Martingale Methods for Control of False Discovery Rate and Expected Number of False Rejections

Inaugural - Dissertation

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

vorgelegt von

Julia Benditkis

aus Berdytschiw

Düsseldorf, März 2015

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

Referent: Prof. Dr. A. Janssen

Koreferent: Apl. Prof. Dr. H. Finner

Tag der mündlichen Prüfung: 23.06.2015

Abstract

In this dissertation we deal with multiple test procedures that control the False Discovery Rate (FDR) and the Expected Number of False Rejections (ENFR) for independent as well as dependent test statistics.

Chapter 1 serves as an introduction to multiple testing and related error rates.

In Chapter 2 we restrict attention to several dependence structures. A new kind of dependence, namely the martingale dependence, is introduced. Among others, useful relationships between different dependence structures are provided.

Chapter 3 is devoted to the Asymptotically Optimal Rejection Curve (AORC) introduced in Finner et al. [2009]. Based on the AORC, we propose some estimates for the FDR that lead to FDR- and ENFR-controlling step down (SD) test procedures.

In Chapter 4 we show how the power of SD procedures can be improved in case the underlying p-values are positive regression dependent on a subset of true null hypotheses or positive orthant dependent. Motivated by the SD procedure in Benjamini and Liu [1999], we propose a new technique for the improvement of the power of SD procedures based on the increase of the smallest critical values.

In Chapter 5 we focus on adaptive SD procedures based on the so-called $\beta-$adjusted AORC. Several new results on FDR- and ENFR-control under martingale dependence are provided. In the case of independent test statistics we prove FDR- and ENFR-control in an alternative way. Moreover, based on results in Chapter 4 we propose an improvement of the procedure introduced in Gavrilov et al. [2009] without loss of FDR-control.

Chapter 6 is devoted to martingale dependent test statistics. We provide sufficient conditions for that kind of dependence. Various examples for martingale dependent statistics are presented.

In Chapter 7 we consider the situation where null hypotheses and alternatives cannot be separated from each other - the case of uniformly distributed p-values under alternatives.

Finally, in Chapter 8 we introduce a necessary condition for FDR-control of SD tests.

Zusammenfassung

In dieser Dissertation beschäftigen wir uns mit multiplen Testprozeduren, die sowohl unter unabhängigen als auch unter abhängigen p-Werten die "False Discovery Rate" (FDR) bzw. die "Expected Number of False Rejections" (ENFR) kontrollieren.
Kapitel 1 ist eine Einleitung in die grundliegenden Konzepte.
In Kapitel 2 werden verschiedene Abhängigkeitsstrukturen betrachtet. Es wird ein neues Abhängigkeitskonzept, nämlich die Martingalabhängigkeit, eingeführt. Außerdem beweisen wir nützliche Beziehungen zwischen unterschiedlichen Abhängigkeitskonzepten.
Kapitel 3 widmet sich der "Asymptotisch Optimalen Ablehnkurve" (AORC), die in Finner et al. [2009] vorgestellt worden ist. Darauf basierend schlagen wir einige Schätzer für die FDR vor, die später zu FDR- bzw. ENFR-kontrollierenden step down (SD) Prozeduren führen.
In Kapitel 4 zeigen wir, wie man die Güte einer FDR-kontrollierenden SD Prozedur verbessern kann, wenn die p-Werte bestimmte Abhändgigkeitsvorausetzungen erfüllen, nämlich "positive regression dependence on a subset" (PRDS) oder "positive orthant dependence" (POD). Motiviert durch die SD Prozedur von Benjamini und Liu [1999] schlagen wir eine neue Technik vor, die eine Verbesserung der Güte durch Vergrößerung der kleinsten kritischen Werte von SD Prozeduren ermöglicht.
In Kapitel 5 betrachten wir die SD Prozedur, die auf der sogenannten $\beta-$adjustierten AORC basiert. Wir beweisen neue Resultate für FDR- und ENFR-Kontrolle unter Martingalabhängigkeit und schlagen alternative Beweise für den Fall unabhängiger p-Werte vor. Basierend auf den Ergebnissen aus Kapitel 4 wird eine Verbesserung der Güte (ohne Verlust der FDR-Kontrolle), von der in Gavrilov et al. vorgestellten SD Prozedur, vorgeschlagen.
In Kapitel 6 studieren wir martingalabhängige Teststatistiken. Es werden hinreichende Bedingungen für die Martingalabhängigkeit vorgeschlagen und bewiesen. Dabei werden verschiedene Beispiele für diese Abhängigkeitsart vorgestellt.
Wir beschäftigen uns in Kapitel 7 mit der Situation, in der man die Nullhypothesen von den Alternativen nicht trennen kann, d.h., mit dem Fall gleichverteilter Alternativen.
Schließlich werden notwendige Bedingungen für die FDR-Kontrolle einer SD Prozedur im Kapitel 8 bewiesen.

Acknowledgments

There are many people who I want to thank for their help. First and foremost, **I deeply thank my parents for their love, understanding and unconditional support**.
I am indebted to Prof. Dr. Arnold Janssen for his guidance, supervision and support of this thesis.
Futhermore, I would like to thank Prof. Dr. Helmut Finner for many fruitful discussions, for fair comments, many helpful suggestions and for his patience.
I would very much like to express my special thanks to Prof. Dr. Klaus Janßen for proof-reading this work and for many important comments.
My colleagues at the chair of mathematical statistics and probability theory at the Heinrich-Heine-University Düsseldorf deserve thanks for friendly and pleasant working atmosphere. In particular, I would also like to thank Dr. Philipp Heesen for proof-reading this work and for the idea of Example 6.16.
Special thanks go to Dr. Veronika Gontscharuk for many helpful discussion and suggestions.
My husband, my brother, my children and all my friends deserve special thanks for their love, support and patience!

Nomenclature

$\xrightarrow{\mathcal{D}}$ Convergence in distribution

$\mathbb{I}(p \leqslant t)$ Indicator function of the event $\{p \leqslant t\}$

$\mathbb{I}_A$ Indicator function of the event $A \subset \Omega$

$\gamma-$FDX False Discovery Exceedance

$\hat{F}_n$ Empirical cumulative distribution function of p-values

$\mathbb{N}$ Set of natural numbers

$\mathbb{R}$ Set of real numbers

$\succcurlyeq_{st}$ Stochastically larger

$a \wedge b$ $\min(a, b)$

I $\{1, ..., n\}$

I_0 $\{i \in I : H_i \text{ is true }\}$

I_1 $\{i \in I : H_i \text{ is false }\}$

$k-$FWER Generalized Family-Wise Error Rate

$N(\mu, \sigma)$ Normal distribution with mean μ and variance σ^2

$R(\tau)$ Number of all rejections of the procedure τ

$S(\tau)$ Number of correct rejections of the procedure τ

$U(0, 1)$ Uniformly distribution on the interval $[0, 1]$

$V(\tau)$	Number of false rejections of the procedure τ
(s)MD	(super-)Martingale Dependent
a.s.	almost surely
AORC	Asymptotically Optimal Rejection Curve
ASDP	Adaptive Step Down Procedure of Gavrilov
BH	Benjamini and Hochberg
BIA	Basic Independence Assumptions
DM	Dirac-Martingale
DU	Dirac-Uniform
ENFR	Expected Number of False Rejections
FDP	False Discovery Proportion
FDR	False Discovery Rate
FWER	Family-Wise Error Rate
i.i.d.	independent identically distributed
iff	if and only if
LFC	Least Favourable Configuration
MD	Martingale Dependent
MTP	Multiple Testing Procedure
PA	Positive Association
POD	Positive Orthant Dependent
PRDS	Positive Regression Dependent on a Subset
SD	Step Down
SU	Step Up

w.l.o.g. without loss of the generality

w.r.t. with respect to

Contents

Preface

An important and challenging problem in multiple testing theory is the construction of valid and, at the same time, powerful multiple testing procedures that control an appropriate Type I error rate. One of the well-known error rates is the so called Family-Wise Error Rate (FWER), that is, the probability to reject at least one true null hypothesis. Thereby, the disadvantage is that test procedures controlling the FWER are often too conservative, especially for larger families of hypotheses, for example, cf. Benjamini and Hochberg [1995], Gordon [2007]. In contrast to tests that control the FWER, procedures controlling the so-called False Discovery Rate (FDR) are typically less restrictive. The FDR originally introduced in Benjamini and Hochberg [1995] is defined to be the expected proportion of false rejections among all rejections. Most of the existing FDR-controlling procedures can be roughly divided into two main groups: procedures which exhaust the pre-chosen significance level α well and operate only with independent test statistics, e.g., cf. step up (SU) tests in Storey et al. [2004], and procedures which allow some dependence structures, e.g., cf. linear SU tests in Benjamini and Hochberg [1995]. However, procedures controlling the FDR under independence as well as under some dependence are typically less powerful than tests controlling the FDR under independence only. Regrettably, independence of the underlying single tests and their test statistics are rare in practice. To develop procedures which control the FDR under some dependence assumptions and exhaust the significance level well, some authors, for instance, Finner et al. [2009], Gontscharuk [2010] and Neuvial [2013], resort to asymptotic considerations. The drawback of this method is that the asymptotic behavior of test statistics and its finite counterpart may differ drastically so that tests controlling an asymptotic error rate may lead to a considerable violation of the pre-chosen level α in the finite setting.

The main focus of this thesis lies on step down (SD) procedures which control the FDR and the so-called Expected Number of False Rejections (ENFR) for a finite number of null hypotheses. The underlying test statistics are assumed to fulfill some special dependence assumptions.

Chapter 1 is an introduction to basic concepts and an outline of some Type I error rates which are relevant for this thesis.

Chapter 2 is devoted to some dependence concepts. Among others, we introduce a new type of dependence, the so-called martingale dependence. We also prove some relationships between different dependence structures which will be useful for later considerations in Chapter 5.

In Chapter 3 we restrict attention to the Asymptotically Optimal Rejection Curve (AORC) proposed by Finner et al. [2009]. It is known that several stepwise test procedures related to that rejection curve exhaust the FDR level α under specific asymptotic models. First, we provide an alternative motivation for the AORC in mixture models. Then we introduce an AORC-based estimator $\hat{\alpha}(t)$ (say) for the FDR related to a deterministic critical value $t \in [0, 1]$ and prove a central limit theorem for this estimator. Note that estimators which are used in Chapter 5 to construct FDR- and ENFR-controlling procedures, are all related to the estimator $\hat{\alpha}(t)$ and, hence, to the AORC.

In Chapter 4 we discuss a possibility to improve the power of SD tests by an increase of the smallest critical value without loss of the FDR-control. The underlying p-values, that is, specific test statistics, must be either positive regression dependent on the subset of true null hypotheses (PRDS) or positive orthant dependent on the subset of true null hypotheses.

In Chapter 5 we expand upon the results in Gavrilov et al. [2010] and Scheer [2012]. We consider models, where the underlying p-values that belong to true null hypotheses are martingale dependent, and prove the ENFR- and FDR-control for corresponding SD tests. Moreover, we introduce a class of SD procedures which control the ENFR and FDR or FDR-related error rates under dependence. Finally, we propose new error rates which are controlled by the aforementioned SD tests under martingale dependence, positive association and PRDS.

Chapter 6 deals with (super-)martingale dependence structures. Here we propose sufficient conditions for martingale dependence. Some non-trivial examples of martingale dependent p-values that are PRDS at the same time are investigated in detail. Thereby, such p-values fulfill the assumptions of the main theorem from Chapter 5. Some numerical examples confirm our theoretical results from Chapter 5.

Several proofs in Chapters 5 and 6 are based on the martingale theory. In the FDR-framework such proofs were first applied in Storey et al. [2004]. It seems that until now martingale methods were applied to SU procedures only. In this thesis we show that the martingale theory can also be used in the area of SD procedures. It looks like martingale approaches may simplify the existing proofs and enable us to get new, generalized results in the field of ENFR- and FDR-control.

In Chapter 7 we restrict our attention to the case where test statistics under alternatives have the same distribution as test statistics under null hypotheses, e.g., the case where null hypotheses and alternatives cannot be separated. We prove that under some basic independence assumptions the linear SU procedure maximizes the FDR if p-values that belong to alternatives are uniformly distributed on $[0, 1]$.

Finally, Chapter 8 presents a discussion concerning some necessary assumptions for FDR-control under general dependence.

Chapter 1

General framework and basic concepts

1.1 Introduction (or how I explained hypothesis testing to my children)

Do you know, last Monday you broke a blue cup, and I laughed and said: "It doesn't matter!" (the cup was old and I wanted to throw it away anyway). So, you could think: "It doesn't matter, if we break a cup! Mama will laugh!" It was your hypothesis H_0. Then you want to carry out an experiment to confirm your hypothesis. What will you do? You break another cup, this time a green one and see how I react. This time I do not laugh and instead reply that it is not funny when one intentionally breaks cups. So you cannot accept your hypothesis H_0, you reject it.
Perhaps you think: "Maybe green was Mama's favorite color and that was why she reacted as she did". You have to take other cups, for instance, green, white, yellow, blue and multicolored one. Now you move to a multiple testing problem. You have five hypotheses:

$$H_{\text{color}}: \qquad \text{mama laughs, if we break a cup}_{\text{color}},$$

$$\text{color=green, white, yellow, blue, multicolored}$$

... I could tell from the look on the faces of my children that they understood the theory well and wanted to move to the practice as soon as possible.

Now things are getting serious - we start with the mathematics.

1.2 Notations

Let $X = (X_1, ..., X_n)$, $n \in \mathbb{N}$, be observations on a probability space $(\Omega, \mathcal{F}, \mathbb{P})$, where $\mathbb{P}$ is a (parametric or nonparametric) family of probability distributions without any special requirement. Let X follow some unknown probability distribution $P \in \mathbb{P}$. We consider a general problem of simultaneously testing n null hypotheses $H_1, \ldots, H_n$. Thereby, null hypotheses H_i, $i \in I \equiv \{1, \ldots, n\}$, are defined as nonempty subfamilies of $\mathbb{P}$ fulfilling $\emptyset \neq H_i \subset \mathbb{P}$, $i \in I$. The corresponding alternatives are given by $\mathbb{P} \setminus H_i$, $i \in I$. We say that a null hypothesis H_i is true if the underlying probability distribution P belongs to H_i. Consequently, H_i is false if $P \notin H_i$. Let $I_0 \equiv I_0(P)$ be the index set of true null hypotheses, that is, $i \in I_0$ if H_i is true. Further, $I_1 \equiv I_1(P) = I \setminus I_0$ denotes the index set of false null hypotheses. The number of true null hypotheses is denoted by $n_0 = |I_0|$ and the number of false ones is $n_1 = n - n_0$.

Remark 1.1
The number of true null hypotheses n_0 and the number of false ones n_1 are typically unknown in practice.

For what follows we assume that testing of an individual null hypothesis H_i is based on a specific test statistic, namely a p-value $p_i \equiv p_i(X)$ fulfilling $p_i : (\Omega, \mathcal{F}) \mapsto ([0,1], \mathcal{B})$, $i \in I$, where $\mathcal{B}$ is the Borel-σ-algebra over the interval $[0,1]$. The vector of p-values is defined by $p \equiv p(X) = (p_1, \ldots, p_n)$. In most cases we assume that p_i, $i \in I_0$ are stochastically greater than uniformly distributed $\mathrm{U}(0,1)$ random variables. Let $\phi = (\phi_1, ..., \phi_n) : [0,1]^n \to \{0,1\}^n$ be a multiple testing procedure (MTP), that is, a rule that makes decisions about each null hypothesis H_i, $i \in I$, in the following way. A null hypothesis H_i is rejected (H_i is significant) if $\phi_i \equiv \phi_i(p) = 1$ and H_i is accepted if $\phi_i = 0$.

Remark 1.2
MTPs can often be represented in the form $\phi_i = \mathbb{I}(p_i \leq \tau)$, $i \in I$, for a suitable threshold $\tau \equiv \tau(p)$ so that a null hypothesis H_i is rejected if the corresponding p-value p_i fulfills $p_i \leqslant \tau$. Thereby, τ can be considered as a stopping rule of the MTP. It is not necessarily that τ is a stopping time in the sense of the theory of stochastic processes.

In this thesis we restrict attention to MTPs in the aforementioned form and, for the sake of simplicity, often denote such MTPs by the related stopping rule τ.

1.3 Stepwise procedures and rejection curves

A lot of existing MTPs belong to the class of stepwise test procedures. These procedures can be defined in terms of critical values or in terms of rejection curves. In this subsection we describe how the most prominent members of that class, namely step down (SD) and step up (SU) tests work.

Let $p_1, \ldots, p_n$ be p-values and let $p_{1:n} \leq \ldots \leq p_{n:n}$ be the corresponding order statistics. The empirical distribution function of the p-values is given by

$$\hat{F}_n(t) = \frac{1}{n} \sum_{i=1}^{n} \mathbb{I}(p_i \leqslant t).$$

For what follows, we use the convention $\max\{\emptyset\} \equiv 0$.

Definition 1.3

The stopping rule of an SD procedure based on a set of critical values $0 \leqslant a_1 \leqslant ... \leqslant a_n \leqslant 1$ is defined by

$$\tau_{SD} = \max\{a_i : p_{j:n} \leqslant a_j, \quad for \quad all \quad j \leqslant i\}$$

and the stopping rule of an SU test based on the same critical values is

$$\tau_{SU} = \max\{a_i : p_{i:n} \leqslant a_i\}.$$

Consequently, SD and SU tests reject all null hypotheses such that related p-values are not larger than the threshold τ_{SD} and τ_{SU}, respectively.

Remark 1.4

Comparing SD and SU procedures based on the same critical values, we observe that an SU procedure rejects at least as many null hypotheses as the corresponding SD test.

A set of critical values $0 \leqslant a_1 \leqslant ... \leqslant a_n \leqslant 1$ can be generated by a rejection function r (say) as follows. Let r by a strictly increasing continuous function fulfilling $r : [0, 1] \to [0, \infty)$, $r(0) = 0$ and $r(1) \geqslant 1$. The related critical values are given by

$$a_i = r^{-1}\left(\frac{i}{n}\right), \quad i \in I. \tag{1.1}$$

Under suitable assumptions stopping rules of SD and SU tests can be rewritten in terms of rejection curves.

Remark 1.5

Setting $a_0 \equiv 0$ and assuming that the multiplicity of each positive p-value is 1 P-a.s., that is, $\hat{F}_n(p_{i:n}) = \frac{i}{n}$ P-a.s. if $p_{i:n} > 0$, the stopping rules of the SD and SU procedures based on a rejection curve r fulfill

$$\tau_{SD} = a_{j^*} \;\; P\text{-}a.s. \;\; \textit{for } j^* = \max\{i \in I : r(p_{j:n}) \leqslant \hat{F}_n(p_{j:n}) \;\; \textit{for all} \;\; j \leqslant i\}$$

and

$$\tau_{SU} = a_{j^*} \;\; P\text{-}a.s. \;\; \textit{for } j^* = \max\{i \in I : r(p_{i:n}) \leqslant \hat{F}_n(p_{i:n})\},$$

respectively.

Note that in Chapter 5 we represent a stopping rule of an SD procedure as the smallest crossing point between the path of a suitable stochastic process $\{\hat{\alpha}(t), t \in (0,1)\}$ with the α-line.

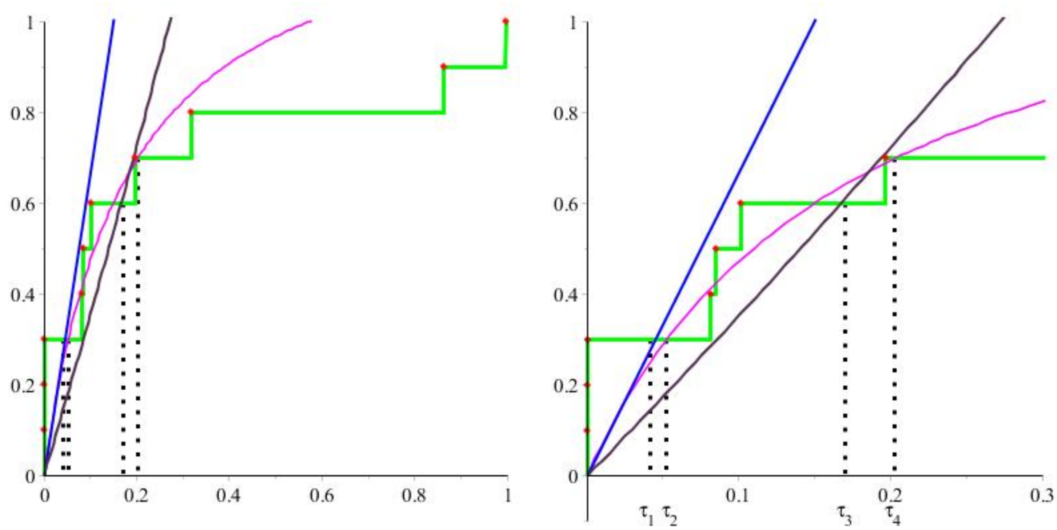

Figure 1.1: The empirical distribution function of $n = 10$ p-values (green curve) together with three rejection curves. Here τ_1 is the stopping rule of the SD and SU tests based on the blue rejection curve, τ_2 is the stopping rule of the SD test related to the magenta rejection curve, τ_3 corresponds to the SD and SU procedures generated by the black curve and τ_4 is the threshold of the SU test related to the magenta curve.

Figure 1.1 illustrates SD and SU procedures based on three different rejection curves. Note that SD and SU procedures based on the same rejection curve may lead to the same threshold and, hence, to the same number of rejections. Obviously, the lower a rejection curve, the more null hypotheses can be rejected by a stepwise test procedure.

The following example shows that the assumption about the multiplicity of positive p-values in Remark 1.5 is necessary for SD procedures if the probability for a tie is positive for the underlying measure $P \in \mathbb{P}$.

Example 1.6

For some fixed $\alpha \in (0,1)$ and $n = 3$ consider an SD procedure with critical values generated by the rejection curve $r(t) = \frac{t}{\alpha}$. Hence, the related critical values are given by

$$a_0 \equiv 0,\ a_1 = \frac{\alpha}{3},\ a_2 = \frac{2\alpha}{3},\ a_3 = \alpha.$$

Let p-values be such that

$$p_1 = p_2 = p_3 = \frac{\alpha}{2}.$$

Then we get by Definition 1.3 $\tau_{SD} = 0$, while Remark 1.5 provides $a_{j^*} = \alpha$. Hence, $\tau_{SD} \neq a_{j^*}$ P-a.s. if the underlying distribution P is such that $P(\alpha/3 < p_1 = p_2 = p_3 < 2\alpha/3) > 0$ for some $i, j \in I$ with $i \neq j$.

Now we show that the condition that a rejection curve is strictly increasing is not necessary, that is, a set of valid critical values for a stepwise procedure can be generated by a non-decreasing rejection function.

Definition 1.7

Let r be a non-decreasing function such that $r : [0,1] \to [0,\infty)$, $r(0) = 0$ and $r(1) \geq 1$. Then the related critical values a_i, $i \in I$, are defined by the right continuous inverse function $\tilde{r}^{-1}$ (say), i.e.,

$$a_i = \sup\left\{t \in [0,1] : r(t) = \frac{i}{n}\right\} \equiv \tilde{r}^{-1}\left(\frac{i}{n}\right),\quad i \in I. \tag{1.2}$$

The next lemma provides an equivalent definition of the stopping rule of an SU and/or SD procedures defined by a non-decreasing rejection curve.

Lemma 1.8

Let $r : [0,1] \to [0,\infty)$ be a continuous non-decreasing function with $r(0) = 0$ and $r(1) \geq 1$.

Let a_i, $i \in I$, be critical values given by (1.2).
(a)Then the stopping rule τ_{SU} from Definition 1.3 fulfills

$$\tau_{SU} = a_{j^*} \;\; P\text{-}a.s. \;\; for \; j^* = \max\{i \in \{1,...,n\} : r(p_{i:n}) \leqslant \hat{F}_n(p_{i:n})\}.$$

(b)If, in addition, $\hat{F}_n(p_{i:n}) = \frac{i}{n}$ P-a.s. for all $i \in I$, then the stopping rule τ_{SD} from Definition 1.3 fulfills

$$\tau_{SD} = a_{j^*} \;\; P\text{-}a.s. \;\; for \; j^* = \max\{i \in I : r(p_{j:n}) \leqslant \hat{F}_n(p_{j:n}) \;\; for \;\; all \;\; j \leqslant i\}.$$

Proof. Obviously by Definition 1.3 it is enough to prove the statement of the lemma for $j \in \{i \in I : \hat{F}_n(p_{i:n}) = \frac{i}{n}\}$ for SU procedures. Hence, we have to show

$$p_{j:n} \leqslant \tilde{r}^{-1}\left(\frac{j}{n}\right) \quad \text{iff} \quad r\left(p_{j:n}\right) \leqslant \hat{F}_n(p_{j:n}) \tag{1.3}$$

for all $j \in \{i \in I : \hat{F}_n(p_{i:n}) = \frac{i}{n}\}$. Let

$$p_{j:n} \leqslant \tilde{r}^{-1}\left(\frac{j}{n}\right) \tag{1.4}$$

for some $j \in \{i \in I : \hat{F}_n(p_{i:n}) = \frac{i}{n}\}$. Since r is continuous, the supremum will be attained. Hence, (1.4) yields

$$r(p_{j:n}) \leqslant r\left(\tilde{r}^{-1}\left(\frac{j}{n}\right)\right) = \frac{j}{n} \equiv \hat{F}_n(p_{j:n}). \tag{1.5}$$

Now we assume that $r\left(p_{j:n}\right) \leqslant \hat{F}_n(p_{j:n})$ is fulfilled for some for $j \in \{i \in I : \hat{F}_n(p_{i:n}) = \frac{i}{n}\}$. Then

$$p_{j:n} \leq \tilde{r}^{-1}(r(p_{j:n})) \leqslant \tilde{r}^{-1}\left(\hat{F}_n(p_{i:n})\right) \equiv \tilde{r}^{-1}\left(\frac{j}{n}\right), \tag{1.6}$$

which completes the proof for the SU procedure.
The part (b) can be proved similar to part (a) with only one difference, here $I = \{i \in I : \hat{F}_n(p_{i:n}) = \frac{i}{n}\}$ is fulfilled, so that the equivalence (1.3) is valid for all $i \in I$. Otherwise, for the cases $\hat{F}_n(p_{i:n}) > \frac{i}{n}$, we get in (1.5)

$$r(p_{j:n}) \leqslant \frac{j}{n} > \hat{F}_n(p_{j:n}),$$

which may injure (1.3) (cf. Example 1.6). ■

The following counter example shows that the right continuous inverse $\tilde{r}^{-1}$ can not be replaced by the left continuous inverse

$$r^{-1}(t) = \inf\left\{t \in [0,1] : r(t) = \frac{i}{n}\right\}$$

in Lemma 1.8.

Example 1.9

Let us consider the following rejection curve (cf. Figure 1.2)

$$r(t) = 2t\mathbb{I}\left(t \leqslant \frac{1}{4}\right) + \frac{1}{2}\mathbb{I}\left(t \in \left(\frac{1}{4}, \frac{3}{4}\right]\right) + \left(\frac{9}{4}t - \frac{19}{16}\right)\mathbb{I}\left(t \in \left(\frac{3}{4}, 1\right]\right).$$

Then the critical values which are defined in terms of right continuous inverse $\tilde{r}^{-1}$ are

$$a_1 = \frac{1}{8},\ a_2 = \frac{3}{4},\ a_3 = \frac{31}{36},\ a_4 = \frac{35}{36}.$$

The critical values which are defined in terms of the left continuous inverse r^{-1} are

$$b_1 = \frac{1}{8},\ b_2 = \frac{1}{4},\ b_3 = \frac{31}{36},\ b_4 = \frac{35}{36}.$$

Now we consider the set of p-values

$$p_1 = \frac{1}{8},\ p_2 = \frac{1}{2},\ p_3 = \frac{31}{32},\ p_4 = \frac{71}{72}.$$

Definition 1.3 provides the stopping rule of the SU procedure

$$\tau_{SU}(a) = a_2 = \frac{3}{4} \text{ and } \tau_{SU}(b) = b_1 = \frac{1}{8}.$$

From the other hand, Lemma 1.8 implies $\tau_{SU} = \frac{3}{4}$ in both cases. Hence, we have $\tau_{SU} = \tau_{SU}(a) > \tau_{SU}(b)$.

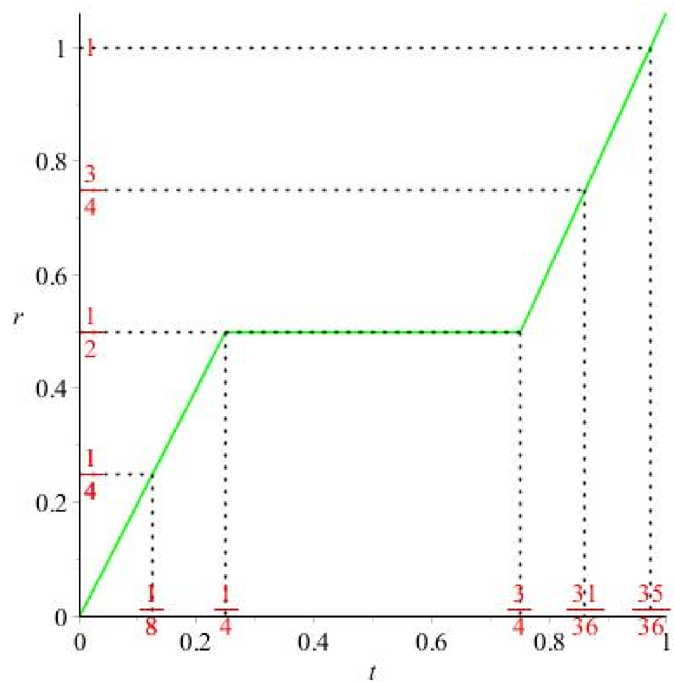

Figure 1.2: Rejection curve from Example 1.9

.

Remark 1.10

- *Critical values of a stepwise procedure can alternatively be defined by a critical value function ρ (say) such that $\rho : [0,1] \to [0,1]$, which is assumed to be continuous non-decreasing function satisfying $\rho(0) = 0$ and $\rho(t) > 0$ for all $t \in (0,1]$. Then the corresponding critical values are given by $a_i = \rho(i/n)$ and the induced rejection curve is defined as a left continuous inverse function*

$$\rho^{-1}(t) = \inf\{s \in [0,1] : \rho(s) = t\},\ t \in [0,1],$$

cf. Finner et al. [2009], p.8 in Gontscharuk [2010] and p.11 in Scheer [2012]. Thereby, rejection curves which are defined in such a way are strictly increasing, but they may be discontinuous, whereas rejection curves from Definition 1.7 may be non-decreasing but continuous.

- *Critical values of the most famous stepwise procedures which control the FDR are generated by strictly increasing rejection curves, so that the corresponding critical value curve is defined by $\rho(t) = r^{-1}(t)$ for all $t \in [0,1]$, cf. Benjamini and Hochberg*

[1995], Storey et al. [2004], Gavrilov et al. [2009] and Finner et al. [2009]. For example, the rejection curve of the Benjamini and Hochberg (BH) procedure is $r(t) = \frac{t}{\alpha}$, the rejection curve of the SD procedure proposed in Gavrilov et al. [2009] is $r(t) = \frac{n+1}{n}\frac{t}{t(1-\alpha)+\alpha}$ and the asymptotically optimal rejection curve in Finner et al. [2009] is $f_\alpha(t) = \frac{t}{t(1-\alpha)+\alpha}$.

1.4 Error rates

In statistical hypothesis testing, a type I error denotes the rejecting a true null hypothesis and a type II error is the failing to reject a false null hypothesis. By testing many null hypotheses, several incorrect decisions, that is, type I and II errors, as well as correct decisions are possible simultaneously. For a stopping rule τ of an MTP for testing n null hypotheses $H_1, \ldots, H_n$ let the number of false rejections, that is, false positives, be denoted by

$$V(\tau) = \sum_{i \in I_0} \phi_i(\tau) = \sum_{i \in I_0} \mathbb{I}(p_i \leqslant \tau)$$

and the number of all rejections, that is, all discoveries, be given by

$$R(\tau) = \sum_{i \in I} \phi_i(\tau) = \sum_{i=1}^{n} \mathbb{I}(p_i \leqslant \tau).$$

Further, we denote the number of true negatives and positives by $U(\tau)$ and $S(\tau)$, respectively, and the number of false negatives, i.e., type II errors, by $T(\tau)$. Table 1.1 provides possible outcomes of an MTP.

Remark 1.11
Due to the fact that the underlying sample $X_1, \ldots, X_n$ follows an unknown distribution $P \in \mathbb{P}$ and $\tau \equiv \tau(p(X))$, outcomes in Table 1.1, among others, $V(\tau)$ and $R(\tau)$, depend on $P \in \mathbb{P}$.

Remark 1.12
Note that the number of all rejections $R(\tau)$ can be observed while the number of false rejections $V(\tau)$ as well as $U(\tau)$, $T(\tau)$ and $S(\tau)$ are typically unknown.

By testing a singular null hypothesis the stopping rule of a statistical test is typically equal to the predefined significance level α so that the probability to accept a true null hypothesis is $1 - \alpha$. Clearly, if we perform n independent individual tests, each of them at

	Null true	Alternative true	Total
not called significant	$U(\tau)$	$T(\tau)$	$n - R(\tau)$
called significant	$V(\tau)$	$S(\tau)$	$R(\tau)$
Total	n_0	n_1	n

Table 1.1: Possible outcomes of a multiple testing procedure.

level α, the probability to accept all true null hypotheses is $(1-\alpha)^{n_0}$ and, consequently, the probability for at least one false rejection is $1-(1-\alpha)^{n_0}$, if the corresponding p-values are uniformly $\mathrm{U}(0,1)$-distributed. For example, for $n_0 = 100$ we get that the probability to reject at least one true null hypotheses is about $99,4\%$. Hence, we need another significance concept and MTPs controlling some other error rates.

In the following subsections we summarize several well-known multiple error rates. An extensive survey of the development of different error criteria is given by Scheer [2012], pp.5-10. Further useful references are Dudoit and van der Laan [2010] and Dickhaus [2014].

1.4.1 Family-Wise Error Rate

Along with the development of the multiple testing theory the ***Family-Wise Error Rate*** (FWER) is the oldest and the most conservative error rate. For a given MTP based on the stopping rule τ the (actual) FWER is defined as the probability to reject at least one true null hypothesis, i.e.,

$$\mathrm{FWER}(\tau) \equiv \mathrm{FWER}(\tau, P) = P(V(\tau) > 0).$$

Definition 1.13

We say that an MTP controls the FWER at level α, if FWER$(\tau) \leqslant \alpha$ for the underlying $P \in \mathbb{P}$.

The most known procedures controlling the FWER were suggested by Bonferroni [1936] and Holm [1979]. These procedures operate without any assumptions about dependence between p-values. In contrast to these procedures Sidak [1967] and Hommel [1986] procedures control the FWER if the underlying p-values are independent.

1.4.2 Generalized Family-Wise Error Rate

An interesting attempt "to reduce the conservatism" of the FWER-concept is the suggestion of a new error rate, namely ***generalized FWER***, or k-FWER. By definition the k-FWER is the probability of at least k false rejections, where $k \in \mathbb{N}$, i.e.,

$$k\text{-FWER}(\tau) \equiv k\text{-FWER}(\tau, P) = P(V(\tau) \geqslant k).$$

Definition 1.14
Let $k \in \mathbb{N}$ be fixed. We say that an MTP controls the k-FWER at level α if the inequality k-FWER$(\tau) \leqslant \alpha$ holds for the underlying probability measure $P \in \mathbb{P}$.

The intuitive point of the criticism of k-FWER control is the fact that k-FWER $= 0$ for $k > n_0$. Since n_0 is typically unknown to the experimenter, it can be difficult to find some fitting k. Nonetheless, MTPs controlling the k-FWER with $k \geqslant 2$ are obviously not as conservative as procedures that control the original FWER. Even though the concept of the generalized FWER was discussed earlier, Victor [1982] for instance has proposed that for given $k \leqslant n_0$ one should allow up to $k-1$ false rejections, the term k-FWER was introduced in Lehmann and Romano [2005]. Lehmann and Romano proposed a class of SD procedures controlling the k-FWER under arbitrary dependence. One year later, Romano and Shaikh [2006] proposed a SU procedure controlling the k-FWER under arbitrary dependence.

1.4.3 False Discovery Rate

Benjamini and Hochberg's [1995] paper has given the theory of multiple tests a new impulse in more liberal direction. This paper proposed the term ***False Discovery Rate*** (FDR). For a given MTP τ the FDP is defined to be the expectation of the proportion of all false rejections among all rejections under the underlying distribution $P \in \mathbb{P}$. Thereby, the proportion of all false rejections among all rejections is called the ***False Discovery Proportion*** (FDP), that is,

$$\text{FDP}(\tau) = \frac{V(\tau)}{R(\tau)} \mathbb{I}(R(\tau) > 0).$$

Clearly, the FDP depends on the underlying probability measure $P \in \mathbb{P}$.

Note that we always use the convention $\frac{0}{0} \equiv 0$.

It follows

$$\mathrm{FDR}(\tau) \equiv \mathrm{FDR}(\tau, P) = \mathbb{E}\left[\mathrm{FDP}(\tau)\right],$$

where $\mathbb{E} \equiv \mathbb{E}(P)$ denotes the expectation under the underlying measure $P \in \mathbb{P}$.

Definition 1.15
We say that an MTP τ controls the FDR at level α, if FDR$(\tau) \leqslant \alpha$ for the underlying measure $P \in \mathbb{P}$.

Benjamini and Hochberg proposed an SU procedure (BH SU) which controls the FDR at level $\alpha n_0/n$ for all $n_0 = 1, \ldots, n$, if the underlying "true" p-values are independent and independent from the "false" ones or if they fulfill the special dependence assumption called positive regression dependence, cf. Benjamini and Yekutieli [2001] which will be treated later. Since the proportion n_0/n may be very small, the power of the BH SU procedure can be also small.

A further approach is to estimate the number of the true null hypotheses n_0 by some appropriate estimator $\hat{n}_0$ by means of the data X to improve the power of the BH SU procedure. Such procedures are called adaptive BH-procedures. The most famous adaptive SU procedures are the procedures of Storey et al. [2004] and Benjamini at al. [2006]. Publications which play an important role in this dissertation are Finner et al. [2009], Gavrilov et al. [2009] and Scheer [2012]. Among others, we provide alternative proofs and generalizations for some of results in Gavrilov et al. [2009] and Scheer [2012].

1.4.4 Expected Number of False Rejections

The ***Expected Number of False Rejections*** (ENFR) is the expectation of the number of false discoveries under the underlying distribution $P \in \mathbb{P}$, that is,

$$\mathrm{ENFR}(\tau) \equiv \mathrm{ENFR}(\tau, P) = \mathbb{E}[V(\tau)],$$

where $\mathbb{E} \equiv \mathbb{E}(P)$ is the expectation for a given $P \in \mathbb{P}$. As was noted in Scheer [2012], there are not many publications that are concerned with the ENFR. Some references related to the ENFR can be found in Finner and Roters [2001], [2002] and Gordon et al. [2007].

Definition 1.16 *(a) We say that an MTP with the stopping rule τ controls the ENFR at some function g fulfilling $g : \{1, ..., n-1\} \to [0, n]$, if ENFR($\tau$) $\leqslant g(n_1)$ for all $n_1 = 0, ..., n-1$.*

(b) An MTP τ controls the ENFR (linearly) at γ, if ENFR(τ) $\leqslant (n_1 + 1)\gamma$ for all $n_1 = 0, ..., n-1$.

Scheer [2012] proposed the aforementioned concept of ENFR-control. He investigated the ENFR for some FDR-controlling procedures and proved ENFR-control for some of them under independence of the underlying p-values.
In this thesis the control of the ENFR plays an important role. We will show for some class of SD procedures that the ENFR is controlled under a special dependence, namely martingale dependence. Furthermore, ENFR-control for these procedures will imply control of the FDR or control of error rate criteria closely related to the FDR.

1.4.5 False Discovery Exceedance

Many authors have noted that the concept of FDR-control works well if the FDP is concentrated around the FDR, for example, cf. Genovese and Wasserman [2004] and Roquain et al. [2011]. In this context a new criterion was suggested, namely the ***False Discovery Exceedance*** (γ-FDX). Thereby, the γ-FDX is defined as the probability that the FDP is greater than some pre-chosen level γ, that is,

$$\gamma\text{-FDX}(\tau) \equiv \gamma\text{-FDX}(\tau, P) = P\left(\frac{V(\tau)}{R(\tau)} > \gamma\right).$$

Definition 1.17
The γ-FDX is said to be controlled at level α by an MTP τ if γ-FDX(τ) $\leqslant \alpha$ for the underlying $P \in \mathbb{P}$.

In the aforementioned work of Lehmann and Romano [2005] some SD procedures controlling the γ-FDX were proposed. One of these SD procedures controls the γ-FDX under arbitrary dependence. One year later Romano and Shaikh [2006] proposed an SU procedure with γ-FDX-control under general dependence.

1.5 Relationships between various error rate

In this section we present some mostly known relationships between the aforementioned error measures. Some of them are obvious and have often been referred in publications such as the second inequality in (1.7), cf. Benjamini and Hochberg [1995] and p.6 in Dickhaus [2014]). Others are simple, but for the sake of clarity we provide proofs.

Lemma 1.18
Let $n_0, n \in \mathbb{N}$ be such that $n_1 = n - n_0 \geqslant 0$ and let τ be a stopping rule of some MTP. Then we have the following relations between related error rates:

$$\frac{FWER(\tau)}{n_1 + 1} \leqslant FDR(\tau) \leqslant FWER(\tau), \tag{1.7}$$

$$ENFR(\tau) = \sum_{k=1}^{n_0} k\text{-}FWER(\tau), \tag{1.8}$$

$$ENFR(\tau) \leqslant n_0 \cdot FWER(\tau), \tag{1.9}$$

$$FDR(\tau) = \int_0^1 P\left(\frac{V(\tau)}{R(\tau)} > \gamma\right) d\gamma = \int_0^1 \gamma\text{-}FDX(\tau)\ d\gamma. \tag{1.10}$$

Proof. The second inequality in (1.7) is trivial and the first one can be proved as follows:

$$\begin{aligned}
\mathrm{FDR}(\tau) &= \mathbb{E}\left[\frac{V(\tau)}{R(\tau)}\mathbb{I}(V(\tau) > 0)\right] \\
&\geqslant \mathbb{E}\left[\frac{V(\tau)}{V(\tau) + n_1}\mathbb{I}(V(\tau) > 0)\right] \\
&\geqslant \frac{1}{n_1 + 1}P(V(\tau) > 0) \\
&= \frac{\mathrm{FWER}(\tau)}{n_1 + 1}.
\end{aligned}$$

Equality (1.8) follows from the fact that $V(\tau)$ is a non-negative discretely distributed random variable. Therefore, the ENFR, that is, the mathematical expectation of $V(\tau)$, can be rewritten in the following way:

$$\begin{aligned}
\mathrm{ENFR}(\tau) &= \sum_{i=1}^{n_0} iP(V(\tau)=i) \\
&= \sum_{i=1}^{n_0}\sum_{k=1}^{i} P(V(\tau)=i) \\
&= \sum_{k=1}^{n_0}\sum_{i=k}^{n_0} P(V(\tau)=i) \\
&= \sum_{k=1}^{n_0} P(V(\tau)\geqslant k) \\
&= \sum_{k=1}^{n_0} \text{k-FWER}(\tau).
\end{aligned}$$

Inequality (1.9) follows immediately from equality (1.8), since obviously

$$P(V(\tau)\geqslant 1)\geqslant P(V(\tau)\geqslant k) \text{ for all } k\geqslant 1.$$

Inequality (1.10) follows directly from the definition of the mathematical expectation via integration by parts. It is an alternative representation of the mathematical expectation of the non-negative random variable $\mathrm{FDP}(\tau) = V(\tau)/R(\tau)\mathbb{I}(R(\tau)>0)$. Let $F_{\mathrm{FDP}(\tau)}$ be the distribution function of the FDP. Then according to Fubini's theorem we get:

$$\begin{aligned}
\mathrm{FDR}(\tau) &= \int_0^1 x dF_{\mathrm{FDP}(\tau)}(x) \\
&= \int_0^1\int_0^x d\gamma dF_{\mathrm{FDP}(\tau)}(x) \\
&= \int_0^1 \left(\int_0^1 \mathbb{I}_{(0,x]}(\gamma) dF_{\mathrm{FDP}(\tau)}(x)\right) d\gamma \\
&= \int_0^1 P(\mathrm{FDP}(\tau)>\gamma)d\gamma \\
&= \int_0^1 \gamma\text{-FDX}(\tau)\ d\gamma.
\end{aligned}$$

■

It should be noted that for all inequalities in Lemma 1.18 we do not require any assumptions about independence or any specific dependence structures between the underlying p-values. The relations remain true if we consider any dependence between p-values.

Chapter 2

Some concepts of dependence

Dependence structures play a crucial role in the FDR-control framework. Many existing multiple testing procedures use independent test statistics under the null and fail if the test statistics are dependent, for instance, the adaptive step-up procedure as proposed by Storey et al. [2002].

2.1 Basic independence assumptions

Let $n \in \mathbb{N}$ and $m \in \mathbb{N}$ with $m \leqslant n$.

Definition 2.1 (Basic independence assumptions (BIA))
We say that random variables $Y = (Y_1, ..., Y_m) : \Omega \to [0,1]^m$ and $X = (X_1, ..., X_{n-m}) : \Omega \to [0,1]^{n-m}$ fulfill the basic independence assumptions if Y and X are mutually independent and $X_1, ..., X_{n-m}$ are i.i.d. $U(0,1)$-distributed random variables (cf. Finner et al [2009], assumptions I1,I2 and D3).

Let $I = \{1, ..., n\}$, $I_1 = I \setminus I_0$ and $i \in I_0$ iff p_i corresponds to the true null hypothesis (cf. p.9).

Remark 2.2 (BIA)
We say that the p-values $p_1, ..., p_n$ fulfill BIA, if $\bar{U} = (p_i)_{i \in I_0}$ and $\bar{f} = (p_i)_{i \in I_1}$ are mutually independent and $p_i, i \in I_0$ are i.i.d. $U(0,1)$- distributed.

2.2 Positive association

Definition 2.3 (Positive association (PA))
Let $X = (X_1, ..., X_n), X_i : \Omega \to \mathbb{R}, i \in \{1, ..., n\}$. The random variables $X_1, ..., X_n$ are said to be positively associated if Cov$(\phi(X), \psi(X)) \geqslant 0$ holds for all component-wise non-decreasing functions $\phi : \mathbb{R}^n \to \mathbb{R}$ and $\psi : \mathbb{R}^n \to \mathbb{R}$, for which Cov$(\phi(X), \psi(X))$ exists.

This concept was introduced by Esary, Proschan and Walkup [1967], although the authors called it not positive association but association. We use the term positive association, like, for example, Jogdeo and Proschan [1981]. Esary et al. [1967] studied different properties of positively associated random variables and gave some equivalent criteria for positive association. We will use just one property of positively associated random variables in this dissertation.

Lemma 2.4 (Esary et al. (1967), p.1467)
Non-decreasing functions of positively associated random variables are positively associated.

A simple, but very useful example of positively associated random variables, are the order statistics $X_{1:n}, ..., X_{n:n}$ of independent random variables $X_1, ..., X_n$ (cf. Esary et al. [1967], pp.1473-1473).
The aforementioned assertion follows directly from Lemma 2.4 and can also be proved by Lemma 3.1 proposed in Hájek [1968] (cf. Hájek [1968], p.331).

2.3 Positive regression dependence on a subset and positive orthant dependence

The next type of dependence we will use is the positive regression dependence on a subset. In the FDR-framework it was firstly considered by Benjamini and Yekutieli [2001] and Sarkar [2002].

Definition 2.5 (Positive regression dependence on a subset (PRDS))
Random variables X_i, $i = 1, ..., n$, with values in $\mathbb{R}$ are said to be positively regression dependent on a subset $J \subset \{1, ..., n\}$ (or are PRDS), if $x \mapsto \mathbb{E}\left[\phi(X_1, ..., X_n) \mid X_i = x\right]$ is increasing (decreasing) in x for each $i \in J$ and any coordinate-wise increasing (decreasing)

integrable function $\phi : \mathbb{R}^n \to \mathbb{R}$.

The PRDS concept is related to the positive regression dependence concept (cf. Lehmann [1966] for the bivariate case, Sarkar T.K. [1969]) as will be defined below, with the difference that the PRDS assumption is required to be fulfilled only on a subset of random variables and the conditioning is on one variable and not on all variables which is the case for positive regression dependence.

Lemma 2.6
Assume that $X_1, ..., X_n$ *are PRDS on* $J \subset \{1, ..., n\}$ *in the sense of Definition 2.5. Let* $F_i : [a, \infty) \to [0, 1]$, $a \in \mathbb{R}$ *be a marginal continuous distribution function of* $X_i, i \in J$ *with* $F_i(a) = 0$. *Then* $x \mapsto \mathbb{E}[\phi(X_1, ..., X_n) \mid X_i \leqslant x]$ *is increasing for each* $i \in J \subset \{1, ..., n\}$ *and all coordinate-wise increasing function* $\phi : [a, \infty)^n \to \mathbb{R}$ *of* $(X_1, ..., X_n)$.

Proof. For $i \in J$ define

$$f_i(x) = \mathbb{E}\left[\phi(X_1, ..., X_n) \mid X_i = x\right],$$

$$g_i(u) = \mathbb{E}\left[\phi(X_1, ..., X_n) \mid X_i \leqslant u\right].$$

Hence, for all $u \in [a, \infty]$ with $F_i(u) > 0$ we get

$$g_i(u) = \frac{\int\limits_a^u f_i(x) dF_i(x)}{F_i(u)}.$$

Now, we obtain for $a \leqslant u_1 \leqslant u_2 < \infty$ and $i \in J$

$$g_i(u_2) - g_i(u_1) = \frac{F_i(u_1) \int\limits_a^{u_2} f_i(x) dF_i(x) - F_i(u_2) \int\limits_a^{u_1} f_i(x) dF_i(x)}{F_i(u_1) F_i(u_2)} \tag{2.1}$$

$$= \frac{(F_i(u_1) - F_i(u_2)) \int\limits_a^{u_1} f_i(x) dF_i(x)}{F_i(u_1) F_i(u_2)} + \frac{F_i(u_1) \int\limits_{u_1}^{u_2} f_i(x) dF_i(x)}{F_i(u_1) F_i(u_2)}. \tag{2.2}$$

Further, due to the mean value theorem for Riemann-Stieltjes integrales (cf. Jie Xiao [2008], p.60), there exist some values $a \leqslant \xi \leqslant u_1 \leqslant \zeta \leqslant u_2$ with

$$\xi F_i(u_1) = \int\limits_a^{u_1} f_i(x) dF_i(x), \quad \zeta(F_i(u_2) - F_i(u_1)) = \int\limits_{u_1}^{u_2} f_i(x) dF_i(x). \tag{2.3}$$

Thereby

$$\inf_{t\in(a,u_1)} f(t) \leqslant \xi \leqslant \sup_{t\in(a,u_1)} f(t) \leqslant \inf_{t\in(u_1,u_2)} f(t) \leqslant \zeta \leqslant \sup_{t\in(u_1,u_2)} f(t)$$

holds, because f is an increasing function by assumptions. Hence, continuing (2.1) we obtain by (2.3)

$$g_i(u_2) - g_i(u_1) = \frac{\xi F_i(u_1)(F_i(u_1) - F_i(u_2)) + \zeta F_i(u_1)(F_i(u_2) - F_i(u_1))}{F_i(u_1)F_i(u_2)} \tag{2.4}$$

$$= \frac{F_i(u_1)(F_i(u_2) - F_i(u_1))(\zeta - \xi)}{F_i(u_1)F_i(u_2)} \geqslant 0, \tag{2.5}$$

hence, the assertion follows. ■

Remark 2.7

The technique of the last proof is similar to the method of proof applied in Lehmann [1966]. The assertion of Lemma 2.6 can also be deduced by applying Wijsman's inequality (Wijsman [1985]), cf. Finner et al. [2009], p.9.

Remark 2.8 (PRDS)

We say that $\bar{U} = (p_i)_{i\in I_0}$ is PRDS if $u \mapsto \mathbb{E}\left[\phi(p_1, ..., p_n) \mid U_i = u\right]$ is increasing (decreasing) in u, $u \in (0,1)$ for each $i \in I_0$ and any coordinate-wise increasing (decreasing) integrable function $\phi : [0,1]^n \to \mathbb{R}$.

Lehmann [1966] introduced a bivariate positive regression concept. The next definition is a generalization of the concept of Lehmann for multivariate cases (cf. Sarkar [1969], Barlow and Proschan [1981]).

Definition 2.9

Random variables $X_1, ..., X_n$ are said to be positively regression dependent if $\mathbb{E}\left(\phi(X_1, ..., X_n) \mid (X_i = x_i)_{i\in J}\right)$ is non-decreasing in $(x_i)_{i\in J}$ for all coordinate-wise increasing integrable functions $\phi : [0,1]^n \to \mathbb{R}$, for all $J \subset I$.

A stricter requirement is the multivariate total positivity of order 2 (MTP_2) property (Karlin and Rinott [1980]).

According to Definition 2.5 the random variables $X_i, i \in \{1, ..., n\}$ are $[0,1]-$valued when we are concerned with PRDS dependence since we have p-values in mind. The subsequent definitions also work for real-valued random variables.

Definition 2.10 (MTP$_2$)
Real-valued random variables $X_1, ..., X_n$ are said to be MTP$_2$ if the random vector $\bar{X} = (X_1, ..., X_n)$ has a density (or probability function) $f : R^n \to [0, \infty)$ satisfying

$$f(x \wedge y) f(x \vee y) \geqslant f(x) f(y) \tag{2.6}$$

with $x \wedge y = \min(x, y)$ and $x \vee y = \max(x, y)$, where min *and* max *are taken coordinate-wise.*

The MTP$_2$ condition (2.6) implies the PRDS property and is easier to verify, that is why it is widely used (cf. Benjamini and Yekutieli [2001], p.1170).
The next dependence type is the oldest of the four dependence structures defined before and was proposed by Lehmann [1966] for the bivariate case. Many authors considered the generalization of this concept for multivariate distributions. For example, Ahmed et al. [1978] and Block and Ting [1981] studied this concept.

Definition 2.11 (Positive orthant dependence (POD))
We say that real-valued random variables $X_1, ..., X_n$ are (upper) positively orthant dependent (POD), if for any set of real values $\{a_1, ..., a_n\}$

$$P(X_1 > a_1, X_2 > a_2, ..., X_n > a_n) \geqslant \prod_{i=1}^{n} P(X_i > a_i).$$

Note that positive regression dependence implies positive association. PRDS does not imply positive association. A simple example confirming this statement is given in Benjamini and Yekutieli [2001] (p.1172). It is the multivariate normal distribution $N(\mu, \Sigma)$, $\Sigma = (\Sigma_{k,l})_{k,l \in I}$, which fulfills the PRDS property on the subset I_0, and on each other subset of I_0, if for all $i \in I_0$ and for all $j \neq i$ $\Sigma_{i,j} \geqslant 0$ is valid, however it is not positively associated if there some indices $k, l \in I$, for which $\Sigma_{k,l} < 0$ holds, exist, cf. Pitt [1982], Tong [1990], p.97, Theorem 5.1.1.

Now we will show that the positive regression dependence on a subset J implies the positive orthant dependence on the same subset J.

Lemma 2.12
If $[0, 1]$-valued random variables $X_1, ..., X_n$ are PRDS on $J \subset \{1, ..., n\}$, then $\{X_i\}_{i \in J}$ are positively orthant dependent.

Proof. W.l.o.g. let us assume that $J = \{1, ..., n\}$ holds. For any other subset J the proof works in the same way.

First, we show the inequality

$$\mathbb{E}\left[\prod_{i=2}^{n}\mathbb{I}(X_i > a_i) \mid X_1 > a_1\right] \geqslant \mathbb{E}\left[\prod_{i=2}^{n}\mathbb{I}(X_i > a_i) \mid X_1 \leqslant a_1\right]. \tag{2.7}$$

This inequality is intuitive and clear. Indeed, the PRDS property means that the knowledge of X_1 becoming greater increases the probability of $\phi(X_1, ..., X_n)$ being greater, if ϕ increases coordinate-wise.

Or more precisely: let F denote the marginal distribution function of X_1, we define $f(u) = \mathbb{E}\left[\prod\limits_{i=2}^{n}\mathbb{I}(X_i > a_i)|X_1 = u\right]$, then we have

$$\mathbb{E}\left[\prod_{i=2}^{n}\mathbb{I}(X_i > a_i)\mathbb{I}(X_1 > a_1)\right] = \int_{a_1}^{1} f(u)dF(u)$$

and

$$\mathbb{E}\left[\prod_{i=2}^{n}\mathbb{I}(X_i > a_i)\mathbb{I}(X_1 \leqslant a_1)\right] = \int_{0}^{a_1} f(u)dF(u).$$

Then (2.7) is equivalent to

$$\frac{\int_{a_1}^{1} f(u)dF(u)}{1 - F(a_1)} \geqslant \frac{\int_{0}^{a_1} f(u)dF(u)}{F(a_1)}. \tag{2.8}$$

From the mean value theorem for Riemann-Stieltjes integrales (cf. Jie Xiao [2008] p.60) we can deduce that there exist some values ξ_1 and ξ_2 with

$$\begin{aligned}
\xi_1 &= \frac{\int_{a_1}^{1} f(u)dF(u)}{1 - F(a_1)}, \qquad \inf_{t\in(a_1,1)} f(t) \leqslant \xi_1 \leqslant \sup_{t\in(a_1,1)} f(t)\\
\xi_2 &= \frac{\int_{0}^{a_1} f(u)dF(u)}{F(a_1)}, \qquad \inf_{t\in(0,a_1)} f(t) \leqslant \xi_2 \leqslant \sup_{t\in(0,a_1)} f(t).
\end{aligned} \tag{2.9}$$

Since f is an increasing function of u, (2.9) yields $\xi_1 \geqslant \xi_2$, hence (2.8).

Further we get:

$$P(X_1 > a_1, X_2 > a_2, ..., X_n > a_n) = \mathbb{E}\left[\prod_{i=1}^{n} \mathbb{I}(X_i > a_i)\right] \tag{2.10}$$

$$= P(X_1 > a_1)\mathbb{E}\left[\prod_{i=2}^{n} \mathbb{I}(X_i > a_i) \mid X_1 > a_1\right] \geqslant P(X_1 > a_1)\mathbb{E}\left[\prod_{i=2}^{n} \mathbb{I}(X_i > a_i)\right] \tag{2.11}$$

$$= P(X_1 > a_1)P(\bigcap_{i=2}^{n}(X_i > a_i)) \geqslant ... \geqslant \prod_{i=1}^{n} P(X_i > a_i). \tag{2.12}$$

The inequalities in (2.11)-(2.12) hold due to the PRDS-assumption since according to the law of total probability we have

$$\mathbb{E}\left[\prod_{i=2}^{n} \mathbb{I}(X_i > a_i)\right] = \mathbb{E}\left[\prod_{i=2}^{n} \mathbb{I}(X_i > a_i) \mid X_1 > a_1\right] - P(X_1 \leqslant a_1)\underbrace{\left(\mathbb{E}\left[\prod_{i=2}^{n} \mathbb{I}(X_i > a_i) \mid X_1 > a_1\right] - \mathbb{E}\left[\prod_{i=2}^{n} \mathbb{I}(X_i > a_i) \mid X_1 \leqslant a_1\right]\right)}_{\geqslant 0 \text{ due to (2.7)}}. \tag{2.13}$$

■

Definition 2.13

We call the set $A \subset \Omega$ increasing on the subset J for $[0,1]$-valued random variables $X_1, ..., X_n$ if the indicator $X_i \mapsto \mathbb{I}_A \equiv \mathbb{I}_A(X_1, ..., X_n)$ is increasing whenever $i \in J$.

Lemma 2.14

If $X_1, ..., X_n$ are PRDS on $J \subset \{1, ..., n\}$, then the following inequalities hold for any increasing on J set A (cf. Definition 2.13), for all $i \in J$ and for all $u \in [0,1]$

(a)

$$\mathbb{E}\left[\mathbb{I}_A \mid X_i > u\right] \geqslant \mathbb{E}\left[\mathbb{I}_A \mid X_i \leqslant u\right], \tag{2.14}$$

(b)

$$\mathbb{E}\left[\mathbb{I}_A \mid X_i > u\right] \geqslant \mathbb{E}\left[\mathbb{I}_A\right], \tag{2.15}$$

(c)

$$\mathbb{E}\left[\mathbb{I}_A \mid X_i \leqslant u\right] \leqslant \mathbb{E}\left[\mathbb{I}_A\right]. \tag{2.16}$$

Proof. The proof of part (a) can be carried out similarly to the proof of inequality (2.12) in the proof of Lemma 2.12.
Part (b) follows directly from (a) since obviously

$$\begin{aligned}\mathbb{E}\left[\mathbb{I}_A\right] &= \mathbb{E}\left[\mathbb{I}_A \mid X_i > u\right] P(X_i > u) + \mathbb{E}\left[\mathbb{I}_A \mid X_i \leqslant u\right] P(X_i \leqslant u) \\ &= \mathbb{E}\left[\mathbb{I}_A \mid X_i > u\right] - P(X_i \leqslant u) \underbrace{\left(\mathbb{E}\left[\mathbb{I}_A \mid X_i > u\right] - \mathbb{E}\left[\mathbb{I}_A \mid X_i \leqslant u\right]\right)}_{\geqslant 0} \\ &\leqslant \mathbb{E}\left[\mathbb{I}_A \mid X_i > u\right]\end{aligned}$$

holds for all $i \in J$.
Part (c) can be proved similar to part (b). We obtain for all $i \in J$

$$\begin{aligned}\mathbb{E}\left[\mathbb{I}_A\right] &= \mathbb{E}\left[\mathbb{I}_A \mid X_i > u\right] P(X_i > u) + \mathbb{E}\left[\mathbb{I}_A \mid X_i \leqslant u\right] P(X_i \leqslant u) \\ &= \mathbb{E}\left[\mathbb{I}_A \mid X_i \leqslant u\right] + P(X_i > u) \underbrace{\left(\mathbb{E}\left[\mathbb{I}_A \mid X_i > u\right] - \mathbb{E}\left[\mathbb{I}_A \mid X_i \leqslant u\right]\right)}_{\geqslant 0},\end{aligned}$$

which completes the proof. ■

Corollary 2.15
Let the following assumptions be fulfilled

- $X_1, ..., X_n$ *are PRDS on* $J \subset \{1, ..., n\}$, *with* $|J| = m$,
- $\mathbb{I}_A \equiv \mathbb{I}_A(X_1, ..., X_n)$ *increases, if* X_i *increases for all* $i \in J$,
- $\phi : [0, 1]^m \to [0, \infty)$ *is coordinate-wise increasing function.*

Then we have for all $i \in J$ *and all* $u \in [0, 1]$*:*

(a)

$$Cov(\mathbb{I}_A, \mathbb{I}(X_i > u)) \geqslant 0, \tag{2.17}$$

(b)

$$Cov(\phi, \mathbb{I}(X_i \geqslant u)) \geqslant 0, \tag{2.18}$$

(c)

$$Cov(\phi, \mathbb{I}(X_i < u)) \leqslant 0. \tag{2.19}$$

Proof. We have to show that

$$P(A \cap \{X_i > u\}) \geqslant P(A)P(\{X_i > u\}) \text{ holds.}$$

The assertion (a) follows directly from Lemma 2.14 (b) since

$$P(A \cap \{X_i > u\}) = P(A|\{X_i > u\})P(X_i > u) \text{ is valid for all } i \in J.$$

(b) For all $i \in J$ and all $u \in [0,1]$ we get by Lemma 2.12 (since PRDS on J implies POD on the same subset)

$$\mathbb{E}\left[\phi \mathbb{I}(X_i \geqslant u)\right] = \int_0^\infty P(\phi \mathbb{I}(X_i \geqslant u) > x)dx$$

$$\geqslant \int_0^\infty P(\phi > x)P(\{X_i \geqslant u\})dx = \mathbb{E}\left[\phi\right] P(X_i \geqslant u).$$

(c) Follows directly from (b). Indeed we have similarly to (b)

$$\mathbb{E}\left[\phi \mathbb{I}(X_i < u)\right] = \mathbb{E}\left[\phi\right] - E\left[\phi \mathbb{I}_{\{X_i \geqslant u\}}\right]$$

$$\leqslant \mathbb{E}\left[\phi\right] - \mathbb{E}\left[\phi\right] P(X_i \geqslant u) = \mathbb{E}\left[\phi\right] P(X_i < u),$$

which completes the proof.

■

2.4 Martingale Dependence

Related to the $[0,1)-$valued random variables $X_1, ..., X_n$ let us define the filtration

$$\mathcal{F}_t = \sigma(\mathbb{I}_{[0,s]}(X_i), 0 \leqslant s \leqslant t, i \in \{1, ..., n\}), \; 0 \leqslant t < 1, \tag{2.20}$$

which contains all information about each $X_i, i = 1, ..., n$, up to the time $t \in [0,1)$. Now we want to introduce a new concept of dependence in multiple testing, the martingale dependence.

Definition 2.16

We say that random variables $X_1, ..., X_n$ are $\mathcal{F}_t$(super-)martingale dependent on $J \subset \{1, ..., n\}$ (or belong to the class M_J on some subset $J \subset \{1, ..., n\}$) ((s)MD), if the stochastic process $M(t) = M_J(t) = \sum_{i \in J} \left(\frac{\mathbb{I}(X_i \leqslant t) - t}{1-t}\right)$, $0 \leqslant t < 1$ is a $\mathcal{F}_t$ (super-)martingale.

(cf. Chapter 6 for examples and more details.)

Remark 2.17
In this dissertation we consider the vector of p-values $p = (p_1, ..., p_n)$, thereby the vector of the p-values which corresponds to true null hypothesis is denoted by $\bar{U} = (U_1, ..., U_{n_0}) = (p_i)_{i \in I_0}$ (cf. p.9). We say that $U_1, ..., U_{n_0}$ are MD, if Definition 2.16 is fulfilled with $(X_1, ..., X_n) = (p_1, ..., p_n)$ and $J = I_0$.

Remark 2.18 (General remark to MD)
If we say that the random variables $X_1, ..., X_n$ are MD w.r.t. the filtration $\mathcal{F}_t$ we always assume that the filtration is complete. If this is not the case then we define $\mathcal{F}_t := \mathcal{F}_t \cup N^{\mathbb{P}}$ where $N^{\mathbb{P}}$ is the set of all $\mathbb{P}$-null-sets in $\mathcal{F}_t$.

2.5 Summary

In this chapter we introduced some well known kinds of dependence, namely PA, POD and PRDS. We showed that PRDS implies POD on the same subset and proved some inequalities for PRDS random variables. Unfortunately, positive association as well as positive regression dependency on a subset are not an easy to verify. Nevertheless, both are widely used in multiple testing theory, especially in the FDR-framework.

We introduced the concept of martingale dependent random variables. The idea of this concept does not appear to be intuitively meaningful at first sight. On the other hand, since the empirical process $(\hat{F}_n(t))_{t \in (0,1)}$ goes well with the theory of semi-martingales, the requirement of super-martingale dependence seems to be natural. The martingale dependence will be very useful for our later considerations in Chapter 5.

In general, the literature about dependency structures is huge. There is much cross over among the many authors who have studied different properties of the dependence structures mentioned above. It is therefore complex to verify the originator of the property.

Chapter 3

Asymptotically optimal rejection curve (AORC)

3.1 Motivation

Let us consider an asymptotic Dirac-uniform model, i.e., the p-values $U_i, i = 1, ..., n_0$, which correspond to true null hypotheses, are i.i.d. U$(0,1)$- distributed random variables and the false ones follow the Dirac distribution with point mass 1 at 0, or which is the same, that $f_1 = ... = f_{n_1} = 0$ $P_{\mathrm{DU}(n_0,n)}$ a.s. Finner et al. [2009] proposed a new rejection curve, called asymptotically optimal rejection curve (AORC). The AORC is defined by

$$f_\alpha(t) = \frac{t}{t(1-\alpha)+\alpha}.$$

This approach was motivated as follows (cf. Finner et al. [2009]).
Let $\zeta_n = \frac{n_0(n)}{n}$ denote the portion of the true null hypotheses with $\lim\limits_{n\to\infty} \zeta_n = \zeta \in [0,1]$. Then by the Glivenko-Cantelli Theorem we get:

$$\hat{F}_n(t) \to F_\infty(t|\zeta) = (1-\zeta) + \zeta t \quad \text{for all} \quad t \in [0,1] \ P_{DU(n_0,n)}\text{a.s.}$$

If for some fixed $t \in (0,1]$ a test procedure rejects all null hypotheses $H_i, i = 1, ..., n$ with $p_i \leqslant t$, then the asymptotic FDR of this procedure is given by

$$FDR(t) = \frac{t\zeta}{(1-\zeta)+t\zeta}. \tag{3.1}$$

If we determine the point t_ζ, for which $FDR(t_\zeta) = \alpha,\ \zeta \in [\alpha, 1)$ holds, we get

$$t_\zeta = \frac{\alpha(1-\zeta)}{\zeta(1-\alpha)}. \tag{3.2}$$

The idea of Finner et al. [2009] was to find a strictly increasing rejection curve $f_\alpha : [0,1] \to [0,1]$, which fulfills the property $f_\alpha(t_\zeta) = F_\infty(t_\zeta)$ for all $\zeta \in [\alpha, 1)$. And $f_\alpha(t) = \frac{t}{t(1-\alpha)+\alpha}$, $t \in [0,1]$ is a such curve.

Now we present a further motivation which is based on a model of independent p-values.
Let $p_1, ..., p_n$ be i.i.d $Q-$distributed random variables with distribution function $F(t) \equiv F_Q(t) \geqslant t,\ t \in [0,1]$. For some fixed $t \in (0,1)$ we may decompose the distribution function F by

$$F(t) = \kappa_t(Q) + t(1 - \kappa_t(Q)), \tag{3.3}$$

where the statistical functional

$$Q \mapsto \kappa_t(Q) = \frac{F(t) - t}{1 - t} \tag{3.4}$$

represents the nonuniform part of Q on the interval $[0,t]$. The maximal relative portion of the uniformity of Q on $[0,t]$ is given by

$$\alpha(t) = 1 - \frac{\kappa_t(Q)}{F(t)} \tag{3.5}$$

when $F(t) > 0$ holds. Then $\alpha(t)$ is the solution of

$$(1 - \alpha(t))F(t) = \frac{F(t) - t}{1 - t} \tag{3.6}$$

and $F(t) = \frac{t}{t(1-\alpha(t))+\alpha(t)}$ holds. In this particular model the control of $\alpha(t)$ yields the asymptotic optimal rejection curve f_α, i.e.,

$$\alpha(t) \leqslant \alpha \ \text{ iff } \ F(t) \geqslant f_\alpha(t) = \frac{t}{t(1-\alpha)+\alpha}. \tag{3.7}$$

Figure 3.1 gives a geometrical interpretation of equality (3.6). The green straight line $g \equiv g(s)$ which connects a fixed point $(t, F(t))$ with the point $(1,1)$, is described by the equality

$$g(s) = \frac{1 - F(t)}{1 - t}s + \frac{F(t) - t}{1 - t}. \tag{3.8}$$

Due to the intercept theorem from elementary geometry, we have the slope of the green line $\frac{x}{t} = \frac{1-F(t)}{1-t}$, where x is a length of the "blue" interval on Figure 3.1, which implies

$x = \frac{t}{1-t}(1 - F(t))$. Consequently, the relative part of x on $[0, t]$ is $\frac{x}{F(t)}$ which corresponds to $\alpha(t)$ in our consideration.

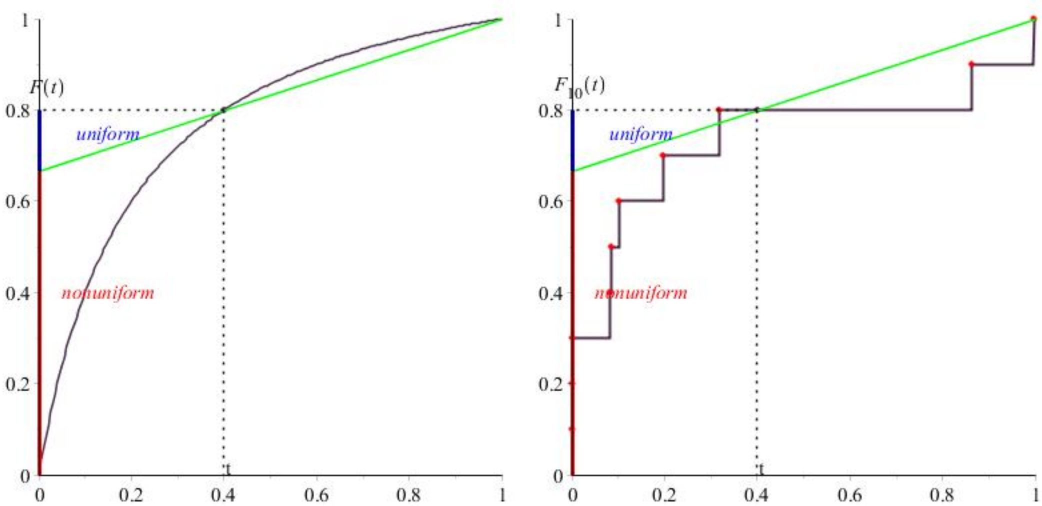

Figure 3.1: Uniform and nonuniform part of Q on $[0, 0.4]$. The black curve on the left graphic is the distribution function of the p-values, the black curve on the right graphic is the empirical distribution function $F_{10}(t)$ of ten realizations $p_1, ..., p_{10}$. The green straight line is described by (3.8). The red line corresponds to the nonuniform part of Q on $[0, 0.4]$, the blue one corresponds to the uniform part of Q on $[0, 0.4]$.

.

3.2 Some FDR-Estimators

Consider the problem of simultaneously testing n null hypotheses $H_1, ..., H_n$. The corresponding p-values are denoted by $p_1, ..., p_n$. The vector of p-values corresponding to true null hypotheses is denoted by $U = (U_1, ..., U_{n_0})$, thereby we have $U = (U_i)_{i \in \{1,...,n_0\}} = (p_j)_{j \in I_0}$. The vector of the p-values which correspond to alternatives is denoted by $f = (f_1, ..., f_{n_1})$. Based on the motivation in Section 3.1, we consider the following empirical version of $\alpha(t)$, which may serve as an estimator of FDR(t) (the FDR of the MTP with $\tau = t$), that is,

$$\hat{\alpha}_n(t) = \frac{t}{1-t} \frac{1 - \hat{F}_n(t)}{\hat{F}_n(t)} \mathbb{I}(\hat{F}_n(t) > 0). \tag{3.9}$$

Subsequently, we will first consider only models, in which $\hat{F}_n(t) \geqslant \frac{1}{n}$ holds for all $t \in [0, 1)$, since by Lemma 4.3 and Lemma 5.3 (in the next Chapters) it will be possible to replace f_1 by 0 for SD-procedures. Consequently, we may drop the indicator $\mathbb{I}(\hat{F}_n(t) > 0)$ in (3.9). The case $n_0 = n$ will be considered separately.

Lemma 3.1
Let $p_1, ..., p_n$ be realizations of i.i.d. random variables, $\hat{F}_n$ be the empirical distribution function of $p_1, ..., p_n$. If for some fixed t with $0 \leqslant t < 1$ we have $\hat{F}_n(t) > 0$, then

$$\mathbb{E}\left[\hat{\alpha}_n(t)\right] \geqslant \alpha(t). \tag{3.10}$$

Proof. Since the empirical distribution function $\hat{F}_n$ is an unbiased estimator of the distribution function F and the function $x \mapsto \frac{1-x}{x}$ is convex, we have by Jensen's inequality

$$\begin{aligned}\mathbb{E}\left[\frac{t}{1-t}\frac{1-\hat{F}_n(t)}{\hat{F}_n(t)}\right] &\geqslant \frac{t}{1-t}\frac{1-\mathbb{E}[\hat{F}_n(t)]}{\mathbb{E}[\hat{F}_n(t)]}\\ &= \frac{t}{1-t}\frac{1-F(t)}{F(t)} = \alpha(t).\end{aligned}$$

■

Remark 3.2
Storey et al.[2004] proposed for some fixed $\lambda \in [0, 1]$ the following estimator for the FDR, that is,

$$F\hat{D}R_\lambda(t) = \frac{t}{\hat{F}_n(t)}\frac{1-\hat{F}_n(\lambda)}{1-\lambda}, \tag{3.11}$$

which is sometimes more convenient for the asymptotical consideration. We can see that for $\lambda = t$

$$F\hat{D}R_t(t)\mathbb{I}(\hat{F}_n(t) > 0) = \hat{\alpha}_n(t),$$

where $\hat{\alpha}_n$ is given in (3.9). In this context the estimator $\hat{\alpha}_n$ can be considered as a modified dynamic version of the estimator (3.11) (cf. Dickhaus [2008], p.81).

Remark 3.3 *1. The process $t \mapsto \hat{\alpha}_n(t)$ is right continuous, adapted to the filtration $\mathcal{F}_t = \sigma(\mathbb{I}_{(0,s]}(p_i),\ 0 \leqslant s \leqslant t,\ i = 1, ..., n), t \in (p_{1:n}, 1),\ n \in \mathbb{N}$.*

2. It is continuous and strictly increasing for $t \in [p_{i:n}, p_{i+1:n})$ with $i < n,\ n \in \mathbb{N}$.

3. *It is lower semi-continuous at each data point* $p_{i:n}$, $i \geqslant 2$, *with strict inequality*

$$\lim_{t \uparrow p_i} \hat{\alpha}_n(t) > \hat{\alpha}_n(p_i) \quad a.s., \; n \in \mathbb{N}.$$

4. *Whenever* p_i *increases, then* $\hat{\alpha}_n(t)$ *increases in that argument for fixed* t, $n \in \mathbb{N}$.

5. $\lim_{n \to \infty} \hat{\alpha}_n(t) = \alpha(t)$ $P-a.s.$

3.3 Asymptotic normality of the estimator $\hat{\alpha}_n(t)$

Theorem 3.4 (Asymptotic Normality of $\alpha_n(t)$)
Let $p_1, ..., p_n : \Omega \to [0,1]$ *be i.i.d. random variables with distribution function* F. *Fix* $t \in [0,1]$ *with* $1 > F(t) > 0$. *Then*

$$\sqrt{n}(\hat{\alpha}_n(t) - \alpha(t)) \xrightarrow{\mathcal{D}} \mathcal{N}\left(0, \frac{\alpha(t)^2}{F(t)(1-F(t))}\right). \tag{3.12}$$

Proof. The empirical distribution function $\hat{F}_n$ of i.i.d random variables is an unbiased estimator for the distribution function F. Its variance is given by

$$\operatorname{Var}(\hat{F}_n(t)) = \frac{1}{n^2} \sum_{i=1}^{n} \left(\mathbb{E}\left[\mathbb{I}^2(p_i \leqslant t)\right] - (\mathbb{E}\left[\mathbb{I}(p_i \leqslant t)\right])^2\right) \tag{3.13}$$

$$= \frac{1}{n}\left(\mathbb{E}\left[\mathbb{I}(p_1 \leqslant t)\right] - (\mathbb{E}\left[\mathbb{I}(p_1 \leqslant t)\right]])^2\right) \tag{3.14}$$

$$= \frac{F(t)(1-F(t))}{n}. \tag{3.15}$$

By the Central Limit Theorem we have

$$\frac{\hat{F}_n(t) - \mathbb{E}\left[\hat{F}_n(t)\right]}{\sqrt{\operatorname{Var}(\hat{F}_n(t))}} = \frac{\sqrt{n}(\hat{F}_n(t) - F(t))}{\sqrt{F(t)(1-F(t))}} \xrightarrow{\mathcal{D}} \mathcal{N}(0,1). \tag{3.16}$$

Consequently by application of the delta method (cf. A.W. van der Vaart pp.25-35) we get

$$\sqrt{n}\left(g(\hat{F}_n(t)) - g(F(t))\right) \xrightarrow{\mathcal{D}} \mathcal{N}\left(0, F(t)(1-F(t))(g'(F(t)))^2\right) \tag{3.17}$$

holds with $g(x) = \frac{t}{1-t}\frac{1-x}{x}\mathbb{I}(x > 0)$. Finally, since $g(\hat{F}_n(t)) = \hat{\alpha}(t)$, $g(F(t)) = \alpha(t)$ and

$$\frac{\partial}{\partial x} g(x)|_{x=F(t)} = -\frac{t}{1-t}\frac{1}{(F(t))^2},$$

we get

$$\begin{aligned}\sqrt{n}(\hat{\alpha}_n(t) - \alpha(t)) &\overset{\mathcal{D}}{\to} \mathcal{N}\left(0, \frac{t^2}{(1-t)^2}\frac{(1-F(t))}{(F(t))^3}\right) \\ &= \mathcal{N}\left(0, \frac{\alpha(t)^2}{F(t)(1-F(t))}\right).\end{aligned}$$

■

3.4 Conclusions

In this chapter we recalled the asymptotically optimal rejection curve (AORC) which was introduced by Finner et al. [2009]. We proposed another motivation for the AORC and provided an estimate for the FDR(t) related to the AORC. Finally, we proved an asymptotic normality of this estimate.

Chapter 4

Improvement of the first critical values under BIA's, PRDS or POD

The literature highlights some methods to characterize the power of a multiple testing procedure - for instance it can be characterized by the expected number of true rejections ENTR= $\mathbb{E}\left[S(\tau)\right]$. In this chapter we concentrate on the power of FDR-controlling procedures. Let us assume that an SD procedure controls the FDR. Is there any possibility to improve the power of this procedure without loss of the FDR-control? This question is the subject of this chapter.

Firstly, let us assume that the p-values are given by

$$p_i = \varepsilon_i V_i + (1-\varepsilon_i) g_i,\ i=1,...,n, \tag{4.1}$$

where $\varepsilon_i \sim B(1,q), i=1,...,n$ are i.i.d. Bernoulli random variable with unknown success parameter q. Suppose that $\bar{\varepsilon}=(\varepsilon_1,...,\varepsilon_n)$, $\bar{V}=(V_1,...,V_n)$ and $\bar{g}=(g_1,...,g_n)$ are mutually independent random vectors. Here $\varepsilon_i=1$ corresponds to the true null H_i with p-value V_i whereas $\varepsilon_i=0$ codes p-value g_i under alternative. Then the p-values which belong to true null hypotheses are defined as

$$U=(U_i)_{i\in\{1,...,\sum\limits_{l=1}^{n}\varepsilon_l\}} = (p_i)_{(i\in\{1,...,n\}:\varepsilon_i=1)} = (V_i)_{(i\in\{1,...,n\}:\varepsilon_i=1)},$$

and the false ones as

$$f = (f_j)_{j\in\{1,\dots,(n-\sum_{l=1}^{n}\varepsilon_l)\}} = (p_j)_{(j\in\{1,\dots,n\}:\varepsilon_j=0)} = (g_j)_{(j\in\{1,\dots,n\}:\varepsilon_j=0)}.$$

The number of true nulls is thereby $n_0 = \sum_{i=1}^{n} \varepsilon_i$ and $n_1 = n - n_0$ is the number of the false ones.

Remark 4.1
In the following chapters we always consider the conditional expectation

$$\mathbb{E}\left[\cdot|\bar{g},\varepsilon\right] = \mathbb{E}_f\left[\cdot\right]$$

conditioned under the false p-values and the status ε. Note that the information about $n_0 = n - n_1$ is also present. Observe that the conditional model given $\bar{g}$ and ε is a part of (4.1) where $\bar{g}$ and ε are deterministic.

Benjamini and Liu [1999] proposed a SD procedure with critical values

$$b_i = 1 - (1 - \alpha\frac{n}{n-i+1} \wedge 1)^{\frac{1}{n-i+1}},\ i = 1, ..., n. \tag{4.2}$$

They proved that this procedure controls the FDR if the underlying test statistics are mutually independent. Sarkar [2002] showed that the FDR of the SD procedure with critical values (4.2) is still controlled at level α if the p-values are PRDS.

Remark 4.2
The critical values (4.2) have the following asymptotic representation.

$$b_i = -\frac{\ln(1-\alpha)}{n} + O(1/n^2), \quad \textit{for all fixed} \ \ i = 1, ..., n \ \ \textit{with} \ \ i \leqslant n(1-\alpha)+1$$

$$b_i = 1, \quad \textit{if} \ \ i > n(1-\alpha)+1.$$

4.1 Increase of the first critical values without loss of the FDR-control

In this chapter we consider a SD-procedure with ordered critical values $(c_1, ..., c_n)$ which controls the FDR at level α, if the true p-values are PRDS or POD. We may do our

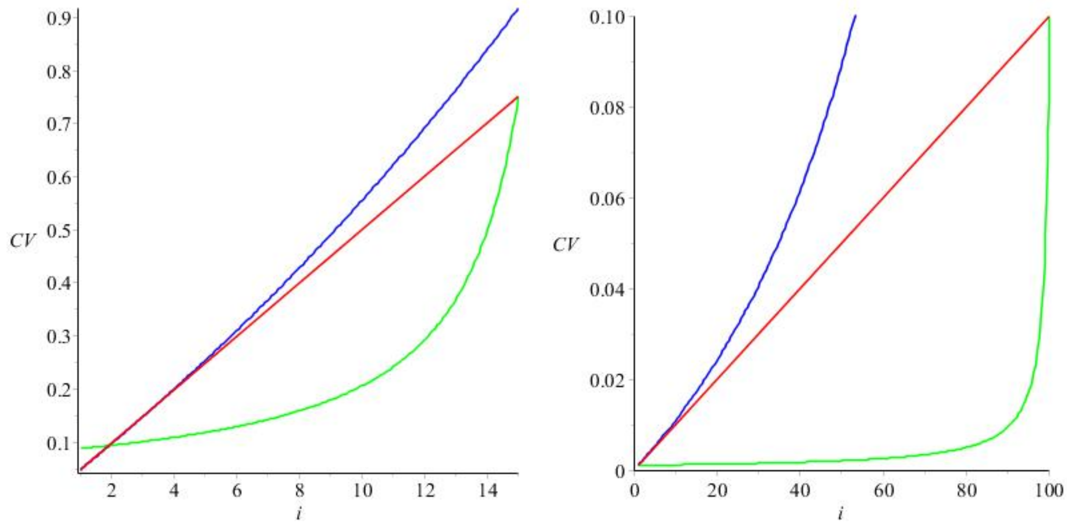

Figure 4.1: Comparison of the critical values $b_i, i = 1, ..., n$ from (4.2) (green line) with critical values of the BH procedure $a_i^{BH} = \frac{i\alpha}{n}$ (red line) and critical values of the SD procedure with $a_i = \frac{i\alpha}{n+1-i(1-\alpha)}$ (blue line). In the left graph the curves correspond to $\alpha = 0.75, n = 15$ in the right one to $\alpha = 0.1, n = 100$.

analysis for the conditional model of Remark 4.1 given $\bar{g}$ and ε.
We define

$$k = \max(i = 1, ..., n : c_j \leqslant b_j \ \text{ for all } \ j \leqslant i), \tag{4.3}$$

where $b_i, i = 1, ..., n$ are given in (4.2).

Lemma 4.3
Let $n_0 \in \mathbb{N}$ be some fixed number. If the following assumptions are fulfilled:

1. *$U_1, ..., U_{n_0} \succcurlyeq_{st} U(0, 1)$ and $(U_1, ..., U_{n_0})$ are PRDS or POD,*
2. *$(U_1, ..., U_{n_0})$ are independent of the false ones,*
3. *the SD-procedure which uses the critical values $c_1 \leqslant ... \leqslant c_n$, controls the FDR under (1-2) for all $f_1, ..., f_{n_1}$ and arbitrary fixed n_0,*

then the SD procedure with critical values

$$d_i = \mathbb{I}_{(i \leqslant k)} b_i + \mathbb{I}_{(i > k)} c_i, \ i = 1, ...n,$$

with k defined in (4.3), controls the FDR under (1-2) at desired level α, too.

Proof. Firstly, we briefly discuss the case $n = n_0$. For this case we have for the SD procedure with critical values $(b_1, ..., b_n)$

$$\text{FDR} = \text{FWER} = P(U_{1:n} \leqslant b_1).$$

We know by Lemma 2.12 that PRDS implies POD consequently we get immediately $P(U_{1:n} \leqslant b_1) \leqslant 1 - (1 - b_1)^n = \alpha$.
Let f be some fixed vector and $f_1 \leqslant ... \leqslant f_{n_1}$ be the ordered coordinates of the vector f. We define $j^* = \max(i = 1, ..., n_1 : f_j \leqslant d_j, \text{ for all } j \leqslant i)$. Since j^* false p-values will be rejected anyway, the FDP for the vector $\bar{p} = (f_1, ..., f_{j^*}, f_{j^*+1}, ..., f_{n_1}, U_1, ..., U_{n_0})$ will be the same as the FDP for the vector $\bar{p}_{j^*} = (0, ...0, f_{j^*+1}, ..., f_{n_1}, U_1, ..., U_{n_0})$.
We still use the convention $\frac{0}{0} \equiv 0$.
We define the FDP of the SD procedure with critical values $d_1, ..., d_n$ by $\frac{V(d)}{R(d)}$ and the FDP of the SD procedure with critical values $c_1, ..., c_n$ by $\frac{V(c)}{R(c)}$, respectively. Let $f^* = (0, ..., 0, f_{j^*+1}, ..., f_{n_1})$.
First, note that

$$\begin{aligned}
&\mathbb{E}_f\left[\frac{V(d)}{R(d)}\mathbb{I}(V(d) > 0)\right] = \\
&= \mathbb{E}_f\left[\mathbb{I}(\{j^* \geqslant k\})\frac{V(d)}{R(d)}\mathbb{I}(V(d) > 0)\right] + \mathbb{E}_f\left[\mathbb{I}(\{j^* < k\})\frac{V(d)}{R(d)}\mathbb{I}(V(d) > 0)\right] \\
&= \mathbb{I}(\{j^* \geqslant k\})\mathbb{E}_f\left[\frac{V(d)}{R(d)}\mathbb{I}(V(d) > 0)\right] + \mathbb{I}(\{j^* < k\})\mathbb{E}_f\left[\frac{V(d)}{R(d)}\mathbb{I}(V(d) > 0)\right],
\end{aligned}$$

with k defined in (4.3).
Further we consider two cases.

1. $j^* \geqslant k$. For this case we have
$$\mathbb{E}_f\left[\frac{V(d)}{R(d)}\mathbb{I}(V(d) > 0)\right] = \mathbb{E}_{f^*}\left[\frac{V(d)}{R(d)}\mathbb{I}(V(d) > 0)\right] = \mathbb{E}_{f^*}\left[\frac{V(c)}{R(c)}\mathbb{I}(V(c) > 0)\right] \leqslant \alpha.$$

2. $j^* < k$. In this case we have
$$\mathbb{E}\left[\frac{V(d)}{R(d)}\mathbb{I}(V(d) > 0)\right] \leqslant \mathbb{E}\left[\frac{V(d)}{j^* + V(d)}\mathbb{I}(V(d) > 0)\right] \tag{4.4}$$
$$\leqslant \frac{n - j^*}{n}\mathbb{E}\left[\mathbb{I}(V(d) > 0)\right] \leqslant \frac{n - j^*}{n}P(U_{1:n_0} \leqslant d_{j^*+1}) \tag{4.5}$$
$$= \frac{n - j^*}{n}P(U_{1:n_0} \leqslant b_{j^*+1}) \leqslant \frac{n - j^*}{n}(1 - (1 - b_{j^*+1})^{n_0}) \tag{4.6}$$
$$\leqslant \frac{n - j^*}{n}(1 - (1 - b_{j^*+1})^{n-j^*}) = \frac{n - j^*}{n}(\alpha\frac{n}{n - j^*} \wedge 1) \leqslant \alpha. \tag{4.7}$$

The inequality in (4.6) holds by POD (since by Lemma 2.12, PRDS implies POD on the same subset) and the equality in (4.7) holds by the choice of b_i's. ■

Remark 4.4

The proof of Lemma 4.3 is similar to the proof of the Theorem in Benjamini and Liu (1999) (cf. Benjamini,Liu [1999], pp.165-166).

Remark 4.5

Let $U_1, ..., U_{n_0}$ be PRDS or POD, then the FDR of the SD-procedure with critical values $c_1 = 1 - \sqrt[n]{1-\alpha} \leqslant c_2 \leqslant ... \leqslant c_n \leqslant 1$ is less or equal to α (equal to α if $U_1, ..., U_{n_0}$ are i.i.d $U(0,1)$-distributed) if $n_0 = n$ holds, i.e., when all hypotheses are true.

Example 4.6

Let us assume that the p-values which correspond to true null hypotheses are i.i.d. $U(0,1)$-distributed and the false ones are i.i.d. $U(\frac{b_1}{2}, \alpha)-$distributed random variables. For $\alpha = 0.1$ we computed the FDR for 5 different procedures:

- the adaptive SD procedure which was proposed by Gavrilov et al [2010] and uses critical values $a_i = \frac{i\alpha}{n+1-i(1-\alpha)}, i = 1, ..., n,$ **(ASDP)**,
- the adaptive SD procedure of Gavrilov with increased first critical values by Theorem 4.3 (**ASDP+**),
- the linear SD which uses the critical values $a_i = \frac{i\alpha}{n}, i = 1, ..., n,$ **(BHsd)**,
- BH SD procedure with increased first critical values by Theorem 4.3 **(BHsd+)**,
- BH SU procedure **(BHsu)**.

The results are summarized in Table 4.1.

n, n_0	ASDP	ADSP+	BHsd	BHsd+	BHsu
$50, 40$	0.071	0.073	0.071	0.074	0.08
$50, 25$	0.037	0.038	0.038	0.039	0.05
$50, 10$	0.017	0.018	0.013	0.013	0.02
$100, 90$	0.08	0.08	0.08	0.08	0.09
$100, 50$	0.036	0.039	0.038	0.04	0.05
$100, 20$	0.0156	0.0157	0.013	0.014	0.02
$300, 250$	0.069	0.072	0.069	0.072	0.083
$300, 150$	0.034	0.035	0.033	0.036	0.05
$300, 100$	0.021	0.022	0.021	0.022	0.033

Table 4.1: Simulation outcome for Example 4.6 based on $L = 10000$ replications, $\alpha = 0.1$. Comparison of the FDR for the adaptive SD-procedure (5.1) (ASDP), for the SD-procedure with the increased first critical values by Theorem 4.3 (ASDP+), for the SD-procedure with Benjamini-Hochberg critical values (BHsd), for BH-SD-procedure with increased critical values based on Theorem 4.3 (BHsd+) and for the linear SU-procedure (BHsu).

As we can see from Table 4.1, the FDR of the SD procedures with the increased critical values is slightly larger than the FDR of the original procedure and it lies below the level α. Obviously the procedure ASDP+ is more powerful than the procedure BHsd+, because the critical values of the procedure ASDP+ are larger than the corresponding critical values of the BHsd+. Nevertheless it does not mean that the ASDP+ exhausts the level α better. For instance, for $n_0 = 40$ and $n = 50$ the FDR of the ASDP+ is less than the FDR of BHsd+ in contrast to the case $n_0 = 10$ and $n = 50$. This result is not surprising, since, in general, FDR(t) is not monotone in $t, t \in (0, 1)$.

Example 4.7

Let us consider the following (extreme) case. The random variables $V_1, ..., V_n$ are i.i.d. U$(0, 1)$-distributed.

(a) The true p-values $U_i = V_i, i = 1, ..., n_0$,

(b) The false p-values $f_i = V_{n_0+i}, i = 1, ..., n_1$.

n, n_0	ASDP	ADSP+	BHsd	BHsd+	BHsu
$50, 40$	0.079	0.083	0.079	0.08	0.08
$50, 25$	0.046	0.048	0.046	0.047	0.05
$50, 10$	0.018	0.019	0.018	0.019	0.02
$100, 90$	0.085	0.09	0.085	0.085	0.09
$100, 50$	0.048	0.05	0.048	0.049	0.05
$100, 20$	0.018	0.019	0.018	0.018	0.02
$300, 250$	0.081	0.086	0.081	0.082	0.083
$300, 150$	0.048	0.05	0.048	0.048	0.05
$300, 100$	0.034	0.036	0.034	0.034	0.033

Table 4.2: Simulation outcome for Example 4.7 based on $L = 10000$ replications, $\alpha = 0.1$. Comparison of the FDR for the adaptive SD-procedure (5.1) (ASDP), for the SD-procedure with the increased first critical values due to the Theorem 4.3 (ASDP+), for the SD-procedure with Benjamini-Hochberg critical values (BHsd), for the BH-SD-procedure with increased critical values according to Theorem 4.3 (BHsd+), for the linear SU-procedure (BHsu).

In Chapter 7 we consider this example in more detail. We will see that under the assumptions (a),(b) the FDR of any SD procedure which uses the first critical value $b_1 = 1 - \sqrt[n]{1-\alpha}$, is less or equal to $\alpha\frac{n_0}{n}$, if all p-values are i.i.d. $U(0,1)$. That is, any SD-procedure which controls the FDR under POD can be modified according to Theorem 4.3 to some procedure which maximizes the FDR under BIA with i.i.d. uniformly $U(0,1)$ distributed alternatives. Later, in Chapter 7 we will prove these statements.

To proceed, let us compare the following three procedures: ASDP with critical values $a_i = \frac{i\alpha}{n+1-i(1-\alpha)}$, the linear SD procedure with critical values $c_i = \frac{i\alpha}{n}$ and the Benjamini and Liu SD procedure with critical values $b_i = 1 - (1 - \alpha\frac{n}{n-i+1} \wedge 1)^{\frac{1}{n-i+1}}, i = 1, ..., n$. We can see that the inequalities

$$a_1 \leqslant b_1 \qquad \text{and} \qquad c_1 \leqslant b_1$$

are valid for all $n \geqslant 1$ and $\alpha \in (0,1)$. In the majority of practical cases, where α is "small" and n is "large", we get $b_2 \leqslant c_2 \leqslant a_2$. But there exist cases, for which $a_2 \leqslant c_2 \leqslant b_2$ holds, as for example $n = 50$ and $\alpha \geqslant 0.76$ or $c_2 \leqslant a_2 \leqslant b_2$ for $n = 2$ and $\alpha = 0.15$. In practice,we can alway increase the smallest critical value of an FDR controlling SD procedure without loss of the FDR control, which can lead to the improvement of power.

4.2 Conclusions

We proved in this chapter that if some SD procedure with critical values $c_1 \leqslant ... \leqslant c_n$, $c_1 \leqslant b_1$, controls the FDR under PRDS oder POD, then we can improve the power of this procedure without loss of FDR-control by replacing the first critical values by $c_i \vee b_i, i \leqslant k$, where $b_i, i = 1, ..., n$, defined in (4.2) and the value k is defined as in (4.3).

Chapter 5

On the adaptive Gavrilov-Benjamini-Sarkar SD procedure

In this chapter we consider the SD procedure which was proposed by Gavrilov, Benjamini and Sarkar [2009] and by Finner et al. [2009] (in context with $\beta-$adjusted AORC). This SD procedure uses the deterministic critical values

$$a_{i:n} = \frac{i\alpha}{n+1-i(1-\alpha)} \tag{5.1}$$

and the SD-stopping rule

$$\tau_1(p) = a_{R:n} = \max\{a_{i:n} : p_{j:n} \leqslant a_{j:n} \ \text{ for all } \ j \leqslant i\}. \tag{5.2}$$

Remark 5.1
Note that by Bernoulli inequality we have

$$\frac{\alpha}{n} \leqslant 1 - \sqrt[n]{(1-\alpha)}$$

for $\alpha \in (0,1)$. This implies that for the smallest critical value from (5.1) we get

$$a_{1:n} = \frac{\alpha}{n+\alpha} < \frac{\alpha}{n} \leqslant 1 - \sqrt[n]{(1-\alpha)}.$$

Remark 5.2
Gavrilov et al. call the procedure with critical values (5.1) "adaptive" SD procedure (ASDP)

and it is not clear why. In the FDR-framework it exists an approach to estimate the number of the true null hypotheses n_0 by some appropriate estimator $\hat{n}_0$ by means of the data X, such procedures are so-called adaptive BH procedures (cf. Subsection 1.4.3). We propose an interpretation of the term adaptive for this procedure in Remark 5.7.

Recall that the p-values which belong to the true null hypotheses are denoted by $U_1, ..., U_{n_0}$. The "false" p-values are denoted by $f_1, ..., f_{n_1}$ and are assumed to be ordered, i.e., $f_1 \leqslant f_2 \leqslant ... \leqslant f_{n_1}$. Further in this chapter we consider the conditional expectation

$$\mathbb{E}\left[\cdot \mid f\right] = \mathbb{E}_f\left[\cdot\right], \qquad f = (f_1, ..., f_{n_1}).$$

The vector of all p-values is denoted by $p = (p_1, ..., p_n)$.

In this chapter we present a new alternative proof of the FDR-control for the adaptive SD procedure with critical values (5.1) under BIA and we prove a new result on the FDR-control under some dependence assumptions (cf. Theorem 5.17 and Theorem 5.23). Moreover, the critical value $a_{1:n}$ in (5.1) can be modified according to Chapter 4.

We start with some useful observations. Let us consider the SD procedure which uses some (deterministic) critical values $d_1 \leqslant d_2 \leqslant ... \leqslant d_n$. We denote the stopping rule of this SD procedure by d, where

$$d = \max\{d_i : p_{j:n} \leqslant d_j \ \text{ for all } \ j \leqslant i\}.$$

Note that we use the convention $\frac{0}{0} \equiv 0$.

Lemma 5.3

Let the following assumptions be fulfilled:

1. $n_1 \geqslant 1$,
2. $d_1 \leqslant 1 - \sqrt[n]{1-\alpha}$,
3. *the random variables $U_1, ..., U_{n_0}$ are POD,*
4. $\mathbb{E}_{f_0}\left[\frac{V(d)}{R(d)}\right] \leqslant \alpha$ *with $f_0 = (0, f_2, ..., f_{n_1})$ for all possible $f_2, ..., f_{n_1}$.*

Then we have $\mathbb{E}\left[\frac{V(d)}{R(d)}\right] \leqslant \alpha$ for all $n_0 \in \mathbb{N}$.

Proof. Note that $\mathbb{E}_f\left[\frac{V(d)}{R(d)}\right] = \mathbb{E}_f\left[\mathbb{I}(f_1 \leqslant d_1)\frac{V(d)}{R(d)} + \mathbb{I}(f_1 > d_1)\frac{V(d)}{R(d)}\right]$ is always valid and let us consider two different cases: (a) $f_1 \leqslant d_1$, (b) $f_1 > d_1$.
(a) Since f_1 will be rejected, the equality $\mathbb{E}_{f_0}\left[\frac{V(d)}{R(d)}\right] = \mathbb{E}_f\left[\frac{V(d)}{R(d)}\right]$ holds. Therefore the statement of the lemma is proved for this case.
(b) If $f_1 > d_1$ holds, we have by POD assumption $P(U_{1:n_0} \leqslant 1 - \sqrt[n]{1-\alpha}) \leqslant \alpha$, which implies

$$\begin{aligned}\mathbb{E}_f\left[\frac{V(d)}{R(d)}\mathbb{I}(V(d)>0)\right] &\leqslant \mathbb{E}_f\left[\mathbb{I}(V(d)>0)\right] = P(U_{1:n_0} \leqslant d_1) \\ &\leqslant P(U_{1:n_0} \leqslant 1 - \sqrt[n]{1-\alpha}) \leqslant \alpha.\end{aligned}$$

Hence, we get $\mathbb{E}_f\left[\frac{V(d)}{R(d)}\right] \leqslant \alpha$ for all possible vectors $f = (f_1, ..., f_{n_1})$. Further, $\mathbb{E}\left[\frac{V(d)}{R(d)}\right] = \mathbb{E}\left[\mathbb{E}_f\left[\frac{V(d)}{R(d)}\right]\right]$, which completes the proof. ■

Remark 5.4

According to Lemma 5.3 we only have to consider the case $f_1 \leqslant d_1$, if we want to prove the FDR-control of the SD procedure with critical values $d_1 \leqslant d_2 \leqslant ... \leqslant d_n$ if $d_1 \leqslant 1 - \sqrt[n]{1-\alpha}$ under PRDS or POD.

Let us return to the SD procedure with critical values (5.1) (ASDP). We define a random variable

$$\tau(p) = \min\{a_{i:n} : p_{i:n} > a_{i:n}\} \wedge a_{n:n}. \tag{5.3}$$

It is easy to see from (5.2) and (5.3) that $\tau(p) \geqslant \tau_1(p)$ a.s. But $R(\tau_1(p)) = R(\tau(p))$ a.s. holds for the number of rejections as we will see later. Let us consider the following estimator for FDR(t), that is,

$$\hat{\alpha}(t) = \frac{t}{1-t}\frac{n-R(t)}{R(t)+1}. \tag{5.4}$$

Remark 5.5

Comparing the estimator $\hat{\alpha}(t)$ from (5.4) with the estimator $\hat{\alpha}_n(t)$ from (3.9), we obtain

$$\frac{R(t)+1}{R(t)}\hat{\alpha}(t) = \hat{\alpha}_n(t)$$

for all t with $R(t) > 0$.

We now show that the SD procedure with critical values (5.1) is related to the estimator (5.4).

Lemma 5.6

The following equality is valid

$$\tau(p) = \inf\{t \in (0,1) : \hat{\alpha}(t) > \alpha\} \wedge a_{n:n} \tag{5.5}$$

for all $\alpha \in (0,1)$, *where* $\tau(p)$ *is defined in (5.3).*

Proof. If there are no points of intersection between the curve $\hat{\alpha}(t)$ and the $\alpha-$line, i.e., the equality $\hat{\alpha}(t) = \alpha$ has no solution on $t \in [0,1)$, then $\tau(p) = a_{n:n}$ by definition and (5.5) is valid. If there exist some points of intersection between the curve $\hat{\alpha}(t)$ and the $\alpha-$line then they are obviously of the form

$$t = \frac{\alpha(R(t)+1)}{n+1-(R(t)+1)(1-\alpha)}.$$

Since $R(t) \in \{1,...,n\}$, for all $t \in (0,1)$, we have at most n points of the intersection, namely $a_{1:n},...,a_{n:n}$ (see also Scheer [2012] (p. 33)). Further we get

$$\begin{aligned}
\tau(p) &= \min\{a_{i:n} : p_{i:n} > a_{i:n}\} \wedge a_{n:n} \\
&= \min\{a_{i:n} : \hat{F}_n(p_{i:n}) > \hat{F}_n(a_{i:n})\} \wedge a_{n:n} \\
&= \min\{a_{i:n} : \hat{F}_n(a_{i:n}) = \frac{i-1}{n}\} \wedge a_{n:n} \\
&= \min\{a_{i:n} : \hat{\alpha}(a_{i:n}) = \alpha\} \wedge a_{n:n} \\
&= \inf\{t \in (0,1) : \hat{\alpha}(t) > \alpha\} \wedge a_{n:n}.
\end{aligned}$$

■

Remark 5.7

The estimator (5.4) can be represented as

$$\begin{aligned}
\hat{\alpha}(t) &= \frac{t}{R(t)} \frac{R(t)}{R(t)+1} \frac{n-R(t)}{1-t} \\
&= \frac{t}{\hat{F}_n(t)} \frac{R(t)}{R(t)+1} \hat{\pi}_0(t), \\
\textit{with} \quad \hat{\pi}_0(t) &= \frac{n-R(t)}{n(1-t)}.
\end{aligned}$$

Thereby $\hat{\pi}_0(t)$, $t \in (0,1)$ *is an estimate of* $\frac{n_0}{n}$, *the proportion of true null hypotheses proposed by Storey et al. (cf. Storey et al. [2004] p.190) with* $\lambda = t$. *Thus, we can* *interpret the SD procedure with critical values (5.1) as an adaptive BH SD procedure with the following estimator* $\hat{\hat{\pi}}_0(t)$ *of* $\frac{n_0}{n}$

$$\hat{\hat{\pi}}_0(t) = \frac{R(t)}{R(t)+1}\hat{\pi}_0(t).$$

Lemma 5.8

Let us consider the values $\tau(p)$ *and* $\tau_1(p)$ *which are defined in (5.2) and (5.3).*

1. *The random variable* $\tau(p)$ *is a stopping-time w.r.t. the filtration* $\mathcal{F}_t = \sigma(\mathbb{I}_{[0,s]}(p_i),\ \forall 0 \leqslant s \leqslant t,\ i = 1, ..., n), t \in [0, 1)$.
2. *On the set* $\bigcup_{i\in\{1,...,n\}} \{\hat{\alpha}(a_{i:n}) > \alpha\}$ *there are no p-values between* $\tau_1(p)$ *and* $\tau(p)$. *Therefore* $\hat{F}_n(\tau(p)) = \hat{F}_n(\tau_1(p)) = \frac{R(\tau(p))}{n}$ *holds.*
3. *On* $\bigcap_{i\in\{1,...,n\}} \{\hat{\alpha}(a_{i:n}) \leqslant \alpha\}$ *we have* $\tau_1(p) = a_{n:n} = \tau(p)$ *and we reject all hypotheses.*
4. *From 5.8.2 and 5.8.3 we get* $R(\tau(p)) = R(\tau_1(p))$ *a.s.*
5. *The functions* $U_i \mapsto \tau(U)$ *and* $U_i \mapsto \tau_1(U)$ *are decreasing for each* $i = 1, ..., n_0$.

Proof. 1. Since $\hat{\alpha}_n$ is the right continuous $\mathcal{F}_t$-adapted process, then by Debut Theorem (cf. Richard F. Bass [2011] p.117), $\tau(p)$ is a stopping time as a first entrance time into an open set.
Let us define $h : (x, y) \mapsto \frac{x}{1-x}\frac{1-y}{y+\frac{1}{n}}$. The implications 2. and 3. follow from the observation that $\hat{\alpha}(t) = h(t, \hat{F}_n(t))$ and $h(a_{i:n}, \frac{i-1}{n}) = \alpha$.
The function $\hat{F}_n(t) \mapsto \hat{\alpha}(t)$ is decreasing which implies that the function $U_i \mapsto \hat{\alpha}(t, \hat{F}_n(t))$ is increasing, so we can prove property 5. ∎

5.1 Control of the ENFR under (s)MD between the true p-values

In the sequel we will often use the Optional Sampling Theorem for bounded stopping times which can be found, for example, in Karatzas [2000] (p. 20, Problem 3.23). The next lemma shows, that the SD-procedure (ASDP) based on the critical values (5.1) controls the ENFR at level $\frac{\alpha}{1-\alpha}(n_1 + 1)$ under the assumption that the p-values $p_1, ..., p_n$ belong to the class M_{I_0}, i.e., the process $\left(\frac{V(t)-n_0 t}{1-t}\right)_{t\in[0,1)}$ is an $\mathcal{F}_t$-martingale (conditionally on f and n_0), where $\mathcal{F}_t$ is defined as in Lemma 5.8.1. (see also Definition 2.16).

Lemma 5.9
If $U = (U_1, ..., U_{n_0}) = (p_i)_{i\in I_0}$ are (s)MD in sense of Definition 2.16 and $P(p_i = 0) = 0$ for all $i \in I_0$, we have

$$\mathbb{E}_f \left[V(\tau(p))\right] \leqslant \frac{\alpha}{1-\alpha}(n_1 + 1). \tag{5.6}$$

Proof. For all $t \in (0, 1)$ we have

$$(1 - \hat{\alpha}(t))(R(t) + 1) = \frac{R(t) - tn}{1-t} + 1 \tag{5.7}$$

$$= \frac{V(t) - n_0 t}{1-t} + \frac{S(t) - n_1 t}{1-t} + 1 \tag{5.8}$$

$$= M(t) + S(t) + 1 + \frac{t}{1-t}(S(t) - n_1) \tag{5.9}$$

$$\leqslant M(t) + S(t) + 1. \tag{5.10}$$

Further, note that the inequality

$$\lim_{t\uparrow p_i} R(t) < R(p_i) \tag{5.11}$$

is valid for all $i \in I$, which implies

$$\lim_{t\uparrow p_i} \hat{\alpha}(t) > \hat{\alpha}(p_i) \tag{5.12}$$

for all $i \in I$. Hence, the process $\hat{\alpha}(t)$ has only negative jumps. Consequently $\hat{\alpha}(\tau) \leqslant \alpha$ by Lemma 5.6. Due to Lemma 5.8, part(4), and the chain of (in)equalities (5.7)-(5.10) we get

$$(1 - \alpha)R(\tau_1(p)) = (1 - \alpha)R(\tau(p)) \tag{5.13}$$

$$\leqslant (1 - \hat{\alpha}(\tau(p)))R(\tau(p)) = M(\tau(p)) + S(\tau(p)) + \hat{\alpha}(\tau(p)) \tag{5.14}$$

$$\leqslant M(\tau(p)) + S(\tau(p)) + \alpha. \tag{5.15}$$

Since $\tau(p)$ is the stopping-time w.r.t. $\mathcal{F}_t$ and $R(\tau(p)) = V(\tau(p)) + S(\tau(p))$, we have from the Optional Stopping Theorem $\mathbb{E}_f \left[M(\tau(p))\right] \leqslant \mathbb{E}_f \left[M(0)\right] \leqslant 0$ (with equality in the case of martingale dependence) and

$$(1 - \alpha)\mathbb{E}\left[V(\tau(p))\right] \leqslant \mathbb{E}\left[M(\tau(p)) + \alpha(S(\tau(p)) + 1)\right] \leqslant \alpha(n_1 + 1). \tag{5.16}$$

This completes the proof of the Lemma. ■

Note that in Lemma 5.9 we do not need the assumption about independence between $(U_1, ..., U_{n_0})$ and $(f_1, ..., f_{n_1})$.

Remark 5.10
Scheer [2012] (pp. 51-55) proved the same result for i.i.d. $U(0,1)$-distributed true p-values. Lemma 5.9 is a generalization of his result, since it allows dependence between the underlying p-values. The proof of Lemma 5.9 is based on a martingale argument which makes the proof more elegant.

Lemma 5.11 (Some exact formulas for the ENFR)
Let $U_1, ..., U_{n_0}$ be MD (cf. Definition 2.16 and Remark 2.17) with $P(U_i = 0) = 0$ for all $i \in \{1, ..., n_0\}$ and $0 < n_0 \leqslant n$ be some fixed number. Then we have:
(a) If $f_i, i = 1, ..., n_1$ follow the Dirac $\delta(0)$ distribution, then

$$\mathbb{E}\left[V(\tau)\right] = \frac{\alpha}{1-\alpha}(n_1+1) - \frac{\alpha}{1-\alpha}(n+1)P(V(\tau) = n_0). \tag{5.17}$$

(b) If $(f_1, ..., f_{n_1})$ belongs to M_{I_1} (cf. Definition 2.16),i.e., $M_1(t) = \left(\frac{S(t)-n_1 t}{1-t}\right)_{t\in[0,1)}$ is an $\mathcal{F}_t-$martingale and $\mathcal{L}(f_i) = \mathcal{L}(U_j)$ for all $i = 1, ..., n_1$ and all $j = 1, ..., n_0$, then

$$\mathbb{E}\left[V(\tau)\right] = \frac{\alpha}{1-\alpha}\frac{n_0}{n} - \frac{\alpha}{1-\alpha}\frac{n+1}{n}P(R(\tau) = n), \tag{5.18}$$

where $\tau \equiv \tau(p)$ is defined in (5.3).

Proof. (a) From the definition of τ and $\hat{\alpha}$ and due to (5.7)-(5.9) we have $\hat{\alpha}(\tau) = \alpha\mathbb{I}_{R(\tau)<n}$ and

$$(1-\hat{\alpha}(\tau))(R(\tau)+1) = (1-\alpha)(R+1)\mathbb{I}_{R(\tau)<n} + (n+1)\mathbb{I}_{R(\tau)=n} \tag{5.19}$$

$$= M(\tau) + n_1 + 1. \tag{5.20}$$

Since $\mathbb{I}_{R(\tau)<n} = 1 - \mathbb{I}_{R(\tau)=n}$ holds, the equalities (5.19)-(5.20) are equivalent to

$$(1-\alpha)(R(\tau)+1) + \alpha(n+1)\mathbb{I}_{R(\tau)=n} = M(\tau) + n_1 + 1, \tag{5.21}$$

which implies the equality

$$(1-\alpha)V(\tau) = M(\tau) + \alpha(n_1+1) - \alpha(n+1)\mathbb{I}_{R(\tau)=n}. \tag{5.22}$$

To complete the proof of this part we take the expectation $\mathbb{E}_f$ and apply the Optional Sampling Theorem.

(b) Analogous to the case (a) we have

$$(1-\alpha)(V(\tau)+S(\tau)+1)+\alpha(n+1)\mathbb{I}_{R(\tau)=n} = M(\tau)+M_1(\tau)+1, \tag{5.23}$$

thereby $M_1(t) = \frac{\sum_{j=1}^{n_1}\mathbb{I}(f_j \leqslant t) - n_1 t}{1-t}$ is a $\mathcal{F}_t-$martingale. Equality (5.23) implies

$$(1-\alpha)\mathbb{E}\left[V(\tau)+S(\tau)\right] = \alpha - \alpha(n+1)P(R(\tau)=n) \tag{5.24}$$

by taking the expectation $\mathbb{E}$ and applying the Optional Sampling Theorem. The equality $\mathbb{E}\left[S(\tau)\right] = \frac{n_1}{n_0}\mathbb{E}\left[V(\tau)\right]$ (which follows from the assumption, that $U_1,...U_{n_0}, f_1,...,f_{n_1}$ are identically distributed) completes the proof of this Lemma. ■

Remark 5.12

The BIA are a special case of the assumptions of Lemma 5.9.

Remark 5.13

Lemma 5.9 remains true for any random variable $\sigma \leqslant \tau(p)$ a.s., i.e., if $U = (U_1,...,U_{n_0}) = (p_i)_{i\in I_0}$ are (s)MD in sense of Definition 2.16 and Remark 2.17 with $P(U_i = 0) = 0$ for all $i \in \{1,...,n_0\}$, we have

$$\mathbb{E}_f\left[V(\sigma)\right] \leqslant \frac{\alpha}{1-\alpha}(n_1+1). \tag{5.25}$$

Remark 5.14

If $U_1,...,U_{n_0}$ are i.i.d uniformly $U(0,1)$-distributed, then the following equalities are valid:
(a)

$$\mathbb{E}\left[V(\tau)\right] = \frac{\alpha}{1-\alpha}(n_1+1) - \frac{\alpha}{1-\alpha}(n+1)F_n(a_{n_1+1:n}, a_{n_1+2:n}, ..., a_{n:n}), \tag{5.26}$$

under the assumptions of Lemma 5.11 (a), i.e., if $f_i, i = 1,...,n_1$ follow the Dirac $\delta(0)$ distribution.
(b)

$$\mathbb{E}\left[V(\tau)\right] = \frac{\alpha}{1-\alpha}\frac{n_0}{n} - \frac{\alpha}{1-\alpha}\frac{n+1}{n}F_n(a_{1:n},...,a_{n:n}), \tag{5.27}$$

under the assumptions of Lemma 5.11 (b), i.e., if $f_1,....,f_{n_1}$ belong to M_{I_1} (cf. Definition 2.16),i.e., and $\mathcal{L}(f_i) = \mathcal{L}(U_j)$ for all $i = 1,...,n_1$ and all $j = 1,...,n_0$.
Thereby F_k is the joint c.d.f. of the order statistics $U_{1:k},...,U_{k:k}$, of k i.i.d. uniformly distributed random variables $U_i, i = 1,...,k$, which can be computed recursively by

$$F_k(\gamma_1,...,\gamma_k) = 1 - \sum_{j=0}^{k-1}\tbinom{k}{j}F_j(\gamma_1,...,\gamma_j)(1-\gamma_{j+1})^{k-j}, \tag{5.28}$$

with $0 \leqslant \gamma_1 \leqslant ... \leqslant \gamma_k$, $k \in \mathbb{N}$ *and* $F_0 \equiv 1$, *cf. Shorack and Wellner [1986] (pp. 366-367). Thereby the stopping rule* τ *is defined in (5.3).*

Remark 5.15

Consider the SD procedure with critical values (5.1). If for some $k \geqslant 1$ *the equalities* $\sum\limits_{i=1}^{n_1} \mathbb{I}(f_i \leqslant a_{1:n}) = n_1 - k$ *and* $\sum\limits_{i=1}^{n_1} \mathbb{I}(f_i > a_{n:n}) = k$ *hold, then we have*

$$\mathbb{E}\left[V(\tau(p))\right] = \frac{\alpha}{1-\alpha}(n_1 - k + 1) \tag{5.29}$$

if $(U_1, ..., U_{n_0})$ *are MD with* $P(U_i = 0) = 0$ *for all* $i \in \{1, ..., n_0\}$.

Proof. The assumption, that $k \geqslant 1$ holds, ensures that some $i = 1, ..., n$ exists with $\hat{F}_n(a_{i:n}) = \frac{i-1}{n}$, which implies that $\hat{\alpha}(\tau(p)) = \alpha$ holds (compare with Lemma 5.8). Hence, we have

$$(1-\alpha)V(\tau) = M(\tau) + \alpha(S(\tau) + 1) = M(\tau) + \alpha(n_1 - k + 1).$$

Applying the Optional Sampling Theorem completes the proof of this Remark. ■

5.2 Control of the FDR under BIA

The next lemma is a technical result which will be used for the proof of FDR-control under BIA of the SD procedure with critical values (5.1) (ASDP). Note that we use the convention $\frac{0}{0} \equiv 0$.

Lemma 5.16

Consider the stopping rule $\tau_1 \equiv \tau_1(p)$ *given in (5.2). Under BIA we have*

$$\mathbb{E}_f\left[\frac{V(\tau_1(p))}{\tau_1(p)}\right] \leqslant n_0. \tag{5.30}$$

Proof. Observe that

$$\mathbb{E}_f\left[\frac{V(\tau_1(p))}{\tau_1(p)}\right] = n_0 \mathbb{E}_f\left[\frac{\mathbb{I}(U_1 \leqslant \tau_1(p))}{\tau_1(p)}\right] \tag{5.31}$$

$$= n_0 \mathbb{E}_f\left[\frac{\mathbb{I}(U_1 \leqslant \tau_1(p))}{\tau_1(p^{(0)})}\right] \leqslant n_0 \mathbb{E}_f\left[\frac{\mathbb{I}(U_1 \leqslant \tau_1(p^{(0)}))}{\tau_1(p^{(0)})}\right] \tag{5.32}$$

$$= n_0, \tag{5.33}$$

where $p^{(0)}$ in (5.32) is given by $p^{(0)} = (f, U^{(0)})$ with $U^{(0)} = (0, U_2, ..., U_{n_0})$. The equality in (5.32) follows from the observation that

$$\tau_1(p)\mathbb{I}(U_1 \leqslant \tau(p)) = \tau_1(p^{(0)})\mathbb{I}(U_1 \leqslant \tau(p)) \quad \text{is valid.}$$

It is easy to see that $\tau_1(p^{(0)}) \geqslant \tau_1(p)$, which implies the inequality in (5.32). The last equality is true due to the Fubini Theorem, since $p^{(0)}$ and U_1 are independent. ■

Remark 5.17

Let $0 < n_0 \leqslant n$ be some fixed number. If $\sigma(p)$ is a stopping time w.r.t. the reverse filtration $\mathcal{RF}_t = \sigma(\mathbb{I}_{(s,1]}(p_i), \forall 0 \leqslant t \leqslant s, i = 1, ..., n), t \in (0, 1]$, then we have $\mathbb{E}\left[\frac{V(\sigma(p))}{\sigma(p)}\right] = n_0$ under BIA, since the process $\left(\frac{V(t)}{t}\right)_{t\in(0,1]}$ is a reverse $\mathcal{RF}_t$-martingale. (cf. Shorack, Wellner [1986] p.136). This fact is not relevant for SD procedures, but it is of great importance for the proof of the FDR-control for some SU procedures, for example the BH SU procedure and the Storey's SU procedure (cf. Storey [2004]).

Now we are able to prove the FDR-control under BIA.

Theorem 5.18

Under BIA we have

$$\mathbb{E}\left[\frac{V(\tau_1(p))}{R(\tau_1(p))}\right] \leqslant \alpha,$$

where $\tau_1 \equiv \tau_1(p)$ is defined in (5.2).

Proof. We note at first that by Lemma 5.3 we only need to prove that $\mathbb{E}_{f_0}\left[\frac{V(\tau_1(p))}{R(\tau_1(p))}\right] \leqslant \alpha$ holds for the vector f_0, with $f_0 = (0, f_2, ..., f_{n_1})$, $f_2 \leqslant f_3 \leqslant ... \leqslant f_{n_1}$. Hence, for the number of all rejections of the procedure $\tau_1(p)$ we get:

$$R(\tau_1) = \frac{(n+1)a_{R:n}}{\alpha + a_{R:n}(1-\alpha)} = \frac{(n+1)\tau_1(p)}{\alpha + \tau_1(p)(1-\alpha)}, \tag{5.34}$$

which implies

$$\mathbb{E}_{f_0}\left[\frac{V(\tau_1(p))}{R(\tau_1(p))}\right] = \mathbb{E}_{f_0}\left[\frac{V(\tau_1)(\alpha + \tau_1(p)(1-\alpha))}{(n+1)\tau_1(p)}\right] \tag{5.35}$$

$$= \frac{(1-\alpha)}{n+1}\mathbb{E}_{f_0}\left[V(\tau_1(p))\right] + \frac{\alpha}{n+1}\mathbb{E}_{f_0}\left[\frac{V(\tau_1(p))}{\tau_1(p)}\right] \tag{5.36}$$

$$\leqslant \alpha\frac{n_1+1}{n+1} + \alpha\frac{n_0}{n+1} = \alpha. \tag{5.37}$$

The inequality in (5.37) follows from Lemma 5.16, Remark 5.12 and Lemma 5.9. ■

Remark 5.19

Gavrilov et al. [2009] proved the same result as in Theorem 5.18. We used an alternative short proof based on a martingale argument and on the control of the ENFR.

5.3 Control of the FDR under PRDS and (s)MD

In this section we continue to consider the SD procedure (5.1)(ASDP) with critical values $a_{i:n} = \frac{i\alpha}{n+1-i(1-\alpha)}$ and the stopping rule $\tau_1 = \max\{a_{i:n} : p_{j:n} \leqslant a_j, \text{ for all } j \leqslant i\}$. Gavrilov et al. [2009] have shown with help of simulations that the FDR-level of the SD procedure (5.1) can exceed α, when the p-values are PRDS (cf. Gavrilov et al. [2009] pp.625-626). They proposed another SD procedure which controls the FDR for this kind of dependence. This new SD procedure uses the critical values $b_{i:n} = \frac{i\alpha}{n+\beta_n-i(1-\alpha)}$, $i = 1, ..., n$ with $\beta_n \geqslant n(1-\alpha)$. Unfortunately, this procedure is very conservative (cf. Gavrilov et al. [2009] p.628). We show that if the p-values are both PRDS and (s)MD (see Chapter 6 for Examples), then the FDR of the SD procedure with critical values (5.1) is controlled at level α. The next lemma is a technical tool for the proof of FDR control under PRDS and MD.

Lemma 5.20

Let the following assumptions be fulfilled

1. $U_1, ..., U_{n_0} \succcurlyeq_{st} U(0,1)$
2. $U_1, ..., U_{n_0}$ *are PRDS in the sense of Definition 2.5 and Remark 2.8.*

Then we have

$$\mathbb{E}_f\left[\frac{V(\tau_1(p))}{\tau_1(p)}\right] \leqslant n_0, \tag{5.38}$$

where $\tau_1 \equiv \tau_1(p)$ is defined in (5.2).

Proof. Let us define $a_{0:n} = 0$ for technical reasons. We obtain the following sequence of (in)equalities:

$$\mathbb{E}_f\left[\frac{V(\tau_1(p))}{\tau_1(p)}\right] = \sum_{j=1}^{n_0} \mathbb{E}_f\left[\frac{\mathbb{I}(U_j \leqslant \tau_1(p))}{\tau_1(p)}\right] \tag{5.39}$$

$$= \sum_{j=1}^{n_0}\sum_{i=1}^{n} \mathbb{E}_f\left[\frac{\mathbb{I}(U_j \leqslant a_{i:n})}{a_{i:n}}\mathbb{I}(\tau_1(p) = a_{i:n})\right] \tag{5.40}$$

$$= \sum_{j=1}^{n_0}\sum_{i=1}^{n} \mathbb{E}_f\left[\frac{\mathbb{I}(U_j \leqslant a_{i:n})}{a_{i:n}}\left(\mathbb{I}(\tau_1(p) \leqslant a_{i:n}) - \mathbb{I}(\tau_1(p) \leqslant a_{i-1:n})\right)\right] \tag{5.41}$$

$$\leqslant \sum_{j=1}^{n_0}\sum_{i=1}^{n} \left(\mathbb{E}_f\left[\mathbb{I}(\tau_1(p) \leqslant a_{i:n})|U_j \leqslant a_{i:n}\right] - \mathbb{E}_f\left[\mathbb{I}(\tau_1(p) \leqslant a_{i-1:n})|U_j \leqslant a_{i:n}\right]\right) \tag{5.42}$$

$$\leqslant \sum_{j=1}^{n_0}\sum_{i=1}^{n} \left(\mathbb{E}_f\left[\mathbb{I}(\tau_1(p) \leqslant a_{i:n})|U_j \leqslant a_{i:n}\right] - E_f\left[\mathbb{I}(\tau_1(p) \leqslant a_{i-1:n})|U_j \leqslant a_{i-1:n}\right]\right) \tag{5.43}$$

$$= \sum_{j=1}^{n_0} \mathbb{E}_f\left[\mathbb{I}(\tau_1(p) \leqslant a_{n:n})|U_j \leqslant a_{n:n}\right] = n_0. \tag{5.44}$$

The inequality in (5.42) is valid, since $U_1, ..., U_{n_0}$ are stochastically greater than $U(0,1)$. The inequality in (5.43) holds, because the function
$x \mapsto \mathbb{I}(\tau_1(p) \leqslant a_{i-1:n} \mid U_i \leqslant x)$ is increasing in x for all $i \in \{1, ..., n_0\}$ and since $U_1, ..., U_{n_0}$ are assumed to be PRDS. Consequently, using the telescoping sum, we obtain the first equality in (5.44). The proof is completed, because $\tau_1(p) \leqslant a_{n:n}$ holds by definition of τ_1.
■

Remark 5.21
The assumptions in Lemma 5.16 are obviously a special case of the assumptions of Lemma 5.20. Since the proof of the Lemma for i.i.d random variables $U_1, ..., U_{n_0}$ can be conducted by simpler arguments, we stated the proof of Lemma 5.16 first for instructive purposes.

Remark 5.22
Lemma 5.20 obviously remains valid for any stopping rule $\sigma(U)$ which is a non-increasing function of U_i, $i = 1, ..., n_0$, and has a finite range of values $\beta_1 < \beta_2 < ... < \beta_n$, with $\beta_i \in (0,1)$.

Lemma 5.23
If the random variables $U_1, ..., U_{n_0}$ are PRDS and MD_{I_0} (cf. Definition 2.5, Remark 2.8

and Remark 2.17) with $P(U_i = 0) = 0$ for all $i \in \{1, ..., n_0\}$, then we have

$$\mathbb{E}\left[\frac{M(\tau)}{\tau}\right] \leqslant 0, \tag{5.45}$$

where $\tau \equiv \tau(p)$ is defined in (5.3).

Proof. Remember that τ is the stopping time w.r.t. the filtration $\mathcal{F}_t = \sigma(\mathbb{I}_{(U_i \leqslant s)}, 0 \leqslant s \leqslant t, i = 1, ..., n_0), t \in (0,1)$ and τ_1 is the classical stopping rule of the SD-procedure (5.1). By Lemma 5.8.4 we have $R(\tau) = R(\tau_1)$. From Lemma 5.20 and due to the Optional Sampling Theorem we conclude

$$\begin{aligned} n_0 \geqslant \mathbb{E}_f\left[\frac{V(\tau_1)}{\tau_1}\right] &\geqslant \mathbb{E}_f\left[\frac{V(\tau)}{\tau}\right] \\ &= \mathbb{E}_f\left[\frac{M(\tau)(1-\tau) + n_0\tau}{\tau}\right] \\ &= \mathbb{E}_f\left[\frac{M(\tau)}{\tau}\right] + n_0, \end{aligned} \tag{5.46}$$

which immediately implies the statement of this lemma. ■

Theorem 5.24

If the random variables $U_1, ..., U_{n_0}$ are PRDS and (s)MD (cf. Definition 2.5, Remark 2.8 and Remark 2.17) with $P(U_i = 0) = 0$ for all $i \in \{1, ..., n_0\}$, then the FDR of the procedure with critical values (5.1) (ASDP) is controlled at the desired level α, i.e.,

$$\mathbb{E}\left[\frac{V(\tau_1(p))}{R(\tau_1(p))}\right] \leqslant \alpha, \tag{5.47}$$

where $\tau_1 \equiv \tau_1(p)$ is defined in (5.2).

Proof. Recall the proof of Theorem 5.18 and use Lemma 5.20 instead of Lemma 5.16.

■

Remark 5.25

Let $U_1, ..., U_{n_0}$ be PRDS and (s)MD (cf. Definition 2.5, Remark 2.8 and Remark 2.17)with $P(U_i = 0) = 0$ for all $i \in \{1, ..., n_0\}$. Based on Remark 5.13 and 5.22 we obtain:

$$E\left[\frac{V(\sigma(p))}{R(\sigma(p))}\right] \leqslant \alpha \tag{5.48}$$

for any stopping rule $\sigma = \sigma(p) \leqslant \tau(p)$ which has a finite range of values and is decreasing in each coordinate of p, where $\tau \equiv \tau(p)$ is defined in (5.3).

We refer to the next chapter for examples of (s)MD and PRDS random variables.

5.4 Another generalization

Let us consider the estimator

$$\hat{\alpha}_\beta(t) = \frac{t}{1-t} \cdot \frac{n - R(t)}{R(t) + \beta} \tag{5.49}$$

for FDR(t) for some real number $\beta \in (0, 1]$ and the SD procedure with stopping rule

$$\tau_\beta = \inf\{t \in (0,1) : \hat{\alpha}_\beta(t) > \alpha\} \wedge \frac{\alpha(n-1+\beta)}{n+\beta-(n-1+\beta)(1-\alpha)}. \tag{5.50}$$

For this τ_β SD procedure we formulate the following theorem.

Theorem 5.26
If the p-values $U_1, ..., U_{n_0}$ which correspond to true null hypothesis are a.s. positive, are PRDS and (s)MD (cf. Definition 2.5, Remark 2.8 and Remark 2.17) and are independent from the p-values $f_1, ..., f_{n_1}$ corresponding to alternatives, then we have

$$\mathbb{E}\left[\frac{V(\tau_\beta)}{R(\tau_\beta) - (1-\beta)}\right] \leqslant \alpha. \tag{5.51}$$

Proof. Firstly, we can see that analogous to Lemma 5.6, it can be easily shown that the SD procedure τ_β is equivalent to the SD procedure with critical values

$$a_i^\beta = \frac{\alpha(i-1+\beta)}{n+\beta-(i-1+\beta)(1-\alpha)}.$$

Consequently, the stopping rule τ_β can be represented as

$$\tau_\beta = \frac{\alpha(R-1+\beta)}{n+\beta-(R-1+\beta)(1-\alpha)},$$

where $R = R(\tau_\beta)$ is the number of rejections of the procedure τ_β. Now, similarly to Theorem 5.24, we get

$$\mathbb{E}\left[\frac{V(\tau_\beta)}{R(\tau_\beta) - (1-\beta)}\right] = \mathbb{E}\left[\frac{\alpha + \tau_\beta(1-\alpha)}{(n+\beta)} \cdot \frac{V(\tau_\beta)}{\tau_\beta}\right] \tag{5.52}$$

$$= \frac{\alpha}{n+\beta}\mathbb{E}\left[\frac{V(\tau_\beta)}{\tau_\beta}\right] + \frac{(1-\alpha)}{n+\beta}\mathbb{E}\left[V(\tau_\beta)\right]. \tag{5.53}$$

For the first summand in (5.53) we can conclude by Remark 5.22 that $\frac{\alpha}{n+\beta}\mathbb{E}\left[\frac{V(\tau_\beta)}{\tau_\beta}\right] \leqslant \frac{\alpha \cdot n_0}{n+\beta}$. In order to make a statement about the second term in (5.53), we obtain from (5.49) and (5.50) the inequality

$$(1-\alpha)(R(\tau_\beta) + \beta) \leqslant \frac{R(\tau_\beta) - \tau_\beta n}{1 - \tau_\beta} + \beta. \tag{5.54}$$

Analogously to Theorem 5.9 the last inequality implies

$$(1-\alpha)\mathbb{E}\left[V(\tau_\beta)\right] \leqslant \mathbb{E}\left[M(\tau_\beta)\right] + \alpha(n_1+\beta). \tag{5.55}$$

Therefore, due to the Optional Sampling Theorem, we finally get

$$\mathbb{E}\left[\frac{V(\tau_\beta)}{R(\tau_\beta)-(1-\beta)}\right] \leqslant \frac{\alpha\cdot n_0}{n+\beta} + \frac{\alpha(n_1+\beta)}{n+\beta} = \alpha.$$

■

Remark 5.27

It can be seen from the proof of Theorem 5.26 that if $U_1, ..., U_{n_0}$ are (s)MD with $P(U_i = 0) = 0$ for all $i \in \{1, ..., n_0\}$, then the SD procedure τ_β controls the ENFR in the sense of Definition 1.16 with $g(n_1) = \frac{\alpha}{1-\alpha}(n_1+\beta)$, i.e.,

$$\mathbb{E}\left[V(\tau_\beta)\right] \leqslant \frac{\alpha}{1-\alpha}(n_1+\beta).$$

5.5 Control of the k-FWER under martingale dependence

Now let us consider a multiple testing procedure $\tilde{\tau}(p)$ which controls the expected number of false rejections at g in the sense of Definition 1.16 with

$$\mathbb{E}\left[V(\tilde{\tau})\right] \leqslant g(n_1) \tag{5.56}$$

for all $n_1 = 0, ..., n-1$. In this section we assume that $n_0 \in \mathbb{N}$ is an arbitrary but fixed number. Since the random variable $V(\tilde{\tau})$ can take only non-negative values $i \in \{0, ..., n_0\}$ we obtain by Lemma 1.18

$$\text{ENFR}(\tau) = \sum_{k=1}^{n_0} k\text{-FWER}(\tau).$$

Thus, if (5.56) holds, we get

$$\sum_{k=1}^{n_0} \text{k-FWER}(\tilde{\tau}) \leqslant g(n_1). \tag{5.57}$$

Obviously, for all $0 < k_1 \leqslant k_2 \leqslant n_0$ and all $k \leqslant n_0$ we have

$$P\left(V(\tau) \geqslant k_2\right) \leqslant P\left(V(\tau) \geqslant k_1\right) \quad \text{and} \quad P\left(V(\tau) = n_0\right) \leqslant P\left(V(\tau) \geqslant k\right).$$

Hence, we obtain the inequality

$$k \cdot \text{k-FWER}(\tilde{\tau}) + (n_0 - k)P(V(\tilde{\tau}) = n_0) \leqslant \sum_{k=1}^{n_0} \text{k-FWER}(\tilde{\tau}) \tag{5.58}$$

for all $k \leqslant n_0$. Consequently, from (5.57) and (5.58) it follows that

$$k \cdot \text{k-FWER}(\tilde{\tau}) + (n_0 - k)P(V(\tilde{\tau}) = n_0) \leqslant g(n_1) \tag{5.59}$$

for $k \leqslant n_0$, which yields an upper bound for the k-FWER$(\tilde{\tau})$, that is,

$$\text{k-FWER}(\tilde{\tau}) \leqslant \frac{g(n_1) - (n_0 - k)P(V(\tilde{\tau}) = n_0)}{k}. \tag{5.60}$$

Let us assume that the false p-values are independent from the true ones. Then the least favorable configuration for the k-FWER (exactly as for the ENFR) is the Dirac-uniform model DU(n_0, n). For this configuration the k-FWER (as well as the ENFR and the FDR) is analytically computable, if the step-wise procedure is defined in terms of the deterministic critical values.

The following lemma yields an upper bound for the k-FWER of the SD procedure with critical values $a_{i:n} = \frac{i\alpha}{n+1-i(1-\alpha)}$ (ASDP). Note that this bound may be large.

Lemma 5.28

If $U_1, ..., U_{n_0}$ are MD (cf. Remark 2.17) with $P(U_i = 0) = 0$ for all $i \in \{1, ..., n_0\}$ and $f_1, ..., f_{n_1}$ are independent from $U_1, ..., U_{n_0}$, then we have

$$\textit{k-FWER}(\tau) \leqslant \frac{\alpha}{1-\alpha}\frac{n_1+1}{k} - P(M(\tau) = n_0)\left(\frac{\alpha}{1-\alpha}(n+1) - (n_0 - k)\right) \tag{5.61}$$

for all $k \in \{1, ..., n_0\}$ and τ which is defined in (5.3).

Proof. We have shown, in the proof of Lemma 5.11 that

$$\mathbb{E}\left[V(\tau)\right] = \frac{\alpha}{1-\alpha}(n_1+1) - \frac{\alpha}{1-\alpha}(n+1)P(M(\tau) = n_0) \tag{5.62}$$

holds, if the p-values, which belong to the false null hypotheses, are all Dirac $\delta(0)-$distributed, i.e., $P_{\mathrm{DU}(n_0,n)}(f_i = 0) = 1$ for all $i = 1, ..., n_1$. ■

Similar to the motivation at the beginning of this section we get from (5.62)

$$k \cdot \text{k-FWER} + (n_0 - k)P(M(\tau) = n_0) \leqslant \sum_{k=1}^{n_0} \text{k-FWER}(\tau) \tag{5.63}$$

$$= \frac{\alpha}{1-\alpha}(n_1+1) - \frac{\alpha}{1-\alpha}(n+1)P(M(\tau) = n_0), \tag{5.64}$$

which implies, that

$$\text{k-FWER}(\tau) \leqslant \frac{\alpha}{1-\alpha}\frac{n_1+1}{k} - P(M(\tau)=n_0)\left(\frac{\alpha}{1-\alpha}(n+1)+(n_0-k)\right) \tag{5.65}$$

remains valid for all $k = 1, ..., n_0$, since $P(V(\tau) = n_0) = P(M(\tau) = n_0)$. The fact that Dirac-uniform configurations are least favorable configurations for the k-FWER and for the ENFR as well, completes the proof. ■

Remark 5.29

Note that if $U_1, ..., U_{n_0}$ are sMD (cf. Definition 2.16 and Remark 2.17) with $P(U_i = 0) = 0$ for all $i \in \{1, ..., n_0\}$ then we have an inequality ($\leqslant$) in (5.62).

5.6 Comparison of the SD procedure 5.1 with BH-SU procedure under total and block-dependence

In Chapter 6 we will study martingale structures and consider some simulated examples for the FDR of different stepwise procedures under the assumption that the true p-values belong to the class M_{I_0}. In this section however we compare at first the FDR of the adaptive SD procedure (5.1) with the FDR of the Benjamini and Hochberg linear SU procedure. It is clear that the BH-procedure is a step-up procedure whereas the procedure (5.1) belongs to the class of step down procedures. But it is interesting from a theoretical point of view to compare the procedures which control the FDR under the same assumptions. This is why we are now considering only the dependence structures, for which it can be easily proved that the FDR of both aforementioned procedures is controlled at α. We begin with the consideration of the following Dirac-Martingale situation (DM1)-(DM2). Let $\alpha \in (0,1)$ be given and assume that $n_0 \leqslant n$ is some fixed positive number in this section.

(DM1) The true p-values are totally dependent and uniformly distributed, that is, $U_1 = ... = U_{n_0} = U \sim$U$(0,1)$-distributed.

(DM2) The false p-values follow a Dirac $\delta(0)$ distribution with point mass 1 at 0.

It is clear that if the p-values fulfill (DM1)-(DM2), then the process $(M(t))_{t\in(0,1)}$ is an $\mathcal{F}_t-$martingale as well as the process $(\frac{V(t)}{t})_{t\in(0,1)}$ is a $\mathcal{RF}_t-$reverse martingale, where $\mathcal{RF}_t = \sigma(\mathbb{I}(U_i > s),\ \text{for all}\ t \leqslant s \leqslant 1)$ is the reverse filtration. Moreover, the process $\tilde{U}(t) = \left(\frac{R(t)-nt}{1-t}\right)_{t\in(0,1)}$ is also a $\mathcal{F}_t-$martingale with $\mathbb{E}\left[\tilde{U}(t)\right] = n_1$.

Since under (DM1)-(DM2) the process $(\frac{V(t)}{t})_{t\in(0,1)}$ is a reverse $\mathcal{RF}_t$ martingale and the stopping-rule of the Benjamini-Hochberg SU procedure τ_{BH} given by

$$\tau_{BH} = \sup\{t \in (0,1) : \frac{t}{\hat{F}_n(t)} \leqslant \alpha\},$$

is a reverse stopping-time w.r.t. $\mathcal{RF}_t$, we get for the FDR of the BH-SU procedure (cf. Heesen, Janssen [2014])

$$\mathbb{E}\left[\frac{V(\tau_{BH})}{R(\tau_{BH})}\right] = \mathbb{E}\left[\frac{V(\tau_{BH})}{\tau_{BH}}\frac{\tau_{BH}}{R(\tau_{BH})}\right]$$
$$= \frac{\alpha}{n}\mathbb{E}\left[\frac{V(\tau_{BH})}{\tau_{BH}}\right] = \frac{n_0}{n}\alpha.$$

Now we compute the FDR of the SD procedure (5.1).

Lemma 5.30
Suppose that (DM1) and (DM2) are fulfilled. Then the FDR of the SD procedure with critical values $0 \leqslant c_1 \leqslant ... \leqslant c_n \leqslant 1$ satisfies

$$\mathbb{E}\left[\frac{V(\tau_c)}{R(\tau_c)}\right] = \frac{n_0}{n}c_{n_1+1}, \tag{5.66}$$

where τ_c is the stopping rule of the SD procedure with critical values $(c_1, ..., c_n)$.

Proof. Obviously τ_c can only take two values, either c_{n_1} or c_n. Then $\mathbb{I}(\tau_c = c_{n_1}) = \mathbb{I}(U > c_{n_1+1})$ and $\mathbb{I}(\tau_c = c_n) = \mathbb{I}(U \leqslant c_{n_1+1})$ hold. Consequently, we get

$$\mathbb{E}\left[\frac{V(\tau_c)}{R(\tau_c)}\right] = E\left[\frac{n_0}{n}\mathbb{I}(U \leqslant c_{n_1+1})\right] = \frac{n_0}{n}c_{n_1+1}.$$

■

Figure (5.1) shows the values of the FDR for the SD procedure with critical values (5.1) and BH SU procedure for different values of α and n_0, where n_0 corresponds to the $x-$axis. For the first picture we have chosen $n = 200$ and for the second one $n = 500$. We can see that FDR of the procedure (5.1) exhausts maximally half of the level α and becomes maximal, when the number of the true nulls is about a quarter of the number of all hypotheses and then it decreases as n_0 increases to n, whereas FDR of the BH SU procedure is equal to α if $n_0 = n$.

Next we consider the case of total dependence in two blocks. Let n_0 be a fixed even number here. We suppose that the following assumption (DM3), that is,

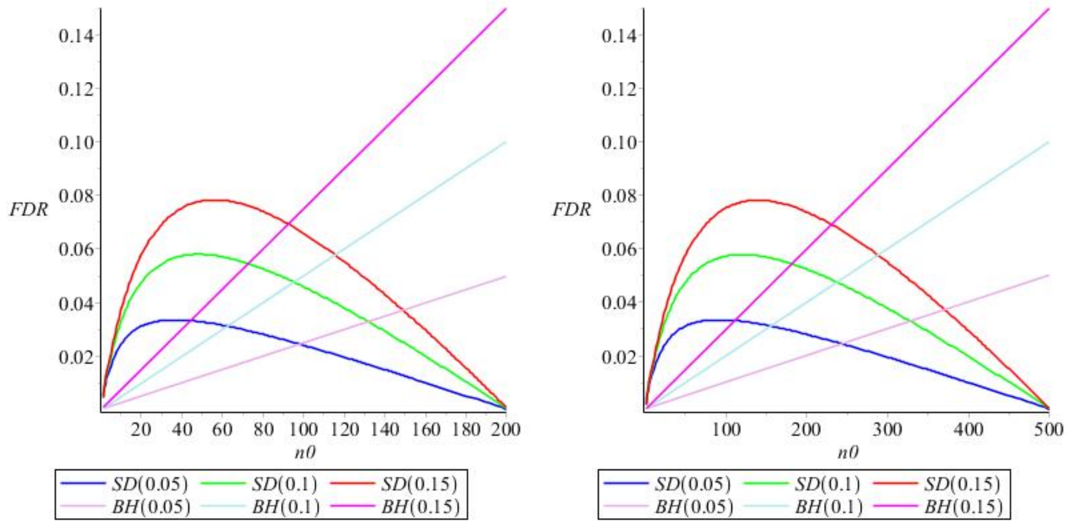

Figure 5.1: Comparison of the FDR for the SD procedure (5.1) and the BH-SU procedure. Here SD(α) corresponds to the FDR of the adaptive SD procedure (5.1) and BH(α) to the FDR of the BH SU procedure.

(DM3) U, V are independent uniformly U$(0,1)$-distributed random variables, $U_1 = U_2 = \ldots = U_l = U \wedge V$, $U_{l+1} = \ldots = U_{n_0} = U \vee V$, $l = \frac{n_0}{2}$,

and the assumption (DM2) is fulfilled. In this case we can compute the FDR of both procedures analytically. The FDR of the BH SU procedure is still $\alpha\frac{n_0}{n}$, since (DM2)-(DM3) model obviously ensures that the process $\left(\frac{V(t)}{t}\right)_{t\in(0,1]}$ is a reverse $\mathcal{RF}_t-$ martingale. And for the FDR of some SD procedure we get the following result.

Lemma 5.31
Assume that (DM2) and (DM3) are fulfilled. Then for the SD procedure with critical values $c_1 < c_2 < \ldots < c_n$ and stopping rule τ_c we have

$$\mathbb{E}\left[\frac{V(\tau_c)}{R(\tau_c)}\right] = 2c_{n_1+1}\frac{l}{n_1+l}\left(1 - c_{n_1+l+1}\right) + \frac{n_0}{n}c_{n_1+1}\left(2c_{n_1+l+1} - c_{n_1+1}\right).$$

Proof. Similarly to the previous lemma we obtain

$$\mathbb{E}\left[\frac{V(\tau)}{R(\tau)}\right] = \frac{l}{n_1+l}P(U \wedge V \leqslant c_{n_1+1}, U \vee V > c_{n_1+l+1}) + \frac{n_0}{n}P(U \wedge V \leqslant c_{n_1+1}, U \vee V \leqslant c_{n_1+l+1})$$

$$= \frac{l}{n_1 + l} P(U \wedge V \leqslant c_{n_1+1}) + P(U \wedge V \leqslant c_{n_1+1}, U \vee V \leqslant c_{n_1+l+1}) \left(\frac{n_0}{n} - \frac{l}{n_1 + l} \right)$$
$$= \frac{l}{n_1 + l} \left(1 - (1 - c_{n_1+1})^2\right) + \left(c_{n_1+l+1}^2 - (c_{n_1+l+1} - c_{n_1+1})^2\right) \left(\frac{n_0}{n} - \frac{l}{n_1 + l} \right)$$
$$= \frac{l}{n_1 + l} \left(2c_{n_1+1} - c_{n_1+1}^2\right) + \left(\frac{n_0}{n} - \frac{l}{n_1 + l} \right) c_{n_1+1} \left(2c_{n_1+l+1} - c_{n_1+1}\right)$$
$$= 2c_{n_1+1} \frac{l}{n_1 + l} \left(1 - c_{n_1+l+1}\right) + \frac{n_0}{n} c_{n_1+1} \left(2c_{n_1+l+1} - c_{n_1+1}\right).$$

■

Figure 5.2 shows the behavior of the FDR of the procedures (5.1) and BH SU for fixed n and α dependent on the number $l = \frac{n_0}{2}$ of the elements in each block. It is known that the FDR of the BH procedure depends linearly on the proporton $\frac{n_0}{n}$ and is equal to $\alpha \frac{n_0}{n} = \alpha \frac{2l}{n}$.

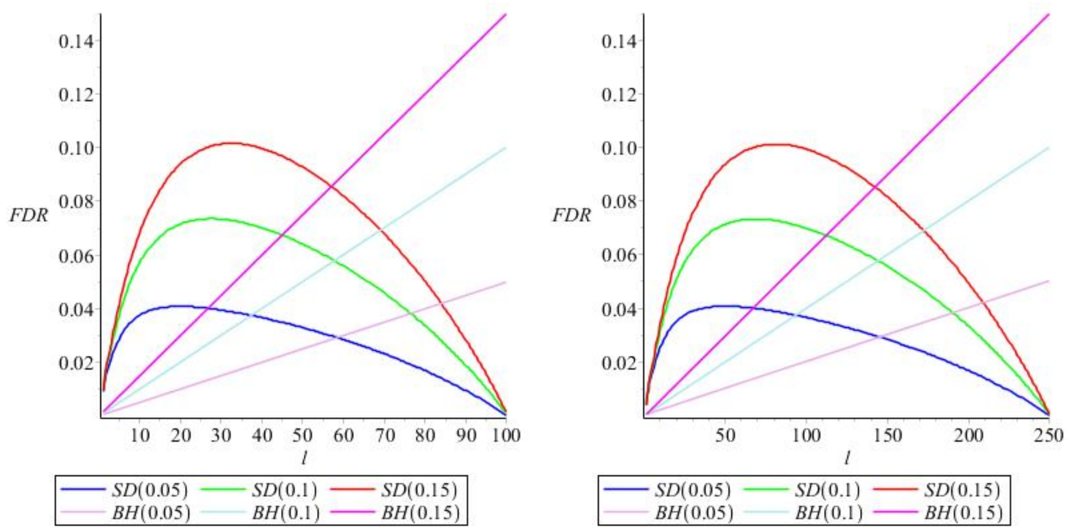

Figure 5.2: Comparison of the FDR of the SD procedure (5.1) and the BH SU procedure. Here SD(α) corresponds to the FDR of the adaptive SD procedure (5.1) and BH(α) to the FDR of the BH SU procedure. The number of the null hypotheses is chosen to be $n = 200$ in the left picture and $n = 500$ in the right one.

We can see that if the proportion of the number of true null hypotheses among all hypotheses $\frac{n_0}{n}$ is small, then the SD procedure exhausts the level α better, than the BH procedure and, if the proportion of the true nulls is large the BH procedure exhausts the level α much better for this kind of dependence.

Finally we want to compare the BH SU procedure with the SD procedure, based on the critical values (5.1) under the DU(n_0, n)-model.

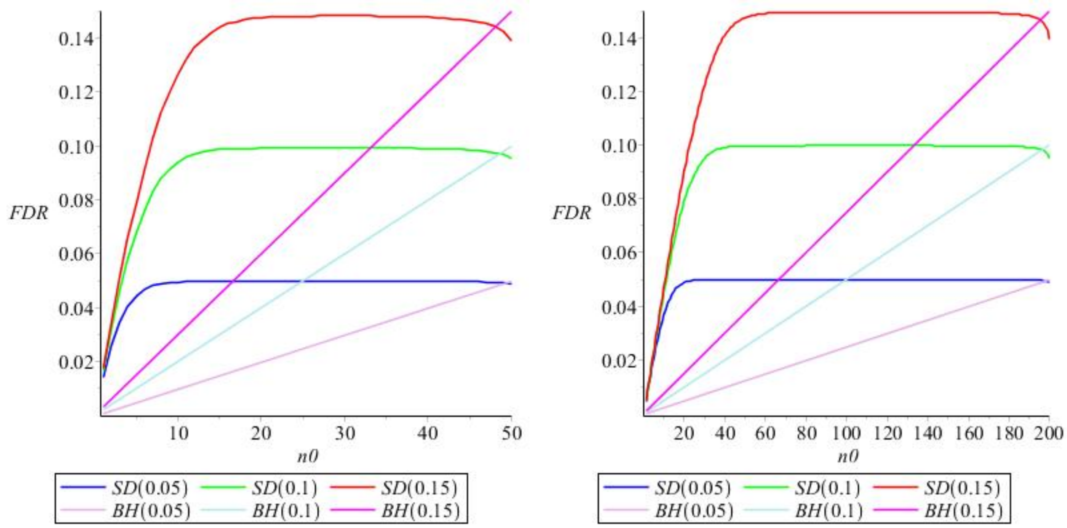

Figure 5.3: Comparison of the FDR for the SD procedure (5.1) and for the BH-SU procedure. Here SD(α) corresponds to the FDR of the adaptive SD procedure (5.1) and BH(α) to the FDR of the BH SU procedure.The number of the null hypotheses is chosen to be $n = 50$ in the first picture and $n = 200$ in the second one.

As we can see from Figure 5.3, the FDR of the SD procedure (5.1) exceeds the FDR of the BH SU procedure for all $n_0 < n$. If all hypotheses are true, i.e., for the case $n_0 = n$, the FDR of the linear SU is slightly larger, than the FDR of the procedure (5.1). This disadvantage can be eliminated, if we increase the first critical value $a_1 = \frac{\alpha}{n+\alpha}$ of the procedure (5.1) (due to Theorem 4.3) to the value $b_1 = 1 - \sqrt[n]{1-\alpha}$ (cf. Figure 5.4). In this way it will be achieved that the FDR of the SD procedure (5.1), as well as the FDR of the BH SU procedure, is equal to α, when $n_0 = n$ holds.

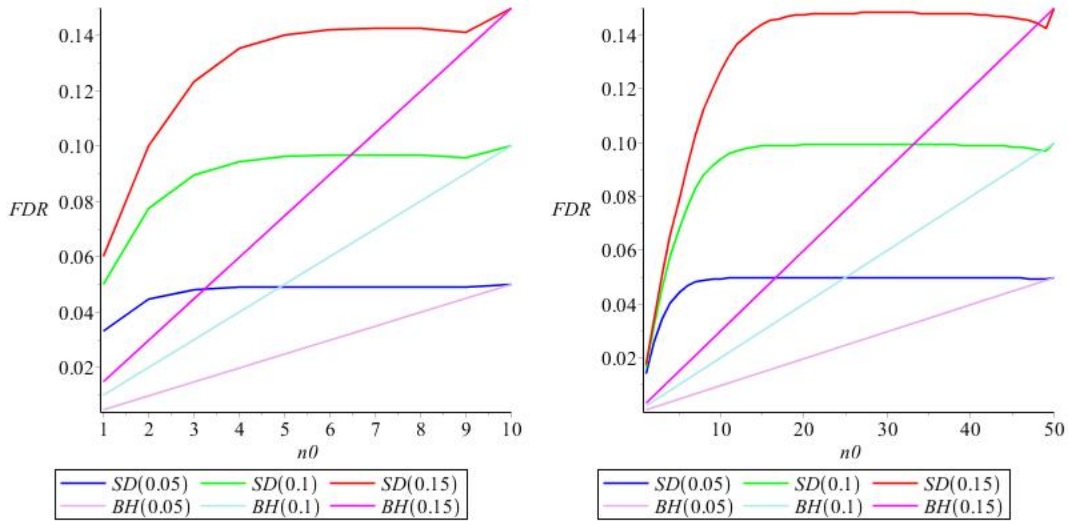

Figure 5.4: Comparison of the FDR for the SD procedure with critical values(5.1) with increased first critical values by Lemma 4.3 and for the BH-SU procedure. Here SD(α) corresponds to the FDR of the adaptive SD procedure (5.1) and BH(α) to the FDR of the BH SU procedure.The number of the null hypotheses is chosen to be $n = 10$ in the first picture and $n = 50$ in the second one.

As we can see from figures 5.2 - 5.4 the FDR of the SD procedure with critical values (5.1) (ASDP) is better exhausted in comparison with the BH SU procedure under BIA. On the other hand the BH SU test is much more robust under total- and total block-dependence.

5.7 What happens under positive association?

In this section we propose the proof of the control of the FDR-related values (see Theorem 5.36). The next three lemmas are technical tools for the proof of Theorem 5.36.

Lemma 5.32

If a.s. positive random variables $U_1, ..., U_{n_0}$ are (s)MD and positively associated (cf. Definition 2.3) conditioned on f, then for any number $k \in \mathbb{R},\ k > 0$ and any coordinate-wise decreasing stopping-time $\sigma = \sigma(U)$, with $0 \leqslant \sigma \leqslant a$, where $a \in [0, 1)$ is a positive real number, the following mathematical expectation is not positive:

$$\mathbb{E}_f \left[\frac{M(\sigma(U))}{S(\sigma(U)) + k} \right] \leqslant 0. \tag{5.67}$$

Proof. Firstly, let us define

$$\tilde{n}_1 = \sum_{i=1}^{n_1} \mathbb{I}_{\{0\}}(f_i). \tag{5.68}$$

Remember that the "false" p-values are assumed to be fixed and ordered, i.e., $f_1 \leqslant f_2 \leqslant ... \leqslant f_{n_1}$. Then we get $S(t) + k = \tilde{S}(t) + (\tilde{n}_1 + k) = \sum_{i=\tilde{n}_1+1}^{n_1} \mathbb{I}_{[0,t]}(f_i) + (\tilde{n}_1 + k)$, with $\tilde{S}(t) = \sum_{i=\tilde{n}_1+1}^{n_1} \mathbb{I}_{[0,t]}(f_i)$, which implies the following identity

$$\frac{1}{S(t)+k} = \sum_{i=\tilde{n}_1+1}^{n_1} b_i \mathbb{I}_{[0,f_i)}(t) + b_{\tilde{n}_1}. \tag{5.69}$$

Here the non-negative deterministic coefficients b_i and $b_{\tilde{n}_1}$ can be computed recursively (see Appendix for more details). From (5.69) we have

$$\frac{M(\sigma(U))}{S(\sigma(U))+k} = \sum_{i=\tilde{n}_1+1}^{n_1} b_i \mathbb{I}_{[0,f_i)}(\sigma(U)) M(\sigma(U)) + b_{\tilde{n}_1} M(\sigma(U)). \tag{5.70}$$

This yields

$$\mathbb{E}_f\left[\frac{M(\sigma(U))}{S(\sigma(U))+k}\right] = \sum_{i=\tilde{n}_1+1}^{n_1} b_i \mathbb{E}_f\left[\mathbb{I}_{[0,f_i)}(\sigma(U)) M(\sigma(U))\right] + b_{\tilde{n}_1} \mathbb{E}_f\left[M(\sigma(U))\right] \tag{5.71}$$

$$\leqslant \sum_{i=\tilde{n}_1+1}^{n_1} b_i \mathbb{E}_f\left[\mathbb{E}\left[\mathbb{I}_{[0,f_i)}(\sigma(U)) M(\sigma(U)) | \mathcal{F}_{f_i}\right]\right] \leqslant \sum_{i=\tilde{n}_1+1}^{n_1} b_i \mathbb{E}_f\left[\mathbb{I}_{[0,f_i)}(\sigma(U)) M(\sigma(U) \wedge f_i)\right] \tag{5.72}$$

$$= \sum_{i=\tilde{n}_1+1}^{n_1} b_i \mathbb{E}_f\left[(1 - \mathbb{I}_{[f_i,1]}(\sigma(U))) M(\sigma(U) \wedge f_i)\right] \leqslant - \sum_{i=\tilde{n}_1+1}^{n_1} b_i \mathbb{E}_f\left[\mathbb{I}_{[f_i,1]}(\sigma(U)) M(f_i)\right] \tag{5.73}$$

$$= - \sum_{j=1}^{n_0} \sum_{i=\tilde{n}_1+1}^{n_1} b_i \mathbb{E}_f\left[\mathbb{I}_{[f_i,1]}(\sigma(U)) \left(\frac{\mathbb{I}(U_j \leqslant f_i) - f_i}{1 - f_i}\right)\right], \tag{5.74}$$

where the (in)equalities (5.72) and (5.73) follow from the Optional Sampling Theorem.

Note that if the random variables U_i $i \in \{1, ...n_0\}$ are positively associated, then we have from (5.74)

$$\mathbb{E}_f\left[\frac{M(\sigma(U))}{S(\sigma(U))+k}\right] \leqslant - \sum_{j=1}^{n_0} \sum_{i=\tilde{n}_1+1}^{n_1} b_i \text{Cov}(\mathbb{I}_{(f_i,1]}(\sigma(U)), m_j(f_i)) \leqslant 0, \tag{5.75}$$

since both of the functions $U_i \mapsto m_j(f_i)$ and $U_i \mapsto \mathbb{I}_{(f_k,1]}(\sigma(U))$ are non-increasing for each $i, j = 1, ..., n_0$ and $k = 1, ..., n_1$. The proof of this Lemma is completed. ■

Lemma 5.33

The statement of Lemma 5.32 remains true, if a.s. positive random variables $U_1, ..., U_{n_0}$ are (s)MD and PRDS.

Proof. The proof follows directly from (5.75) and Corollary 2.15. ■

Now we formulate a generalization of Lemma 5.32.

Lemma 5.34

Let the following assumptions be fulfilled.

(a) $U_1, ..., U_{n_0} \in M(U, n_0)$ *are positively associated, conditioned on* f*, and stochastically independent from* $f_1, ..., f_{n_0}$ *with* $P(U_i = 0) = 0$ *for all* $i \in \{1, ..., n_0\}$*,*

(b) $f = f(\sigma, \bar{f}) : [0, 1] \times [0, 1]^{n_1} \to (0, \infty)$ *is some positive and increasing function,*

(c) $\sigma = \sigma(U, f)$ *is an* $\mathcal{F}_{d_i}-$*stopping time, with finite range* $\{d_1, ..., d_m\}$*,* $m \in \mathbb{N}$*,* $d_1 \leqslant ... \leqslant d_m$*,*

(d) $U_i \mapsto \sigma(U, f)$ *is decreasing for all* $i = 1, ..., n_0$*.*

Then we have

$$\mathbb{E}_f \left[\frac{M(\sigma)}{f(\sigma)} \right] \leqslant 0. \tag{5.76}$$

Proof. We have

$$\mathbb{E}_f \left[\frac{M(\sigma)}{f(\sigma)} \right] = \mathbb{E}_f \left[\sum_{i=1}^{m} \frac{M(d_i)}{f(d_i)} \mathbb{I}(\sigma = d_i) \right] \tag{5.77}$$

$$= \mathbb{E}_f \left[\sum_{i=1}^{m} \frac{M(d_i)}{f(d_i)} \left(\mathbb{I}(\sigma \leqslant d_i) - \mathbb{I}(\sigma \leqslant d_{i-1}) \right) \right] \tag{5.78}$$

$$= \sum_{i=1}^{m} \left(\mathbb{E}_f \left[\frac{M(d_i)}{f(d_i)} \mathbb{I}(\sigma \leqslant d_i) \right] - \mathbb{E}_f \left[\frac{M(d_i)}{f(d_i)} \mathbb{I}(\sigma \leqslant d_{i-1}) \right] \right) \tag{5.79}$$

$$= \sum_{i=1}^{m} \left(\mathbb{E}_f \left[\frac{M(d_i)}{f(d_i)} \mathbb{I}(\sigma \leqslant d_i) \right] - \mathbb{E}_f \left[\mathbb{E} \left[\frac{M(d_i)}{f(d_i)} \mathbb{I}(\sigma \leqslant d_{i-1}) | \mathcal{F}_{d_{i-1}} \right] \right] \right) \tag{5.80}$$

$$= \sum_{i=1}^{m} \left(\mathbb{E}_f \left[\frac{M(d_i)}{f(d_i)} \mathbb{I}(\sigma \leqslant d_i) \right] - \mathbb{E}_f \left[\frac{M(d_{i-1})}{f(d_i)} \mathbb{I}(\sigma \leqslant d_{i-1}) \right] \right) \tag{5.81}$$

$$\leqslant \sum_{i=1}^{m} \left(\frac{\mathbb{E}_f \left[M(d_i) \mathbb{I}(\sigma \leqslant d_i) \right]}{f(d_i)} - \frac{\mathbb{E}_f \left[M(d_{i-1}) \mathbb{I}(\sigma \leqslant d_{i-1}) \right]}{f(d_{i-1})} \right) \tag{5.82}$$

$$= \frac{\mathbb{E}_f \left[M(d_m) \mathbb{I}(\sigma \leqslant d_m) \right]}{f(d_m)} = 0. \tag{5.83}$$

The above relations hold due to the following reasons: equality (5.81) is valid, because σ is a stopping time, and hence, $\mathbb{I}(\sigma \leqslant d_{i-1})$ is $\mathcal{F}_{d_{i-1}}$-measurable. Because of the positive association, conditioned on f, we have $\mathbb{E}_f \left[M(d_{i-1}) \mathbb{I}(\sigma \leqslant d_{i-1}) \right] \leqslant 0$ and, due to assumption (b), we have inequality in (5.82). Finally, the first equality in (5.83) is valid by the telescopic sum argument. The application of the Optional Sampling Theorem completes the proof, since $\sigma \leqslant d_m$ a.s. ■

Remark 5.35

The assumption (a) of Lemma 5.34 can be replaced by the requirment that $U_1, ..., U_{n_0}$ are PRDS. Because of Corollary 2.15 we get

$$\mathbb{E}_f \left[M(d_j) \mathbb{I}(\sigma \leqslant d_i) \right] = Cov_f(M(d_j), \mathbb{I}(\sigma \leqslant d_i)) \leqslant 0 \tag{5.84}$$

under assumption (d) of Lemma 5.34. Hence, the chain of (in)equalities (5.77)-(5.83) remains true.

As a consequence of Lemma 5.32 we propose the next theorem. It yields some functionals which can be controlled by the procedure 5.1 under the martingale dependence and positive association.

Theorem 5.36

If a.s. positive random variables $U_1, ..., U_{n_0} \in M(U, n_0)$ are positively associated and stochastically independent from $f_1, ..., f_{n_0}$, then for the SD procedure with critical values (5.1) the following bounds are valid:

(a)

$$\mathbb{E}_f \left[\frac{V(\tau_1)}{S(\tau_1) + 1} \right] \leqslant \frac{\alpha}{1 - \alpha},$$

(b)

$$\mathbb{E}_f\left[\frac{V(\tau_1)}{R(\tau_1)+1}\right] \leqslant \alpha,$$

(c) let $n_1 \geqslant 1$ *then*

$$\mathbb{E}_{f_0}\left[\frac{V(\tau_1)-\alpha}{R(\tau_1)}\right] \leqslant \alpha,$$

where f_0 *is given by* $f_0 = (0, f_2, ..., f_{n_1})$, $0 \leqslant f_2 \leqslant f_3 \leqslant ... \leqslant f_{n_1} \leqslant 1$.

Proof. Firstly, remember that $R(\tau_1) = R(\tau)$ and $V(\tau_1) = V(\tau)$ hold (cf. Lemma 5.8. 4). Similarly to (5.13) we have:

$$(1-\alpha)R(\tau) \leqslant M(\tau) + S(\tau) + \alpha, \tag{5.85}$$

which is equivalent to

$$(1-\alpha)V(\tau) \leqslant M(\tau) + \alpha(S(\tau)+1). \tag{5.86}$$

Dividing (5.86) by $S(\tau)+1$ yields

$$(1-\alpha)\frac{V(\tau)}{S(\tau)+1} \leqslant \frac{M(\tau)}{S(\tau)+1} + \alpha. \tag{5.87}$$

Taking $\mathbb{E}_f$ in (5.87) and applying Lemma 5.32 for $k=1$ proves part (a) of the Theorem. Based on part (a) and due to Jensen's inequality we obtain

$$\begin{aligned}\mathbb{E}_f\left[\frac{V(\tau)}{R(\tau)+1}\right] &= \mathbb{E}_f\left[\frac{\frac{V(\tau)}{S(\tau)+1}}{\frac{V(\tau)}{S(\tau)+1}+1}\right] \\ &\leqslant \frac{\mathbb{E}_f\left[\frac{V(\tau)}{S(\tau)+1}\right]}{\mathbb{E}_f\left[\frac{V(\tau)}{S(\tau)+1}\right]+1} \leqslant \alpha,\end{aligned}$$

which proves part (b).

Now we prove part (c). Firstly note, that the function $x \mapsto \frac{1}{1+x}$ is convex on $x \in (-1, \infty)$. From (5.85) we get

$$\frac{M(\tau)}{S(\tau)+\alpha} \geqslant (1-\alpha)\frac{R(\tau)}{S(\tau)+\alpha} - 1 > -1 \quad \text{on} \quad \{f_1 = 0\}.$$

Hence, from (5.86) and the Jensen's inequality we obtain

$$\begin{aligned}\mathbb{E}_{f_0}\left[\frac{V(\tau)-\alpha}{R(\tau)}\right] &= 1-\mathbb{E}_{f_0}\left[\frac{S(\tau)+\alpha}{R(\tau)}\right]\\ &\leqslant 1-(1-\alpha)\mathbb{E}_{f_0}\left[\frac{S(\tau)+\alpha}{M(\tau)+S(\tau)+\alpha}\right]\\ &\leqslant 1-(1-\alpha)\frac{1}{\mathbb{E}_{f_0}\left[\frac{M(\tau)}{S(\tau)+\alpha}\right]+1}.\end{aligned}$$

Finally, Lemma 5.32 with $k=\alpha$ provides the desired inequality. ■

5.8 Conclusions

In this chapter we considered a well-known SD procedure (5.1) (ASDP) with critical values $a_{i:n}=\frac{i\alpha}{n+1-i(1-\alpha)}$. This procedure was studied by Finner et al. [2009] in context with the $\beta-$adjusted asymptotically optimal rejection curve. Two years later Gavrilov et al. [2010] proved that this procedure controls the FDR if the underlying p-values are independent. Scheer [2012] showed that the procedure (5.1) controls the ENFR linearly. We proved the ENFR-control under new dependence assumptions - under the martingale dependence. We also proposed a new martingale based proof for FDR-control of the procedure (5.1) under the BIA assumptions. We generalized the result of Gavrilov et al. [2010] by the proof of the FDR-control under PRDS and (s)MD. Moreover, we proposed a class of SD procedures $\tau_\beta,\ \beta\in(0,1]$, that control ENFR and FDR (or FDR-related values) under martingale dependence and PRDS.

Chapter 6

(Super-)Martingale Structures

In this chapter we study structures, for which the process

$$M(t) = \frac{\sum_{i=1}^{n_0} \mathbb{I}(U_i \leqslant t) - n_0 t}{1-t} = \sum_{i=1}^{n_0} \frac{\mathbb{I}(U_i \leqslant t) - t}{1-t} = \sum_{i=1}^{n_0} m_i(t) \tag{6.1}$$

is a (super-)martingale w.r.t. the filtration $\mathcal{F}_t = \sigma(\mathbb{I}_{(U_i \leqslant s)}, 0 \leqslant s \leqslant t, \forall i = 1, ..., n_0), t \in [0,1)$.

Remember that the vector of p-values is denoted by $p = (p_1, ..., p_n)$, $U = (p_i, i \in I_0) = (U_1, ..., U_{n_0})$ is the vector of true p-values, $f = (p_i, i \in I_1) = (f_1, ..., f_{n_1})$, is the vector of the false ones, thereby we assume that $f_1 \leqslant f_1 \leqslant ... \leqslant f_{n_1}$ holds. The number of true nulls $n_0 \leqslant n$ is assumed to be an arbitrary fixed natural number in this chapter.

The following Lemma can be found in Shorack and Wellner [1986], p.133.

Lemma 6.1

If $U_1, ..., U_{n_0}$ are i.i.d uniformly distributed on $(0,1)$, then $\{M(t), t \in [0,1)\}$ is a $\mathcal{F}_t$-martingale.

Proof. Since U_i, $i = 1, ..., n_0$, are i.i.d random variables it is sufficient to prove that for all $i = 1, ..., n_0$, $m_i(t)$ is a $\mathcal{F}_t^i$-martingale with $\mathcal{F}_t^i = \sigma(\mathbb{I}_{(U_i \leqslant s)}, 0 \leqslant s \leqslant t), t \in [0,1)$. Obviously, $m_i(t)$ is adapted on the filtration $\mathcal{F}_t^i$ and we have for every fixed $t \in [0,1)$

$$\mathbb{E}\left[\mid m_i(t) \mid\right] \leqslant \frac{\mathbb{E}\left[\mathbb{I}(U_i \leqslant t)\right] + t}{1-t} = \frac{2t}{1-t} < \infty. \tag{6.2}$$

It remains to prove the martingale property $\mathbb{E}\left[m_i(t)|\mathcal{F}_s^i\right] = m_i(s)$, for all $s \leqslant t$. For all fixed $t \in [0,1)$ and for all $s \leqslant t$ we have:

$$\mathbb{E}\left[\frac{\mathbb{I}(U_i \leqslant t) - t}{1-t}|\mathcal{F}_s^i\right] =$$

$$\frac{\mathbb{E}\left[\mathbb{I}(U_i \leqslant t)|U_i \leqslant s\right]\mathbb{I}(U_i \leqslant s) + \mathbb{E}\left[\mathbb{I}(U_i \leqslant t)|U_i > s\right]\mathbb{I}(U_i > s) - t}{1-t} =$$

$$\frac{\mathbb{I}(U_i \leqslant s) + \mathbb{I}(U_i > s)\frac{t-s}{1-s} - t}{1-t} = \frac{\mathbb{I}(U_i \leqslant s) - s}{1-s} = m_i(s).$$

■

Remark 6.2

The process $\{\frac{V(t)}{t}, t \in (0,1]\}$ *is a reverse martingale w.r.t. the reverse Filtration* $\mathcal{RF}_t = \sigma(\mathbb{I}_{(s,1)}(U_i), \forall 0 < t \leqslant s \leqslant 1, i = 1, ..., n_0)$ *under the assumptions of Lemma 6.1. This statement is proved in Shorack and Wellner [1986], p.136. The reversal* $t \mapsto 1-t$ *yields the statement of Lemma 6.1.*

Lemma 6.3

If $U_1, ..., U_{n_0}$ *are i.i.d. random variables and the distribution function* F *of* $U_i, i = 1, ..., n_0$, *is convex on* $[0,1]$, *then* $\{M(t), t \in [0,1)\}$ *is an* $\mathcal{F}_t$*-supermartingale.*

Proof. By the convexity of F we have

$$P(U_i \leqslant t|U_i > s) = \frac{F(t) - F(s)}{1 - F(s)} \leqslant \frac{t-s}{1-s}, \ s \leqslant t,$$

which implies that

$$\mathbb{E}\left[m_i(t)|\mathcal{F}_s^i\right] = \frac{\mathbb{I}(U_i \leqslant s) + \mathbb{I}(U_i > s)P(U_i \leqslant t|U_i > s) - t}{1-t} \leqslant m_i(s)$$

holds, and consequently $\{M(t), t \in [0,1)\}$ is a super-martingale. ■

Remark 6.4

Note that the assumption from Lemma 6.3 is not necessary, but sufficient. We can formulate the neccesary and sufficient assumption as follows.

For the distribution function F *of i.i.d. random variables* $U_1, ..., U_{n_0}$ *the following inequality holds*

$$\frac{F(t) - F(s)}{1 - F(s)} \leqslant \frac{t-s}{1-s}, \ 0 \leqslant s \leqslant t,$$

$$iff$$

$\{M(t), t \in [0,1)\}$ *is a* $\mathcal{F}_t$*−super-martingale.*

6.1 Sufficient conditions for MD

Theorem 6.5

If the two following conditions are fulfilled, then the process $\{M(t), t \in [0,1)\}$ from (6.1) is a supermartingale w.r.t. the filtration $\mathcal{F}_t$.

1. **(Markov Property)** *The elementary Markov property holds, i.e., for all $i = 1, ..., n_0$, for all $m \in \mathbb{N}$, for all $0 < s_1 < ... < s_m \leqslant s \leqslant t$, for all $l_j^1, ...l_j^m \in \{0,1\}$, for which the intersection $\bigcap_{j=1}^{n_0} \bigcap_{k=1}^{m} \{\mathbb{I}_{(0,s_k]}(U_j) = l_j^k\}$ is not empty, we get*
$$P\left(U_i \leqslant t \mid \bigcap_{j=1}^{n_0} \bigcap_{k=1}^{m} \{\mathbb{I}_{(0,s_k]}(U_j) = l_j^k\}\right) = P\left(U_i \leqslant t \mid \bigcap_{j=1}^{n_0} \{\mathbb{I}_{(0,s_m]}(U_j) = l_j^m\}\right).$$
2. **((Super-)Martingale Property)** *For all $i \in \{1, ..., n_0\}$, for all $l_j \in \{0,1\}, j = 1, ..., n_0$, with $j \neq i$, $s \leqslant t$ either the following inequality (a) or (b) holds:*

   (a)
$$\mathbb{E}\left[\mathbb{I}(U_i \leqslant t) \mid (U_i > s) \cap \bigcap_{\substack{j=1 \\ j \neq i}}^{n_0} \{\mathbb{I}_{(0,s]}(U_j) = l_j\}\right] \leqslant \frac{t-s}{1-s}, \tag{6.3}$$

   (b)
$$\mathbb{E}\left[\sum_{i=1}^{n_0} \mathbb{I}(U_i \leqslant t) \mid (U_i > s) \cap \bigcap_{\substack{j=1 \\ j \neq i}}^{n_0} \{\mathbb{I}_{(0,s]}(U_j) = l_j\}\right] \leqslant n_0 \frac{t-s}{1-s}. \tag{6.4}$$

Remark 6.6

The integrability assumption $\mathbb{E}[|M(t)|] < \infty$ is always fulfilled, because obviously the inequality
$$-n_0 \frac{t}{1-t} \leqslant M(t) \leqslant n_0$$
is valid for all $t \in [0,1)$ and fixed $n_0 \in \mathbb{N}$.

Proof. In order to verify that $\{M(t), t \in [0,1)\}$ is a $\mathcal{F}_t$-(super-)martingale, we have to show that the (super-)martingale property $\mathbb{E}[M(t) \mid \mathcal{F}_s] \leqslant M(s)$ for all $s \leqslant t$ holds. The integrability of $M(t)$ is fulfilled (see Remark 6.6). By assumption 1 we have
$$\mathbb{E}[M(t) \mid \mathcal{F}_s] = \mathbb{E}\left[M(t) \mid \mathbb{I}_{(0,s]}(U_1), ..., \mathbb{I}_{(0,s]}(U_{n_0})\right]. \tag{6.5}$$

Further, by assumption 2 (a) we get

$$\begin{aligned}
&\mathbb{E}\left[\mathbb{I}_{[0,t]}(U_i) \mid \mathbb{I}_{[0,s]}(U_1), ..., \mathbb{I}_{[0,s]}(U_{n_0})\right] = \mathbb{I}(U_i \leqslant s) \\
&+ \mathbb{I}(U_i > s) \sum_{\substack{l_j \in \{0,1\} \\ j=1,...,n_0, j\neq i}} \left(\prod_{\substack{j=1 \\ j\neq i}}^{n_0} \mathbb{I}(\mathbb{I}_{(0,s]}(U_j) = l_j) \mathbb{E}\left[\mathbb{I}_{[0,t]}(U_i) \mid \{U_i > s\} \cap \bigcap_{\substack{j=1 \\ j\neq i}}^{n_0} \{\mathbb{I}_{(0,s]}(U_j) = l_j\} \right]\right) \\
&\leqslant \mathbb{I}(U_i \leqslant s) + \mathbb{I}(U_i > s) \sum_{\substack{l_j \in \{0,1\} \\ j=1,...,n_0, j\neq i}} \left(\prod_{\substack{j=1 \\ j\neq i}}^{n_0} \mathbb{I}(\mathbb{I}_{(0,s]}(U_j) = l_j) \right) \frac{t-s}{1-s},
\end{aligned} \tag{6.6}$$

thereby the summation is over all vectors $(l_1, ..., l_{i-1}, l_{i+1}, ..., l_{n_0})$ from $\{0,1\}^{n_0-1}$.
Since

$$\sum_{\substack{l_j \in \{0,1\} \\ j=1,...,n_0, j\neq i}} \prod_{\substack{j=1 \\ j\neq i}}^{n_0} \mathbb{I}(\mathbb{I}_{(0,s]}(U_j) = l_j) = 1$$

holds, we have

$$\mathbb{E}\left[\mathbb{I}_{[0,t]}(U_i) \mid \mathbb{I}_{[0,s]}(U_1), ..., \mathbb{I}_{[0,s]}(U_{n_0})\right] \leqslant \mathbb{I}(U_i \leqslant s) + \mathbb{I}(U_i > s)\frac{t-s}{1-s},$$

which implies that $\{M(t), t \in [0,1)\}$ is an $\mathcal{F}_t-$super-martingale.
The case of assumption 2 (b) can be proved in the same way. ■

Remark 6.7

Note that the following structures obviously fulfill the assumptions of Theorem 6.5.

(i) Total dependence between true p-values, i.e., $U_1 = ... = U_{n_0} = U$ *and* U *has a convex distribution function. If the random variable* U *is uniformly* $U(0,1)$*-distributed, then the total dependence between true p-values belongs to the class of martingale dependence.*

(ii) Block-total dependence between true p-values, that is, a partition $I_0 = \sum_{k=1}^{d} J_k$ *of pairwise disjoint sets* J_k *of the index set* I_0 *of true p-values exist such that* $U_i = U_j$ *whenever* $\{i,j\} \subset J_k$ *holds and the vectors*

$$(U_{i_1})_{i_1 \in J_1},\ (U_{i_2})_{i_2 \in J_2}, ..., (U_{i_d})_{i_d \in J_d}$$

are independent.

(iii) *The set of null p-values satisfies the Joint Null Criterion (cf. Leek and Storey [2011]), that is, the joint distribution of the order statistics $U_{1:n_0}, ..., U_{n_0:n_0}$ is equal to the joint distribution of $U^*_{1:n_0}, ..., U^*_{n_0:n_0}$, where $U^*_1, ...U^*_{n_0}$ are i.i.d $U(0,1)$-distributed random variables.*

Let for some interval $T \subset [0, \infty)$, $(\Omega, \mathcal{G}, (\mathcal{G}_t)_{t \in T}, P)$ be some filtered probability space which satisfies the "usual conditions", i.e.,

- the probability space $(\Omega, \mathcal{G}, P)$ is complete, i.e., for all the null-sets $N \in \mathcal{G}$, with $P(N) = 0$ and all $\tilde{N} \subset N$ we have $\tilde{N} \in \mathcal{G}$,
- the filtration $\mathcal{G}_t$ is complete, i.e., every σ-algebra $\mathcal{G}_t$ contains all the null-sets in $\mathcal{G}$,
- the filtration $(\mathcal{G}_t)_{t \in T}$ is is right-continuous, i.e., for every (non-maximal) $t \in T$ the $\sigma-$algebra $\mathcal{G}_{t+} = \cap_{s>t} \mathcal{G}_s$ is equal to $\mathcal{G}_t$.

Lemma 6.8 *1. If $\{X(t), t \in T\}$ is a real valued $\mathcal{G}_t-$super-martingale and $\phi : \mathbb{R} \to \mathbb{R}$ is a concave, increasing function, so that $\phi(X(t))$ is integrable for each $t \in [0, 1)$, then the process $Y(t) = \phi(X(t))$ is also a $\mathcal{G}_t-$super-martingale.*

2. If $X(t)$ is a real valued $\mathcal{G}_t-$martingale and $\phi : \mathbb{R} \to \mathbb{R}$ is a concave, $\phi(X_t)$ is integrable function (not necessarily increasing), then $Y(t) = \phi(X(t))$ is a $\mathcal{G}_t-$super-martingale.

Proof. The proof is a simple consequence of Jensen's inequality for conditional expectations (cf. Hoffman-Jørgensen, [1994], pp.492-493). ■

Let us denote the space of cadlag-$\mathcal{G}_t-$martingales by $\mathcal{M}$ and the space of cadlag-$\mathcal{G}_t-$ super-martingales by $\mathcal{M}s$.

Remark 6.9 *(a) The space $\mathcal{M}$ of $\mathcal{G}_t-$martingales is a real vector space.*

(b) The space $\mathcal{M}s$ of $\mathcal{G}_t-$ super martingales is closed with respect to addition, multiplication with positive constants and the finite Infimum.

Lemma 6.10 (Optional Switching)
Let $(M_1(t), \mathcal{G}_t)_{t \in (0,1)}$ and $(M_2(t), \mathcal{G}_t)_{t \in (0,1)}$ be cadlag-martingales and $\tau \in (0, 1)$ be a bounded $\mathcal{G}_t-$stopping time. If $M_1(\tau) = M_2(\tau)$ holds a.s., then the following process

$$M(t) = \begin{cases} M_1(t), t \leqslant \tau \\ M_2(t), t > \tau \end{cases} \tag{6.7}$$

is also a $\mathcal{G}_t$-martingale.

Proof. By the definition of the process $(M(t))_{t\in(0,1)}$ we have

$$\begin{aligned} M(t) &= \mathbb{I}_{t\leqslant\tau}M_1(t) + \mathbb{I}_{t>\tau}M_2(t) \\ &= M_1(t\wedge\tau)(1-\mathbb{I}_{t>\tau}) + \mathbb{I}_{t>\tau}M_2(t) \\ &= M_1(t\wedge\tau) + \mathbb{I}_{t>\tau}(M_2(t) - M_1(t\wedge\tau)) \\ &= M_1(t\wedge\tau) + \mathbb{I}_{t>\tau}(M_2(t) - M_2(t\wedge\tau)) \end{aligned}$$

For $s \leqslant t$ we can compute $\mathbb{E}\left[M(t) \mid \mathcal{G}_s\right]$. Applying the Optional Sampling Theorem brings us to:

$$\mathbb{E}\left[M(t) \mid \mathcal{G}_s\right] = M_1(s\wedge\tau) + \mathbb{I}_{\tau<s}(M_2(s) - M_2(s\wedge\tau)) + \mathbb{E}\left[\mathbb{I}_{s\leqslant\tau<t}(M_2(t) - M_2(t\wedge\tau)) \mid \mathcal{G}_s\right]. \tag{6.8}$$

Now, we observe that $\{\tau \geqslant s\} = \{\tau < s\}^c$ and consequently $\mathbb{I}_{s\leqslant\tau}$ is $\mathcal{G}_s$-measurable.

Let us consider the third term in (6.8). Firstly, we decompose $\mathbb{I}_{s\leqslant\tau<t}$ in the following way $\mathbb{I}_{s\leqslant\tau<t} = \mathbb{I}_{s\leqslant\tau} - \mathbb{I}_{\tau>t}$, further by applying the Optional Sampling Theorem we get

$$\begin{aligned} &\mathbb{E}\left[\mathbb{I}_{s\leqslant\tau<t}(M_2(t) - M_2(t\wedge\tau)) \mid \mathcal{G}_s\right] \\ &= \mathbb{E}\left[\mathbb{I}_{s\leqslant\tau}(M_2(t) - M_2(t\wedge\tau)) \mid \mathcal{G}_s\right] - \mathbb{E}\left[\mathbb{I}_{\tau>t}(M_2(t) - M_2(t\wedge\tau)) \mid \mathcal{G}_s\right] = \\ &= \mathbb{I}_{s\leqslant\tau}(M_2(s) - M_2(s\wedge\tau)) - \mathbb{E}\left[\mathbb{I}_{\tau>t}(M_2(t) - M_2(t)) \mid \mathcal{G}_s\right] = 0 \end{aligned}$$

Hence, we have shown

$$\mathbb{E}\left[M(t) \mid \mathcal{G}_s\right] = M_1(s\wedge\tau) + \mathbb{I}_{\tau<s}(M_2(s) - M_2(s\wedge\tau)) \tag{6.9}$$

$$= M_1(s\wedge\tau) + \mathbb{I}_{\tau<s}(M_2(s) - M_1(s\wedge\tau)) = M(s), \tag{6.10}$$

which implies that $(M(t), \mathcal{G}_t)$ is a martingale. ■

The next lemma shows that martingale dependence and uniform distribution are in a certain sense related concepts.

Lemma 6.11

Let random variables $X_1, ..., X_n$ be MD on $\{1, ..., n\}$ in the sense of Definition 2.16 and let $\sigma : \{1, ..., n\} \to \{1, ..., n\}$ be some random permutation of the index-set $\{1, ..., n\}$ which is independent of $X_i, i = 1, ..., n$.

(a) If

$$M(t) = \sum_{i=1}^{n} \frac{\mathbb{I}(X_i \leqslant t) - t}{1-t} \quad \textit{is} \quad \textit{a} \quad \mathcal{F}_t - \textit{martingale},$$

$$\textit{with} \quad \mathcal{F}_t = \sigma(\mathbb{I}_{[0,t]}(X_i), 0 \leqslant s \leqslant t, i = 1, ..., n),$$

and if additional $\mathbb{E}[M(t)] = 0$ *is valid, then the random variable* $Y_i = X_{\sigma(i):n}, i = 1, ..., n,$ *is* $U(0,1)$*-distributed.*

(b) If the random variables $X_1, ..., X_n$ *are sMD on* $\{1, ..., n\}$ *(i.e., that the process* $M(t)$ *is* $\mathcal{F}_t$*−super-martingale) with* $E[M(0)] \leqslant 0$*, then we have* $Y_i \succcurlyeq_{st} U(0,1), i = 1, ..., n.$

Proof. Firstly note, that since σ is an independent permutation, the random variables $Y_i, i = 1, ..., n$ are exchangeable. This implies

$$\mathbb{E}[\mathbb{I}(Y_i \leqslant t)] = \mathbb{E}\left[\frac{1}{n}\sum_{i=1}^{n} \mathbb{I}(Y_i \leqslant t)\right] \quad \text{holds.} \tag{6.11}$$

Moreover, we get

$$\mathbb{E}\left[\frac{1}{n}\sum_{i=1}^{n} \mathbb{I}(Y_i \leqslant t)\right] = \mathbb{E}\left[\frac{1}{n}\sum_{i=1}^{n} \mathbb{I}(X_i \leqslant t)\right] \tag{6.12}$$

$$= \frac{(1-t)}{n}\mathbb{E}[M(t)] + t \leqslant t. \tag{6.13}$$

Note that we have an (in)equality in (6.13), if $(M(t), \mathcal{F}_t)$ is a (super-)martingale. ■

Lemma 6.12

Let $X_1, ..., X_n$ *be MD on* $J \in \{1, ..., n\}$*, i.e., the process* $M(t) = \left(\frac{\sum_{i\in J} \mathbb{I}(X_i \leqslant t) - |J|t}{1-t}\right)_{t\in(0,1)}$ *is an* $\mathcal{F}_t$*−martingale. If additionally* $X_i, \in J$*, are exchangeable, then the processes* $m_i(t) = \left(\frac{\mathbb{I}(X_i \leqslant t) - t}{1-t}\right)_{t\in(0,1)}$ *are* $\mathcal{F}_t$*- martingales for all* $i \in J$*.*

Proof. By exchangeability of $X_1, ..., X_n$ on J we get for $0 < s \leqslant t$

$$\begin{aligned} \mathbb{E}\left[\frac{\sum_{i\in J} \mathbb{I}(X_i \leqslant t) - |J|t}{1-t} | \mathcal{F}_s\right] &= \sum_{i\in J} \mathbb{E}\left[\frac{\mathbb{I}(X_i \leqslant t) - t}{1-t} | \mathcal{F}_s\right] \\ = |J|\mathbb{E}\left[\frac{\mathbb{I}(X_i \leqslant t) - t}{1-t} | \mathcal{F}_s\right] &= |J|\mathbb{E}[m_i(t)|\mathcal{F}_s] \quad \text{a.s. for all } i \in J. \end{aligned} \tag{6.14}$$

On the other hand by the martingale dependence of $X_1, ..., X_n$ on J we have

$$\mathbb{E}\left[\frac{\sum\limits_{i\in J}\mathbb{I}(X_i \leqslant t) - |J|t}{1-t}|\mathcal{F}_s\right] = \tag{6.15}$$
$$\sum_{i\in J}\mathbb{E}\left[\frac{\mathbb{I}(X_i \leqslant s) - s}{1-s}|\mathcal{F}_s\right] = \sum_{i\in J}\mathbb{E}\left[m_s(t)\right] \text{ a.s.}$$

From (6.14) and (6.15) follows that for all $i \in J$

$$\mathbb{E}\left[m_i(t)|\mathcal{F}_s\right] = \frac{\sum\limits_{i\in J} m_i(s)}{|J|} = m_i(s) \text{ a.s.}$$

and, consequently $(m_i(t))_{t\in(0,1)}$ is an $\mathcal{F}_t-$martingale for all $i \in J$. ■

The following lemma shows a relation between the martingale dependence of the individual p-value (see RV model of Heesen [2014], p.22) with MD on J.

Lemma 6.13 *(a) Let $X_1, ..., X_n$ be MD on J ($J \subset \{1, ..., n\}$) and $\sigma : J \to J$ be an independent permutation of the index set J. Then the random variable $Y_i = X_{\sigma(i)}$, $i \in J$ is MD on $J_{\sigma(i)} = \{i\}$, i.e., $(\tilde{m}_i(t))_{t\in(0,1)} = \left(\frac{\mathbb{I}(Y_i\leqslant t)-t}{1-t}\right)_{t\in(0,1)}, i \in J$ is an $\mathcal{F}_t-$martingale.*

(b) If $X_1, ..., X_n$ are MD on each $J_i = \{i\}$, $i \in J$, i.e., $(m_i(t))_{t\in(0,1)}$ is an $\mathcal{F}_t-$martingale for each $i \in J$, then $X_1, ..., X_n$ are MD on J.

Proof. The proof of part (a) follows directly from Lemma 6.12. The part (b) is obvious. ■

Lemma 6.14

Let $E_1, ..., E_{n+1}$ be i.i.d. standard exponential distributed. Set $V_i = \frac{\sum_{j=1}^{i} E_j}{\sum_{j=1}^{n+1} E_j}$, $i = 1, ..., n$. The random variables $V_1, ..., V_n$ are martingale dependent random variables, i.e., the process $M(t) = \frac{\sum_{i=1}^{n}\mathbb{I}(V_i\leqslant t)-nt}{1-t}$ is a martingale w.r.t. the filtration $\mathcal{F}_t = \sigma(\mathbb{I}(V_i \leqslant s) : 0 \leqslant s \leqslant t < 1, i = 1, ..., n), t \in (0, 1)$.

Proof. The proof is based on the fact that the random variables $V_i, i = 1, ..., n$, have the same distribution as the order statistics $U_{1:n}, ..., U_{n:n}$ of a sample of size n from the uniform distribution $U(0,1)$ (cf. Shorack,Wellner [1986] pp.335-336). Therefore we have

$$\mathbb{E}\left[\frac{\sum_{i=1}^{n}\mathbb{I}(V_i \leqslant t) - nt}{1-t}\Big|\mathcal{F}_s\right] = \mathbb{E}\left[\frac{\sum_{i=1}^{n}\mathbb{I}(U_{i:n} \leqslant t) - nt}{1-t}\Big|\mathcal{F}_s\right]$$

for all $0 \leqslant s \leqslant t < 1$. ■

Remark 6.15

The random variables $V_1, ..., V_n$ obviously fulfill the Joint Null Criterion (cf. Leek and Storey [2011]), (see also Remark 6.7 (iii)).

6.2 Some Examples

In this section we consider more complicated examples of martingale dependent random variables. We give the proofs of the martingale dependence for all of them. We will soon see that the class of random variables which are martingale dependent and fulfill the PRDS assumption at the same time is greater than merely i.i.d U$(0,1)$-distributed random variables. There exist some nontrivial dependence structures which fulfill the assumptions of Lemma 5.9 and Theorem 5.24, as well as Lemma 5.28 and Theorem 5.36.

Example 6.16 (martingale dependence)

Let us consider the random variables $V_0, V_1, ..., V_{n_0}$ which are i.i.d uniformly distributed on $(0,1)$. Then the random variables

$$U_i = 1 - (1 - V_0 \wedge V_i)^2 \tag{6.16}$$

are U$(0,1)$-distributed and MD, i.e., the process $M(t) = \frac{\sum_1^{n_0}\mathbb{I}(U_i \leqslant t) - n_0 t}{1-t}$ is an $\mathcal{F}_t$-martingale with $\mathcal{F}_t = \sigma(\mathbb{I}_{(0,s]}(U_i), i = 1, ..., n_0, 0 \leqslant s \leqslant t),\ 0 \leqslant t < 1$.

Proof. Since V_0 and $V_i,\ i = 1, ...n_0$ are i.i.d. U$(0,1)$-distributed random variables, we have

$$P(U_i \leqslant t) = P(1 - (1 - V_0 \wedge V_i)^2 \leqslant t) = P(V_0 \wedge V_i \leqslant 1 - \sqrt{1-t})$$
$$= 1 - P(V_0 > 1 - \sqrt{1-t})P(V_i > 1 - \sqrt{1-t}) = t.$$

To show that the random variables $U_1, ..., U_{n_0}$ are MD random variables, we have to check the martingale property

$$\mathbb{E}\left[M(t)|\mathcal{F}_s\right] = M(s), \quad \text{for} \quad s \leqslant t. \tag{6.17}$$

Let us assume that the elementary Markov property holds, i.e.,

$$\mathbb{E}\left[\mathbb{I}(U_i \leqslant t)|\mathcal{F}_s\right] = \mathbb{E}\left[\mathbb{I}(U_i \leqslant t)|\mathbb{I}(U_1 \leqslant s), \mathbb{I}(U_2 \leqslant s), ..., \mathbb{I}(U_{n_0} \leqslant s)\right]. \tag{6.18}$$

We have

$$\mathbb{E}\left[\mathbb{I}_{[0,t]}(U_i)|\mathbb{I}_{[0,s]}(U_1), ..., \mathbb{I}_{[0,s]}(U_{n_0})\right] = \mathbb{I}(U_i \leqslant s)$$

$$+\mathbb{I}(U_i > s) \sum_{\substack{l_j \in \{0,1\} \\ j=1,...,n_0, j\neq i}} \left(\prod_{\substack{j=1 \\ j\neq i}}^{n_0} \mathbb{I}(\mathbb{I}_{(0,s]}(U_j) = l_j)\, \underbrace{\mathbb{E}\left[\mathbb{I}_{[0,t]}(U_i) \mid (U_i > s), \bigcap_{\substack{j=1 \\ j\neq i}}^{n_0} \{\mathbb{I}_{(0,s]}(U_j) = l_j\}\right]}_{p_{(l_1,...,l_{n_0})}} \right), \tag{6.19}$$

thereby the summation in the second term corresponds to the one in the proof of Theorem 6.5. We will see that $p_{(l_1,...,l_{n_0})}$ only takes one value for each possible $(l_1, ..., l_{n_0}) \in \{0, 1\}^{n_0 - 1}$. To proceed, let us delve closely into the structure of $U_1, ..., U_{n_0}$.
If we define $f(x) = 1 - \sqrt{1 - x}$, we get

$$\begin{aligned} U_i > s \Leftrightarrow 1 - (1 - V_0 \wedge V_i)^2 > s \Leftrightarrow V_i \wedge V_0 > f(s) \\ \Leftrightarrow \mathbb{I}(V_0 > f(s))\mathbb{I}(V_i > f(s)) = 1. \end{aligned} \tag{6.20}$$

The above chain of equivalences implies that for all $\omega \in \{\omega \in \Omega : U_j > s\}$

$$\mathbb{I}_{(0,s]}(U_j) = \mathbb{I}_{(0,f(s)]}(V_j), \ j = 1, ..., n_0 \tag{6.21}$$

must hold. Thus, we have

$$p_{(l_1,\dots,l_{n_0})} = P(U_i \leqslant t \mid (U_i > s) \cap \bigcap_{\substack{j=1 \\ j\neq i}}^{n_0} (\mathbb{I}_{(0,s]}(U_j) = l_j)) \tag{6.22}$$

$$= 1 - P(U_i > t \mid (U_i > s) \cap \bigcap_{\substack{j=1 \\ j\neq i}}^{n_0} (\mathbb{I}_{(0,s]}(U_j) = l_j)) \tag{6.23}$$

$$= 1 - P(V_0 > f(t), V_i > f(t) \mid V_0 > f(s) \cap V_1 > f(s) \cap \bigcap_{\substack{j=1 \\ j\neq i}}^{n_0} (\mathbb{I}_{(0,s]}(U_j) = l_j)) \tag{6.24}$$

$$= 1 - P(V_0 > f(t), V_i > f(t) \mid V_0 > f(s) \cap V_1 > f(s)) = 1 - \frac{(1-f(t))^2}{(1-f(s))^2} \tag{6.25}$$

$$= \frac{t-s}{1-s}. \tag{6.26}$$

The equalities in (6.25) hold, because the random variables $V_1, \dots, V_{n_0}$ are i.i.d. $U_{(0,1)}$. The equality (6.19), the chain of equalities (6.22) - (6.26) and the fact that

$$\sum_{\substack{l_j \in \{0,1\} \\ j=1,\dots,n_0, j\neq i}} \prod_{\substack{j=1 \\ j\neq i}}^{n_0} \mathbb{I}(\mathbb{I}_{(0,s]}(U_j) = l_j) = 1 \text{ holds,}$$

bring us the following result

$$\mathbb{E}\left[M(t)|\mathcal{F}_s\right] = \sum_{i=1}^{n_0} \left(\frac{\mathbb{I}(U_i \leqslant s) + \mathbb{I}(U_i > s)\frac{t-s}{1-s} - t}{1-t} \right) = M(s). \tag{6.27}$$

Therefore, assuming that the property (6.18) holds, we have shown that $U_1, \dots, U_{n_0}$ are MD. It remains to show that the Markov property (6.18) is fulfilled. Or equivalently: we have to show that for all $0 < s_1 < \dots < s_n = s$ and for all possible $l_j^k \in \{0,1\},\ j = 1, \dots, n_0,\ k = 1, \dots, n$ the following equality is valid

$$\begin{gathered} \overbrace{P(U_i \leqslant t \mid \bigcap_{k=1}^{n} \{\mathbb{I}_{(0,s_k]}(U_i) = l_i^k\} \cap \bigcap_{\substack{j=1 \\ j\neq i}}^{n_0} \bigcap_{k=1}^{n} \{\mathbb{I}_{(0,s_k]}(U_j) = l_j^k\})}^{:=P_1} = \\ \underbrace{P(U_i \leqslant t \mid \{\mathbb{I}_{(0,s_n]}(U_i) = l_i^n\} \cap \bigcap_{\substack{j=1 \\ j\neq i}}^{n_0} \{\mathbb{I}_{(0,s_n]}(U_j) = l_j^n\})}_{:=P_2}. \end{gathered} \tag{6.28}$$

Note that we have to consider only the cases, for which $l_j^1 \leqslant l_j^2, ... \leqslant l_j^n$ for all $j = 1, ..., n_0$ holds, since for the other cases the probability of the condition in (6.28) is equal to zero. Let us consider following two cases

1.Case $l_i^n = 1$, i.e., $U_i \leqslant s_n$ holds. Then we have $(U_i \leqslant t) \cap (U_i \leqslant s) = (U_i \leqslant s)$, and hence,

$$P_1 = \frac{P\left((U_i \leqslant t) \cap \bigcap_{k=1}^{n} \{\mathbb{I}_{(0,s_k]}(U_i) = l_i^k\} \cap \bigcap_{\substack{j=1\\ j\neq i}}^{n_0} \bigcap_{k=1}^{n} \{\mathbb{I}_{(0,s_k]}(U_j) = l_j^k\}\right)}{P\left(\bigcap_{k=1}^{n} \{\mathbb{I}_{(0,s_k]}(U_i) = l_i^k\} \cap \bigcap_{\substack{j=1\\ j\neq i}}^{n_0} \bigcap_{k=1}^{n} \{\mathbb{I}_{(0,s_k]}(U_j) = l_j^k\}\right)} = 1,$$

$$P_2 = \frac{P\left((U_i \leqslant t) \cap (U_i \leqslant s) \cap \bigcap_{\substack{j=1\\ j\neq i}}^{n_0} \{\mathbb{I}_{(0,s_n]}(U_j) = l_j^n\}\right)}{P\left((U_i \leqslant s) \cap \bigcap_{\substack{j=1\\ j\neq i}}^{n_0} \{\mathbb{I}_{(0,s_n]}(U_j) = l_j^n\}\right)} = 1.$$

2.Case $l_i^n = 0$, i.e., $U_i > s_n$ holds. With the same argumentation as in (6.20) we get

$$\mathbb{I}_{(0,s_n]}(U_i) = \mathbb{I}_{(0,f(s_n)]}(V_i) \quad \text{for all} \quad \omega \in \{\omega \in \Omega : U_i > s\}, \tag{6.29}$$

which together with the fact that $\{V_j > f(t)\} \cap \{V_j > f(s_n)\} = \{V_j > f(t)\}$ for all

$j = 1, ..., n_0$ holds, implies analogously to (6.22)-(6.26) the following equations:

$$P_1 = 1 - \frac{P\left((U_i > t) \cap (U_i > s_n) \cap \bigcap_{\substack{j=1 \\ j \neq i}}^{n_0} \bigcap_{k=1}^{n} \{\mathbb{I}_{(0,s_k]}(U_j) = l_j^k\}\right)}{P\left((U_i > s_n) \cap \bigcap_{\substack{j=1 \\ j \neq i}}^{n_0} \bigcap_{k=1}^{n} \{\mathbb{I}_{(0,s_k]}(U_j) = l_j^k\right)} \tag{6.30}$$

$$= 1 - \frac{P\left((V_0 > f(t)) \cap (V_i > f(t)) \cap \bigcap_{\substack{j=1 \\ j \neq i}}^{n_0} \bigcap_{k=1}^{n} \{\mathbb{I}_{(0,f(s_k)]}(V_j) = l_j^k\})\right)}{P\left((V_i > f(s_n)) \cap (V_0 > f(s_n)) \cap \bigcap_{\substack{j=1 \\ j \neq i}}^{n_0} \bigcap_{k=1}^{n} \{\mathbb{I}_{(0,f(s_k)]}(V_j) = l_j^k\}\right)} \tag{6.31}$$

$$= 1 - \frac{P\left((V_0 > f(t)) \cap (V_i > f(t))\right)}{P\left((V_i > f(s_n)) \cap (V_0 > f(s_n))\right)}. \tag{6.32}$$

$$= 1 - \frac{P\left((V_0 > f(t)) \cap (V_i > f(t)) \cap (V_0 > f(s_n)) \cap \bigcap_{\substack{j=1 \\ j \neq i}}^{n_0} \{\mathbb{I}_{(0,f(s_n)]}(V_j) = l_j^n\}\right)}{P\left((V_i > f(s_n)) \cap (V_0 > f(s_n)) \cap \bigcap_{\substack{j=1 \\ j \neq i}}^{n_0} \{\mathbb{I}_{(0,f(s_n)]}(V_j) = l_j^n\}\right)} \tag{6.33}$$

$$= P(U_i \leqslant t \mid (U_i > s) \cap \bigcap_{\substack{j=1 \\ j \neq i}}^{n_0} \{\mathbb{I}_{(0,s_n]}(U_j)\}) = P_2. \tag{6.34}$$

The equality in (6.32) holds, because $V_0, ..., V_{n_0}$ are independent. Hence, we have shown that the elementary Markov property is valid and consequently by (6.27) $U_1, ..., U_{n_0}$ are MD-random variables.

∎

Of particular interest is the fact that the random variables from this example are PRDS.

Lemma 6.17

The random variables $U_1, ..., U_{n_0}$ from Example 6.16 are PRDS.

Proof. Firstly note, that the structure of $(U_1, ..., U_{n_0})$ belongs to the unidimensional latent variable model, cf. Holland and Rosenbaum [1986]. We denote this model by (U, V_0). The components of the vector $\bar{U} = (U_1, ..., U_{n_0})$ are conditional independent given $V_0 = v$. Each component U_i is increasing in V_0 and $V_i, i = 1, ..., n$. Hence, the latent variable model (U, V_0) is monotone, satisfies the conditions of latent conditional independence and latent unidimensionality, cf. Holland and Rosenbaum [1996]. Due to Theorem 6 of Holland, Rosenbaum [1986], pp.1533-1534, it implies, that the random variables $U_1, ..., U_{n_0}$ are conditionally associated. Consequently $(U_1, ..., U_{n_0})$ are PRDS on I_0 (and on each subset of I_0), cf. Benjamini and Yekutieli [2001], p.1173. ■

Example 6.18 (martingale dependence)
We consider random variables V_0, V_1, V_2, V_3 which are i.i.d. $U(0, 1)$. Then the random variables

$$U_i = xV_0\mathbb{I}(V_3 \leqslant x) + (1 - (1 - x)V_i)\mathbb{I}(V_3 > x), \quad \text{for some fixed } x \in (0, 1), \ i = 1, 2$$

are U$(0, 1)$-distributed and MD random variables.

Proof. Since V_0, V_1, V_2, V_3 are i.i.d U$(0, 1)$-distributed random variables, we have

$$\begin{aligned}
P(U_i \leqslant t) &= P(xV_0\mathbb{I}(V_3 \leqslant x) + (1 - (1 - x)V_i)\mathbb{I}(V_3 > x) \leqslant t) \\
&= P(xV_0 \leqslant t)x + (1 - x)P(1 - (1 - x)V_i \leqslant t) \\
&= t\mathbb{I}(t \leqslant x) + x\mathbb{I}(t > x) + \mathbb{I}(t > x)(1 - x)\left(1 - \frac{1 - t}{1 - x}\right) = t.
\end{aligned}$$

Let us assume, as before, that the Markov property (6.18) holds. Then we have to show that the martingale property $\mathbb{E}\left[\sum_{i=1}^{2} \frac{\mathbb{I}(U_i \leqslant t) - t}{1 - t} | \mathbb{I}_{(0,s]}(U_1), \mathbb{I}_{(0,s]}(U_2)\right] = \sum_{i=1}^{2} \frac{\mathbb{I}(U_i \leqslant s) - s}{1 - s}$ is valid. Similarly to the previous example, we have

$$\mathbb{E}\left[\mathbb{I}_{[0,t]}(U_1)|\mathbb{I}_{[0,s]}(U_1), \mathbb{I}_{[0,s]}(U_2)\right] = \mathbb{I}(U_1 \leqslant s) \tag{6.35}$$

$$+ \mathbb{I}(U_1 > s)\mathbb{I}(U_2 > s) \underbrace{\mathbb{E}\left[\mathbb{I}_{[0,t]}(U_1)|(U_1 > s) \cap (U_2 > s)\right]}_{:=p_1} \tag{6.36}$$

$$+ \mathbb{I}(U_1 > s)\mathbb{I}(U_2 \leqslant s) \underbrace{\mathbb{E}\left[\mathbb{I}_{[0,t]}(U_1)|(U_1 > s) \cap (U_2 \leqslant s)\right]}_{:=p_2}. \tag{6.37}$$

Further we get

$$\begin{aligned}
p_1 &= \frac{P(U_1 \leqslant t, U_1 > s, U_2 > s | V_3 \leqslant x)x + P(U_1 \leqslant t, U_1 > s, U_2 > s | V_3 > x)(1-x)}{P(U_1 > s, U_2 > s | V_3 \leqslant x)x + P(U_1 > s, U_2 > s | V_3 > x)(1-x)} \\
&= \frac{P(xV_0 \leqslant t, xV_0 > s)x + P(V_1 \geqslant \frac{1-t}{1-x}, V_1 < \frac{1-s}{1-x}, V_2 < \frac{1-s}{1-x})(1-x)}{P(xV_0 > s)x + P(V_1 < \frac{1-s}{1-x}, V_2 < \frac{1-s}{1-x})(1-x)} \\
&= \frac{\mathbb{I}_{s \leqslant t < x}(t-s) + \mathbb{I}_{s < x \leqslant t}(t-s) + \mathbb{I}_{x \leqslant s \leqslant t}\frac{(t-s)(1-s)}{1-x}}{\mathbb{I}_{s<x}(1-s) + \mathbb{I}_{x \leqslant s}\frac{(1-s)^2}{1-x}} \\
&= \mathbb{I}_{s \leqslant t < x}\frac{t-s}{1-s} + \mathbb{I}_{s < x \leqslant t}\frac{t-s}{1-s} + \mathbb{I}_{x \leqslant s \leqslant t}\frac{t-s}{1-s} = \frac{t-s}{1-s}.
\end{aligned}$$

$$\begin{aligned}
p_2 &= \frac{P(U_1 \leqslant t, U_1 > s, U_2 \leqslant s | V_3 \leqslant x)x + P(U_1 \leqslant t, U_1 > s, U_2 \leqslant s | V_3 > x)(1-x)}{P(U_1 > s, U_2 \leqslant s | V_3 \leqslant x)x + P(U_1 > s, U_2 \leqslant s | V_3 > x)(1-x)} \\
&= \frac{(1-x)P(V_1 \geqslant \frac{1-t}{1-x}, V_1 < \frac{1-s}{1-x}, V_2 \geqslant \frac{1-s}{1-x})}{P(V_1 < \frac{1-s}{1-x}, V_2 \geqslant \frac{1-s}{1-x})(1-x)} = \frac{t-s}{1-s}\mathbb{I}_{x \leqslant s \leqslant t}.
\end{aligned}$$

Now we have to prove the Markov property (6.18) which is equivalent to

$$\begin{aligned}
&P(U_i \leqslant t | (\bigcap_{k=1}^{n} \{\mathbb{I}_{(0,s_k]}(U_1) = l_1^k) \cap (\bigcap_{k=1}^{n} \{\mathbb{I}_{(0,s_k]}(U_2) = l_2^k)\}) \\
&= P(U_i \leqslant t | (\mathbb{I}_{(0,s_n]}(U_1) = l_1^n) \cap (\mathbb{I}_{(0,s_n]}(U_2) = l_2^n))
\end{aligned} \tag{6.38}$$

for this example.

1.CASE $s \leqslant x$. Note that in this case $l_1^k = l_2^k$ holds for $k = 1, ..., n$. Otherwise, if, for example, $l_1^k = 1$ and $l_2^k = 0$ holds for some $k = 1, ..., n$, we would have $\mathbb{I}(U_1 \leqslant s_k) = 1$. It would imply $\mathbb{I}(U_1 \leqslant x) = 1$. The last identity yields $U_1 = U_2$, and consequently $\mathbb{I}(U_2 \leqslant s_k) = 1$ must hold. Thus, we got $l_1^k = l_2^k = 1$. Hence, we have to consider only two cases: $\begin{pmatrix} l_1^n \\ l_2^n \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \end{pmatrix}$ and $\begin{pmatrix} l_1^n \\ l_2^n \end{pmatrix} = \begin{pmatrix} 1 \\ 1 \end{pmatrix}$

In the first case we have $U_1 > s_n$ and $U_2 > s_n$, so we get

$$\{\bigcap_{k=1}^{n} \{\mathbb{I}_{(0,s_k]}(U_1) = l_1^k) \cap (\bigcap_{k=1}^{n} \{\mathbb{I}_{(0,s_k]}(U_2) = l_2^k)\}\} = \{U_1 > s_n, U_2 > s_n\} \tag{6.39}$$

$$= \{(\mathbb{I}_{(0,s_n]}(U_2) = l_1^n) \cap (\mathbb{I}_{(0,s_n]}(U_2) = l_2^n)\}, \tag{6.40}$$

which implies (6.18) for this case. In the second case we have $U_1 \leqslant s_n$ and $U_2 \leqslant s_n$ and consequently we obtain

$$(U_i \leqslant t) \cap (U_i \leqslant s_n) = (U_i \leqslant s_n) \text{ for } i = 1, 2,$$

which implies the identity (6.28).

2. CASE $x < s$ (w.l.o.g. we can assume that $x \leqslant s_n$, otherwise see CASE 1 of this example). Here we have the four following possibilities: $\begin{pmatrix} l_1^n \\ l_2^n \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \end{pmatrix}$, $\begin{pmatrix} l_1^n \\ l_2^n \end{pmatrix} = \begin{pmatrix} 0 \\ 1 \end{pmatrix}$ and $\begin{pmatrix} l_1^n \\ l_2^n \end{pmatrix} = \begin{pmatrix} 1 \\ 0 \end{pmatrix}$, $\begin{pmatrix} l_1^n \\ l_2^n \end{pmatrix} = \begin{pmatrix} 1 \\ 1 \end{pmatrix}$.

W.l.o.g let $i = 1$. The two first cases, when $U_1 \leqslant s_n$ holds, were considered in the Example 6.16, (**1.CASE**).
Let us consider the case $U_1 > s_n$, $U_2 \leqslant s_n$. We can transform

$$P(U_1 \leqslant t | (\bigcap_{k=1}^{n} \{\mathbb{I}_{(0,s_k]}(U_1) = l_1^k) \cap (\bigcap_{k=1}^{n} \{\mathbb{I}_{(0,s_k]}(U_2) = l_2^k)\})$$

$$= P(U_1 \leqslant t | (U_1 > s) \cap (U_2 \leqslant s_n) \cap (\bigcap_{k=1}^{n-1} \{\mathbb{I}_{(0,s_k]}(U_2) = l_2^k)\}))$$

$$= \frac{P(V_3 \geqslant x) P(V_1 \geqslant \frac{1-t}{1-x}, V_1 < \frac{1-s}{1-x}, \overbrace{V_2 < \frac{1-s}{1-x}, (\bigcap_{k=1}^{n-1} \{\mathbb{I}_{[\frac{1-s_k}{1-x}, s1)}(V_2) = l_2^k)\}))}^{\perp V_1}}{P(V_3 \geqslant x) P(V_1 < \frac{1-s}{1-x}, \underbrace{V_2 < \frac{1-s}{1-x}, (\bigcap_{k=1}^{n-1} \{\mathbb{I}_{[\frac{1-s_k}{1-x}, s1)}(V_2) = l_2^k)\}))}_{\perp V_1}}$$

$$= \frac{P(V_1 \geqslant \frac{1-t}{1-x}, V_1 < \frac{1-s}{1-x})}{P(V_1 < \frac{1-s}{1-x})} = \frac{P(V_1 \geqslant \frac{1-t}{1-x}, V_1 < \frac{1-s}{1-x}, V_2 < \frac{1-s}{1-x})}{P(V_1 < \frac{1-s}{1-x}, V_2 < \frac{1-s}{1-x})} = P(U_1 \leqslant t | U_1 > s, U_2 \leqslant s),$$

where $\perp$ signifies stochastic independence. Thus we have shown the Markov property for this case.
It remains to consider the case $\begin{pmatrix} l_1^n \\ l_2^n \end{pmatrix} = \begin{pmatrix} 1 \\ 1 \end{pmatrix}$, i.e., $U_1 > s_n$ and $U_2 > s_n$. Since $s_1 < s_2 < \ldots < s_n$, we have

$$\bigcap_{j=1}^{2} \bigcap_{k=1}^{n} (U_j > s_k) = (U_1 > s_n) \cap (U_2 > s_n). \tag{6.41}$$

Hence, the Markov property

$$P(U_1 \leqslant t \mid \bigcap_{j=1}^{2}\bigcap_{k=1}^{n}(U_j > s_k)) = P(U_1 \leqslant t \mid (U_1 > s_n) \cap (U_2 > s_n)) \tag{6.42}$$

holds for this case and, consequently, overall. ■

6.3 A counter example

The examples for the martingale dependent random variables considered before were all designed as functions of the i.i.d. random variables. For all of these structures the Markov property was fulfilled. Hence, the following question seems to be reasonable: Is the elementary Markov property always fulfilled for functions of the i.i.d. random variables? The next counter example provides the answer.

Example 6.19

For independent, identically U$(0,1)-$distributed random variables Y_0, Y_1, Y_2 we define the random variables X_1, X_2 as

$$X_i = g(Y_0, Y_i) = \begin{cases} \frac{Y_i}{5}, & \text{if } \mathbb{I}(Y_0 \leqslant \frac{1}{5})\mathbb{I}(Y_i \leqslant \frac{1}{2}) = 1, \\ \frac{1}{2} + \frac{2Y_i - 1}{10}, & \text{if } \mathbb{I}(Y_0 \leqslant \frac{1}{5})\mathbb{I}(Y_i > \frac{1}{2}) = 1, \\ \frac{1+8Y_i}{10}, & \text{if } \mathbb{I}(Y_0 > \frac{1}{5})\mathbb{I}(Y_i \leqslant \frac{1}{2}) = 1, \\ \frac{6}{10} + \frac{8Y_i - 4}{10}, & \text{if } \mathbb{I}\left(Y_0 > \frac{1}{5}\right)\mathbb{I}(Y_i > \frac{1}{2}) = 1. \end{cases} \qquad i \in \{1,2\}.$$

Now we show that the process $Z(t) = \begin{pmatrix} \mathbb{I}_{([0,t])}(X_1) \\ \mathbb{I}_{([0,t])}(X_2) \end{pmatrix}$ does not fulfill the Markov property, since

$$P\left(Z(0.6) = \binom{1}{1} \middle| Z(0.5) = \binom{1}{0}, Z(0.1) = \binom{1}{0}\right) \neq P\left(Z(0.6) = \binom{1}{1} \middle| Z(0.5) = \binom{1}{0}\right) \tag{6.43}$$

holds. More precisely:

$$\begin{aligned} \{Z(0.1) = \tbinom{1}{0}\} &= \{X_1 \leqslant 0.1\} \cap \{X_2 > 0.1\} \\ &\subset \{Y_0 \leqslant \frac{1}{5}\} \cap \{Y_1 \leqslant \frac{1}{2}\} \cap \{Y_2 > \frac{1}{2}\} \\ &\subset \{X_1 \leqslant 0.6\} \cap \{X_2 \leqslant 0.6\} = \{Z(0.6) = \tbinom{1}{1}\}, \end{aligned}$$

which implies that $P\left(Z(0.6) = \binom{1}{1} \middle| Z(0.5) = \binom{1}{0}, Z(0.1) = \binom{1}{0}\right) = 1$ is valid. On the other hand the inequality $P\left(Z(0.6) = \binom{1}{1} \middle| Z(0.5) = \binom{1}{0}\right) < 1$ obviously holds which implies the inequality in (6.43).

6.4 (Super-)martingale dependent test statistics

We propose the following example in answer to the question - what kind of dependence must be between the test statistics to guarantee the martingale dependence between the corresponding p-values.

Let us consider the following multiple test with components:

$$\phi_i = \begin{cases} 1, & \text{if } D_i < a_i \\ 0, & \text{if } D_i \geqslant a_i \end{cases}, \; i = 1, ..., n. \tag{6.44}$$

Thereby the test statistics $D_i, i = 1, ..., n,\ D_i : \Omega \to R$ have a common continuous and strictly monotone distribution function G under the null hypothesis and $a_i, i = 1, ..., n$ are some critical values.

Now we can define the p-values as

$$p_i = G(D_i) \tag{6.45}$$

which are uniformly distributed on $[0, 1]$. Then we have:

Lemma 6.20
If the process $\tilde{M}(x) = \frac{\sum_{i=1}^{n} \mathbb{I}(D_i \leqslant x) - G(x)}{1 - G(x)}$ *is a* $\mathcal{G}_x-$ *martingale, with* $\mathcal{G}_x = \sigma(\mathbb{I}(D_i \leqslant s), s \leqslant x,\ i = 1, ..., n)$, *then* $p_i^{\prime}s$, *defined as in (6.45), belong to the class* $MD(P, n)$.

Proof. With the transformation $x = G^{-1}(t)$ and because of the continuity and strict monotonicity of the distribution function G we get:

$$\tilde{M}(x) = \frac{\sum_{i=1}^{n} \mathbb{I}(D_i \leqslant x) - nG(x)}{1 - G(x)} \tag{6.46}$$

$$= \frac{\sum_{i=1}^{n} \mathbb{I}(D_i \leqslant G^{-1}(t)) - nG(G^{-1}(t))}{1 - G(G^{-1}(t))} \tag{6.47}$$

$$= \frac{\sum_{i=1}^{n} \mathbb{I}(p_i \leqslant t) - nt}{1 - t} =: M(t), \tag{6.48}$$

which implies that $M(t)$ is a $\mathcal{F}_t-$martingale. ■

The following example is analogous to Example 6.16 with only one difference. Here, we have the martingale structure of the test statistics which implies by Lemma 6.20 the martingale dependence between the corresponding p-values $P_i, i = 1, ..., n$.

Example 6.21 (martingale dependence of test statistics)
Let us consider the random variables $D_0, D_1, ..., D_{n_0}$, which are i.i.d, continuous distributed with a strictly monotone distribution function F. Then for the random variables $\tilde{D}_i = D_0 \bigwedge D_i, i = 1, ..., n_0$ the process $M_{\tilde{D}}(x) = \frac{\sum_{i=1}^{n_0} \mathbb{I}(\tilde{D}_i \leqslant x) - n_0 \tilde{F}(x)}{1-\tilde{F}(x)}$ is a $\mathcal{G}_x-$ martingale, where $\tilde{F}$ is the common distribution function of $\tilde{D}_i, i = 1, ...n$.

6.5 Simulation example

We conducted a simulation study to investigate the FDR level under martingale dependence of adaptive procedures numerically. Therefore we compare the SD procedure (5.1) with the magnified critical values by Lemma 4.3, the linear Benjamini Hochberg SU procedure and the adaptive $\lambda-$based Storey's SU procedure. We have set the number of tests to be $n = 100, 500$ and 800. The fraction of the true null hypotheses n_0/n was set at $(n-1)/n, 1/4, 1/2$, and $3/4$, the tuning parameter $\lambda = 0.8$ and the FDR-controlling level $\alpha = 0.15$. Our computations are based on $L = 10000$ replications. We have investigated the following three configurations. Note that the first two structures belong to the martingale model, the last one belongs to the super-martingale model.

(1) The first configuration is based on the following multiple testing situation.

The random variables $X_1, ..., X_n$ are independent, 2-parameter exponentially distributed with scale parameter $\lambda = 1$ and location parameter $\vartheta_i \leqslant 5, i = 1, ..., n$. X_0 is independent from each $X_i, i = 1, ..., n$ and is exponential distributed with scale parameter $\lambda = 1$ and location parameter $\vartheta_0 = 5$. We consider the test problem

$$H_i : \vartheta_i = 5 \qquad \text{vs.} \qquad H_i^c : \vartheta_i < 5. \tag{6.49}$$

The location parameter $\vartheta_i, i \in I$ is generated as

$$\vartheta_i = 5\mathbb{I}(i \in I_0) + \zeta_i \mathbb{I}(i \in I_1), i \in I, \quad \text{with} \quad \zeta_i \quad \text{i.i.d.} \quad U(0, 0.5).$$

The p-values are generated in the following way. At first we defined the test statistics

$$D_i = X_0 \wedge X_i,$$

and then we defined the corresponding p-values as

$$p_i = G_{H_0}(D_i).$$

We have

$$\begin{aligned} G_{H_0}(t) &= P_{H_0}(X_0 \wedge X_i \leqslant t) \\ &= 1 - P_{H_0}(X_i > t)P(X_0 > t) \\ &= 1 - (\mathbb{I}(t < 5) + \mathbb{I}(5 \leqslant t)\exp(-(t-5)))^2 \\ &= \mathbb{I}(5 \leqslant t)(1 - \exp(-2(t-5))). \end{aligned}$$

(2) The second configuration referes to the four blocks total dependence.
For X_1, X_2, X_3, X_4 i.i.d $\mathcal{N}(0,1)$, $\varepsilon_i = 1 - \mathbb{I}(i \in \{1, ..., n_0)$ and $\mu_i \sim U(1,5)$, distributed i.i.d random variables independent of $X = (X_1, ..., X_4)$, $i = 1, ..., n$, we have defined the p-values in the following way

$$p_i = 1 - \Phi(X_j + \varepsilon_i \mu_i),\ i = 1, ..., n, \tag{6.50}$$

with $j = 1$ for $i = 1, ..., k_1 - 1$, $j = 2$ for $i = k_1, ..., k_2 - 1$, $j = 3$ for $i = k_2, ..., k_3 - 1$ and $j = 4$ for $i = k_3, ..., n$ $1 \leqslant k_1 < k_2 < k_3 < n$. Thereby Φ is the c.d.f. of $N(0,1)$.

(3) The third configuration corresponds to an n_0-dimensional version of Example 6.18, i.e., for i.i.d. random variables $V_0, ..., V_{n_0}, V_{n_0+1}$, $V_i \sim U(0,1)$, $i = 0, ..., n_0 + 1$, we consider

$$U_i = xV_0\mathbb{I}(V_{n_0+1} \leqslant x) + (1 - (1-x)V_i)\mathbb{I}(V_{n_0+1} > x)$$

for some fixed $x \in (0,1)$, $i = 1, ..., n_0$. We set $x = 0.05$ in our simulated example. The false p-values are independent of the true ones and are uniformly U$(0, \frac{\alpha}{n+\alpha})$-distributed.

We summerize the results of the aforementioned sumulations in the following three tabels.

Table 6.1: Comparison of the FDR for the SD-procedure (5.1) (ASDP), for the adaptive $\lambda-$ Storeys SU-procedure (StSU), for the linear SD-procedure (BHsd) and for the linear SU-procedure (BHsu) under the martingale dependence (1) for $\alpha = 0.15$

n	ξ	ASDP	BHsd	BHsu	StSu
$n = 100$	99/100	0.0734	0,0728	0.1417	0.39
	3/4	0.093	0.071	0.1148	0.306
	1/2	0.103	0.057	0.074	0.212
	1/4	0.102	0.033	0.037	0.124
$n = 500$	499/500	0.0782	0.0781	0.152	0.404
	3/4	0.091	0.069	0.108	0.302
	1/2	0.105	0.058	0.079	0.216
	1/4	0.104	0.033	0.037	0.125
$n = 800$	799/800	0.074	0.073	0.145	0.395
	3/4	0.093	0.072	0.11	0.3
	1/2	0.104	0.057	0.074	0.207
	1/4	0.103	0.033	0.037	0.125

As we can see from Table 6.1 the linear SU exhausts the level α well, when almost all hypotheses are true, as in the BIA case. When the portion of the true nulls is smaller than $\frac{3}{4}$, the ASDP seems to exhaust the level α better than the BH SU. The FDR of the linear SD is smaller than the FDR of the ASDP for all values of $n_0(n)$. The FDR of the adaptive $\lambda-$based procedure of Storey lies above the level α in most cases, hence, this procedure can not be used for such kind of dependence.

Table 6.2: Comparison of the FDR under the martingale dependence (2) for $\alpha = 0.15$

n	ξ	ASDP	BHsd	BHsu	StSu
$n = 100$	96/100	0.011	0, 011	0.13	0.32
	72/100	0.097	0.072	0.11	0.27
	52/100	0.012	0.075	0.097	0.195
	24/100	0.011	0.037	0.042	0.132
$n = 200$	196/200	0.036	0.036	0.13	0.27
	152/200	0.066	0.055	0.105	0.23
	1/2	0.093	0.05	0.07	0.21
	52/200	0.105	0.037	0.039	0.15
$n = 800$	792/800	0.006	0.006	0.15	0.264
	3/4	0.063	0.053	0.11	0.231
	1/2	0.1	0.06	0.075	0.202
	1/4	0.1	0.034	0.038	0.139

The results from Table 6.2 again show that the Storey SU procedure is not suitable for such kind of dependence because the FDR level essentially exceeds the pre-chosen level α. The BH SU procedure still has the largest FDR level when most of the hypotheses are true. When the portion of the true null hypotheses is smaller or equal to the half of all hypotheses, the FDR of the SD procedure (5.1) is larger as the FDR of the linear SU procedure. It is interesting that the FDR of the BH SU procedure behaves similarly to the BIA case (as well as the reverse martingale case), i.e., it gets closer to $\frac{\alpha n_0}{n}$, even though the dependence type 2 does not belong to the class of reverse-martingale dependent random variables.

Table 6.3: Comparison of the FDR dependence (3) for $\alpha = 0.15$

n	ξ	ASDP	BHsd	BHsu	StSu
$n = 100$	99/100	0.	0	0.06	0.18
	3/4	0.04	0	0.04	0.07
	1/2	0.12	0.05	0.05	0.13
	1/4	0.15	0.05	0.05	0.14
$n = 200$	199/200	0	0	0.05	0.17
	3/4	0.023	0.023	0.023	0.06
	1/2	0.12	0.053	0.053	0.11
	1/4	0.13	0.02	0.02	0.16
$n = 800$	799/800	0	0	0.07	0.13
	3/4	0.04	0.04	0.03	0.06
	1/2	0.12	0.045	0.045	0.12
	1/4	0.141	0.037	0.038	0.14

Table 6.3 shows that both SD procedures, ASDP and BH SD as well, have the FDR equal to zero if almost all hypotheses are true. The FDR of the Storey's SU procedure lies slightly above the level α for these cases. For the other cases the FDR of the Storey's SU is smaller, than α. The FDR of the ASDP and Storey's SU as well get closer to the level α if the portion of the true null hypotheses is equal to 1/4. Note that the FDR of the BH SU procedure is not longer equal (or approximately equal) to $\alpha\frac{n_0}{n}$ in this example.

6.6 Concluding remarks

In this chapter we considered different dependence structures which belong to the (super-) martingale dependence structures. Theorem 6.5 yields sufficient conditions for the martingale dependence. We have shown that some non-trivial examples of the martingale dependent random variables exist which are PRDS or PA at the same time, and consequently fulfill the assumptions of the main theorems from Chapter 5. Finally, we proposed some numerical examples which confirm our theoretical statements from Chapter 5. Furthermore as we can see from Tables 6.1 - 6.3, only knowledge of $U_i, i = 1, ..., n_0$ being martingale dependent does not give us the information about which kind of procedures is more powerful in terms of the FDR-level.

Chapter 7

On uniformly distributed alternatives

In this section we devote our attention to the situation when null hypotheses and alternatives are not really separable. These alternatives can be seen as a boundary case of possible alternatives. Note that Dirac-alternatives, where p-values under alternatives are equal to zero, can also be seen as a boundary case. In comparison to Dirac-alternatives, uniformly distributed p-values under alternatives lie on the opposite side of the space of possible alternatives for a given test. Results related to boundary cases of alternatives yield some insight in the structure of MTPs.

(U) Let us assume that all (both true and false) p-values are i.i.d. uniformly U$[0,1]$-distributed. We label the first n_0 of p-values as true and the last n_1 as false:

$$\bar{p} = (\underbrace{U_1, ..., U_{n_0}}_{\text{true}}, \underbrace{U_{n_0+1}, ..., U_n}_{\text{false}})$$

Theorem 7.1

Let $\tau = \tau(p_{i:n}), i \leqslant n$ be an arbitrary stopping procedure (not necessary SD) which rejects all p-values with $p_{i:n} \leqslant \tau$. Suppose that assumption **U** *is fulfilled.*

If

$$FWER(\tau) \leqslant \alpha \tag{7.1}$$

holds for $n_0 = n$, then

$$FDR(\tau) = \mathbb{E}\left[\frac{V(\tau)}{R(\tau)}\mathbb{I}(R(\tau) > 0)\right] \leqslant \frac{n_0}{n}\alpha \tag{7.2}$$

is valid for any $n_0 \leqslant n$. *If FWER*$(\tau) = \alpha$ *for* $n_0 = n$, *then FDR*$(\tau) = \frac{n_0}{n}\alpha$ *for all* $n_0 \leqslant n$.

Proof. For $i \in \{1, ..., n\}$ we consider the following values

$$c_i = \begin{cases} 1, & i \leqslant n_0 \\ 0, & i > n_0 \end{cases} \tag{7.3}$$

Let D_i be the anti-rank of $U_{i:n}$ (cf. Definition 8.12 in Appendix). Then it follows from (7.3)

$$c_{D_i} = \begin{cases} 1, & \text{if } U_{i:n} \text{ is true,} \\ 0, & \text{if } U_{i:n} \text{ is false.} \end{cases} \tag{7.4}$$

Consequently, we get

$$V(\tau) = \sum_{i=1}^{n_0} \mathbb{I}(U_i \leqslant \tau) = \sum_{i=1}^{R(\tau)} c_{D_i}. \tag{7.5}$$

Since U_i, $i = 1, ..., n$ are i.i.d. random variables we have from Basu's Theorem, that $(D_i)_{i \leqslant n}$ and $(U_{i:n})_{i \leqslant n}$ are stochastically independent (cf. Ghosh [2002]). Thus, by the assumption that $\tau = \tau(p_{i:n}) = \tau(U_{i:n}), i \leqslant n$, the random variables c_{D_i} and $R(\tau)$ are also independent. Consequently we get

$$\mathbb{E}\left[\frac{V(\tau)}{R(\tau)}\mathbb{I}(R(\tau) > 0)\right] = \mathbb{E}\left[\mathbb{E}\left[\frac{V(\tau)}{R(\tau)}\mathbb{I}(R(\tau) > 0)|R(\tau)\right]\right] \tag{7.6}$$

$$= \sum_{r=1}^{n} \mathbb{E}\left[\frac{V(\tau)}{R(\tau)}|R(\tau) = r\right] P(\{R(\tau) = r\}) \tag{7.7}$$

$$= \sum_{r=1}^{n} \mathbb{E}\left[\frac{1}{r}\sum_{i=1}^{r} c_{D_i}\right] P(R(\tau) = r) \tag{7.8}$$

$$= \sum_{r=1}^{n} (\frac{1}{r}\sum_{i=1}^{r} \mathbb{E}\left[c_{D_i}\right])]P(R(\tau) = r) \tag{7.9}$$

$$= \sum_{r=1}^{n} \frac{n_0}{n} P(R(\tau) = r) = \frac{n_0}{n} P(R(\tau) > 0) \tag{7.10}$$

$$= \frac{n_0}{n} \text{FWER}_{n=n_0}(\tau) \leqslant \alpha\frac{n_0}{n}, \tag{7.11}$$

thereby $\text{FWER}_{n=n_0}(\tau)$ is defined to be $\text{FWER}(\tau)$ under global null hypothesis, i.e., if $n_0 = n$. We can see from (7.11) that $\text{FWER}_{n=n_0}(\tau) = \alpha$ implies the corresponding equality for $\text{FDR}(\tau)$. ■

Remark 7.2 *(a) From Theorem 7.1 it follows that the BH-procedure maximizes the FDR if all p-values are i.i.d., $U(0,1)$-distributed, as actually any other procedure for which under global null hypothesis FWER= α.*

(b) By Lemma 4.3 an SD-procedure with critical values $a_i, i = 1, ..., n$ controlling the FDR at level α under BIA, can be modified to an SD-procedure, for which $FWER = \alpha$ for $n_0 = n$. Theorem 7.1 implies that for such modified SD-procedure FDR= $\frac{n_0}{n}\alpha$ if all p-values are i.i.d., $U(0,1)$-distributed.

(c) From (a) and (b) it follows that if all p-values are i.i.d., $U(0,1)$-distributed, then any SD-procedure which controls the FDR at α under BIA, can be modified to the procedure which attains the FDR of the BH SU procedure, by an adjustment of the first coefficients due to Lemma 4.3.

Remark 7.3
In comments on the publication of Romano et al. [2008] Yekutieli stated that if p-values are exchangeable, then it is easy to see that the FDR for any value of n_0 is $FDR = \frac{n_0}{n} FDR_0$, if all p-values are marginally $U(0,1)$-distributed. Thereby FDR_0 is the FDR under the complete null hypothesis, $n_0 = n$, although it is stated without proof (cf. Yekutieli [2008], p.459). This statement coincides with the statement of Theorem 7.1 for i.i.d $U(0,1)$ p-values.

The next theorem is a weakened version of Theorem 5.3 from Benjamini and Yekutieli [2001], p.1181.

Theorem 7.4
Consider the problem of testing of n hypotheses with the SU test which uses the critical values $0 \equiv c_0 \leqslant c_1 \leqslant c_2 \leqslant ... \leqslant c_n$. Assume that

1. *the false p-values $f_1, ..., f_{n_1}$ are i.i.d. P_1-distributed, with $U(0,1) \succcurlyeq_{st} P_1$,*
2. *the true p-values $U_1, ..., U_{n_0}$ are i.i.d $U(0,1)$-distributed,*
3. *vectors $(f_1, ..., f_{n_1})$ and $(U_1, ..., U_{n_0})$ are mutually independent.*
   (a) If the ratio $\frac{c_k}{k}$ is increasing in k, as the distribution P_1 increases stochastically then the FDR decreases.
   (b) If the ratio $\frac{c_k}{k}$ is decreasing in k, as the distribution P_1 decreases stochastically then the FDR increases.

Corollary 7.5

Assume that the assumptions 1-3 from Theorem 7.4 are fulfilled. Further, assume that under global null hypothesis $FWER(\tau_c) = \alpha$ holds. Then, we have

1. *If the ratio $\frac{c_k}{k}$ is increasing in k, then $FDR(\tau_c) \geqslant \frac{n_0}{n}\alpha$.*

2. *If the ratio $\frac{c_k}{k}$ is decreasing in k, then we have $FDR(\tau_c) \leqslant \frac{n_0}{n}\alpha$.*

Thereby τ_c is the stopping rule of the SU procedure which uses critical values $0 \equiv c_0 \leqslant c_1 \leqslant c_2 \leqslant ... \leqslant c_n$.

Proof. By Theorem 7.4 the $\mathrm{FDR}(\tau_c)$ becomes minimal (maximal) under $P_1 = U(0,1)$, if the ratio $\frac{c_k}{k}$ decreases (increases, respectively). Further, due to Theorem 7.1, we have $\mathrm{FDR}(\tau_c) = \alpha\frac{n_0}{n}$ under assumption of Corollary 7.5 which completes the proof. ■

Remark 7.6

In case of the BH critical values $c_i = \frac{i\alpha}{n}$, $i = 1, ..., n$, Theorem 7.1 already yields: since for $n_0 = n$ $FWER(\tau_c) = \alpha$ is valid, we get $FDR(\tau_c) = \frac{n_0}{n}\alpha$ if the underlying p-values are i.i.d. $U(0,1)$-distributed.

In this chapter we discussed the case of uniformly distributed alternatives. We proved that if FWER of an MTP is controlled at α under the global null hypothesis, i.e., when $n = n_0$ holds, it implies that the FDR of this procedure is controlled at level $\alpha' \equiv \alpha'(n_0) = \frac{\alpha n_0}{n}$. This means also that BH SU procedure maximizes the FDR if underlying p-values are i.i.d. $\mathrm{U}(0,1)$-distributed.

Chapter 8

Upper bounds for the critical values for SD-procedures and maximal procedures

In this chapter the following assumptions are used.

U1 The vector of the true p-values $U = (U_1, .., U_{n_0}) \in \tilde{U}(n_0)$, with $\tilde{U}(m) = \{V : V = (V_1, ..., V_m), V_i \sim U(0,1), i = 1, ..., m\}$, i.e., U_i's , $i = 1, ..., n_0$, are uniformly distributed (with no assumptions about independence).

U2 $(f_1, ..., f_{n_1})$ and $(U_1, ..., U_{n_0})$ are independent random vectors.

U3 We consider an SD-procedure which uses the deterministic critical values $0 \leqslant a_1 \leqslant a_2 \leqslant ... \leqslant a_n \leqslant 1$, $a = (a_1, ..., a_n)$.

U4 The FDR and the FWER of the above procedure are denoted by FDR(a) and FWER(a).

For $y \in [0, 2]$ let us define the "mod 1"-operation as follows

$$\text{“}y \mod 1\text{”} = y - \mathbb{I}(y > 1).$$

The next example motivates the main theorem of this Chapter.

Example 8.1
(a) Let us consider a SD-procedure using critical values $a_1 \leqslant a_2 \leqslant ... \leqslant a_n$ and a multiple testing problem with $n_1 = 0$. Then by the Bonferroni inequality we have

$$\begin{aligned}\mathrm{FDR}(a) &= \mathrm{FWER}(a) = P(U_{1:n} \leqslant a_1) \\ &= P\left(\bigcup_{i=1}^{n}\{U_i \leqslant a_1\}\right) \leqslant n a_1.\end{aligned} \tag{8.1}$$

If all the sets $\{U_i \leqslant a_1\},\ i = 1, ..., n$, are disjoint, then we have an equality in (8.1). We will see that there exist some configurations of the p-values, so that due to the Bonferroni (in)equality in (8.1) $\mathrm{FDR}(a) = na_1$ is valid.

We want to construct some uniformly distributed random variables for which Bonferroni inequality (8.1) is sharp under the assumption $n_1 = 0$. We consider for $i = 1, ..., n$, and $U \sim U(0, 1)$ the following random variables

$$U_i = \left(U + \frac{i}{n}\right) \bmod 1 = U + \frac{i}{n} - \mathbb{I}(U + \frac{i}{n} > 1). \tag{8.2}$$

Firstly, note that $U_i,\ i = 1, ..., n$, are U(0,1)-distributed. Indeed, for $t \in [0, 1]$ we have

$$\begin{aligned}P\left(U_i \leqslant t\right) &= P\left(U \leqslant t - \frac{i}{n}, U \leqslant 1 - \frac{i}{n}\right) + P\left(U > 1 - \frac{i}{n}, U \leqslant 1 + t - \frac{i}{n}\right) \\ &= \left(t - \frac{i}{n}\right)\mathbb{I}(t > \frac{i}{n}) + \frac{i}{n}\mathbb{I}(t > \frac{i}{n}) + t\mathbb{I}(t \leqslant \frac{i}{n}) = t.\end{aligned}$$

Let us calculate the distance $\mid U_k - U_i \mid$. W.l.o.g. assume that $k > i$. We get

$$\mid U_k - U_i \mid = \left| \frac{k-i}{n} + \underbrace{\mathbb{I}(U + \frac{i}{n} > 1) - \mathbb{I}(U + \frac{k}{n} > 1)}_{\in\{-1,0\}} \right| \geqslant \frac{1}{n}. \tag{8.3}$$

Consequently if $a_1 \leqslant \frac{1}{n}$ holds, then all the sets $\{U_i \leqslant a_1\}, i = 1, ...n$, in (8.1) are disjoint which implies that $\mathrm{FDR}(a) = a_1 n$ must hold. And with the choice $a_1 = \frac{\alpha}{n}$, we have

$$\mathrm{FDR}(a) = nP(U_i \leqslant \frac{\alpha}{n}) = \alpha$$

We summarize it as Remark.

Remark 8.2

If some SD procedure with critical values $0 \leqslant a_1 \leqslant a_2 \leqslant ... \leqslant a_n \leqslant 1$ controls the FDR at level α for all possible parameter configurations then

$$a_1 \leqslant \frac{\alpha}{n}$$

must hold.

Example 8.1 (continued)

(b) Our second example is a problem of the testing of n, $n \geqslant 2$ hypotheses, thereby the number of alternatives is $n_1 = 1$. Further, let us consider an SD procedure which controls the FDR with $a_1 = \frac{\alpha}{n}$. We are interested in the case $f_1 \leqslant a_1$. For the case $f_1 > a_1$ we have, conditioned on f,

$$\mathrm{FDR}(a) \leqslant \mathrm{FWER}(a) \leqslant P(U_{1:n_0} \leqslant a_1) \leqslant \frac{n-1}{n}\alpha \leqslant \alpha.$$

Otherwise, for the case $f_1 \leqslant a_1$, we get:

$$\alpha \geqslant \mathbb{E}_f\left[\frac{V(a)}{R(a)}\mathbb{I}(V(a)>0)\right] = \mathbb{E}_f\left[\frac{V(a)}{V(a)+1}\mathbb{I}(V(a)>0)\right] \geqslant \frac{1}{2}\mathbb{E}_f\left[\mathbb{I}(V(a)>0)\right]. \quad (8.4)$$

Analogously to (8.2) we consider now for $n \geqslant 2$ the U(0,1)-distributed random variables V_i

$$V_i = \left(U + \frac{i}{n-1}\right) \bmod 1,\ i = 1, ..., n-1 \quad (8.5)$$

For such “true” p-values V_i we have analogously to (8.3):

$$\mid V_k - V_i \mid \geqslant \frac{1}{n-1}. \quad (8.6)$$

Consequently if $a_2 \leqslant \frac{1}{n-1}$ is valid, we get:

$$\alpha \geqslant \frac{1}{2}P(V>0) = \frac{1}{2}P\left(\bigcup_{i=1}^{n-1}\{V_i \leqslant a_2\}\right) = \frac{(n-1)a_2}{2}. \quad (8.7)$$

Hence, we can discuss $a_2 = \frac{2\alpha}{n-1}$ if $2\alpha \leqslant 1$.

(c) Now let us consider the case $n_0 = n-k$, $f_1 = f_2 = ... = f_k = 0$, $k \in \{1, ..., n\}$. In the same way as before we get

$$\alpha \geqslant \frac{1}{k+1}\max_{U_i \in \tilde{U}} P(U_{1:n-k} \leqslant a_{k+1}). \quad (8.8)$$

The consideration of the uniformly distributed random variables of the form

$$W_i = \left(U + \frac{i}{n-k}\right) \bmod 1, i = 1, ..., n-k \quad (8.9)$$

leads to

$$\frac{1}{k+1}P(W_{1:n_0} \leqslant a_{k+1}) = \frac{1}{k+1}(n-k)a_{k+1} = \alpha \text{ for } a_{k+1} = \frac{(k+1)\alpha}{n-k} \quad (8.10)$$

if $\alpha \leqslant \frac{1}{k+1}$ holds.

Now we are able to prove the following theorem which gives us a necessary condition for FDR-control.

Theorem 8.3 *(a) Let us assume that $n\alpha < 1$ and $n_1 \leqslant n$ hold. If the SD-procedure using critical values $b_1, ..., b_n$ controls the FDR at α for all configurations under assumptions U1-U2, then the critical values fulfill*

$$b_{k:n} \leqslant \frac{k\alpha}{n+1-k} = a_{k:n} \equiv a_{k:n}(\alpha), \;\; k = 1, ..., n. \tag{8.11}$$

(b) Let us assume that $f_1 = ... = f_{n_1} = 0$ and $\alpha \leqslant \frac{1}{n_1+2}$ hold. Then for all $n_0 > 1$ there exists some $U \in \tilde{U}(n_0)$, so that $FDR(a) = \alpha$ holds, where $a = (a_{k:n}(\alpha))_{k\in(1,...,n)}$.

Proof. (a) We have

$$\alpha \geqslant \mathbb{E}\left[\frac{V(b)\mathbb{I}(V(b)>0)}{R(b)}\right] \geqslant \mathbb{E}\left[\frac{V(b)\mathbb{I}(V(b)>0)}{n_1+V(b)}\right] \geqslant \frac{1}{n_1+1}P(V(b)>0) \tag{8.12}$$

The chain of inequalities (8.12) holds for all possible $U = (U_1, ..., U_{n_0})$, which belong to the class $\tilde{U}$. So we consider the p-values, for which the probability in (8.12) becomes maximal. With the choice

$$U_i = U + \frac{i}{n-n_1} \mod 1, \;\; i = 1, ..., n_0 \tag{8.13}$$

$$f_j = 0, j = 1, ..., n_1 \tag{8.14}$$

we get by (8.12):

$$\alpha \geqslant \frac{n_0}{n_1+1} b_{n_1+1}. \tag{8.15}$$

This implies $b_{n_1+1} \leqslant \frac{\alpha(n_1+1)}{n-n_1}$.

The number n_1 of the false null hypothesis is, in general, unknown and can take values from $\{0, ..., n\}$ and by the assumptions of this theorem it can take the values from the set $\{0, ..., n-1\}$. Therefore, if it is claimed that the SD-procedure with critical values $b_1 \leqslant ... \leqslant b_n$ controls the FDR at α for all possible distributions which belong to the class $\tilde{U}$, then we have

$$b_k \leqslant \frac{k\alpha}{n+1-k} =: a_k = a_k(\alpha). \tag{8.16}$$

Thus, we have proved part (a).
(b) If $f_1 = ... = f_{n_1} = 0$ and $U = (U_1, ..., U_{n_0}) \in \tilde{U}$ hold, then we have

$$\mathrm{FDR}(a) = \mathbb{E}\left[\frac{V(a)}{n_1 + V(a)}\mathbb{I}(U_{1:n} \leqslant a_{n_1+1})\right]. \tag{8.17}$$

Analogously to (8.13), we consider the random variables

$$U_i = U + \frac{i}{n_0 - 1} \mod 1,\ i = 1, ..., n_0. \tag{8.18}$$

Due to the condition $\alpha \leqslant \frac{1}{n_1+2}$, we get

$$a_{n_1+2} = \frac{\alpha(n_1+2)}{n+1-(n_1+2)} \leqslant \frac{1}{n_0-1} \leqslant \mid U_i - U_j \mid,\ i \neq j. \tag{8.19}$$

This implies that $V(a)$ in (8.17) can be either 0 or 1. Hence, it follows from (8.17):

$$\mathrm{FDR}(a) = \mathbb{E}\left[\frac{1}{n_1+1}\mathbb{I}(U_{1:n} \leqslant a_{n_1+1})\right] = \frac{1}{n_1+1}P(\bigcup_{j=1}^{n_0}\{U_j \leqslant a_{n_1+1}\}) = \alpha \tag{8.20}$$

by the choice of a_i's, $i = 1, ..., n$ and U_j's, $j = 1, ..., n_0$. ■

Remark 8.4

Finner and Gontscharuk [2013] have also considered configurations of the p-values which are weakly dependent and for which the Bonferroni inequality is sharp. As a consequence, they found a model of dependent p-values for which the FDR of the BH procedure is greater than α, even in the asymptotic sense.

Definition 8.5

We say that the procedure $\bar{\tau}$ is $\alpha-$maximal in sense of the FDR-control in some class Υ if

- *FDR($\bar{\tau}(p)$) $\leqslant \alpha$ holds, whenever the p-values belong to the class Υ,*
- *there exists some parameter configuration $\bar{p} \in \Upsilon$, so that FDR($\bar{\tau}(\bar{p})$) $= \alpha$ holds.*

Lemma 8.6 *(a) The SD procedure which uses the critical values $a_{k:n},\ k = 1, ..., n$, from (8.11) is $\alpha-$maximal for the class*

$$\begin{aligned}\Upsilon =&\{U \in \tilde{U}(n),\ \ \textit{the sets}\ \ \{U_i \leqslant a_{n_1+1}\},\ i = 1, ..., n_0,\ \ \textit{are disjoint},\\ &U\ \ \textit{and}\ \ f\ \ \textit{are independent random vectors}\}.\end{aligned}$$

(b) *For* $n = 2$ *the SD procedure which uses the critical values* $c_1 = 1 - \sqrt[2]{1-\alpha}, c_2 = a_{2:n} = 2\alpha$ *is* $\alpha-$*maximal, if the p-values belong to the class* $\Upsilon = \{U_i$ *are PRDS,* $U_i \sim U(0,1), i = 1, ..., n_0,$ U *and* f *are independent random vectors*$\}$.

(c) *The BH SU procedure is* $\alpha'-$*maximal,* $\alpha' \equiv \alpha'(n_0) = \alpha\frac{n_0}{n}$, *for the following class*

$$\Upsilon = \left\{ U = (U_1, ..., U_{n_0}) : \text{ the process } \left(\frac{V(t)}{t}\right)_{t\in(0,1)} \text{ is a reverse } \mathcal{RF}_t\text{-martingale} \right\}.$$

Remember that the reverse filtration $\mathcal{RF}_t$ *is defined as*

$$\mathcal{RF}_t = \sigma(\mathbb{I}_{(s,1)}(p_i), \forall 0 < t \leqslant s \leqslant 1, i = 1, ..., n).$$

Proof. (a) see proof of the Theorem 8.3.
(b) Let $n_0 = 1$. Consider two cases $f \leqslant c_1$ and $f > c_1$. For the first case we have $\text{FDR}(c) = \mathbb{E}_f\left[\frac{\mathbb{I}(U\leqslant c_2)}{2}\right] = \alpha$. For the second one $\text{FDR}(c) \leqslant \text{FWER}(c) = 1 - \sqrt[2]{1-\alpha} \leqslant \alpha$ holds.
If all hypotheses are true, i.e., $n_0 = n = 2$ holds, then by Lemma 2.12 (which states that the positive regression dependence on the subset implies the positive orthant dependence on the same subset) we have

$$\text{FDR}(c) = \text{FWER}(c) = P(U_{1:2} \leqslant c_1) \leqslant 1 - P(U_1 > c_1)P(U_2 > c_1) = \alpha. \tag{8.21}$$

Note that if the random variables U_1 and U_2 are independent, then we have an equality in (8.21).
(c) Let τ_{BH} denote the stopping rule of the BH procedure, i.e. $\tau_{BH} = \frac{j^*\alpha}{n}$, with $j^* = \max\{i, i = 0, ..., n : p_{i:n} \leqslant \frac{i\alpha}{n}\}$, $p_{0:n} \equiv 0$. We define $\tilde{\tau}_{BH} = \tau_{BH} \vee \frac{\alpha}{n}$. Note that $\tilde{\tau}_{BH}$ is a positive reverse stopping time. Then we get

$$\begin{aligned}
\mathbb{E}\left[\frac{V(\tau_{BH})}{R(\tau_{BH}) \vee 1}\right] &= \mathbb{E}\left[\frac{V(\tilde{\tau}_{BH})}{R(\tilde{\tau}_{BH}) \vee 1}\right] \\
&= \mathbb{E}\left[\frac{V(\tilde{\tau}_{BH})}{\tilde{\tau}_{BH}} \frac{\tilde{\tau}_{BH}}{R(\tilde{\tau}_{BH}) \vee 1}\right] \\
&= \frac{\alpha}{n}\mathbb{E}\left[\frac{V(\tilde{\tau}_{BH})}{\tilde{\tau}_{BH}}\right] = \frac{\alpha n_0}{n}.
\end{aligned}$$

The last equality is valid due to the Optional Sampling Theorem for reverse martingales (cf. Heesen, Janssen [2015]). ∎

In this chapter we proposed some necessary assumptions for FDR-control under general dependence. We proposed and discussed the consept of maximal procedures. In particular, based on results in Chapter 7 we showed that there exists a class of the distribution of p-values for which the BH SU procedure is maximal.

Appendix

Let S be a separable metric space. Suppose $\mathcal{F}_t$ is a filtration satisfying the usual conditions and X is a stochastic process taking values in S.
The following version of the optional sampling theorem can be found in Bauer [2002] (p. 456).

Theorem 8.7 (Optional sampling theorem for right-continuous super-martingales)
Let $X = (X_t)_{t\geqslant 0}$ be a cadlag $\mathcal{F}_t$-super-martingal and $p \in \mathbb{N}$ some finite number. Further let $\tau_1, ...\tau_p$ be bounded $\mathcal{F}_t-$stopping times with $\tau_1 \leqslant \tau_2 \leqslant ... \leqslant \tau_p$. Then $(X_{\tau_j}),\ j = 1, ..., p$ is a right-continuous $\mathcal{F}_{\tau_j}-$ super-martingale.

If B is a Borel subset of a metric space S, let

$$U_B = \inf\{t > 0 : X_t \in B\}.$$

U_B is called a first entry time of B.
The next Theorem is a variant of the Debut Theorem and can be found, for example, in Richard F. Bass [2011] (p.117).

Theorem 8.8 (Debut Theorem)
If X is a cadlag process taking values in S and B is a borel subset of S, then U_B is a stopping time.

Definition 8.9 (Comonotonicity)
Two real-valued measurable functions f and g are comonotone on $S \subset \Omega$ iff for any ω_1 and $\omega_2 \in S$,

$$(f(\omega_1) - f(\omega_2))(g(\omega_1) - g(\omega_2)) \geqslant 0.$$

Definition 8.10 (Class of comonotone functions)
The family of measurable real-valued functions $(f_i)_{i\in I=\{1,2,...,n\}}$ is a class of comonotone

functions if for all $i,j \in I$, for any ω_1 and $\omega_2 \in S$, $S \subset \Omega$,

$$(f_i(\omega_1) - f_i(\omega_2))(f_j(\omega_1) - f_j(\omega_2)) \geqslant 0.$$

Remark 8.11

The set of increasing (without strong increasing) functions is a comonotone class.

Definition 8.12

Let $x = (x_1, ..., x_n)$ be a point in $\mathbb{R}^n$ with pairwise distinct coordinates $x_i, i \in \{1, ..., n\}$. Let $x_{1:n} < x_{2:n} < ... < x_{n:n}$ be ordered coordinates of x.

- *For $i \in \{1, ..., n\}$ we call the value*

$$r_i = \sum_{j=1}^{n} \mathbb{I}(x_j \leqslant x_i)$$

  the rank of x_i. The vector $r = (r_1, ..., r_n)$ is called the rank vector of x.

- *The inverse permutation $d(x) = r^{-1}(x)$ is called the anti-rank vector of x, thereby $d(x) = (d_1(x), ..., d_n(x))$. The value $d_i(x)$ is called the anti-rank of i (the index which corresponds to the i. smallest observation).*

Example 8.13 (How to compute the coefficients b_i in proof of Lemma 5.32.)
From the formula

$$\frac{1}{\tilde{S}(t) + \tilde{n}_1 + k} = \sum_{i=\tilde{n}_1+1}^{n_1} b_i \mathbb{I}_{[0,f_i)}(t) + b_{\tilde{n}_1} \tag{8.22}$$

we get $b_{\tilde{n}_1} = \frac{1}{n_1+k}$ by the substitution $t = f_{n_1}$ in (8.22). Further, if the multiplicity of each $f_j, j \geqslant \tilde{n}_1 + 1, ..., n_1$ is equal to 1, we obtain

$$b_{n_1+1-l} = \frac{1}{(n_1 - l + k)(n_1 + 1 - l + k)}, \qquad l = 1, ..., n_1 - \tilde{n}_1, \tag{8.23}$$

due to the sequential substitution of $t = f_{n_1-l},\ l = 1, ..., n_1 - \tilde{n}_1$, in (8.22).

If for some $j = \tilde{n}_1 + 1, ..., n_1$ the relations $f_{j-1} < f_j = ... = f_{j+\mathrm{mult}(f_j)} < f_{j+\mathrm{mult}(f_j)+1}$ hold, where $\mathrm{mult}(f_j)$ is die multiplicity of f_j, we set $b_j = b_{j+1} = ... = b_{j+\mathrm{mult}(f_j)} = 0$ and compute $b_{j+\mathrm{mult}(f_j)+1}$ at the same way as in (8.23) by the substitution of $t = f_{j+\mathrm{mult}(f_j)}$ and $t = f_{j+\mathrm{mult}(f_j)+1}$ into (8.22). Consequently, we get

$$b_{j+\mathrm{mult}(f_j)+1} = \frac{\mathrm{mult}(f_{j+\mathrm{mult}(f_j)+1})}{(S(f_{j+\mathrm{mult}(f_j)}) + \tilde{n}_1 + k)(S(f_{j+\mathrm{mult}(f_j)}) + \mathrm{mult}(f_{j+\mathrm{mult}(f_j)+1}) + \tilde{n}_1 + k)},$$

The way of the computation of the coefficients $b_i, i = \tilde{n}_1, ..., n_1$ seems to be complex, but in reality it is just the comparison of the formula (8.23) at the points f_{j^*} and f_{j^*+1} for $j^* = \tilde{n}_1, ..., n_1$.

Literature

[1] Ahmed, A. N., Langberg, N. A., Leon, R. and Proschan, F. (1978) *Two concepts of positive dependence, with applications in multivariate analysis.* Tech. Report 78-6, Department of Statistics, Florida State University.

[2] Barlow, R. D. and Proschan, R. (1975) *Statistical theory of reliability and life testing: probability models.* Holt, Rinehart and Winston, New York.

[3] Bass, R. F. (2011) *Stochastic processes.* Cambridge University Press.

[4] Benjamini, Y. and Hochberg, Y. (1995) *Controlling the false discovery rate: a practical and powerful approach to multiple testing.* J. Roy. Statist. Soc. B **57** , 289-300.

[5] Benjamini, Y., Krieger, A. M. and Yekutieli, D. (2006) *Adaptive linear step-up procedures that control the false discovery rate.* Biometrika **93** , 491-507.

[6] Benjamini, Y. and Liu, W. (1999) *Step-down multiple hypothesis procedure that controls the false discovery rate under independence.* J. Statist. Planng Inf. **82**, 163-170.

[7] Benjamini, Y. and Yekutieli, D. (2001) *The control of the false discovery rate in multiple testing under dependency.* Ann. Statist. **29** , 1165-1188.

[8] Bonferroni, C. E. (1936) *Teoria statistica delle classi e calcolo delle probabilita.* Pubblicazioni del R Istituto Superiore di Scienze Economiche e Commerciali di Firenze **8**, 3-62.

[9] Dickhaus, T. (2008) *False discovery rate and asymptotics.* Dissertation, Heinrich-Heine-Universität Düsseldorf.

[10] Dickhaus, T. (2014) *Simultaneous statistical inference with applications in the life sciences.* Springer.

[11] Dudoit, S. and van der Laan, M. J. (2008) *Multiple testing procedures with applications to genomics.* Springer Series in Statistics.

[12] Finner, H., Dickhaus, T. and Roters, M. (2009) *On the false discovery rate and an asymptotically optimal rejection curve.* Ann. Statist. **37**, 596-618.

[13] Finner, H. and Gontscharuk, V. (2009) *Controlling the familywise error rate with plug-in estimator for the proportion of true null hypotheses.* J. R. Statist. Soc. B **71**, 1-18.

[14] Finner, H. and Roters, M. (2001) *On the false discovery rate and expected type I errors.* Biom. J. **43**, 985-1005.

[15] Finner, H. and Roters, M. (2002) *Multiple hypotheses testing and expected number of type I errors.* Ann. Statist. **30**, 220-238.

[16] Gavrilov, Y., Benjamini, Y. and Sarkar, S. K. (2009) *An adaptive step-down procedure with proven FDR control under independence.* Ann. Statist. **37**, 619-629.

[17] Genovese, C. and Wasserman, L. (2004) *A stochastic process approach to false discovery control.* Ann. Statist. **32**, 1035-1061.

[18] Ghosh, M. (2002) *Basu's theorem with applications: a personalistic review.* Indian J. Statist. **64**, 509-531.

[19] Gontscharuk, V. (2010) *Asymptotic and exact results on FWER and FDR in multiple hypotheses testing.* Dissertation, Heinrich Heine Universität Düsseldorf.

[20] Gontscharuk, V. and Finner, H. (2013) *Asymptotic FDR control under weak dependence: a counterexample.* Statist. Probab. Lett. **83**, 1888-1893.

[21] Gordon, A. Y. (2007) *Explicit formulas for generalized family-wise error rates and unimprovable step-down multiple testing procedures.* J. Statist. Planng Inf. **137**, 3497-3512.

[22] Gordon, A., Glazko, G., Qiu, X. and Yakovlev, A. (2007) *Control of the mean number of false discoveries, Bonferroni and stability of multiple testing.* Ann. Appl. Statist. **1**, 179-190.

[23] Hájek, J., (1968) *Asymptotic normality of simple linear rank statistics under alternatives.* Ann. Statist. **39**, 325-246.

[24] Heesen, P. and Janssen, A. (2014) *Inequalities for the false discovery rate (FDR) under dependence.* Preprint.

[25] Hochberg, Y. (1988) *A sharper Bonferroni procedure for multiple tests of significance.* Biometrika **75**, 800-802.

[26] Hochberg, Y. and Benjamini, Y. (1990) *More powerful procedures for multiple significance testing.* Statist. Med. **9**, 811-818.

[27] Hoffman-Jørgensen, J. (1994) *Probability with a view towards statistics.* Chapman and Hall, Probability series, Volume 1.

[28] Holland, P. and Rosenbaum, P. (1986) *Conditional association and unidimensionality in monotone latent variable models.* Ann. Statist. **14**, 1523-1543.

[29] Holm, S. (1979) *A simple sequentially rejective multiple test procedure.* Scand. J. Stat. **6**, 65-70.

[30] Hommel, G. (1988) *A stagewise rejective multiple test procedure based on a modified Bonferroni test.* Biometrica **75**, 383-386.

[31] Jogdeo, K. and Proschan, F. (1981) *Negative association of random variables, with applications.* FSU Statistics Report M-590, AFOSR Technical Report No. 81-127.

[32] Karlin, S. and Rinott, Y. (1980) *Classes of orderings of measures and related correlation inequalities. I. Multivariate totally positive distributions.* J. Multivariate Anal. **10**, 467?498.

[33] Lehmann, E. L. (1966) *Some concept of dependence.* Ann. Math. Statist. **37**, 1137-1153.

[34] Lehmann, E. L. and Romano, J. P. (2005) *Generalizations of the familywise error rate.* Ann. Statist. **37**, 1138-1154.

[35] Leek, J.T. and Storey, J. D. (2011) *The joint null criterion for multiple hypothesis tests.* Stat. Appl. Genet. Mol. Biol. 10-28.

[36] Neuvial, P. (2013) *Asymptotic results on adaptive false discovery rate controlling procedures based on kernel estimators.* JMLR **14**, 1423-1459.

[37] Neuvial, P. and Roquain, E. (2012) *On false discovery rate thresholding for classification under sparsity.* Ann. Statist. **40**, 2572-2600.

[38] Pitt, L.D. (1982) *Positively correlated normal variables are associated.* Ann. Probab. **10**, 496-499.

[39] Romano, J. P. and Shaikh, A. M. (2006) *Stepup procedures for control of generalizations of the familywise error rate.* Ann. Statist. **34**, 1850-1873.

[40] Roquain, E. and Villers, F. (2011) *Exact calculations for false discovery proportion with application to least favorable configurations.* Ann. Statist. **39**, 584-612.

[41] Sarkar, T.K. (1969) *Some lower bounds of reliability.* Technical report 124. Department of operations research and department of statistics, Stanford University.

[42] Sarkar, S. K. (1998) *Some probability inequalities for ordered MTP2 random variables: a proof of the Simes conjecture.* Ann. Statist. **26**, 494-504.

[43] Sarkar, S. K. (2002) *Someresults on false discovery rate in stepwise multiple testing procedures.* Ann. Statist. **30**, 239-257.

[44] Sarkar, S. K. and Guo, W. (2009) *On a generalized false discovery rate.* Ann. Statist. **37**, 1545-1565.

[45] Scheer, M. (2012) *Controlling the number of false rejections in multiple hypotheses testing.* Dissertation, Heinrich Heine Universität Düsseldorf.

[46] Shaked, M. and Spizzichino, F. (1998) *Positive dependence properties of conditionally independent random variables.* Math Oper Res **23** , 944-959.

[47] Shorack, G. P. and Wellner, J. A. (1986) *Empirical processes with applications to statistics.* John Wiley and Sons. Inc., New York.

[48] Sidak, Z. K. (1967) *Rectangular confidence regions for the means of multivariate normal distributions.* J. Am. Statist. Assoc. **62**, 626-633.

[49] Simes, R. J. (1986) *An improved Bonferroni procedure for multiple tests of significance.* Biometrika **73**, 751-754.

[50] Storey, J. D. (2002) *A direct approach to false discovery rates.* J. R. Statist. Soc. B **64**, 479-498.

[51] Storey, J. D., Taylor, J. E. and Siegmund, D. (2004) *Strong control, conservative point estimation and simultaneous conservative consistency of false discovery rates: a unified approach.* J. R. Statist. Soc. B **66**, 187-205.

[52] Tong, Y. L.(1990) *Multivariate Normal Distribution.* Springer Series in Statistics.

[53] van der Vaart, A. W. (2000) *Asymptotic statistics.* Cambridge Series in Statistical and Probabilistic Mathematics.

[54] Victor, N. (1982) *Exploratory data analysis and clinical research.* Methods. Inf. Med., **21**, 53-4.

[55] Xiao, J. (2008) *Integral and functional analysis.* Nova Science Publishers, New York.

[56] Yekutieli, D. (2008) *False discovery rate control for non-positively regression dependent test statistics.* J. Statist. Plann. Inference **138**, 405-415.

Die hier vorgelegte Dissertation habe ich eigenständig und ohne unerlaubte Hilfe angefertigt. Die Dissertation wurde in der vorgelegten oder ähnlichen Form noch bei keiner anderen Institution eingereicht. Ich habe bisher keine erfolgslosen Promotionsversuche unternommen.

Düsseldorf, den 12.02.2015

Julia Benditkis