)$ perform well if the false $p$-values are likely to lie below $\lambda_1$ and if the true $p$-values, which are stochastically much larger than $U(0,1)$, are likely to lie above $\gamma_1$. Moreover, the larger $\gamma_1 - \lambda_1$ the smaller the variance of these estimators. It is often assumed that the false $p$-values are likely to lie below 0.5 and the tuning parameter $\lambda$ is often chosen close to 0.5. In this situation, we thus advise to choose a small fixed $\lambda_1 = \lambda \approx 0.5$. As said in Section 3.2, the choice of $\gamma_1$ is quite more complex, but here the dynamic approach of Theorem 4.11 comes into play. We now construct a data dependent selection of $\gamma_1$ which we denote by $\gamma^*$. For the adaptive SU test we then use the estimator

$$\hat{m} = \hat{m}(\lambda, \gamma^*). \tag{4.40}$$

The construction of $\gamma^*$ is done in the sense of Theorem 4.11.

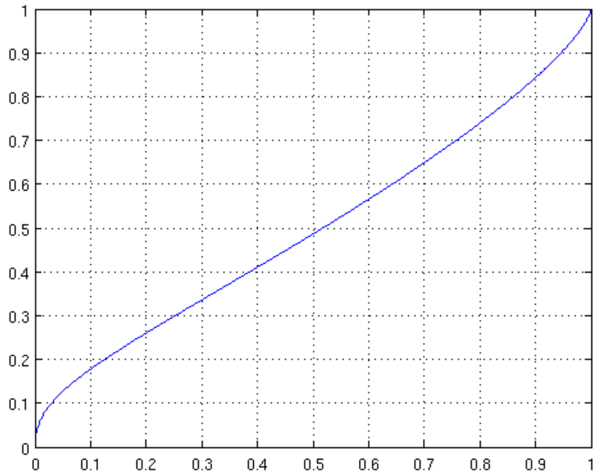



Figure 4.3: Limit ecdf of $p$-values in a one sided normal mean testing problem.

Before we introduce a suggestion for a data dependent $\gamma^*$ we begin by an analysis of likely ecdfs for some meaningful examples. At first, let us recall the situation of Example 4.2 (b). The ecdf of Figure 4.2 looks almost smooth and in the following we would like to talk about derivatives. But since the derivative of every ecdf is almost everywhere equal to zero, we have to turn to an idealized model. Thus, by Glivenko-Cantelli arguments and since we are considering very large $n$ let us consider the $p$-value df

$$F(t) = \frac{3}{20}(1 - \Phi(\Phi^{-1}(1-t) - 2)) + \frac{12}{20}t + \frac{5}{20}(1 - \Phi(\Phi^{-1}(1-t) + 1))$$

which is a possible and nearby limit ecdf of an appropriate sequence of $p$-value models. $F$ is displayed in Figure 4.3. Observe that the limit ecdf is almost linear on the interval $[0.5, 0.8]$. Due to the convexity of the df of true $p$-values, it is likely that there are too many true $p$-values located in the interval $[0.8, 1]$ which disturbs the estimation in the sense of Example 3.4. The adaptive SU test would reject more $p$-values when the estimator $\hat{m}(\lambda, \gamma^*)$ gets smaller since then the critical values $\hat{\alpha}_{i:n}$ get larger. Thus, $\hat{m}(\lambda, \gamma^*)$ applied to $F$ is preferably small when it is applied to the linear part of $F$, i.e. for the choice of $\gamma^* = 0.8$. Up to this point, this contemplation also holds for the primary ecdf $\hat{F}_n$ of Figure 4.2. Observe that a linear part of a function $F$ exhibits $F'' \equiv 0$ restricted to this part. Thus, in this idealized situation $\gamma^*$ should be chosen as the largest $t$ such that $F''(t) \approx 0$ holds since this would indicate the end of the linear part.

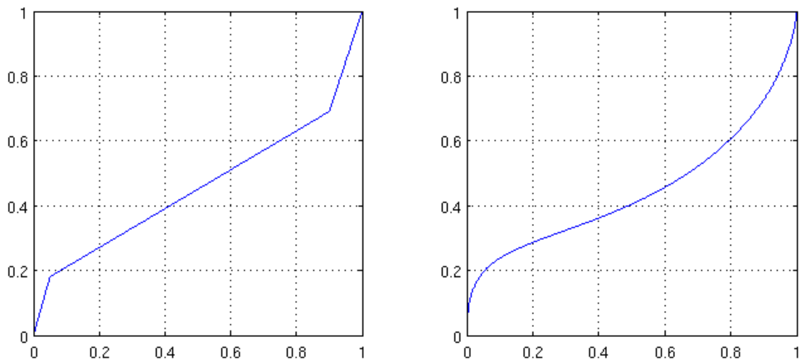
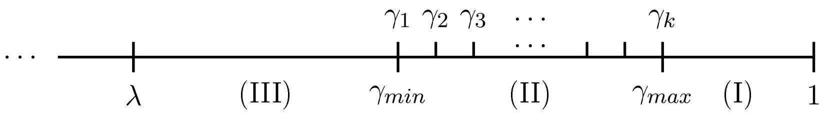



Figure 4.4: Other possible limit ecdf.

Figure 4.4 shows two limit ecdfs for other possible situations, where we forgo their explicit form in the following consideration. Observe that the recently proposed choice of $\gamma^*$ does not work well for the limit ecdf on the left hand side since the limit ecdf already consists of linear parts. Hence, this choice just yields $\gamma^* = 1$. But again, the estimator $\hat{m}(\lambda, \gamma^*)$ is preferably small when applied to the central linear part, i.e. when applied to the interval $[0.5, 0.9]$. This corresponds to the choice of $\gamma^* = 0.9$. The probability of a true $p$-value in $[0.9, 1]$ is increased and the first derivative on this interval is too large. Thus, similar as before $\gamma^*$ should also be chosen as largest $t$ such that $F'(t)$ is close to the slope of the central linear part. But in general, this slope is unknown. We merely know that the slope of such a linear part is bounded from above by 1. This bound is obtained in the BI Model for fixed $N_0 = n$. Hence, in the idealized situation, $\gamma^*$ should also fulfill $F'(\gamma^*) \leq 1$. But notice that this choice of $\gamma^*$ may still be far away from an optimal choice. But observe that it is not guaranteed that there exists a central linear part of the limit ecdf. This is the case for the picture on the right hand side in Figure 4.4. Hence, there should be a lower bound $\gamma_{min}(\lambda)$ for $\gamma^*$. This ensures that $\gamma^* - \lambda \geq \gamma_{min}(\lambda) - \lambda$ is not too small.

Based on the previous considerations we now introduce an explicit adaptive estimator $\hat{m}$. Therefore, let $\lambda < \gamma_{min} = \gamma_1 < \ldots < \gamma_k = \gamma_{max} < 1$ be inspection points and consider the situation of Theorem 4.11 with $\hat{m}(i) = \hat{m}(\lambda, \gamma_i)$, $i = 1, \ldots, k$. We will choose a data dependent $\gamma^*(\omega) \in \{\gamma_1, \ldots, \gamma_k\}$ which corresponds to setting $\hat{\beta}_i(\omega) = 1$ and hence setting all other $(\hat{\beta}_j(\omega))_{j \neq i}$ to zero iff




(I) minimal interval for adjusting $\gamma^*$
(II) interval for adjusting $\gamma^*$ and estimation by $\hat{m}(\lambda, \gamma^*)$
(III) minimal interval for the estimation by $\hat{m}(\lambda, \gamma^*)$

Figure 4.5: Areas of the estimation process.

$\gamma^*(\omega) = \gamma_i$, see Figure 4.5 for a graphical explanation. For each inspection point $\gamma_i$ we basically need the first and second derivative of an appropriate limit ecdf at $\gamma_i$ by our previous consideration. But this is not realistic. Therefore, we fit a polynomial smoothing function of degree 3 by least squares to the ecdf $\hat{F}_n$ on the interval $[\gamma_i, \gamma_i + \delta_i]$, where $\delta_i$ is a fixed constant with $0 < \delta_i < 1 - \gamma_i$. Then we take the derivatives of the smoothing function at $\gamma_i$. Since we are assuming convex dfs of true $p$-values, ties of $p$-values in $[\gamma_i, \gamma_i + \delta_i]$ can be neglected. However, the probability of ties at 1 may be positive. Hence $\delta_i = 1 - \gamma_i$ should be excluded. A polynomial smoothing function of at least degree 3 seems to be appropriate since linear changes of the second derivative may be recognized. Let $j_{1,i} < \ldots < j_{r,i}$ be the indices of the $p$-values $p_j \in [\gamma_i, \gamma_i + \delta_i]$. Then the least squares smoothing function

$$g_i(t) = b_{0,i} + b_{1,i} \cdot t + b_{2,i} \cdot t^2 + b_{3,i} \cdot t^3 \tag{4.41}$$

is given by the coefficients

$$\begin{pmatrix} b_{0,i} \\ \vdots \\ b_{3,i} \end{pmatrix} = (A^T A)^{-1} A^T \begin{pmatrix} \hat{F}_n(p_{j_{1,i}}) \\ \vdots \\ \hat{F}_n(p_{j_{r,i}}) \end{pmatrix}, \tag{4.42}$$

with

$$A = \begin{pmatrix} 1 & p_{j_{1,i}} & p_{j_{1,i}}^2 & p_{j_{1,i}}^3 \\ \vdots & \vdots & \vdots & \vdots \\ 1 & p_{j_{r,i}} & p_{j_{r,i}}^2 & p_{j_{r,i}}^3 \end{pmatrix} \in \mathbb{R}^{r \times 4}. \tag{4.43}$$

In addition the derivatives at $\gamma_i$ are given by

$$g_i'(\gamma_i) = b_{1,i} + 2b_{2,i}\gamma_i + 3b_{3,i}\gamma_i^2, \tag{4.44}$$

$$g_i''(\gamma_i) = 2b_{2,i} + 6b_{3,i}\gamma_i. \tag{4.45}$$

Of course, this only holds if $rank(A) = 4$ which is known to be true iff there are at least 4 different $p$-values located in $[\gamma_i, \gamma_i + \delta_i]$. If this is not the case, we advise in the following to set $g_i'(\gamma_i) = 1$ and $g_i''(\gamma_i) = 0$ so that $\gamma^* \geq \gamma_i$ automatically holds. Let $\epsilon > 0$ be a tuning parameter. Altogether, we define

$$\hat{\beta}_i = \mathbb{1}_{[0,1]\times(-\infty,\epsilon]} \left(g_i'(\gamma_i), g_i''(\gamma_i)\right) \prod_{j=i+1}^{k} \mathbb{1}_{([0,1]\times(-\infty,\epsilon])^c} \left(g_j'(\gamma_j), g_j''(\gamma_j)\right) \quad (4.46)$$

for $i = 2, \ldots, k$ and

$$\hat{\beta}_1 = \prod_{j=2}^{k} \mathbb{1}_{([0,1]\times(-\infty,\epsilon])^c} \left(g_j'(\gamma_j), g_j''(\gamma_j)\right). \quad (4.47)$$

In other words, we set $\hat{\beta}_i = 1$ iff $g_i'(\gamma_i) \in [0,1]$ and $g_i''(\gamma_i) \in (-\infty, \epsilon]$ for the last time and hence take $\gamma^* = \gamma_i$ in (4.40). Otherwise, we set $\hat{\beta}_1 = 1$ and take $\gamma^* = \gamma_{min} = \gamma_1$. According to the construction of $g_i$, the data dependent weights $\hat{\beta}_i$ are obviously measurable with respect to $\sigma((n\hat{F}_n(t))_{t\geq\gamma_i})$ and fulfill the assumptions of Theorem 4.11. Thus, the error rate control of $\mathbb{E}\left(\frac{V}{\rho(R)}\right)$ by $\alpha$ of the adaptive SU test with critical values (4.10) and estimator

$$\hat{m} = \hat{m}(\lambda, \gamma^*) = \sum_{i=1}^{k} \hat{\beta}_i \hat{m}(\lambda, \gamma_i) \quad (4.48)$$

follows by Lemma 4.7 and Theorem 4.11.

**Remark 4.14**

(a) Consider some true $p$-values which are distributed according to some strictly convex df and an arbitrary subinterval $[a, b] \subset [0, 1]$. Observe that the $p$-values which fall into $[a, b]$, tend to lie in the second half of the interval. Thus, we expect that the best fit of the least squares smoothing functions $g_i$ in the practical guide are attained at $\gamma_i + \delta_i$. Due to our construction of the adaptive estimator, we would like to have a good fit of $g_i$ at $\gamma_i$. Therefore, a weighted lest squares approach may perform better.

(b) A different approach for the data dependent choice of $\gamma$ may be based on the spacings $p_{i+h:n} - p_{i:n}$.

(c) Standard methods (like kernel density estimation, see Silverman [57] for instance) do not work in the situation of Theorem 4.11 since they often require the entire information of the ecdf $\hat{F}_n$. Many estimators perform poorly at the border and some of them do not even reconstruct the convexity of the df which is important for the proceeding of the Practical Guide.

Liang and Nettleton [36] already proposed a data dependent choice of $\lambda$ for the Storey estimator $\hat{n}_0(\lambda)$. Their approach is based on the filtration $\sigma_t = \sigma(\mathbb{1}\{p_i \leq s\} : 0 \leq s \leq t, i = 1, \ldots, n)$, $0 \leq t \leq 1$, whereas our approach is based on the reverse filtration $\mathcal{F}_t$, $0 \leq t \leq 1$.

Finally, we give two concluding remarks and an outlook on future work.

**Remark 4.15** (Discrete distributions)
Under the assumptions of Theorem 4.11, consider the estimators $\hat{m}(i) = \hat{m}(\lambda, \gamma_i)$, $i = 1, \ldots, k$. By Lemma 4.7, it follows that (4.38) is fulfilled. Furthermore, a reinspection of Lemma 4.7 and the proof of Theorem 4.11 yields that Theorem 4.11 also holds for this choice of $\hat{m}(i)$ if (4.28) is only fulfilled for all $x \in \{\lambda, \gamma_1, \ldots, \gamma_k\}$ instead of all $x \in [\lambda, 1]$. Note that a df of a discrete distribution cannot satisfy (4.28) for all $x \in [\lambda, 1]$, but such a df may fulfill (4.28) for all $x \in \{\lambda, \gamma_1, \ldots, \gamma_k\}$. Observe that (4.28) can only be fulfilled for all $x \in \{\lambda, \gamma_1, \ldots, \gamma_k\}$ if every possible df $F_{0,H,i}$ of a discrete distribution has jump discontinuities at all points $\lambda$ and $\gamma_1, \ldots, \gamma_k$. Thus, a reasonable choice of $\lambda, \gamma_1, \ldots, \gamma_k$ and the just described homogeneity of the discrete distributions are crucial for the treatment of dynamic adaptive SU tests under discrete distributions. Discrete distributions do not exclude the simultaneous consideration of continuous distributions, i.e. some of the dfs $F_{0,H,i}$ may be continuous and some may have jump discontinuities.

**Remark 4.16**
In a couple of works, Donoho and Jin [13], Hall and Jin [27], Jager and Wellner [30] and Jin [31] essentially considered the Higher-Criticism statistic and related statistics as test statistic for the global intersection null hypothesis

$$\mathcal{H}_0 : \ p_i \sim U(0,1), \quad 1 \leq i \leq n,$$

versus several sparse alternative hypotheses. The recently announced adaptive SU tests with estimator $\hat{m}(\lambda_1, \gamma_1)$ and the dynamic adaptive SU tests may have some advantage when the true $p$-values may be stochastically larger than the uniform distribution. For an asymptotic sparse normal mean testing problem, Donoho and Jin [13] already showed that the so-called detection region of the BH test is only slightly smaller than the detection region of the Higher Criticism statistic which is the maximum possible detection region in this asymptotic setting.

Future work may include the analysis of other estimators, like the estimators of Section 3.4. So far, it is not clear if some of these estimators (or modified versions) fulfill (4.24). Moreover, an adequate choice of the quantities $k$, $\gamma_1, \ldots, \gamma_k$, $\delta_1, \ldots, \delta_k$ and $\epsilon$ of the Practical Guide is not clear and may be analyzed. Obviously, the number of inspection points $k$ should depend on the sample size $n$.

## 4.4 Simulation

Here we present a small simulation study based on Example 4.2. Therefore, let $n = 1000$ and $X = (X_1, \ldots, X_n) \sim N(\mu, I_n)$ with fixed vector $\mu = (\mu_1, \ldots, \mu_n) \in \mathbb{R}^n$ and identity matrix $I_n$. Let us consider the multiple testing problem

$$\mathcal{H}_i = \{\mu_i \leq 0\} \quad \text{versus} \quad \mathcal{K}_i = \{\mu_i > 0\}, \quad i = 1, \ldots, n,$$

and define the $p$-values $p_i = 1 - \Phi(X_i)$, where $\Phi$ denotes the df of the standard normal distribution. For our simulation study, we consider $\mu$ with

$$\mu_i = \begin{cases} \vartheta, & 1 \leq i \leq k, \\ 0, & k < i \leq 800, \\ 3, & 800 < i \leq 1000, \end{cases} \qquad i = 1, \ldots, n,$$

for several choices of $k$ and $\vartheta \leq 0$. Note that we have fixed $N_0 = n_0 = 800$ in all cases. Table 4.1 shows the result of a Monte-Carlo simulation with 10.000 iterations for the FDR of several multiple tests with predetermined level $\alpha = 0.05$. Therefore, $FDR^{BH}$ denotes the FDR of the BH test with critical values (1.19), $FDR^{\hat{n}_0(\lambda)}$ denotes the FDR of the adaptive SU test of Storey with critical values (1.29) and estimator $\hat{n}_0(\lambda)$, and $FDR^{\hat{m}(\lambda,\gamma)}$ denotes the FDR of the adaptive SU test with critical values (4.10) and estimator $\hat{m}(\lambda, \gamma)$. Moreover, let $FDR^{BH,o}$ be the FDR of the oracle BH test which is based on the critical values $\alpha_{i:n}^{BH,o} = \frac{i}{n_0}\alpha$, $i = 1, \ldots, n$. The oracle BH test includes the exact value $n_0$ which the Storey estimator $\hat{n}_0(\lambda)$ tries to estimate.

For $k = 0$ and $\vartheta = 0$ the FDR of the BH test is just given by the BH Theorem and does not exhaust the predetermined FDR level $\alpha$. In contrast, the other tests exhaust the predetermined FDR level $\alpha$ very well. Note that slight variations are due to the Monte-Carlo simulation. For the other cases of

| $k$ | | 0 | 200 | 200 | 400 | 400 |
|---|---|---|---|---|---|---|
| $\vartheta$ | | 0 | -1 | -2 | -1 | -3 |
| $FDR^{BH}$ | | 0.04 | 0.03 | 0.03 | 0.21 | 0.0201 |
| $FDR^{\hat{n}_0(\lambda)}$ | $\lambda = 0.5$ | 0.05 | 0.032 | 0.03 | 0.019 | 0.016 |
| $FDR^{BH,o}$ | | 0.05 | 0.038 | 0.0377 | 0.026 | 0.0248 |
| $FDR^{\hat{m}(\lambda,\gamma)}$ | $\gamma = 0.95$ $\lambda = 0.5$ | 0.049 | 0.035 | 0.04 | 0.023 | 0.041 |
| $FDR^{\hat{m}(\lambda,\gamma)}$ | $\gamma = 0.9$ $\lambda = 0.5$ | 0.05 | 0.037 | 0.043 | 0.024 | 0.045 |

Table 4.1: Monte-Carlo Simulation of the FDR of several multiple tests.

$k$ and $\vartheta$ the exhaustion of the FDR in case of the BH test and adaptive SU test of Storey deteriorates and for $k = 400$ and $\vartheta = -1, -3$ the FDR of the adaptive SU test of Storey is even lower than the FDR of the BH test. Furthermore, observe that the exhaustion of the FDR in case of the oracle BH test also deteriorates for the other cases of $k$ and $\vartheta$. In contrast to the BH test and adaptive SU test of Storey, the new adaptive SU tests, based on $\hat{m}(\lambda, \gamma)$, yield a completely better exhaustion of the FDR. For $k = 200, \vartheta = -2$ and $k = 400, \vartheta = -3$ the new tests even outperform the oracle BH test. In particular, in the extreme situation with $\vartheta = -3$ the exhaustion of the FDR works very well in comparison to the other tests. As already mentioned, the deeper inside the null hypotheses, the better the new tests work.

Finally, note that a better exhaustion of the predetermined FDR level $\alpha$ comes along with an increased power of the tests.

# Chapter 5

# FDR control of adaptive SU tests under dependence

In Section 2.6 we were concerned with the asymptotic FDR of adaptive SU tests for several dependence structures and in Chapter 4 we dealt with finite sample FDR control for adaptive SU tests under independence. This chapter is devoted to finite sample FDR control of adaptive SU tests under dependence. The treatment of adaptive SU tests for dependent $p$-values is known to be far more difficult and there are only a few references in the literature. For example, Blanchard and Roquain [7, 8] considered two-stage adaptive SU tests, where the number of true null hypotheses is estimated by the number of rejections of a previously conducted multiple test. The previously conducted tests are based on the same $p$-values and exhibit some error rate control. Unfortunately, the predetermined levels of both tests add up, in some sense. Furthermore, several asymptotic results can be found in the literature, see Farcomeni [15] for instance. Here, we will focus on the RM Model, see Model 2.2 for a definition. We will show that finite sample FDR control actually requires further restrictions and obtain control under a block dependence structure for Storey type estimators $\hat{n}_0(\lambda, \kappa)$, see (1.32), by a careful choice of the tuning parameter $\kappa$. The material of this chapter is presented in Section 5 of Heesen and Janssen [29].

The next lemma provides an exact FDR formula and imposes a condition on the estimators of adaptive SU tests for the RM Model which is the same condition as in Theorem 4.4 and again ensures FDR control.

**Lemma 5.1** (cf. Heesen and Janssen [29] Lemma 5.1)
*Consider the RM Model and the adaptive SU test based on the critical values (4.10) with $\rho = id$ and estimator (4.9). Then we have*

$$FDR = \frac{\alpha}{\lambda} \cdot \mathbb{E}\left(V(\lambda) \cdot \min\left\{\frac{1}{\hat{n}_0}, \frac{\lambda}{n\hat{F}_n(\lambda)\alpha}\right\}\right) \leq \frac{\alpha}{\lambda} \cdot \mathbb{E}\left(\frac{V(\lambda)}{\hat{n}_0}\right). \tag{5.1}$$

**Proof**. The proof is very similar to the proof of Theorem 4.4. Therefore, let us only sketch it and use the notation of the proof of Theorem 4.4. Conditioned under $\mathcal{F}_\lambda = \sigma(H, \mathbb{1}\{p_i \leq s\} : s \geq \lambda, 1 \leq i \leq n)$ we may again construct new rescaled $p$-values $q_i$, $i = 1, \ldots, n(\lambda) = n\hat{F}_n(\lambda)$. These rescaled $p$-values again fulfill the requirement of the RM Model by use of the reverse filtration $\mathcal{F}_t^q = \sigma\big(\sigma(\mathbb{1}\{q_i \leq s\} : s \geq t, 1 \leq i \leq n(\lambda)), \mathcal{F}_\lambda\big)$. Here, we only have to consider two cases for $\alpha' = \frac{n(\lambda)}{\lambda \hat{n}_0}\alpha$. 1. If $\alpha' \geq 1$, then every $p$-value $q_i$, $i = 1, \ldots, n(\lambda)$, is rejected by the present SU test for the $q$'s and we obtain $\frac{V_q}{R_q \vee 1} = \frac{V(\lambda)}{n(\lambda)}$. 2. As other case suppose that $\alpha' < 1$. Then Lemma 2.5 for the RM Model implies $\mathbb{E}\left(\frac{V_q}{R_q \vee 1}\Big|\mathcal{F}_\lambda\right) = \frac{V(\lambda)}{n(\lambda)}\alpha'$. Hence, we have

$$\mathbb{E}\left(\frac{V_q}{R_q \vee 1}\Big|\mathcal{F}_\lambda\right) = \frac{V(\lambda)}{n(\lambda)} \min\{\alpha', 1\}.$$

By the same arguments as before, we obtain $R = R_q$ and $V = V_q$ conditioned under $\mathcal{F}_\lambda$ and it follows that

$$\mathbb{E}\left(\frac{V}{R \vee 1}\right) = \mathbb{E}\left(\frac{V(\lambda)}{n(\lambda)} \min\{\alpha', 1\}\right)$$
$$= \frac{\alpha}{\lambda} \cdot \mathbb{E}\left(V(\lambda) \cdot \min\left\{\frac{1}{\hat{n}_0}, \frac{\lambda}{n\hat{F}_n(\lambda)\alpha}\right\}\right).$$

□

In Chapter 3 and 4, we considered estimators which try to underestimate $N_0$ in specific situations. Here, we will only focus on estimators for $N_0$ since these situations do not occur in the RM Model. It is an open question if the estimation concept of Chapter 3 may be applied to the reverse super martingale model, described in Remark 2.3 (b).

Although Lemma 5.1 basically gives the same condition as Theorem 4.4, we note that this condition is more difficult to fulfill under the RM Model. The following proposition contains a negative result. It shows that the usability of adaptive SU tests is limited under dependence and a restriction of the RM Model is necessary to allow reasonable estimators with FDR control.

**Proposition 5.2** (cf. Heesen and Janssen [29] Proposition 5.2)
*Under the assumptions of Lemma 5.1, let $p_i \mapsto \hat{n}_0$ be non-decreasing in each p-value $p_i$ and $\alpha < \lambda$. If $FDR_P \leq \alpha$ for all $P \in \mathcal{P}_{RM}$, then $\hat{n}_0 \geq n$ follows.*

**Proof**. Let $n = N_0 > 2$ and $p_1 = \ldots = p_n = U \sim U(0,1)$ which gives $V(\lambda) = n\hat{F}_n(\lambda) = n\mathbb{1}\{U \leq \lambda\}$. The exact FDR formula in (5.1) yields

$$\alpha \geq FDR = \frac{\alpha}{\lambda} n\lambda \cdot \min\left\{\frac{1}{\hat{n}_0(\vec{1})}, \frac{\lambda}{n\alpha}\right\},$$

where $\hat{n}_0(\vec{1})$ stands for the value of the estimator when $\hat{F}_n(x) = 1$ for all $x \geq \lambda$. The minimum is attained by the first argument since otherwise, $\alpha < \lambda$ gives a contradiction. Thus, we obtain $n \leq \hat{n}_0(\vec{1}) \leq \hat{n}_0$. □

Note that $\hat{n}_0 \geq n$ implies that the adaptive critical values $\hat{\alpha}_{i:n}$, $i = 1, \ldots, n$, are dominated by the BH critical values. Thus, to control the FDR under the complete RM Model, the BH test should be used instead.

**Example 5.3**
Consider the distribution of the last proof which is a possible case of the RM and PRDS Model. Observe that (2.68) is violated for the Storey type estimators $\hat{n}_0(\lambda, \kappa)$ with $\kappa = \frac{1}{n}$ and Proposition 5.2 applies. The exact FDR is then given by

$$FDR = \frac{\alpha}{\lambda} \cdot \mathbb{E}\left(n\mathbb{1}\{U \leq \lambda\} \cdot \min\left\{1 - \lambda, \frac{\lambda}{n\alpha}\right\}\right) = \min\{\alpha n(1-\lambda), \lambda\},$$

cf. Blanchard and Roquain [8, Proposition 17]. Hence, $FDR > \alpha$ if $\lambda > \alpha$ and $n(1 - \lambda) > 1$.

Due to the drawback, let us introduce the following submodel of the RM Model.

**Model 5.4** (Block model with martingale structure)
*Consider the RM Model and suppose that the p-value vector $p$ can be divided by*

$$\{p_1, \ldots, p_n\} = \bigcup_{i=1}^{k} G_i \tag{5.2}$$

*into $k$ disjoint groups (blocks) $G_i = \{p_i : p_i$ belongs to group $i\}$, $i = 1, \dots, k$, of $p$-values. Let $\widetilde{G}_i = \{p_i \in G_i : H_i = 0\} \subset G_i$ be the subsets of true $p$-values which are uniformly distributed. Note that the division into the $G_i$'s is deterministic, whereas the $\widetilde{G}_i$'s depend on the random variable $H$. Moreover, assume that the groups $\widetilde{G}_i$, $i = 1, \dots, k$, are conditionally independent given $H$.*

Gontscharuk [24] considered a similar model, called block dependent $p$-values, where the $p$-values consist of independent groups, but with arbitrary dependence structure within each group. Gontscharuk [24] then introduces further conditions so that the $p$-values are weakly dependent and gives asymptotic results for the FWER and FDR which have already been discussed in Section 2.6. Briefly, the $p$-values are weakly dependent if the Glivenko-Cantelli Theorem holds for the empirical cumulative distribution function of true $p$-values. In contrast to this, the next theorem works for finite $n$.

In practice, the model may have the following meaning for genome data. Each $p$-value may be formulated for a specific gene and each group $G_i$ may stand for all those genes which come from one specific chromosome. It is often assumed that $p$-values of different chromosomes are independent, see also the motivation of Gontscharuk [24, page 91] for block dependent $p$-values. Then $\widetilde{G}_i$, $i = 1, \dots, k$, stand for independent portions of true $p$-values which come from different chromosomes. The $p$-values of $G_i$ and $\widetilde{G}_i$, respectively, may be reverse martingale dependent. Some of them may be equal, for instance.

It will be shown that a Storey type estimator

$$\hat{n}_0(\lambda, \kappa) = n \frac{1 - \hat{F}_n(\lambda) + \frac{\kappa}{n}}{1 - \lambda}, \quad 0 < \lambda < 1,\ \kappa \geq 1, \tag{5.3}$$

yields FDR control in a submodel of Model 5.4. Observe that the larger $\kappa$, the more conservative the estimator and hence the more conservative the adaptive SU test. The control under the complete Model 5.4 is not possible in general and there is a trade-off between the conservativeness of the estimator and the maximal size of the controlled submodel. To be more precise, the control of the FDR essentially depends on an appropriate choice of $\kappa$. Therefore, introduce the maximal group size and the residual

$$M = \max_{i \leq k} |G_i| \quad \text{and} \quad r = Mk - n, \tag{5.4}$$

respectively. Furthermore, assume that the number of true $p$-values is lower bounded by

$$N_0 \geq n_{min} \quad a.s. \tag{5.5}$$

for some constant $n_{min} > 0$. Then

$$\kappa > M + r + (n - n_{min}) \tag{5.6}$$

is sufficient for FDR control under the present distribution. If the groups are balanced, i.e. if $|G_1| = \ldots = |G_k|$, then $r$ vanishes and the best fit is obtained. Of course, $\hat{n}_0(\lambda, \kappa) > n$ may happen for large $r$ in the unbalanced case. As before, the BH test is then preferable.

**Theorem 5.5** (cf. Heesen and Janssen [29] Theorem 5.3)
*Consider Model 5.4 and the adaptive SU test based on (4.10) with $\rho = id$ and Storey type estimator (5.3). If (5.4)-(5.6) hold, then the adaptive SU test has finite sample FDR control by $\alpha$, i.e. we have $FDR \leq \alpha$.*

**Proof**. By Lemma 5.1 it suffices to prove that $\mathbb{E}\left(\frac{V(\lambda)}{\hat{n}_0(\lambda,\kappa)}\right) \leq \lambda$. For the $i$-th group, let us introduce the quantities

$$V_i(\lambda) = \sum_{j\,:\,p_j \in \widetilde{G}_i} \mathbb{1}\{p_j \leq \lambda\} \quad \text{and} \quad N_{0,i} = |\widetilde{G}_i|, \tag{5.7}$$

where $V(\lambda) = \sum_{i=1}^k V_i(\lambda)$. Thus, we obtain

$$\mathbb{E}\left(\frac{V(\lambda)}{\hat{n}_0(\lambda,\kappa)}\Big| H\right) = (1-\lambda)\cdot\mathbb{E}\left(\frac{\sum_{i=1}^k V_i(\lambda)}{n - n\hat{F}_n(\lambda) + \kappa}\Big| H\right) \tag{5.8}$$

$$\leq (1-\lambda)\cdot\mathbb{E}\left(\frac{\sum_{i=1}^k V_i(\lambda)}{N_0 - \sum_{i=1}^k V_i(\lambda) + \kappa}\Big| H\right). \tag{5.9}$$

Let us keep the condition under $H$. Whenever $|\widetilde{G}_i| > 0$, let $\widetilde{p}_i \in \widetilde{G}_i$ be a fixed true $p$-value which we may arbitrarily select. Observe that the $\widetilde{p}_i$'s are conditionally independent given $H$. Without restrictions, let us assume that $|\widetilde{G}_i| > 0$ holds for all groups. Otherwise, the groups with $|\widetilde{G}_i| = 0$ can be omitted. By our assumptions, there is at least one group which contains a true $p$-value. In the next step we are going to condition under $\sum_{i=2}^k V_i(\lambda)$. By

Lemma 6.6 of the appendix, a substitution of $V_1(\lambda)$ by $N_{0,1}\mathbb{1}\{\widetilde{p}_i \leq \lambda\}$ gives

$$\mathbb{E}\left(\frac{V_1(\lambda)+\sum_{i=2}^k V_i(\lambda)}{N_0 - V_1(\lambda)+\sum_{i=2}^k V_i(\lambda)+\kappa}\Big| H, \sum_{i=2}^k V_i(\lambda)\right)$$
$$\leq \mathbb{E}\left(\frac{N_{0,1}\mathbb{1}\{\widetilde{p}_i \leq \lambda\}+\sum_{i=2}^k V_i(\lambda)}{N_0 - N_{0,1}\mathbb{1}\{\widetilde{p}_i \leq \lambda\}+\sum_{i=2}^k V_i(\lambda)+\kappa}\Big| H, \sum_{i=2}^k V_i(\lambda)\right)$$

since $\mathbb{E}(V_1(\lambda)|H) = N_{0,1}\lambda = \mathbb{E}(N_{0,1}\mathbb{1}\{\widetilde{p}_i \leq \lambda\}|H)$ obviously implies that the distribution of $N_{0,1}\mathbb{1}\{\widetilde{p}_i \leq \lambda\}$ is just given by $P'$ in Lemma 6.6. If we proceed in this way, we arrive at the upper bound and obtain

$$(5.9) \quad \leq \quad (1-\lambda)\cdot\mathbb{E}\left(\frac{\sum_{i=1}^k N_{0,i}\mathbb{1}\{\widetilde{p}_i \leq \lambda\}}{N_0 - \sum_{i=1}^k N_{0,i}\mathbb{1}\{\widetilde{p}_i \leq \lambda\}+\kappa}\Big| H\right) \tag{5.10}$$

$$\leq \quad (1-\lambda)\cdot\mathbb{E}\left(\frac{\sum_{i=1}^k M\mathbb{1}\{\widetilde{p}_i \leq \lambda\}}{N_0 - \sum_{i=1}^k M\mathbb{1}\{\widetilde{p}_i \leq \lambda\}+\kappa}\Big| H\right) \tag{5.11}$$

$$= \quad (1-\lambda)\cdot\mathbb{E}\left(\frac{\sum_{i=1}^k \mathbb{1}\{\widetilde{p}_i \leq \lambda\}}{\frac{N_0}{M} - \sum_{i=1}^k \mathbb{1}\{\widetilde{p}_i \leq \lambda\}+\frac{\kappa}{M}}\Big| H\right), \tag{5.12}$$

because $x \mapsto \frac{x}{N_0 - x + \kappa}$ is increasing for $x \in [0, N_0+\kappa)$ and

$$M\sum_{i=1}^k \mathbb{1}\{\widetilde{p}_i \leq \lambda\} \leq kM \leq kM + M + N_0 - n_{min} \leq N_0 + \kappa$$

holds by (5.4)-(5.6). Finally,

$$\frac{N_0+\kappa}{M} \geq \frac{N_0 - n_{min} + M + r + n}{M} = \frac{N_0 - n_{min} + M + Mk}{M} \geq k+1$$

and Lemma 6.5 of the appendix imply

$$(5.12) \leq (1-\lambda)\cdot\mathbb{E}\left(\frac{\sum_{i=1}^k \mathbb{1}\{\widetilde{p}_i \leq \lambda\}}{k+1-\sum_{i=1}^k \mathbb{1}\{\widetilde{p}_i \leq \lambda\}}\Big| H\right) = \lambda(1-\lambda^k) \leq \lambda.$$

□

Lemma 5.1 and Theorem 5.5 may be extended to the control of the generalized error rates $\mathbb{E}\left(\frac{V}{\rho(R)}\right)$ of Meskaldji et al. [39] with $\rho$ as in (4.8).

# Chapter 6

# Appendix

This appendix contains some technical lemmas which are used in the proofs of the previous chapters. These lemmas are used a few times or would disturb the narrative flow.

The next lemma applies when a true $p$-value is set to zero. Therefore, let $p^{(i)} = (p_1, \ldots, p_{i-1}, 0, p_{i+1}, \ldots, p_n)$ be the vector of $p$-values, where the $i$-th $p$-value is set to zero, $R^{(i)} = R(p^{(i)})$ and $\hat{\alpha}^{(i)}_{i:n} = \hat{\alpha}_{i:n}(p^{(i)})$, $i = 1, \ldots, n$. Obviously, $R = R(p)$ and $\hat{\alpha}_{i:n} = \hat{\alpha}_{i:n}(p)$ are functions of the $p$-value vector $p$, but let us also write $\hat{\alpha}_{i:n} = \hat{\alpha}_{i:n}((\hat{F}_n(t))_{t\geq\lambda})$ if we want to accentuate that $\hat{\alpha}_{i:n}$ may also be defined as function of $(\hat{F}_n(t))_{t\geq\lambda}$. The random variables $R$ and $R^{(i)}$ refer to SU tests, whereas $R_{SD}$, $R^{(i)}_{SD}$, $R_{SUD(\lambda)}$ and $R^{(i)}_{SUD(\lambda)}$ refer to the corresponding SD and SUD tests, respectively. For a definition of SD and SUD test see Definition 2.28.

**Lemma 6.1**

*(a) Consider an adaptive SU test with data dependent critical values* $0 < \hat{\alpha}_{0:n} = \hat{\alpha}_{1:n} \leq \ldots \leq \hat{\alpha}_{n:n} \leq \lambda < 1$ *given by measurable functions*

$$\hat{\alpha}_{i:n} = \hat{\alpha}_{i:n}((\hat{F}_n(t))_{t\geq\lambda}), \quad i = 1, \ldots, n. \tag{6.1}$$

*Then we have*

$$\{p_i \leq \hat{\alpha}_{R:n}\} = \left\{p_i \leq \hat{\alpha}^{(i)}_{R^{(i)}:n}\right\}, \tag{6.2}$$

*and* $R = R^{(i)}$ *holds on* $\{p_i \leq \hat{\alpha}_{R:n}\}$.

*(b) For an adaptive SD test with same critical values as in (a) we get*

$$\{p_i \leq \hat{\alpha}_{R_{SD}:n}\} \subseteq \left\{p_i \leq \hat{\alpha}^{(i)}_{R^{(i)}_{SD}:n}\right\}, \tag{6.3}$$

*and* $R_{SD} = R_{SD}^{(i)}$ *holds on* $\{p_i \le \hat{\alpha}_{R_{SD}:n}\}$.

**Proof**. (a) Since $p_i \le \lambda$ holds on every set which is considered below we always have $\hat{\alpha}_{j:n} = \hat{\alpha}_{j:n}^{(i)}$, $j = 1, \dots, n$, on those sets by (6.1). Let $p_{1:n}^{(i)}, \dots, p_{n:n}^{(i)}$ be the order statistics of $p^{(i)}$. Then for $r = 1, \dots, n-2$, observe that

$$\begin{aligned}
&\{(p_1, \dots, p_n) \,:\, p_i \le \hat{\alpha}_{R:n}, R = r\} \\
&\quad = \Big\{(p_1, \dots, p_n) \,:\, p_i \le \hat{\alpha}_{r:n},\, p_{r:n} \le \hat{\alpha}_{r:n},\, p_{r+1:n} > \hat{\alpha}_{r+1:n},\, \dots,\, p_{n:n} > \hat{\alpha}_{n:n}\Big\} \\
&\quad = \Big\{(p_1, \dots, p_n) \,:\, p_i \le \hat{\alpha}_{r:n}^{(i)},\, p_{r:n} \le \hat{\alpha}_{r:n}^{(i)},\, p_{r+1:n} > \hat{\alpha}_{r+1:n}^{(i)},\, \dots,\, p_{n:n} > \hat{\alpha}_{n:n}^{(i)}\Big\} \\
&\quad = \Big\{(p_1, \dots, p_n) \,:\, p_i \le \hat{\alpha}_{r:n}^{(i)},\, p_{r:n}^{(i)} \le \hat{\alpha}_{r:n}^{(i)},\, p_{r+1:n}^{(i)} > \hat{\alpha}_{r+1:n}^{(i)},\, \dots,\, p_{n:n}^{(i)} > \hat{\alpha}_{n:n}^{(i)}\Big\} \\
&\quad = \{(p_1, \dots, p_n) \,:\, p_i \le \hat{\alpha}_{R:n}^{(i)}, R^{(i)} = r\}
\end{aligned}$$

since $i \mapsto \hat{\alpha}_{i:n}$ is non-decreasing. By similar arguments, we also obtain

$$\{(p_1, \dots, p_n) \,:\, p_i \le \hat{\alpha}_{R:n}, R = r\} = \Big\{(p_1, \dots, p_n) \,:\, p_i \le \hat{\alpha}_{R:n}^{(i)}, R^{(i)} = r\Big\}$$

for $r = n-1$ and $r = n$. However, there is nothing to show for $r = 0$ since then $\{p_i \le \hat{\alpha}_{R:n}, R = 0\} \subseteq \{p_i \le \hat{\alpha}_{1:n}, R = 0\} = \emptyset \; = \{R^{(i)} = 0\}$. Combining the above results for $r = 0, \dots, n$ gives (6.2) and $R = R^{(i)}$ holds on $\{p_i \le \hat{\alpha}_{R:n}\}$.
(b) The statement follows by an analogue argument for SD tests since

$$\begin{aligned}
&\{p_i \le \hat{\alpha}_{R_{SD}:n},\, R_{SD} = r\} \\
&\quad = \{p_i \le \hat{\alpha}_{r:n},\, p_{1:n} \le \hat{\alpha}_{1:n}, \dots, p_{r:n} \le \hat{\alpha}_{r:n},\, p_{r+1:n} > \hat{\alpha}_{r+1:n}\} \\
&\quad = \Big\{p_i \le \hat{\alpha}_{r:n}^{(i)},\, p_{1:n} \le \hat{\alpha}_{1:n}^{(i)}, \dots, p_{r:n} \le \hat{\alpha}_{r:n}^{(i)},\, p_{r+1:n} > \hat{\alpha}_{r+1:n}^{(i)}\Big\} \\
&\quad \subseteq \Big\{p_i \le \hat{\alpha}_{r:n}^{(i)},\, p_{1:n}^{(i)} \le \hat{\alpha}_{1:n}^{(i)}, \dots, p_{r:n}^{(i)} \le \hat{\alpha}_{r:n}^{(i)},\, p_{r+1:n}^{(i)} > \hat{\alpha}_{r+1:n}^{(i)}\Big\} \\
&\quad = \Big\{p_i \le \hat{\alpha}_{R_{SD}^{(i)}:n}^{(i)},\, R_{SD}^{(i)} = r\Big\}
\end{aligned}$$

and the cases $r = 0$ and $r = n$ has to be considered similar to (a). □

**Remark 6.2**

(a) Lemma 6.1 applies generally without any distributional assumption and also for false $p$-values.
(b) For non adaptive SU and SD tests with critical values $0 < \alpha_{1:n} \le \dots \le$

$\alpha_{n:n} < 1$, which are included in the adaptive case, Lemma 6.1 reads as

$$\{p_i \leq \alpha_{R:n}\} = \{p_i \leq \alpha_{R^{(i)}:n}\} \tag{6.4}$$

for the SU test and

$$\{p_i \leq \alpha_{R_{SD}:n}\} \subseteq \left\{p_i \leq \alpha_{R_{SD}^{(i)}:n}\right\},$$

for the SD test, respectively.
(c) Under the assumptions of (b) and for $SUD(\lambda)$ tests it is also easy to verify that

$$\{p_i \leq \alpha_{R_{SUD(\lambda)}:n}\} \subseteq \left\{p_i \leq \alpha_{R_{SUD(\lambda)}^{(i)}:n}\right\} \tag{6.5}$$

since $R_{SUD(\lambda)}^{(i)}$ is non increasing in each $p$-value.

Remark 6.2 (b) is a special case of Lemma 6.1. But note that a proof of Remark 6.2 (b) is actually dispensable.

**Lemma 6.3**
*Let $(\Omega, \mathcal{F}, \{\mathcal{M}(n, p_1, p_2, 1 - p_1 - p_2) : (p_1, p_2) \in A\})$ be a statistical space for some set $A \subseteq \{(p_1, p_2) : 0 \leq p_1, p_2 \leq 1, p_1 + p_2 \leq 1\}$ and let $(M_1, M_2, M_3)$ be distributed according to the multinomial distribution $\mathcal{M}(n, p_1, p_2, 1 - p_1 - p_2)$. If the interior of $A$ is not empty, then $(M_1, M_2)$ is a complete statistic.*

**Proof**. Let $f : \{(m_1, m_2) : m_1, m_2 \in \mathbb{N}, 0 \leq m_1 + m_2 \leq n\} \to \mathbb{R}$ be an arbitrary function with $\mathbb{E}_{(p_1,p_2)}(f(M_1, M_2)) = 0$ for all $(p_1, p_2) \in A$ where $\mathbb{E}_{(p_1,p_2)}$ denotes the expectation according to $\mathcal{M}(n, p_1, p_2, 1-p_1-p_2)$. Furthermore, let $\mathring{A}$ be the interior of $A$ on which $p_1, p_2, 1 - p_1 - p_2 > 0$ hold. For $(p_1, p_2) \in \mathring{A}$ we observe

$$\begin{aligned} 0 &= (1 - p_1 - p_2)^{-n} \cdot \mathbb{E}_{(p_1,p_2)}(f(M_1, M_2)) \\ &= \sum_{\substack{0 \leq m_1, m_2 \leq n \\ 0 \leq m_1 + m_2 \leq n}} \frac{f(m_1, m_2) \cdot n!}{m_1! m_2! (n - m_1 - m_2)!} p_1^{m_1} p_2^{m_2} (1 - p_1 - p_2)^{-m_1 - m_2} \\ &= \sum_{(m_1+m_2)=0}^{n} \sum_{m_2=0}^{m_1+m_2} \frac{f(m_1, m_2) \cdot n!}{m_1! m_2! (n - m_1 - m_2)!} \left(\frac{p_1}{1 - p_1 - p_2}\right)^{m_1+m_2} \cdot \left(\frac{p_2}{p_1}\right)^{m_2}. \end{aligned}$$

Consider the continuously differentiable reparametrization

$$g : (p_1, p_2) \mapsto \left(\frac{p_1}{1 - p_1 - p_2}, \frac{p_2}{p_1}\right)$$

defined for $(p_1, p_2) \in \mathring{A}$. Observe that the determinant of the derivation satisfies

$$\det \mathrm{D}g_{|(p_1,p_2)} = \det \begin{pmatrix} \frac{1-p_2}{(1-p_1-p_2)^2} & \frac{p_1}{(1-p_1-p_2)^2} \\ -\frac{p_2}{p_1^2} & \frac{1}{p_1} \end{pmatrix} = \frac{-1}{p_1(1-p_1-p_2)^2} \neq 0$$

and hence by the inverse function theorem there exists an open set $A_0 \subseteq \mathring{A}$ and an open set $B_0$ such that $g : A_0 \to B_0$ is bijective. Moreover, for $(q_1, q_2) \in B_0$ we have

$$\sum_{(m_1+m_2)=0}^{n} \sum_{m_2=0}^{m_1+m_2} \frac{f(m_1,m_2) \cdot n!}{m_1!m_2!(n-m_1-m_2)!} q_1^{m_1+m_2} q_2^{m_2} = 0. \tag{6.6}$$

Since every derivative of the polynomial in (6.6) is equal to zero on $B_0$ it can easily be obtained by Taylor's formula that every coefficient of the polynomial is equal to zero. Hence, $f \equiv 0$ and $(M_1, M_2)$ is a complete statistic. □

**Lemma 6.4**
*Let $0 < \lambda < \gamma \leq 1$ and $\hat{n}_0$ be an estimator as in (4.9). Consider the gBI Model and assume that*

$$F_{0,H,i}(t\gamma) \leq t \cdot F_{0,H,i}(\gamma), \quad t \in [0,1], \ i \in I_0. \tag{6.7}$$

*If the functions*

$$p_i \mapsto \hat{n}_0(p_1, \ldots, p_n) \quad \textit{are non decreasing on } [0,\gamma] \tag{6.8}$$

*for $i = 1, \ldots, n$, then we have*

$$\mathbb{E}\left(\frac{V(\lambda)}{\hat{n}_0}\Big|(n\hat{F}_n(t))_{t\geq\gamma}\right) \leq \mathbb{E}_{BI}\left(\frac{V(\lambda)}{\hat{n}_0}\Big|(n\hat{F}_n(t))_{t\geq\gamma}\right) \tag{6.9}$$

*and*

$$\mathbb{E}\left(\sum_{i\in I_0} \frac{1}{\hat{n}_0^{(i)}}\Big|(n\hat{F}_n(t))_{t\geq\gamma}\right) \leq \mathbb{E}_{BI}\left(\sum_{i\in I_0} \frac{1}{\hat{n}_0^{(i)}}\Big|(n\hat{F}_n(t))_{t\geq\gamma}\right). \tag{6.10}$$

*$\mathbb{E}_{BI}$ denotes the expectation with respect to the BI Model, where the distribution of $(H, \xi)$ stays the same, i.e. conditioned under $H$ the true p-values are exchanged by independent on $[0,1]$ uniformly distributed p-values.*

**Proof**. Let us condition under $\mathcal{F}_\gamma = \sigma(H, \mathbb{1}\{p_i \leq s\} : s \geq \gamma, 1 \leq i \leq n)$. Under this condition, let $\{i_1 < \ldots < i_k\} \subset I_0$ be the index set of true $p$-values which are less than or equal to $\gamma$. Those true $p$-values $p_{i_1}, \ldots, p_{i_k}$ are still independent and their marginal conditional distribution is stochastically larger than the uniform distribution on $[0, \gamma]$ since

$$P(p_{i_j} \leq t | \mathcal{F}_\gamma) = P(p_{i_j} \leq t | p_{i_j} \leq \gamma, H_{i_j} = 0)$$
$$= \frac{F_{0,H,i_j}(t)}{F_{0,H,i_j}(\gamma)} \leq \frac{t}{\gamma}, \quad t \in [0, \gamma], \ j = 1, \ldots, k,$$

follows by (6.7). Now let $\widetilde{p}_{i_1}, \ldots, \widetilde{p}_{i_k}$ be independent and uniformly distributed on $[0, \gamma]$. By Müller and Stoyan [41, Theorem 3.3.8] it follows that

$$(p_{i_1}, \ldots, p_{i_k}) \overset{st}{\geq} (\widetilde{p}_{i_1}, \ldots, \widetilde{p}_{i_k}) \tag{6.11}$$

since both random vectors have the independence copula under the above condition and where $\overset{st}{\geq}$ denotes the usual multivariate stochastically larger property, see Müller and Stoyan [41, Definition 3.3.1 (a)] for a definition. Furthermore, since $\frac{V(\lambda)}{\hat{n}_0}$ and $\sum_{i \in I_0} \frac{1}{\hat{n}_0^{(i)}}$ are non increasing in each $p$-value $p_{i_1}, \ldots, p_{i_k}$ on $[0, \gamma]$, the inequalities (6.9) and (6.10) just follow by the definition of the multivariate stochastically larger property and by integration. □

**Lemma 6.5** (cf. Lemma 6.1 in Janssen and Heesen [28])
*Let $(V_1, V_2, V_3)$ be distributed according to the multinomial distribution $\mathcal{M}(n, p_1, p_2, p_3)$ with $p_3 > 0$ and $n \geq 0$. Then we have*

$$E\left(\frac{V_1}{n + 1 - V_1 - V_2}\right) = \frac{p_1}{p_3}(1 - (p_1 + p_2)^n). \tag{6.12}$$

**Proof**. A simple calculation gives

$$E\left(\frac{V_1}{n + 1 - V_1 - V_2}\right) \tag{6.13}$$
$$= \sum_{\substack{k_1 + k_2 \leq n \\ k_1 > 0, k_2 \geq 0}} \frac{n!}{(k_1 - 1)! k_2! (n + 1 - k_1 - k_2)!} p_1^{k_1} p_2^{k_2} p_3^{n - k_1 - k_2} \tag{6.14}$$
$$= \frac{p_1}{p_3} \cdot \sum_{\substack{j + k_2 \leq n - 1 \\ j, k_2 \geq 0}} \frac{n!}{j! k_2! (n - j - k_2)!} p_1^j p_2^{k_2} p_3^{n - j - k_2}, \tag{6.15}$$

where the last equality follows from the substitution $j = k_1 - 1$. Observe that the last term adds the probabilities of the multinomial distribution times a constant factor. Thus, extending the missing probabilities yields

$$(6.15) \quad = \frac{p_1}{p_3} \cdot \left(1 - \sum_{j+k_2=n} \frac{n!}{j!k_2!(n-j-k_2)!} p_1^j p_2^{k_2} p_3^{n-j-k_2}\right)$$

$$= \frac{p_1}{p_3} \cdot \left(1 - \sum_{j=0}^{n} \frac{n!}{j!(n-j)!} p_1^j p_2^{n-j}\right) = \frac{p_1}{p_3} \cdot \left(1 - (p_1+p_2)^n\right).$$

□

**Lemma 6.6** (cf. Heesen and Janssen [29] Lemma 6.3)
*Let $f : \{0, \ldots, m\} \to \mathbb{R}$ be a convex function and $P = \sum_{j=0}^{m} p_j \epsilon_j$ be a distribution on $\{0, 1 \ldots, n\}$, where $\epsilon_j$ denotes the Dirac distribution on $\{j\}$. Furthermore, introduce*

$$P' = \left(1 - \sum_{j=1}^{m} \frac{j}{m} p_j\right) \cdot \epsilon_0 + \sum_{j=1}^{m} \frac{j}{m} p_j \cdot \epsilon_m. \tag{6.16}$$

*Then we have $\mathbb{E}_P(f) \leq \mathbb{E}_{P'}(f)$ and $\mathbb{E}_P(id) = \mathbb{E}_{P'}(id)$ for the identity function.*

**Proof**. Obviously, the last assertion holds by the definition of $P'$. Moreover, by the convexity of $f$, we obtain

$$\mathbb{E}_P(f) = \sum_{j=0}^{m} f(j) p_j \leq \sum_{j=0}^{m} \left(\frac{m-j}{m} f(0) p_j + \frac{j}{m} f(m) p_j\right) = \mathbb{E}_{P'}(f).$$

□

# List of Abbreviations and Symbols

| | |
|---|---|
| $A^c$ | Complement of the set $A$ |
| $\|A\|, \#A$ | Cardinality of the set $A$ |
| $\mathring{A}$ | Interior of the set $A$ |
| AORC | Asymptotically Optimal Rejection Curve |
| $a.s.$ | Almost surely |
| $a.e.$ | Almost everywhere |
| $\alpha_{1:n}, \ldots, \alpha_{n:n}$ | Critical values of a stepwise multiple test |
| $\hat{\alpha}_{1:n}, \ldots, \hat{\alpha}_{n:n}$ | Data dependent critical values |
| $\alpha_{1:n}^{BH}, \ldots, \alpha_{n:n}^{BH}$ | Critical values of the Benjamini Hochberg test |
| BI Model | Basic Independence Model |
| $B(n, p)$ | Binomial distribution with parameters $n$ and $p$ |
| det | Determinant |
| df | Distribution function |
| DU | Dirac-Uniform |
| $DU(n, n_0)$ | Dirac-Uniform configuration with parameters $n$ and $n_0$ |
| ecdf | Empirical cumulative distribution function |
| ENFR | Expected number of false rejections |
| $\hat{F}_n$ | Ecdf of $p_1, \ldots, p_n$ |
| $F_0$ | Marginal df of true $p$-values in the mixture models |
| $F_1$ | Marginal df of false $p$-values in the mixture models |
| $F_{0,H,i}$ | Marginal df of the $i$-th true $p$-value in the gBI Model |
| FDP | False discovery proportion |
| FDR | False discovery rate |

| | |
|---|---|
| $FDR(t)$ | FDR of the single step test with fixed threshold $t$ |
| $\widehat{FDR}(t)$ | Estimator for $FDR(t)$ |
| FWER | Family Wise Error Rate |
| gBI Model | Generalized Basic Independence Model |
| $H$ | Vector of the status of the null hypotheses |
| $\mathcal{H}_i$ | $i$-th null hypothesis |
| $H_i = 0$ | $i$-th null hypothesis is true |
| $H_i = 1$ | $i$-th null hypothesis is false |
| $I_0$ | Index set of true null hypotheses / $p$-values |
| $I_1$ | Index set of false null hypotheses / $p$-values |
| $id$ | Identity function |
| iff | If and only if |
| $i.i.d.$ | Independent and identical distributed |
| $\text{IM}(f)$ | Image of a mapping $f$ |
| LFC | Least favorable parameter configuration |
| $\lambda\!\!\lambda_{\mid(0,1)}$ | Lebesgue measure restricted to $(0, 1)$ |
| $\mathcal{L}(X)$ | distribution of the random variable $X$ |
| $\mathcal{L}(X\|P)$ | distribution of the random variable $X$ under the probability measure $P$ |
| $\mathcal{L}(P\|A)$ | conditional distribution given $A$ |
| $\mathcal{M}(n, p_1, \ldots, p_k)$ | Multinomial distribution with parameters $n, k$ and $p_1, \ldots, p_k$ |
| $n$ | Number of hypotheses / $p$-values |
| $N_0$ | Random number of true null hypotheses |
| $n_0$ | Fixed number of true null hypotheses |
| $\hat{n}_0$ | Estimator for $N_0, n_0$ and related terms |
| $\hat{n}_0(\lambda)$ | Storey estimator |
| $\mathbb{N}$ | $\{1, 2, \ldots\}$ |
| $N(\mu, \sigma)$ | Normal distribution with mean $\mu$ and variance $\sigma$ |
| $N(\mu, \Sigma)$ | Multivariate normal distribution with mean vector $\mu$ and covariance matrix $\Sigma$ |
| NRDS | Negatively Regression Dependence on a Subset |
| $\hat{m}$ | Estimator for the effective number of true null hypotheses which concern the FDR |

| | |
|---|---|
| $\hat{m}(\lambda, \gamma)$ | Generalized Storey estimator |
| $p$ | Vector of $p$-values |
| $p_i$ | $i$-th $p$-value |
| $p_{i:n}$ | $i$-th order statistic of $p_1, \ldots, p_n$ |
| $p_{I_0}$ | Vector of true $p$-values |
| $p_{I_1}$ | Vector of false $p$-values |
| $P^n$ | $n$-fold product measure of $P$ |
| $P^X = \mathcal{L}(X\|P)$ | distribution of the random variable $X$ under the probability measure $P$ |
| $\mathcal{P}$ | Family of probability measures |
| $\mathcal{P}_{n,\cdot}$ | Distributions of the BI, gBI or RM Model for fixed $n$ |
| $\mathcal{P}^C_{n,gBI}$ | Distributions of the gBI Model for fixed $n$ with convex dfs $F_{0,H,i}$ |
| pFDR | Positive FDR |
| $\pi_0$ | (Expected) proportion of true null hypotheses |
| $\hat{\pi}_0$ | Estimator for $\pi_0$ |
| $\hat{\varpi}(t, \cdot)$ | Estimator for $\pi_0 F_0(t)$ |
| PRDS | Positively Regression Dependence on a Subset |
| $\Phi$ | Cumulative distribution function of the $N(0,1)$ distribution |
| $R$ | Number of rejections of the multiple test |
| $R(t)$ | Number of $p$-values less than or equal to $t$ |
| $\mathbb{R}$ | Real numbers |
| $\bar{\mathbb{R}}$ | $\mathbb{R} \cup \{-\infty, \infty\}$ |
| RM Model | Reverse Martingale Model |
| $r(x)$ | Rejection curve |
| SD test | Step-down test |
| SU test | Step-up test |
| SUD test | Step-up-down test |
| $\sigma(X)$ | $\sigma$-algebra generated by $X$ |
| $U(a, b)$ | Uniform distribution on the interval $[a, b]$ |
| $V$ | Number of false rejections of the multiple test |
| $V(t)$ | Number of true $p$-values less than or equal to $t$ |

| | |
|---|---|
| w.r.t. | with respect to |
| $X \sim P, F$ | $X$ is distributed according to the measure $P$ and distribution function $F$, respectively |
| $\lfloor x \rfloor$ | $\max\{m \in \mathbb{Z} : m \leq x\}$ |
| $x \vee y$ | $\max(x, y)$ |
| $x \wedge y$ | $\min(x, y)$ |
| $(\xi)_{i \leq n}$ | Vector of possible false $p$-values |
| $\mathbb{Z}$ | $\{0, \pm 1, \pm 2, \ldots\}$ |
| $(\zeta)_{i \leq n}$ | Vector of possible true $p$-values |
| $\mathbb{1}A$, $\mathbb{1}_A$ | Indicator function of $A$ |
| $\xrightarrow[n \to \infty]{a.s.}$ | Almost sure convergence |
| $\emptyset$ | Empty set |
| $\overset{st}{\geq}$ | Stochastically larger property |

# List of Figures


2.1 $FDR_{DU(n,n_0)}$ plot with $n = 300$ for (2.41) (blue curve), (2.39) with $k = 283, 250, 223$ based on (2.41) (black curves in decreasing order), BH test (increasing red straight line) and lower bound (2.26) (green curve). . . . . . . . . . . . . . . . . . . . . . . . . . 40

4.1 Distribution functions of true $p$-values in a one sided normal mean testing problem. . . . . . . . . . . . . . . . . . . . . . . . . . . . . 82

4.2 Realization of an ecdf of $p$-values in a one sided normal mean testing problem. . . . . . . . . . . . . . . . . . . . . . . . . . . . . 83

4.3 Limit ecdf of $p$-values in a one sided normal mean testing problem. 96

4.4 Other possible limit ecdf. . . . . . . . . . . . . . . . . . . . . . . . 97

4.5 Areas of the estimation process. . . . . . . . . . . . . . . . . . . . . 98

# List of Tables


2.1 Worst case FDR for different choices of $k$ in (2.39) based on (2.41). 41

3.1 Expectation of the original Storey estimator. . . . . . . . . . . . 62

4.1 Monte-Carlo Simulation of the FDR of several multiple tests. . . 102

# Bibliography


[1] Y. Benjamini. Discovering the false discovery rate. *J. R. Stat. Soc. Ser. B Stat. Methodol.*, 72(4):405–416, 2010.

[2] Y. Benjamini and Y. Hochberg. Controlling the false discovery rate: a practical and powerful approach to multiple testing. *J. Roy. Statist. Soc. Ser. B*, 57(1):289–300, 1995.

[3] Y. Benjamini and Y. Hochberg. On the adaptive control of the false discovery rate in multiple testing with independent statistics. *J. Educ. Behav. Statist.*, 25(1):60–83, 2000.

[4] Y. Benjamini, A. M. Krieger, and D. Yekutieli. Adaptive linear step-up procedures that control the false discovery rate. *Biometrika*, 93(3):491–507, 2006.

[5] Y. Benjamini and D. Yekutieli. The control of the false discovery rate in multiple testing under dependency. *Ann. Statist.*, 29(4):1165–1188, 2001.

[6] G. Blanchard, T. Dickhaus, E. Roquain, and F. Villers. On least favorable configurations for step-up-down tests. arXiv:1108.5262v1, August 2011.

[7] G. Blanchard and E. Roquain. Two simple sufficient conditions for FDR control. *Electron. J. Stat.*, 2:963–992, 2008.

[8] G. Blanchard and E. Roquain. Adaptive false discovery rate control under independence and dependence. *J. Mach. Learn. Res.*, 10:2837–2871, 2009.

[9] A. Celisse and S. Robin. A cross-validation based estimation of the proportion of true null hypotheses. *J. Statist. Plann. Inference*, 140(11):3132–3147, 2010.

[10] X. Chen and R. W. Doerge. Generalized estimators for multiple testing: proportion of true nulls and false discovery rate. Technical report, Department of Statistics, Purdue University, West Lafayette, USA, April 2012.

[11] T. Dickhaus. Randomized $p$-values for multiple testing of composite null hypotheses. *J. Statist. Plann. Inference*, 143(11):1968–1979, 2013.

[12] S. Döhler. A sufficient criterion for control of generalised error rates in multiple testing. arXiv:1307.4394v1, July 2013.

[13] D. Donoho and J. Jin. Higher criticism for detecting sparse heterogeneous mixtures. *Ann. Statist.*, 32(3):962–994, 2004.

[14] B. Efron, R. Tibshirani, J. D. Storey, and V. Tusher. Empirical Bayes analysis of a microarray experiment. *J. Amer. Statist. Assoc.*, 96(456):1151–1160, 2001.

[15] A. Farcomeni. Some results on the control of the false discovery rate under dependence. *Scand. J. Statist.*, 34(2):275–297, 2007.

[16] H. Finner, T. Dickhaus, and M. Roters. Dependency and false discovery rate: asymptotics. *Ann. Statist.*, 35(4):1432–1455, 2007.

[17] H. Finner, T. Dickhaus, and M. Roters. On the false discovery rate and an asymptotically optimal rejection curve. *Ann. Statist.*, 37(2):596–618, 2009.

[18] H. Finner and V. Gontscharuk. Controlling the familywise error rate with plug-in estimator for the proportion of true null hypotheses. *J. R. Stat. Soc. Ser. B Stat. Methodol.*, 71(5):1031–1048, 2009.

[19] H. Finner, V. Gontscharuk, and T. Dickhaus. False discovery rate control of step-up-down tests with special emphasis on the asymptotically optimal rejection curve. *Scand. J. Stat.*, 39(2):382–397, 2012.

[20] H. Finner and M. Roters. On the false discovery rate and expected type I errors. *Biom. J.*, 43(8):985–1005, 2001.

[21] Y. Gavrilov, Y. Benjamini, and S. K. Sarkar. An adaptive step-down procedure with proven FDR control under independence. *Ann. Statist.*, 37(2):619–629, 2009.

[22] C. Genovese and L. Wasserman. Operating characteristics and extensions of the false discovery rate procedure. *J. R. Stat. Soc. Ser. B Stat. Methodol.*, 64(3):499–517, 2002.

[23] C. Genovese and L. Wasserman. A stochastic process approach to false discovery control. *Ann. Statist.*, 32(3):1035–1061, 2004.

[24] V. Gontscharuk. *Asymptotic and Exact Results on FWER and FDR in Multiple Hypotheses Testing.* PhD thesis, Heinrich-Heine-Universität Düsseldorf, Germany, 2010,
http://docserv.uni-duesseldorf.de/servlets/DocumentServlet?id=16990.

[25] W. Guo and M. B. Rao. On control of the false discovery rate under no assumption of dependency. *J. Statist. Plann. Inference*, 138:3176–3188, 2008.

[26] W. Guo and M. B. Rao. On optimality of the Benjamini-Hochberg procedure for the false discovery rate. *Statist. Probab. Lett.*, 78(14):2024–2030, 2008.

[27] P. Hall and J. Jin. Innovated higher criticism for detecting sparse signals in correlated noise. *Ann. Statist.*, 38(3):1686–1732, 2010.

[28] P. Heesen and A. Janssen. Dynamic adaptive multiple tests with finite sample FDR control [1]. Preprint, October 2013.

[29] P. Heesen and A. Janssen. Inequalities for the false discovery rate (FDR) under dependence. Preprint, February 2014.

[30] L. Jager and J. A. Wellner. Goodness-of-fit tests via phi-divergences. *Ann. Statist.*, 35(5):2018–2053, 2007.

[31] J. Jin. Detecting a target in very noisy data from multiple looks. In *A festschrift for Herman Rubin*, volume 45 of *IMS Lecture Notes Monogr. Ser.*, pages 255–286. Inst. Math. Statist., Beachwood, OH, 2004.

[1] Note added when the dissertation was under revision on October 23, 2014. After the dissertation was submitted, a new version was published, see arXiv:1410.6296v1, October 2014

[32] M. Langaas, B. H. Lindqvist, and E. Ferkingstad. Estimating the proportion of true null hypotheses, with application to DNA microarray data. *J. R. Stat. Soc. Ser. B Stat. Methodol.*, 67(4):555–572, 2005.

[33] E. L. Lehmann and G. Casella. *Theory of point estimation*. Springer Texts in Statistics. Springer, New York, second edition, 1998.

[34] E. L. Lehmann and J. P. Romano. Generalizations of the familywise error rate. *Ann. Statist.*, 33(3):1138–1154, 2005.

[35] E. L. Lehmann and J. P. Romano. *Testing statistical hypotheses*. Springer Texts in Statistics. Springer, New York, third edition, 2005.

[36] K. Liang and D. Nettleton. Adaptive and dynamic adaptive procedures for false discovery rate control and estimation. *J. R. Stat. Soc. Ser. B. Stat. Methodol.*, 74(1):163–182, 2012.

[37] N. Meinshausen and J. Rice. Estimating the proportion of false null hypotheses among a large number of independently tested hypotheses. *Ann. Statist.*, 34(1):373–393, 2006.

[38] D. E. Meskaldji. *Multiple Comparison Procedures for Large Correlated Data with Application to Brain Connectivity Analysis*. PhD thesis, École Polytechnique Fédérale de Lausanne, Switzerland, 2013, http://infoscience.epfl.ch/record/188425.

[39] D. E. Meskaldji, J.-P. Thiran, and S. Morgenthaler. A comprehensive error rate for multiple testing. arXiv:1112.4519v4, July 2013.

[40] D. E. Meskaldji, J.-P. Thiran, and S. Morgenthaler. Optimality in multiple comparison procedures. arXiv:1307.2614v1, July 2013.

[41] A. Müller and D. Stoyan. *Comparison methods for stochastic models and risks*. Wiley Series in Probability and Statistics. John Wiley & Sons Ltd., Chichester, 2002.

[42] J. Pfanzagl. *Parametric statistical theory*. de Gruyter Textbook. Walter de Gruyter & Co., Berlin, 1994. With the assistance of R. Hamböker.

[43] S. Pounds and C. Cheng. Improving false discovery rate estimation. *Bioinformatics*, 20(11):1737–1745, 2004.

[44] S. Pounds and C. Cheng. Robust estimation of the false discovery rate. *Bioinformatics*, 22(16):1979–1987, 2006.

[45] A. W. Roberts, J. Roberts, and D. E. Varberg. *Convex functions*. Academic Press [A subsidiary of Harcourt Brace Jovanovich, Publishers], New York-London, 1973. Pure and Applied Mathematics, Vol. 57.

[46] J. P. Romano and A. M. Shaikh. On stepdown control of the false discovery proportion. In *Optimality*, volume 49 of *IMS Lecture Notes Monogr. Ser.*, pages 33–50. Inst. Math. Statist., Beachwood, OH, 2006.

[47] J. P. Romano and A. M. Shaikh. Stepup procedures for control of generalizations of the familywise error rate. *Ann. Statist.*, 34(4):1850–1873, 2006.

[48] J. P. Romano and M. Wolf. Control of generalized error rates in multiple testing. *Ann. Statist.*, 35(4):1378–1408, 2007.

[49] E. Roquain and F. Villers. Supplement to "Exact calculations for false discovery proportion with application to least favorable configurations", 2010.

[50] E. Roquain and F. Villers. Exact calculations for false discovery proportion with application to least favorable configurations. *Ann. Statist.*, 39(1):584–612, 2011.

[51] S. K. Sarkar. Some results on false discovery rate in stepwise multiple testing procedures. *Ann. Statist.*, 30(1):239–257, 2002.

[52] S. K. Sarkar. On methods controlling the false discovery rate. *Sankhyā*, 70(2, Ser. A):135–168, 2008.

[53] S. K. Sarkar. Two-stage stepup procedures controlling FDR. *J. Statist. Plann. Inference*, 138(4):1072–1084, 2008.

[54] M. Scheer. *Controlling the number of false rejections in multiple hypotheses testing*. PhD thesis, Heinrich-Heine-Universität Düsseldorf, Germany, 2012 http://docserv.uni-duesseldorf.de/servlets/DocumentServlet?id=23691.

[55] T. Schweder and E. Spjøtvoll. Plots of p-values to evaluate many tests simultaneously. *Biometrika*, 69(3):493–502, 1982.

[56] G. R. Shorack and J. A. Wellner. *Empirical processes with applications to statistics*. Wiley Series in Probability and Mathematical Statistics: Probability and Mathematical Statistics. John Wiley & Sons Inc., New York, 1986.

[57] B. W. Silverman. *Density estimation for statistics and data analysis*. Monographs on Statistics and Applied Probability. Chapman & Hall, London, 1986.

[58] R. J. Simes. An improved bonferroni procedure for multiple tests of significance. *Biometrika*, 73(3):751–754, 1986.

[59] J. D. Storey. A direct approach to false discovery rates. *J. R. Stat. Soc. Ser. B Stat. Methodol.*, 64(3):479–498, 2002.

[60] J. D. Storey. The positive false discovery rate: a Bayesian interpretation and the $q$-value. *Ann. Statist.*, 31(6):2013–2035, 2003.

[61] J. D. Storey, J. E. Taylor, and D. Siegmund. Strong control, conservative point estimation and simultaneous conservative consistency of false discovery rates: a unified approach. *J. R. Stat. Soc. Ser. B Stat. Methodol.*, 66(1):187–205, 2004.

[62] J. D. Storey and R. Tibshirani. Statistical significance for genomewide studies. *PNAS*, 100(16):9440–9445, 2003.

[63] J. W. Tukey. T13 n: The higher criticism. Course Notes, Statistics 411, Princeton Univ., 1976.

[64] H. Witting. *Mathematische Statistik. I.* B. G. Teubner, Stuttgart, 1985. Parametrische Verfahren bei festem Stichprobenumfang. [Parametric methods for fixed sample size].

[65] A. Zeisel, O. Zuk, and E. Domany. FDR control with adaptive procedures and FDR monotonicity. *Ann. Appl. Stat.*, 5(2A):943–968, 2011.

# Erklärung

Ich versichere an Eides Statt, dass die Dissertation von mir selbständig und ohne unzulässige fremde Hilfe unter Beachtung der "Grundsätze zur Sicherung guter wissenschaftlicher Praxis an der Heinrich-Heine-Universität Düsseldorf" erstellt worden ist. Die Dissertation wurde in der vorgelegten oder in ähnlicher Form noch bei keiner anderen Institution eingereicht. Ich habe bisher keine erfolglosen Promotionsversuche unternommen.

Philipp Heesen                Düsseldorf, den 08.05.2014