Asymptotic and Exact Results on FWER and FDR in Multiple Hypotheses Testing

Inaugural - Dissertation

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

vorgelegt von

Veronika Gontscharuk

aus Charkow

Düsseldorf, Oktober 2010

Aus dem Institut für Biometrie und Epidemiologie des
Deutschen Diabetes-Zentrums, Leibniz-Institut an
der Heinrich-Heine-Universität Düsseldorf

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

Referent: Apl. Prof. Dr. H. Finner

Koreferenten: Prof. Dr. A. Janssen
Prof. Dr. G. Blanchard

Tag der mündlichen Prüfung: 27. Oktober 2010

Contents

Abstract

Nowadays, multiple hypotheses testing has become a promising area of statistics. In medicine, biology, pharmacology, epidemiology and even marketing, many hypotheses often have to be tested simultaneously. In some applications like genome-wide association studies, there may be several hundreds of thousands hypotheses to be tested.

An important concept in multiple testing is controlling a suitable Type I error rate. The Family-Wise Error Rate (FWER) is a classical error rate criterion and denotes the probability of one or more false rejections. Unfortunately, the FWER is often too restrictive if the number of hypotheses is very large. In 1995, Benjamini and Hochberg introduced an alternative error rate called the False Discovery Rate (FDR). The FDR denotes the expected proportion of falsely rejected hypotheses among all rejections. Typically, multiple test procedures controlling the FDR are more powerful than multiple tests controlling the FWER. However, if the number of true hypotheses is large and almost all hypotheses are true, procedures controlling the FWER may be a good alternative to tests controlling the FDR.

In this work we deal with multiple test procedures that control one of the aforementioned multiple error rates for independent test statistics and dependent ones as well. In the case of dependent test statistics, asymptotic considerations play a decisive role. Chapter 1 is an introduction into basic concepts and problems concerning multiple hypotheses testing.

In Chapter 2 we discuss a possibility to improve the power of some classical multiple tests controlling the FWER by applying a plug-in estimate for the number of true null hypotheses. We investigate several plug-in estimates and prove FWER control of Bonferroni, Šidàk and so-called step-down plug-in multiple test procedures. Moreover, we obtain some asymptotic results and compare the power of plug-in tests with the power of the corresponding classical procedures.

In Chapter 3 we restrict our attention to exact control of the FDR for step-up-down (SUD) test procedures. We give a recursive scheme which allows to calculate critical values such that the corresponding FDR equals the pre-specified FDR bounding curve. This scheme is numerically extremely sensitive so that computation of feasible solutions remains a challenging problem. We introduce alternative FDR bounding curves and study their connection to rejection curves as well as the existence of valid sets of critical values leading to these FDR bounding curves. In order to compute feasible critical values two further approaches are presented.

In Chapter 4 we focus on situations where some kind of weak dependence occurs. We consider models where the empirical cumulative distribution function of p-values corresponding to true null hypotheses is asymptotically bounded by the distribution function of a uniform variate. Important examples of weak dependence like block-dependence of test statistics and pairwise comparisons are investigated in more detail. We prove that large classes of plug-in tests and SUD procedures control the corresponding error rate under weak dependence at least asymptotically. Various numerical examples illustrate our theoretical results.

Zusammenfassung

In den letzten Jahrzehnten ist multiples Hypothesentesten ein vielversprechender Bereich der Statistik geworden. In der Medizin, Biologie, Pharmakologie, Epidemiologie und sogar im Bereich Marketing handelt es sich bei vielen Fragestellungen um multiple Testprobleme. Zum Beispiel werden in genomweiten Assoziationsstudien manchmal viele Hunderttausende von SNPs auf Assoziation mit einer Erkrankung getestet.

Ein wichtiges Konzept multiplen Hypothesentestens ist die Kontrolle eines geeigneten multiplen Fehlerkriteriums. Die bekannteste Fehlerrate ist die sogenannte Family Wise Error Rate (FWER). Damit wird die Wahrscheinlichkeit bezeichnet, dass mindestens eine Nullhypothese fälschlicherweise abgelehnt wird. Ist die Anzahl von Tests groß, so sind die meisten FWER kontrollierenden multiplen Testverfahren sehr konservativ. Im Jahr 1995 haben Benjamini und Hochberg vorgeschlagen, die False Discovery Rate (FDR) zu kontrollieren, d.h. den erwarteten Anteil fälschlich abgelehnter Nullhypothesen bzgl. aller abgelehnten Hypothesen. Typischerweise lehnen FDR kontrollierende Verfahren mehr Hypothesen ab als Prozeduren, die die FWER kontrollieren. Dennoch, die letzteren können eine gute Alternative zu FDR kontrollierenden Verfahren darstellen, falls die Anzahl der Tests groß ist und fast alle Hypothesen wahr sind.

In dieser Arbeit untersuchen wir multiple Testverfahren, die die FWER oder die FDR kontrollieren, sowohl für unabhängige als auch abhängige Teststatistiken. In dem abhängigen Fall spielen asymptotische Betrachtungen eine entscheidende Rolle. In Kapitel 1 werden Grundkonzepte und Problemstellungen des multiplen Testens eingeführt.

In Kapitel 2 wird die Güte einiger klassischer FWER kontrollierender Tests verbessert, indem die Anzahl aller Tests durch die geschätzte Anzahl wahrer Hypothesen bei der Berechnung kritischer Werte ersetzt wird. Wir untersuchen einige Schätzer für die Anzahl wahrer Hypothesen und beweisen FWER Kontrolle für Bonferroni, Šidàk und sogenannte step-down plug-in Tests. Wir präsentieren asymptotische Ergebnisse und vergleichen Güten von neuen und klassischen Tests.

In Kapitel 3 wird der Fokus auf step-up-down Testsverfahren gelegt, die die FDR kontrollieren. Wir präsentieren ein rekursives Schema zur Berechnung zulässiger kritischer Werte, die zu vorher festgesetzten Schranken für die FDR führen. Das Schema ist numerisch sehr sensibel, so dass die Existenz einer zulässigen Lösung ein anspruchsvolles Problem ist. Wir führen neue sogenannte FDR beschränkende Kurven ein und untersuchen sowohl deren Zusammenhang zu Ablehnkurven als auch die Lösbarkeit des rekursiven Schemas für diese FDR beschränkende Kurven. Außerdem werden weitere Verfahren zur Berechnung zulässiger kritischer Werte vorgestellt.

Kapitel 4 widmet sich abhängigen Teststatistiken, die eine sogenannte "weak dependence" Bedingung erfüllen. Wir betrachten Modelle, bei denen die empirische Verteilungsfunktion von p-Werten unter Nullhypothesen asymptotisch nicht oberhalb der Winkelhalbierenden verläuft. Blockabhängigkeit von Teststatistiken und Paarvergleiche sind die bedeutendsten Beispiele für "weak dependence" und werden ausführlich untersucht. Wir prüfen FWER und FDR Kontrolle für große Klassen von plug-in und SUD Tests. Verschiedene numerische Beispiele veranschaulichen die theoretischen Ergebnisse.

Acknowledgments

There are many people who I would like to thank for their support during the preparation process of this thesis.

First and foremost, I want to express my sincere appreciation to my advisor Prof. Dr. Helmut Finner for his invaluable support and continuous encouragement over the last years. Many extensive discussions filled with helpful suggestions made it possible for me to complete this work.

Warm thanks are due to many colleagues at the German Diabetes Center, especially Klaus Straßburger and Marsel Scheer, who always had an open door for discussing problems, Thorsten Dickhaus, meanwhile working at Humboldt-University Berlin, for exchanging ideas and some fruitful joint work, and Sandra Landwehr for her careful proof reading.

Special thanks are due to the Director of the Institute of Biometrics and Epidemiologie, Prof. Dr. Guido Giani, and also to Prof. Dr. Arnold Janssen and Prof. Dr. Gilles Blanchard for writing the referee reports on this thesis. I am also very grateful for the financial support of the Deutsche Forschungsgemeinschaft (DFG).

Finally, I thank my family for their love and understanding.

List of Abbreviations and Symbols

AORC	*Asymptotically Optimal Rejection Curve*
$a \vee b$	$\max(a, b)$
BPI	*Bonferroni plug-in*
cdf	*Cumulative distribution function*
F_{t_ν}	*Cdf of a univariate (central)* t*-distribution with* ν *degrees of freedom*
Cov	*Covariance*
DU	*Dirac-uniform*
ecdf	*Empirical cumulative distribution function*
$F_\infty(t\|\zeta)$	$1 - \zeta + \zeta t$
$\hat{F}_n$	*Ecdf of* p*-values*
$\hat{F}_{n,0}$	*Ecdf of* p*-values corresponding to true null hypotheses*
$\hat{F}_{n,1}$	*Ecdf of* p*-values corresponding to alternatives*
FDR	*False Discovery Rate*
Φ	*Standard Gaussian cdf*
ϕ	*Standard Gaussian pdf*
FWER	*Family-Wise Error Rate*
I_n	$\{1, \ldots, n\}$
$I_{n,0}$	$\{i \in I_n : \ H_i \text{ is true}\}$
$I_{n,1}$	$\{i \in I_n : \ H_i \text{ is false}\}$
$I(p \leq t)$	*Indicator function of the event* $\{p \leq t\}$
iid	*independent and identically distributed*
$\lceil x \rceil$	*Largest integer smaller than or equal to* x
LFC	*Least Favourable Configuration*

$\lfloor x \rfloor$	*Smallest integer larger than or equal to* x
LSU	*Linear step-up*
$N(\mu, \sigma^2)$	*Normal distribution with mean* μ *and variance* σ^2
$\mathbb{N}$	*Set of natural numbers*
pdf	*Probability density function*
PRDS	*Positive Regression Dependency on Subset*
R_n	$\#\{i \in I_n : \ H_i \text{ is rejected}\}$
$R_n(t)$	$\#\{i \in I_n : \ p_i \leq t\}$
$O(g(n))$	$\{f(n) \ : \ \exists\, C > 0 : \ \exists\, N_0 \in \mathbb{N} : \ \forall\, n \geq N_0 : \ 0 \leq f(n) \leq Cg(n)\}$
$o(g(n))$	$\{f(n) \ : \ \forall\, C > 0 : \ \exists\, N_0 \in \mathbb{N} : \ \forall\, n \geq N_0 : \ 0 \leq f(n) \leq Cg(n)\}$
OB	*Oracle Bonferroni*
$\mathbb{R}$	*Set of real numbers*
SD	*Step-down*
SDPI	*Step-down plug-in*
SU	*Step-up*
SUD	*Step-up-down*
U([0,1])	*Uniform distribution on the interval* $[0, 1]$
V_n	$\#\{i \in I_{n,0} : \ H_i \text{ is rejected}\}$
$V_n(t)$	$\#\{i \in I_{n,0} : \ p_i \leq t\}$
WD	*Weak dependence*

Overview

In various applications of statistics, simultaneous testing of a large number of hypotheses is everyday life. For example, in multiple endpoints studies in clinical trials, a new treatment has to be compared with an existing one in terms of a number of measurements (endpoints). In genome-wide association studies, sometimes hundreds of thousands of single-nucleotide polymorphisms (SNPs) have to be tested simultaneously. Other applications in multiple testing can be found in medicine, biology, pharmacology, epidemiology, bioinformatics and even marketing.

Typically, one is not interested in whether or not all null hypotheses are true. It is important to make decisions about individual hypotheses, that is, we want to decide which hypotheses are false. Clearly, if we carry out many statistical tests simultaneously, the probability of making false rejections increases with the number of tests. The aim of a multiple test procedure is to control a suitable Type I error rate and to maximise the number of correct rejections at the same time. Note that a single test controls the probability of a false rejection (Type I error). In the multiple case, the Type I error rate can be generalised in different ways.

One of the well-known multiple error measures is the so-called Family-Wise Error Rate (FWER), that is, the probability of falsely rejecting at least one true null hypothesis. Up to a few years ago, the FWER was the most used error rate criterion. Unfortunately, multiple test procedures controlling the FWER require that individual tests are performed at a lower level than the pre-specified FWER-level, which often results in a low power. Instead of controlling the FWER, one can control the False Discovery Rate (FDR) introduced in Benjamini and Hochberg [1995]. The FDR is the expected proportion of falsely rejected null hypotheses among all rejected hypotheses. Since the FDR is less restrictive than the FWER, the FDR has become an attractive error measure especially if the number of hypotheses is large. On the other hand, if the number of null hypotheses increases and the proportion of true null hypotheses converges to 1, multiple test procedures controlling the FWER may be good alternatives to multiple tests controlling the FDR.

In this dissertation we deal with both types of multiple test procedures, that is, multiple tests controlling the FWER and others controlling the FDR. We consider independent test statistics and dependent ones as well, where the latter often occur in applications. Moreover, because of massive multiplicity appearing in many applications, asymptotic investigations feature prominently in this work. This dissertation is organised as follows.

Chapter 1 serves as an introduction for this treatise. A general multiple-testing problem and possible error rate criteria are presented. We consider various classical multiple test procedures and show under which conditions these tests control the corresponding error rate. We give some notations and definitions and describe the problems that are considered in further chapters.

In Chapter 2 we discuss a special approach of improving the power of some classical multiple test procedures controlling the Family-Wise Error Rate (FWER). This approach is based on plug-in estimates for the number of true null hypotheses. Although, the idea of plug-in multiple test procedures is not new, cf. e.g. Schweder and Spjøtvoll [1982], Hochberg and Benjamini [1990] or Benjamini and Hochberg [2000], no theoretical results seem to be available until recently. In this chapter we investigate several plug-in estimates and prove FWER control of Bonferroni and so-called Šidàk plug-in multiple tests. Moreover, we show that suitable plug-in step-down tests also yield FWER control. Thereby, we obtain some asymptotic results and provide some power considerations. Some of the main results of this chapter are published in Finner and Gontscharuk [2009]. Independently, similar findings concerning FWER control of special plug-in tests with respect to a specific mixture model were obtained in Guo [2009].

Chapter 3 deals with exact control of the False Discovery Rate (FDR) for step-up-down (SUD) test procedures related to the Asymptotically Optimal Rejection Curve (AORC). The AORC was introduced in Finner *et al.* [2009] and has the property to exhaust the pre-specified FDR level α under extreme parameter configurations, at least asymptotically. Since SUD procedures based on this curve do not control the FDR for a finite number of hypotheses, we propose various methods for the computation of critical values leading to finite FDR control. Finner *et al.* [2009] propose an upper bound for the FDR of an SUD test which is exact for an SU test in so-called Dirac-uniform models. We give a recursive scheme which allows to calculate critical values such that the corresponding FDR equals the pre-specified FDR bounding curve and discuss its solvability. Another interesting approach, which yields a set of critical values such that the corresponding FDR is close to α, is given by an iterative method based on the fixed point theorem. The main results in this chapter are submitted for publication.

In Chapter 4 we investigate multiple test procedures based on dependent test statistics. We introduce a modified version of weak dependence and present a simple condition that is equivalent to some boundary case of this modified version of weak dependence. We show that plug-in procedures and SUD tests control the corresponding error rate under weak dependence at least asymptotically. Assuming some type of weak dependence between p-values, one of the main problems with respect to asymptotic FDR control occurs if the proportion of rejected hypotheses tends to 0. We prove asymptotic FDR control for a broad class of step-wise multiple tests with respect to some restrictions on a given parameter space guaranteeing that the proportion of rejected hypotheses is asymptotically bounded away from 0. An important boundary case of weak dependence is given by dependent p-values such that the asymptotic empirical distribution function (ecdf) of those p-values that correspond to true null hypotheses, coincides with the asymptotic ecdf of independently uniformly distributed p-values. This case of weak dependence is asymptotically least favourable for the FWER of suitable multiple tests. Moreover, if in addition to this kind of weak dependence, p-values under alternatives follow a Dirac distribution with point mass in 0, these p-values are asymptotically least favourable for the FDR of special step-wise procedures satisfying some power requirement. We consider different types of dependence ensuring weak dependence. Block-dependence of test statistics and pairwise comparisons will be investigated in more detail. Thereby, various numerical examples illustrate our theoretical results.

Some definitions of different types of convergence and relevant theorems are summarised in an Appendix.

Most issues investigated in this treatise except the plug-in methods in Chapter 2 were raised in a research project sponsored by the Deutsche Forschungsgemeinschaft (DFG), grant No. FI 524/3-1, under the responsibility of my advisor Apl. Prof. Dr. Helmut Finner and Prof. Dr. Guido Giani.

Chapter 1

General framework for multiple testing

In this chapter we briefly introduce the multiple testing framework and some basic concepts. Section 1.1 describes the general setup and provides basic definitions and notation. In Section 1.2 we review the concept of the Family-Wise Error Rate (FWER) and introduce some well known elementary multiple test procedures. Moreover, we introduce the concept of rejection curves and critical value functions as a useful tool in multiple testing. Section 1.3 is concerned with the false discovery rate (FDR) criterion introduced by Benjamini and Hochberg [1995]. We discuss different multiple test procedures controlling some error rates and show how multiple tests can be defined in terms of rejection curves and crossing points. In Section 1.4 we introduce a set of possible assumptions for deriving theoretical results and define Dirac-uniform models which provide least favourable parameter configurations with respect to different error rates under several conditions.

1.1 Introduction to basic concepts

First of all, we introduce the notation of our general setup which applies in this work.

Notation 1.1 (General setup)
For some statistical experiment $(\Omega, \mathcal{A}, \{\mathbb{P}_\vartheta : \vartheta \in \Theta\})$ we consider the general problem of simultaneously testing a finite number of hypotheses H_i, $i \in I_n$, where $I_n = \{1, \ldots, n\}$. Hypotheses are interpreted as subsets of the underlying parameter space Θ, and it will be assumed that $\emptyset \neq H_i \subset \Theta$, $i \in I_n$. The corresponding alternatives are given by $\Theta \setminus H_i$. Let p_i, $i \in I_n$, be p-values for testing H_i. Suppose $p_i : (\Omega, \mathcal{A}) \longrightarrow ([0,1], \mathcal{B})$, $i \in I_n$, where $\mathcal{B}$ denotes the Borel-σ-field over $[0,1]$. For $\vartheta \in \Theta$, $\mathbb{P}_\vartheta$ denotes the underlying probability measure. As usual, let a p-value p_i satisfy $0 < \mathbb{P}_\vartheta(p_i \leq x) \leq x$ for all $\vartheta \in H_i$, $i \in I_n$, and $x \in (0,1]$, i.e. p-values under null hypotheses are uniformly distributed or stochastically larger than a uniform variate. Let $n_0 = n_0(n, \vartheta)$ denote the number of true null hypotheses and $I_{n,0} = I_{n,0}(\vartheta) = \{i \in I_n : \vartheta \in H_i\}$ and $I_{n,1} = I_{n,1}(\vartheta) = I_n \setminus I_{n,0} = \{i \in I_n : \vartheta \notin H_i\}$ denote the index set of true and false null hypotheses, respectively. Furthermore, $n_0 = |I_{n,0}(\vartheta)|$. Let $n_1 = n_1(n, \vartheta)$ be the number of false

	Test decision		
Hypothesis	0	1	
true	U_n	V_n	n_0
false	T_n	S_n	n_1
	$n - R_n$	R_n	n

Table 1.1: Outcomes in testing n hypotheses.

hypotheses, i.e. $n_1 = n - n_0 = |I_{n,1}(\vartheta)|$*. Below, we write* n_0*,* $n_0(n)$ *or* $n_0(\vartheta)$ *(and* n_1*,* $n_1(n)$ *or* $n_1(\vartheta)$*, resp.) depending on which parameter dependence we would like to point out. Finally, let* $\varphi = (\varphi_i : i \in I_n)$ *denote a non-randomised multiple test procedure for* H_i*,* $i \in I_n$*. For* $i \in I_n$*, a hypothesis* H_i *is rejected if and only if* $\varphi_i = 1$*.*

Table 1.1 shows the possible outcomes in testing n hypotheses. The number of all rejections is given by R_n, the number of false (true) rejections is denoted by V_n (S_n, resp.) and the number of correctly (falsely) accepted hypotheses is given by U_n (T_n, resp.). Note that V_n, S_n, U_n and T_n are not observable and, typically, n_0, n_1 are unknown.

By testing a single hypothesis, the probability of a false rejection (Type I error) has to be controlled while we are looking for a test that possibly minimises the probability of a false rejection (Type II error).

In the multiple testing case, if we perform each individual test φ_i, $i \in I_n$, at level α, the corresponding multiple test $\varphi = (\varphi_i : i \in I_n)$ can reject a huge number of true null hypotheses. For example, when testing $n = 500000$ null hypotheses at level $\alpha = 0.05$ (e.g., in genome-wide association studies, several hundreds of thousands of single-nucleotide polymorphisms (SNPs) have to be tested simultaneously), around $V_n = 25000$ false rejections are expected if almost all hypotheses are true. In real applications, this is completely out of the question.

The Type I error rate can be generalised for multiple testing in different ways. Typically, all generalisations involve the number of false rejections V_n. First, we consider those error rate criteria which are only based on the distribution of V_n. One of the classical multiple error rates is the Family-Wise Error Rate (FWER), i.e. the probability of at least one false rejection, i.e.

$$\text{FWER} = \mathbb{P}_\vartheta(V_n \geq 1).$$

In the next section, the FWER will be considered in detail.

One can generalise the FWER as follows. For a fixed $k \in \mathbb{N}$ the generalised FWER denotes the probability of rejecting at least k true null hypotheses, that is,

$$g\text{FWER}(k) = \mathbb{P}_\vartheta(V_n \geq k).$$

Obviously, the case $k = 1$ reduces to the usual FWER.

Another possibility is to control the False Discovery Proportion (FDP), which is defined as the number of false rejections V_n divided by the number of all rejections R_n and we set $\text{FDP} = 0$

if $R_n = 0$, i.e.

$$\text{FDP} = \frac{V_n}{R_n \vee 1}.$$

For a given $\gamma \in (0,1)$, one wishes to control $\mathbb{P}_\vartheta(\text{FDP} > \gamma)$ at some pre-specified level α. More information about $g\text{FWER}(k)$ and FDP control can be found in Lehmann and Romano [2005].

There is no doubt that the latter error measure was motivated by the False Discovery Rate (FDR) introduced in Benjamini and Hochberg [1995]. The FDR is defined as the expected FDP, i.e.

$$\text{FDR} = \mathbb{E}_\vartheta[\text{FDP}].$$

When all null hypotheses are true, i.e. $n = n_0$, controlling the FWER and the FDR are equivalent. In that case either FDP $= 0$ (if $V_n = 0$) or FDP $= 1$ (if $V_n > 0$, since all rejections are false), and the expected ratio is equal to the probability of any false rejection. However, if $n_1 > 0$ and the number S_n of truly rejected hypotheses is greater than 0, the FDP is either 0 (if $V_n = 0$) or $0 <$ FDP < 1 (if $V_n > 0$), and the expected ratio is smaller than the probability of at least one false rejection. In those cases the FDR is smaller than the FWER, and controlling the FDR at a pre-specified level α can result in fewer Type II errors than controlling the FWER at the same level α. The power increases when more alternative hypotheses are true.

There are many other possibilities to generalise the Type I error rate in the multiple case, see, for instance, Sarkar and Guo [2009]. In this work, however, we restrict our attention to the FWER and the FDR.

1.2 Family-Wise Error Rate

As mentioned before, by testing $n \geq 2$ null hypotheses quite a few false rejections (Type I errors) are possible. The probability for at least one false rejection among H_i, $i \in I_n$, is given by the so-called **Family-Wise Error Rate** (**FWER**), which is a well-known error rate criterion. For a fixed $\vartheta \in \Theta$ and a given test φ we define the number of false rejections by

$$V_n = V_n(\varphi) = \#\{i \in I_{n,0} : \ H_i \text{ is rejected}\}.$$

Note that V_n is typically unknown. The actual FWER of a multiple test φ, given a $\vartheta \in \Theta$, can formally be expressed by

$$\text{FWER}_\vartheta(\varphi) = \mathbb{P}_\vartheta\left(V_n \geq 1\right).$$

A multiple test φ controls the FWER at pre-specified level $\alpha \in (0,1)$ if

$$\sup_{\vartheta \in \Theta} \text{FWER}_\vartheta(\varphi) \leq \alpha.$$

The **Bonferroni** test is a classical multiple test procedure controlling the FWER. Thereby all individual tests φ_i, $i \in I_n$, are performed at level α/n, that is, a H_i is rejected if and only if $p_i \leq \alpha/n$. Since

$$\text{FWER}_\vartheta(\varphi) \leq \sum_{i \in I_{n,0}} \mathbb{P}_\vartheta\left(p_i \leq \frac{\alpha}{n}\right) \leq \frac{n_0}{n}\alpha,$$

the Bonferroni test always controls the FWER at level α under the general setup, that is, p-values under nulls are uniformly distributed or stochastically larger than a uniform variate, and it does not matter whether the p-values are independent or not. Unfortunately, the threshold α/n is very small if the number of hypotheses n is large. Obviously, this results in low power for individual hypotheses of the Bonferroni test.

A possible improvement of the classical Bonferroni test is the **oracle Bonferroni (OB)** test, where each φ_i, $i \in I_n$, is carried out at level α/n_0. Clearly, the oracle Bonferroni test also controls the FWER under the same assumptions.

If p-values are independent, then for a fixed threshold $\alpha' \in (0,1)$ we get

$$\mathbb{P}_\vartheta\Big(\bigcap_{i\in I_{n,0}} \{p_i > \alpha'\}\Big) = \prod_{i\in I_{n,0}} \mathbb{P}_\vartheta(p_i > \alpha') \geq (1-\alpha')^{n_0}.$$

The expression $\mathbb{P}_\vartheta(\bigcap_{i\in I_{n,0}}\{p_i > \alpha'\})$ can be interpreted as $1 - \mathrm{FWER}_\vartheta(\varphi)$, where φ is the multiple test such that each φ_i, $i \in I_n$, is performed at level α'. Then φ controls the FWER at level α if $1-\alpha \leq (1-\alpha')^{n_0}$, which is equivalent to $\alpha' \leq 1-(1-\alpha)^{1/n_0}$. Thus if p-values corresponding to true null hypotheses are independent, the **Šidàk** test, which rejects each hypothesis H_i if $p_i \leq 1-(1-\alpha)^{1/n}$ for $i \in I_n$, controls the FWER at level α. Moreover, if all hypotheses are true and the corresponding p-values are iid uniformly distributed, then the FWER for the Šidàk test is exactly α. Similar to the Bonferroni test case, the **oracle Šidàk** test with the threshold $1-(1-\alpha)^{1/n_0}$ controls the FWER under the same condition as the Šidàk test.

The disadvantage of the considered oracle tests is that the number of true null hypotheses n_0 is typically unknown. In Chapter 2 we introduce **Bonferroni plug-in (BPI)** procedures related to the OB test or the oracle Šidàk test based on an estimator for n_0. It will be shown that the FWER of a BPI test is controlled under suitable assumptions.

The test procedures described before provide examples of single-parameter adjustment procedures, meaning that a hypothesis is rejected if its corresponding p-value is not greater than the common threshold (which is α/n for the Bonferroni case and α/n_0 for the OB test). Now we briefly describe some stepwise multiple test procedures, which are often uniformly more powerful than their single-parameter counterparts. Firstly, we introduce **step-down (SD)** test procedures. An SD procedure for testing n hypotheses can be defined in terms of n critical values

$$0 < \alpha_{1:n} \leq \ldots \leq \alpha_{n:n} < 1 \tag{1.1}$$

and works as follows. Let $p_{1:n} \leq \ldots \leq p_{n:n}$ be the ordered p-values and denote the corresponding hypotheses by $H_{(1)}, \ldots, H_{(n)}$. Then a hypothesis $H_{(i)}$, $i \in I_n$, is rejected if and only if $p_{j:n} \leq \alpha_{j:n}$ for all $j \leq i$, otherwise it cannot be rejected. In other words, the SD procedure starts with the most significant p-value (i.e. $p_{1:n}$) by comparing it with the smallest critical value (i.e. $\alpha_{1:n}$). If $p_{1:n} > \alpha_{1:n}$, then all hypotheses are accepted, otherwise we reject $H_{(1)}$ and compare $p_{2:n}$ with $\alpha_{2:n}$. If $p_{2:n} > \alpha_{2:n}$, then $H_{(2)}, \ldots, H_{(n)}$ are accepted, otherwise we reject $H_{(2)}$ and compare $p_{3:n}$ with $\alpha_{3:n}$ and so on.

One example for an SD procedure is the **Bonferroni–Holm** step-down test with critical values $\alpha_{i:n} = \alpha/(n-i+1)$, $i \in I_n$. It controls the FWER at level α. As in the case of the Bonferroni test, control of the FWER of the Bonferroni–Holm procedure is guaranteed for any type of dependence of p-values. Moreover, it is well-known that the Bonferroni–Holm SD procedure is uniformly more powerful than the classical Bonferroni single-parameter procedure.

A further type of stepwise procedures are **step-up (SU)** tests starting with the least significant p-value ($p_{n:n}$). For a given set of critical values (1.1), reject all hypotheses if $p_{n:n} \leq \alpha_{n:n}$. Otherwise, for $i \in I_n$ reject hypotheses $H_{(1)}, \ldots, H_{(i)}$ if $p_{i:n} \leq \alpha_{i:n}$ and $p_{j:n} > \alpha_{j:n}$ for all $j \geq i+1$. Note that an SU test rejects at least as many hypotheses as the corresponding SD test with the same set of critical values.

The **Hochberg** test is an SU test with critical values $\alpha_{i:n} = \alpha/(n-i+1)$, $i \in I_n$, i.e. an SU test with the same critical values as in the Bonferroni–Holm test, cf. Hochberg [1988]. Obviously, the Hochberg SU procedure is more powerful than the Bonferroni–Holm SD test. On the other hand, the Hochberg procedure controls the FWER under more restrictive assumptions, for example, if test statistics are independent or multivariate totally positive of order 2 or a scale mixture thereof, cf. Sarkar [1998]. A further example for an SU procedure is the **Simes** test with critical values $\alpha_{i:n} = i\alpha/n$, $i \in I_n$. Simes [1986] showed that his procedure controls the FWER for independent test statistics under the global null hypothesis, that is, $H_0 = \bigcap_{i=1}^{n} H_i$.

Now we introduce the notation of rejection curves and show that various multiple tests can be implemented in terms of crossing points between the corresponding rejection curve and the empirical distribution function of p-values. Let φ be a multiple test defined in terms of critical values (1.1). Thereby, the critical values may be defined in terms of a **critical value function** $\rho : [0,1] \to [0,1]$ such that ρ is non-decreasing and continuous, $\rho(0) = 0$ and $\alpha_{i:n} = \rho(i/n)$, $i \in I_n$. Moreover, r defined by $r(t) = \inf\{u : \rho(u) = t\}$ for $t \in [0,1]$, will be called a **rejection curve**. For example, $r(t) = (t(n+1) - \alpha)/(nt)$ is the rejection curve of the Bonferroni-Holm and Hochberg test procedures.

Denoting the empirical cumulative distribution function (ecdf) of the p-values by

$$\hat{F}_n(t) = \sum_{i=1}^{n} I(p_i \leq t),$$

Sen [1999] mentioned the following relationship

$$p_{i:n} \leq \alpha_{i:n} \text{ if and only if } \hat{F}_n(p_{i:n}) \geq r(p_{i:n}).$$

We say a point $t = \alpha_{i:n}$ is a **crossing point** between $\hat{F}_n$ and r, if it satisfies $\hat{F}_n(p_{i:n}) \geq r(p_{i:n})$ and $\hat{F}_n(p_{i+1:n}) < r(p_{i+1:n})$ for $i \in I_{n-1}$ or $\hat{F}_n(p_{n:n}) \geq r(p_{n:n})$ for $i = n$. If we define t^* as the smallest (or largest) crossing point between $\hat{F}_n$ and r, it follows that t^* is a random threshold of the SD (or SU) procedure based on r. Thereby, this SD (or SU) test rejects all H_i, $i \in I_n$ with $p_i \leq t^*$. Note that SU and SD procedures belong to the class of step-up-down (SUD) procedures which will be introduced and investigated in Chapter 3.

1.3 False Discovery Rate

When the number of hypotheses n is in the tens or hundreds of thousands, control of the FWER becomes too rigorous so that individual tests φ_i, $i \in I_n$, have little chance to reject any hypothesis. A radical weakening of the FWER is the **False Discovery Rate (FDR)**, which was proposed by Benjamini and Hochberg [1995] as follows. For a fixed $\vartheta \in \Theta$ and a given test φ let

$$R_n = R_n(\varphi) = \#\{i \in I_n : \ H_i \text{ is rejected}\}$$

be the number of all rejections. Define the **false discovery proportion** as

$$\text{FDP}_\vartheta(\varphi) = \frac{V_n}{R_n \vee 1}.$$

The actual FDR is given by

$$\text{FDR}_\vartheta(\varphi) = \mathbb{E}_\vartheta[\text{FDP}_\vartheta(\varphi)] = \mathbb{E}_\vartheta\left[\frac{V_n}{R_n \vee 1}\right].$$

Alternatively, the actual FDR can be expressed as

$$\text{FDR}_\vartheta(\varphi) = \mathbb{E}_\vartheta\left[\frac{V_n}{R_n}|R_n > 0\right] \cdot \mathbb{P}_\vartheta(R_n > 0).$$

We say that φ controls the FDR at level $\alpha \in (0,1)$ if

$$\sup_{\vartheta \in \Theta} \text{FDR}_\vartheta(\varphi) \leq \alpha.$$

When all hypotheses are true, that is, $n = n_0$, we obtain

$$\text{FDR}_\vartheta(\varphi) = \mathbb{P}_\vartheta(R_n > 0) = \mathbb{P}_\vartheta(V_n > 0) = \text{FWER}_\vartheta(\varphi).$$

In general, since $V_n/(R_n \vee 1) \leq 1$, we get $V_n/(R_n \vee 1) \leq I(V_n \geq 1)$ and consequently

$$\text{FDR}_\vartheta(\varphi) \leq \text{FWER}_\vartheta(\varphi),$$

and typically this inequality is strict except when all hypotheses are true. If a test procedure φ controls the FWER, then φ implies FDR control. On the other hand, if $\text{FDR}_\vartheta(\varphi) \leq \alpha$ the FWER may be greater than α. Thereby, FDR control allows more false rejections (i.e. the number of true null hypotheses which are rejected) than FWER control especially if the number of true rejections (i.e. the number of rejected false hypotheses) is large so that the FDR is more liberal (in the sense of permitting more rejections) than the FWER.

One of the best known multiple-testing procedures controlling the FDR is the **linear step-up (LSU)** procedure proposed and investigated in Benjamini and Hochberg [1995]. The original LSU procedure $\varphi_{(n)}^{\text{LSU}}$ (say) rejects H_i, $i \in I_n$, if and only if $p_i \leq m\alpha/n$, where $m = \max\{i \in I_n : p_{i:n} \leq \alpha_{i:n}^{\text{LSU}}\}$ with $\alpha_{i:n}^{\text{LSU}} = i\alpha/n$, $i \in I_n$ (i.e. Simes' critical values), cf. Simes [1986]. Now let $\vartheta \in \Theta$ and suppose that p_i, $i \in I_{n,0}(\vartheta)$, are iid uniformly distributed on $[0,1]$ and that

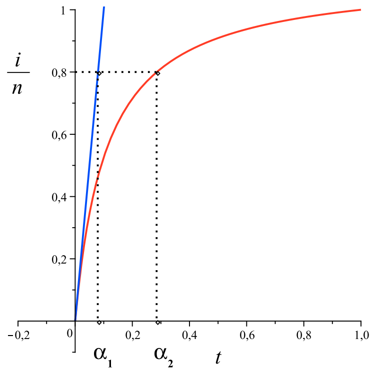

Figure 1.1: AORC with $\alpha = 0.1$ (curve) and the rejection curve corresponding to the LSU procedure with $\alpha = 0.1$ (straight line). Here α_1 denotes the ith critical value $\alpha_{i:n}^{\mathrm{LSU}}$ corresponding to the LSU test and α_2 denotes the ith critical value induced by the AORC.

$(p_i : i \in I_{n,0})$ and $(p_i : i \in I_{n,1})$ are independent random vectors. Then one of the most interesting results for the LSU procedure is that

$$\mathrm{FDR}_{\vartheta}(\varphi_{(n)}^{\mathrm{LSU}}) = \frac{n_0}{n}\alpha.$$

Different proofs of this equality can be found, for instance, in Benjamini and Yekutieli [2001], Finner and Roters [2001], Sarkar [2002] or Storey *et al.* [2004].

The fact that the FDR is bounded by $n_0\alpha/n$, that is, the FDR is distinctively smaller than α for smaller n_0-values, raised hope that improvements of the LSU procedure should be possible. For example, Finner *et al.* [2009] proposed a non-linear **asymptotically optimal rejection curve (AORC)**. For a fixed $\alpha \in (0, 1)$, the AORC is defined by

$$f_\alpha(t) = \frac{t}{t(1-\alpha)+\alpha}, \ t \in [0, 1]. \tag{1.2}$$

Figure 1.1 displays the AORC with $\alpha = 0.1$ (curve) and the rejection curve of the LSU procedure with $\alpha = 0.1$ (straight line). Larger critical values $\alpha_{i:n}$ induced by the AORC are considerably greater than the corresponding Simes' critical values $\alpha_{i:n}^{\mathrm{LSU}}$. This may result in a larger number of rejected hypotheses. In the picture, α_1 denotes the ith critical value $\alpha_{i:n}^{\mathrm{LSU}}$ corresponding to the LSU test and α_2 denotes the ith critical value induced by the AORC.

The idea behind the AORC is as follows. Consider models such that p-values corresponding to true null hypotheses are iid uniformly distributed and p-values under alternatives are equal to 0. Moreover, let the proportion of true null hypotheses converge to a $\zeta \in (\alpha, 1)$ with $\alpha \in (0, 1)$. Then the limiting ecdf of p-values converges to $1 - \zeta + t\zeta$ denoted by $F_\infty(t|\zeta)$. Let $\varphi^{\mathrm{SS}}(t)$ be a single-parameter procedure, which rejects hypotheses with p-values not greater than t. Thereby, the asymptotic FDR of $\varphi^{\mathrm{SS}}(t)$ in the considered models is given by

$$\mathrm{FDR}_\infty(\varphi^{\mathrm{SS}}(t)|\zeta) = \frac{t\zeta}{1-\zeta+t\zeta}.$$

By setting $\mathrm{FDR}_\infty(\varphi^{\mathrm{ss}}(t)|\zeta) \equiv \alpha$ we obtain a solution for t depending on ζ, i.e.

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

We are looking for a curve r such that the crossing point between r and the limiting ecdf $F_\infty(\cdot|\zeta)$ is t_ζ, that is,

$$r\left(\frac{\alpha(1-\zeta)}{\zeta(1-\alpha)}\right) = F_\infty\left(\frac{\alpha(1-\zeta)}{\zeta(1-\alpha)}\middle|\, \zeta\right) = \frac{1-\zeta}{1-\alpha}.$$

Noting that

$$t = \frac{\alpha(1-\zeta)}{\zeta(1-\alpha)} \text{ if and only if } \zeta = \zeta(t) = \frac{\alpha}{(1-\alpha)t+\alpha},$$

we get $r(t) = f_\alpha(t)$ given in (1.2). Note that for $\zeta \in [0, \alpha]$ we can set $t_\zeta \equiv 1$, which implies that all hypotheses are rejected and $\mathrm{FDR}_\infty(\varphi^{\mathrm{ss}}(1)|\zeta) = \zeta \leq \alpha$. Below, we will show that the described models, which will be called Dirac-uniform models, are least favourable for certain SU procedures (cf. Theorem 1.2 in Section 1.4). The AORC f_α is in some sense asymptotically optimal since the FDR level α is exhausted in this least favourable case, cf. Finner *et al.* [2009]. In Chapter 3 we present different methods how to construct multiple tests related to the AORC. Moreover, in Chapter 4 we introduce a modified version of weak dependence and show that a large class of step-up-down (SUD) procedures controls the FDR under weak dependence at least asymptotically. This result is in a line with recent investigations concerning FDR control of the LSU procedure under dependence, for example, in Benjamini and Yekutieli [2001], Finner *et al.* [2007] or Sarkar [2002].

1.4 General assumptions and Dirac-uniform models

As mentioned in the previous sections, FDR and/or FWER control for certain multiple test procedures, especially for those which exhaust the corresponding error rate level, is usually guaranteed under special conditions on the distribution function of p-values like

(D1) $\forall\, \vartheta \in \Theta : \forall\, i \in I_{n,0}(\vartheta) : \, p_i \sim U([0,1])$,

(I1) $\forall\, \vartheta \in \Theta : \, p_i, \, i \in I_{n,0}(\vartheta)$, are independent,

(I2) $\forall\, \vartheta \in \Theta : \, (p_i, i \in I_{n,0}(\vartheta))$ and $(p_i, i \in I_{n,1}(\vartheta))$ are independently distributed random vectors.

For example, if (I1) is fulfilled, then the Šidàk test controls the FWER at level α. Conditions (D1), (I1) and (I2) are sufficient for FDR control of the LSU test. We will use these assumptions or at least a few of them for deriving theoretical results in the following chapters.

One possible way to construct multiple tests controlling one of the error rates for all $\vartheta \in \Theta$ is to find a **least favourable parameter configuration (LFC) for** Θ, i.e. a parameter ϑ_0 such that under ϑ_0 the corresponding error rate is larger than under each $\vartheta \in \Theta$. An LFC ϑ_0 does not have to belong to Θ and it is not necessarily unique. Obviously, the FWER/FDR is controlled for all parameters $\vartheta \in \Theta$ if the FWER/FDR is controlled in an LFC. For example, let Θ be a parameter space such that condition (I1) is fulfilled and $n_0(\vartheta, n) = n_0(n)$ for all $\vartheta \in \Theta$ and some $n_0(n) < n$. Then each ϑ_0 such that $n_0(\vartheta_0) = n_0(n)$ and p-values corresponding to true null hypotheses are independently uniformly distributed on $[0, 1]$ is an LFC for the Šidàk test.

Condition (D1) mostly serves as an LFC for further investigations so that the main results of this work apply if p-values under nulls are stochastically larger than a uniform variate. However, in the next theorem (D1) is a necessary condition.

The next theorem shows the behaviour of the FDR for an SU procedure under specific assumptions on the corresponding critical values.

Theorem 1.2 (Benjamini and Yekutieli [2001])
Suppose that (D1), (I1) and (I2) are fulfilled. Then an SU procedure with critical values satisfying (1.1) *has the following properties:*
(a) If the ratio $\alpha_{i:n}/i$ is increasing in i, as $(p_i : i \in I_{n,1})$ increases stochastically, the FDR decreases.
(b) If the ratio $\alpha_{i:n}/i$ is decreasing in i, as $(p_i : i \in I_{n,1})$ increases stochastically, the FDR increases.

In the case of the LSU procedure $\alpha_{i:n}/i$ equals α so that the FDR of the LSU test is independent of the distribution of p-values under alternatives. The condition that $\alpha_{i:n}/i$ is increasing in i can be equivalently expressed in terms of a rejection curve ρ corresponding to the given critical values (1.1), that is,

(A1) $\rho(t)/t$ is non-decreasing for $t \in (0, 1]$.

Note that condition (A1) is equivalent to the property that $r(t)/t$ is non-increasing for $t \in (0, 1]$, where $r = \rho^{-1}$.

It follows from Theorem 1.2 that under (D1), (I1), (I2) and (A1) LFCs for an SU test are obtained in one of the so-called **Dirac-uniform (DU) models**. Thereby, $\mathbb{P}_{n,n_0}$ denotes a situation, where (D1) and (I1) are fulfilled and p_i, $i \in I_{n,1}$, follow a Dirac distribution with point mass 1 at 0. This implies that condition (I2) is fulfilled. We refer to this setting as DU(n, n_0). Note that $\mathbb{P}_{n,n_0}$ does not necessarily belong to the model $\{\mathbb{P}_\vartheta : \vartheta \in \Theta\}$.

It will be shown that DU models are LFCs for the FWER of a BPI test, cf. Chapter 2. Moreover, for a broad class of SU tests, DU models are LFCs for the FDR, cf. Chapter 3. Unfortunately, so far it is not known whether DU models are LFCs for an SD procedure. However, Finner *et al.* [2009] constructed upper bounds for the FDR of an SUD test and showed that these upper bounds are the largest in DU models. In Chapter 3 we utilise theses bounds to construct various SUD tests controlling the FDR.

Moreover, Chapters 3 and 4 deal with asymptotic control of the FWER and/or FDR, where useful tools are so-called asymptotic DU models. These are defined in the following way. Consider DU(n, n_0) models with $n_0/n \to \zeta$ for some $\zeta \in [0, 1]$. The Extended Glivenko-Cantelli Theorem (cf. Shorack and Wellner [1986], p.105) yields that the ecdf $\hat{F}_n(t) = \sum_{i=1}^{n} I(p_i \leq t)$ of all p-values converges almost surely and uniformly on $[0, 1]$ to the limiting function given by

$$F_\infty(t) = F_\infty(t|\zeta) = 1 - \zeta + \zeta t.$$

This limiting DU model with infinite number of p-values, where ζ is the proportion of true null hypotheses, is called the **asymptotic DU model**.

Chapter 2

Plug-in procedures controlling the FWER

In this chapter we deal with control of the Family-Wise Error Rate (FWER) of some multiple test procedures based on an estimator for the number of true null hypotheses n_0. In Section 2.1 we consider Bonferroni and Šidàk procedures with plug-in estimates. We call these tests Bonferroni plug-in (BPI) tests and show that a BPI procedure controls the FWER under the assumption that p-values are independent random variables under true null hypotheses, i.e. condition (I1) given in Chapter 1 is assumed to be fulfilled. In Section 2.2 we investigate the asymptotic behaviour of BPI test procedures and derive the asymptotic distribution of the number of false rejections V_n. Section 2.3 deals with plug-in tests related to the Bonferroni-Holm and Šidàk-Holm multiple-testing procedures. In Section 2.4 we evaluate the power of BPI tests for normally distributed test statistics. BPI tests for dependent test statistics will be discussed in Chapter 4. In Section 2.5 some concluding remarks will be given.

As mentioned in the previous chapter, although Bonferroni-type test procedures (for example, Bonferroni or Šidàk tests) control the FWER at a pre-specified level α, they typically have extremely low power if the number n of all hypotheses is large. If the number n_0 of true null hypotheses is known, then the corresponding oracle procedures, where the number of all hypotheses n is replaced by the number of true null hypotheses n_0, typically control the FWER. Thus, if n_0 is distinctively smaller than n, it should be possible to test the individual hypotheses at a higher level than a corresponding classical procedure does, which results in more power.

Unfortunately, the number n_0 of true null hypotheses is mostly unknown. To overcome this problem, we can replace n_0 in thresholds of oracle tests by an estimator for the number of true null hypotheses denoted by $\hat{n}_0$. This idea is not new. For example, Schweder and Spjøtvoll [1982] considered a pairwise comparisons problem with 17 means, i.e. $n = 136$ pair hypotheses. They estimated n_0 by a visual fit of a line to the larger p-values (i.e. to the least significant p-values) in a p-value plot and mentioned that in their specific example there might be about 25 true null hypotheses, so that the level $\alpha/25$ should be used for the individual tests. However, Schweder and Spjøtvoll [1982] did not give any proof for FWER control. Moreover, it seems that there have been no theoretical results concerning strong control of the FWER of a Bonferroni procedure with a plug-in estimate for the number of true null hypotheses until recently. The main results of this chapter are published in Finner and Gontscharuk [2009]. Independently and at the same time some similar findings concerning FWER control of adaptive Bonferroni and Holm procedures with respect to a specific mixture model were obtained in Guo [2009]. He proved that a special version of an adaptive Bonferroni procedure controls the FWER in finite samples while the corresponding adaptive Holm test controls it asymptotically.

Applications of plug-in estimators can be found in the literature on FDR procedures. For example, Storey [2002] proposed a **plug-in linear step-up (plug-in LSU)** procedure using an estimator for the proportion of true null hypotheses $\pi_0 = n_0/n$ depending on a tuning parameter $\lambda \in (0,1)$. Thereby, the critical values $\alpha_{i:n} = i\alpha/n$, $i \in I_n$, of the LSU test are replaced by $\hat{\alpha}_{i:n} = i\alpha/(n\hat{\pi}_0)$, $i \in I_n$, where $\hat{\pi}_0$ denotes an estimator for π_0. The critical values $\alpha_{i:n} = i\alpha/(n\pi_0)$, $i \in I_n$, correspond to the "oracle LSU" procedure. The plug-in LSU test can be interpreted as an LSU test with a random level $\alpha/\hat{\pi}_0$. Let

$$R_n(t) = \#\{i \in I_n : p_i \leq t\}$$

denote the number of p-values that are less than or equal to t for $t \in [0,1]$. Then the empirical cumulative distribution function (ecdf) $\hat{F}_n$ of all p-values can be expressed as $\hat{F}_n(t) = R_n(t)/n$, $t \in [0,1]$. Storey [2002] proposed to estimate π_0 by

$$\hat{\pi}_0 = \frac{n - R_n(\lambda)}{(1-\lambda)n} = \frac{1 - \hat{F}_n(\lambda)}{(1-\lambda)}, \tag{2.1}$$

where λ is a tuning parameter. The corresponding estimate for the number of true hypotheses can be found in Schweder and Spjøtvoll [1982] and is given by

$$\hat{n}_0 = \frac{n - R_n(\lambda)}{1-\lambda} = \frac{1 - \hat{F}_n(\lambda)}{(1-\lambda)} n$$

for some fixed λ. Obviously, $\hat{\pi}_0 = \hat{n}_0/n$. The following consideration shows why these estimators work. If p-values corresponding to true null hypotheses are iid uniformly distributed, then the number of true p-values which are greater than λ is about $(1-\lambda)n_0$. Assuming that p-values corresponding to false hypotheses are "false enough", i.e. p_i, $i \in I_{n,1}$, are small enough, only a few of them are expected to be greater than λ. Consequently, $n - R_n(\lambda)$ is also about $(1-\lambda)n_0$ or perhaps somewhat larger. Figure 2.1 illustrates this estimation method for $n = 50$ and $n_0 = 30$, where the p-values are generated with independent normal variables (mean 0 for true null hypotheses and mean 1 for false hypotheses).

In Storey *et al.* [2004] it was shown that under suitable assumptions concerning the joint distribution of the p-values the estimate (2.1) can be used in the plug-in LSU procedure, resulting in asymptotic FDR control. Moreover, Storey *et al.* [2004] proposed a slightly modified version of (2.1), that is,

$$\hat{\pi}_0^1 = \frac{n - R_n(\lambda) + 1}{(1-\lambda)n} = \frac{1 - \hat{F}_n(\lambda) + 1/n}{(1-\lambda)}, \tag{2.2}$$

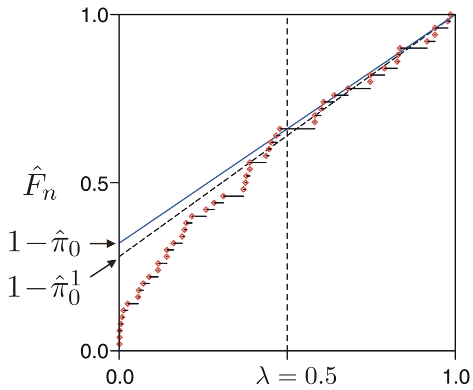

Figure 2.1: Estimation of π_0: illustration of Schweder and Spjøtvoll's idea. Here $\hat{\pi}_0$ corresponds to (2.1) and $\hat{\pi}_0^1$ to (2.2). The ecdf $\hat{F}_n$ of p-values is generated by $n = 50$ p-values with $n_0 = 30$.

which ensures finite FDR control.

In this chapter we replace the constant 1 in the plug-in estimator in formula (2.2) by a suitable parameter $\kappa > 0$. The parameter κ will be chosen such that the FWER of a BPI test is not larger than a pre-specified α-level. We also consider an alternative estimator of n_0, which was proposed in Benjamini and Hochberg [2000], that is,

$$\hat{n}_0 = \frac{n - k + 1}{1 - p_{k:n}}, \tag{2.3}$$

where $k \in I_n$ is fixed and $p_{k:n}$ is the kth smallest p-value.

2.1 Bonferroni plug-in procedure

Consider the general problem of multiple-testing defined in Notation 1.1. We first require that for all parameter configurations $\vartheta \in \Theta$ p-values are independent random variables under the corresponding null hypotheses, that is, (I1) is fulfilled. Note that we do not require any assumptions concerning the joint distribution of the p-values under alternatives, i.e. the p_i, $i \in I_{n,1}$, may be mutually dependent and may depend on p_i, $i \in I_{n,0}$. As mentioned in Chapter 1 an important tool for theoretical investigations are Dirac-uniform (DU) configurations, that is, p-values corresponding to true null hypotheses are independently uniformly distributed on $[0, 1]$, whereas p-values under the alternatives follow a Dirac distribution with point mass in 0. In this case we write $\mathbb{P}_{n,n_0}$ and FWER_{n,n_0} instead of $\mathbb{P}_\vartheta$ and FWER_ϑ, respectively.

We now give a formal definition of a Bonferroni-type plug-in procedure in terms of estimators $\hat{n}_0$ for n_0.

Definition 2.1

Let $\hat{n}_0 : [0, 1]^n \to [0, \infty)$ *be an estimator of* n_0 *and let* $\hat{\alpha} : [0, \infty] \to [0, 1]$ *be non-increasing. Then the random quantity* $\hat{\alpha} = \hat{\alpha}(\hat{n}_0)$ *will be called a plug-in threshold. A multiple-test procedure* *which rejects all hypotheses H_i with $p_i \leq \hat{\alpha}$, $i \in I_n$, will be called* ***Bonferroni plug-in (BPI)*** *test (based on $\hat{n}_0$).*

In this section we consider two types of thresholds $\hat{\alpha}$, that is,

$$\hat{\alpha}_1 = \alpha/\hat{n}_0, \tag{2.4}$$

$$\hat{\alpha}_2 = 1-(1-\alpha)^{1/\hat{n}_0}, \tag{2.5}$$

where equation (2.4) is in line with a Bonferroni correction and equation (2.5) is in line with a Šidàk correction. Similarly as in (2.1) and (2.2), we consider the following class of estimators for the number of true null hypotheses n_0, that is,

$$\hat{n}_0 = \frac{n - R_n(\lambda) + \kappa}{1-\lambda},\ \kappa \geq 0, \tag{2.6}$$

where $\lambda \in (0,1)$ is a pre-specified tuning parameter. In what follows, the parameter $\kappa \in \mathbb{R}$ will be chosen such that FWER is controlled by the corresponding BPI procedure. Thereby, the estimator $\hat{n}_0$ may take values in $[0,\infty)$ and not necessarily in $\mathbb{N}$. Since an estimator given in (2.6) is constructed by assuming that most of the p-values greater than λ belong to true null hypotheses, it is natural to reject only p-values smaller than λ. Requiring $\hat{\alpha}_i \leq \lambda$, $i = 1, 2$, we get the following restriction on κ, that is,

$$\kappa \geq \frac{\alpha(1-\lambda)}{\lambda} \tag{2.7}$$

in the case of equation (2.4) and

$$\kappa \geq (1-\lambda)\frac{\log(1-\alpha)}{\log(1-\lambda)} \tag{2.8}$$

in the case of equation (2.5). It will be shown that BPI procedures with thresholds (2.4) and (2.5) based on the estimator (2.6) control the FWER.

Estimators given in (2.3) yield a further class of estimators for the number of true null hypotheses n_0. This class is given by

$$\hat{n}_0 = \frac{n-k+\kappa}{1-p_{k:n}},\ \kappa \geq 0, \tag{2.9}$$

where $p_{k:n}$ is the kth smallest p-value and $k \in I_n$ is pre-specified. Again, we will choose the parameter κ such that the FWER is controlled.

The following lemma shows that under weak assumptions concerning an estimator $\hat{n}_0$ the FWER of a BPI test becomes largest if p-values corresponding to true null hypotheses are independently uniformly distributed on $[0,1]$ and p-values under alternatives are set to 0, that is, in a DU model. This is an important fact because FWER under $\mathbb{P}_{n,n_0}$ can be calculated exactly.

Lemma 2.2
Let $\vartheta \in \Theta$ be such that (I1) is fulfilled. Let $\hat{n}_0 : [0,1]^n \to [0,\infty)$ be a symmetric function of n

arguments such that $\hat{n}_0(x_1, \dots, x_n)$ *is non-decreasing in each* x_i*. Then a BPI test based on* $\hat{n}_0$ *satisfies*

$$\mathbb{P}_\vartheta(V_n \geq r) \leq \mathbb{P}_{n,n_0}(V_n \geq r) \ \textit{for all} \ \ r \in I_{n_0} = \{1, \dots, n_0\},$$

and

$$\mathit{FWER}_\vartheta \leq \mathit{FWER}_{n,n_0}, \tag{2.10}$$

i.e. Dirac-uniform configurations are least favourable for the FWER.

Proof: Note that $\hat{\alpha}(\hat{n}_0(x_1, \dots, x_n))$ is symmetric and non-increasing in each x_i. Setting

$$\tilde{\alpha}(x_1, \dots, x_{n_0}) = \hat{\alpha}(\hat{n}_0(x_1, \dots, x_{n_0}, 0, \dots, 0)) \ \text{ for } \ (x_1, \dots, x_n) \in [0,1]^n,$$

we get

$$\forall \ (x_1, \dots, x_n) \in [0,1]^n \ : \ \hat{\alpha}(\hat{n}_0(x_1, \dots, x_n)) \leq \tilde{\alpha}(x_1, \dots, x_{n_0}).$$

Let $p^0_{1:n_0}, \dots, p^0_{n_0:n_0}$ denote the order statistic of p_i, $i \in I_{n,0}$. Then

$$\mathbb{P}_\vartheta(V_n \geq r) = \mathbb{P}_\vartheta(p^0_{r:n_0} \leq \hat{\alpha}(\hat{n}_0)) \leq \mathbb{P}_\vartheta(p^0_{r:n_0} \leq \tilde{\alpha}(p^0_{1:n_0}, \dots, p^0_{n_0:n_0})), \tag{2.11}$$

where $\hat{\alpha}(\hat{n}_0) = \hat{\alpha}(\hat{n}_0(p_1, \dots, p_n))$. For the given $\mathbb{P}_\vartheta$, let $U_{i:n_0}$, $i \in I_{n_0}$, denote the ith order statistic of n_0 iid uniformly distributed on $[0,1]$ random variables. Since $p^0_{r:n_0}$ is stochastically not smaller than $U_{r:n_0}$ and $\tilde{\alpha}(p^0_{1:n_0}, \dots, p^0_{n_0:n_0})$ is stochastically not larger than $\tilde{\alpha}(U_{1:n_0}, \dots, U_{n_0:n_0})$, Lemma A.11 yields

$$\mathbb{P}_\vartheta(p^0_{r:n_0} \leq \tilde{\alpha}(p^0_{1:n_0}, \dots, p^0_{n_0:n_0})) \leq \mathbb{P}_\vartheta(U_{r:n_0} \leq \tilde{\alpha}(U_{1:n_0}, \dots, U_{n_0:n_0})). \tag{2.12}$$

Noting that p_i, $i \in I_{n,0}$ are iid uniformly distributed on $[0,1]$ under the measure $\mathbb{P}_{n,n_0}$, we get

$$\mathbb{P}_\vartheta(U_{r:n_0} \leq \tilde{\alpha}(U_{1:n_0}, \dots, U_{n_0:n_0})) = \mathbb{P}_{n,n_0}(p^0_{r:n_0} \leq \hat{\alpha}(\hat{n}_0)).$$

The latter and the inequalities (2.11), (2.12) complete the proof. ■

Remark 2.3
Lemma 2.2 implies that DU models are LFCs for each Θ such that for all parameter configurations $\vartheta \in \Theta$ p-values are independent random variables under the corresponding null hypotheses.

For an arbitrary but fixed $t \in [0,1]$ the number of p-values corresponding to true null hypotheses which are not greater than t is denoted by

$$V_n(t) = \#\{i \in I_{n,0} : p_i \leq t\}.$$

Since DU models are least favourable parameter configurations for the FWER of a BPI test, it is an interesting question which of the estimators of n_0 are unbiased in DU models. The next lemma provides formulas for the expectation of $\hat{n}_0$ with respect to (2.6) and (2.9) in DU(n, n_0) models. Let $n_1 = n_1(n)$ denote the number of false null hypotheses, i.e. $n_1 = n - n_0$.

Lemma 2.4

In a DU(n, n_0) model the expected value of the estimator in (2.6) *is given by*

$$\mathbb{E}_{n,n_0}[\hat{n}_0] = \mathbb{E}_{n,n_0}\left[\frac{n_0 - V_n(\lambda) + \kappa}{1-\lambda}\right] = n_0 + \frac{\kappa}{1-\lambda} \quad (\textit{independent of } n).$$

In case of (2.9) *we get*

$$\hat{n}_0 = n - k + \kappa \ \textit{ almost surely } \ \textit{for} \ \ k \leq n_1,$$

and

$$\mathbb{E}_{n,n_0}[\hat{n}_0] = n_0 + \frac{\kappa}{1-(k-n_1)/n_0} \ \ \textit{for} \ \ k > n_1.$$

Proof: Since $\mathbb{E}_{n,n_0}[V_n(\lambda)] = n_0\lambda$, the formula for $\mathbb{E}_{n,n_0}[\hat{n}_0]$ in case of (2.6) is obvious. In case of (2.9), we first note that $p_{k:n} = 0$ almost surely for $k \leq n_1$ in a DU(n, n_0) model, which yields the second formula of this lemma. In case of $k > n_1$, define $s = k - n_1$. Then $p_{k:n} = p^0_{s:n_0}$ is the sth smallest p-value corresponding to the true null hypotheses. The pdf of $p^0_{s:n_0}$, denoted by f_s, is given by

$$f_s(x) = n_0\binom{n_0-1}{s-1}x^{s-1}(1-x)^{n_0-s}.$$

It holds

$$\begin{aligned}\mathbb{E}_{n,n_0}\left[\frac{n-k+\kappa}{1-p_{k:n}}\right] &= n_0\binom{n_0-1}{s-1}\int_0^1 \frac{n_0-s+\kappa}{1-p}p^{s-1}(1-p)^{n_0-s}dp \\ &= n_0\binom{n_0-1}{s-1}(n_0-s+\kappa)\int_0^1 p^{s-1}(1-p)^{n_0-s-1}dp \\ &= (n_0-s+\kappa)\frac{n_0}{n_0-s} = n_0 + \frac{\kappa}{1-s/n_0}.\end{aligned}$$

The substitution $s = k - n_1$ completes the proof.

■

Remark 2.5

Lemma 2.4 implies that estimators given in (2.9) are always larger than n_0 if $k < n_1$, while estimators given in (2.6) have a fixed bias $\kappa/(1-\lambda)$. Therefore, estimators given in (2.6) seem to be preferable. Moreover, estimators given in (2.6) and those given in (2.9) with $k \geq n_1$ are unbiased for $\kappa = 0$. Clearly, it is tempting to try $\kappa = 0$ in a BPI test. Unfortunately, this does not work. For example, for $n = n_0 = 2$, $\alpha = 0.05$ and $\lambda = 0.5$ a BPI test with $\hat{\alpha}_1 = \alpha/\hat{n}_0$ based on $\hat{n}_0$ given in (2.6) does not control the FWER under $\mathbb{P}_{n,n_0}$. In what follows it will be shown that $\kappa = 1$ is always a reasonable choice.

The next theorem yields explicit formulas for the FWER and the distribution of the number of false rejections V_n with respect to a BPI test with critical values (2.4) and (2.5) based on the estimator (2.6) under $\mathbb{P}_{n,n_0}$. If $V_n(\lambda) = s$, $s \in I_{n_0} \cup \{0\}$, then

$$c_1(s) = \frac{\alpha(1-\lambda)}{n_0 - s + \kappa} \ \text{ and } \ c_2(s) = 1 - (1-\alpha)^{(1-\lambda)/(n_0-s+\kappa)}$$

denote the realised thresholds under $\mathbb{P}_{n,n_0}$ according to $\hat{\alpha}_1$ and $\hat{\alpha}_2$, respectively.

Theorem 2.6
Let $\alpha \in (0,1)$ and $\lambda \in (0,1)$ such that κ satisfies conditions (2.7) and (2.8), respectively. In the DU(n, n_0) model it holds for a BPI test with thresholds $\hat{\alpha}_i$, $i = 1, 2$, based on the estimator (2.6), *that*

$$\mathbb{P}_{n,n_0}(V_n = r) = \sum_{s=r}^{n_0} \binom{n_0}{s}\binom{s}{r}(1-\lambda)^{n_0-s} c_i(s)^r (\lambda - c_i(s))^{s-r} \tag{2.13}$$

for $r \in I_{n_0} \cup \{0\}$. Moreover,

$$\mathit{FWER}_{n,n_0} = 1 - \sum_{s=0}^{n_0} \binom{n_0}{s}(1-\lambda)^{n_0-s}(\lambda - c_i(s))^s. \tag{2.14}$$

Note that $\mathbb{P}_{n,n_0}(V_n = r)$ and FWER_{n,n_0} are independent of n.

Proof: For notational simplicity, we denote p-values corresponding to true null hypotheses by $p_1^0, \ldots, p_{n_0}^0$ and for ordered p-values we write $p_{1:n_0}^0, \ldots, p_{n_0:n_0}^0$. By noting that

$$\mathbb{P}_{n,n_0}(V_n = r) = \sum_{s=r}^{n_0} \mathbb{P}_{n,n_0}(V_n = r, V_n(\lambda) = s)$$

and setting $p_{n_0+1:n_0}^0 \equiv 1$ we obtain

$$\begin{aligned}
&\mathbb{P}_{n,n_0}(V_n = r, V_n(\lambda) = s) \\
= \;& \mathbb{P}_{n,n_0}(p_{r:n_0}^0 \le \hat{\alpha}_i, p_{r+1:n_0}^0 > \hat{\alpha}_i, V_n(\lambda) = s) \\
= \;& \mathbb{P}_{n,n_0}\left(V_n(\hat{\alpha}_i) = r, V_n(\lambda) = s\right) \\
= \;& \mathbb{P}_{n,n_0}\left(V_n(c_i(s)) = r, V_n(\lambda) = s\right) \\
= \;& \binom{n_0}{s}\binom{s}{r}\mathbb{P}_{n,n_0}\left(p_1^0, \ldots, p_r^0 \le c_i(s), p_{r+1}^0, \ldots, p_s^0 \in (c_i(s), \lambda], p_{s+1}^0, \ldots, p_{n_0}^0 > \lambda\right) \\
= \;& \binom{n_0}{s}\binom{s}{r}(1-\lambda)^{n_0-s} c_i(s)^r (\lambda - c_i(s))^{s-r}.
\end{aligned}$$

Since

$$\text{FWER}_{n,n_0} = \mathbb{P}_{n,n_0}(V_n \ge 1) = 1 - \mathbb{P}_{n,n_0}(V_n = 0),$$

formula (2.14) is obvious by choosing $r = 0$ in (2.13). ■

Remark 2.7
If the conditions (2.7) and/or (2.8) are not fulfilled, the probability of exactly r rejections, i.e. $\mathbb{P}_{n,n_0}(V_n = r)$, cannot be calculated with formula (2.13). As a consequence, FWER_{n,n_0} cannot be calculated with (2.14) in this case.

The next theorem yields the FWER of a BPI procedure with critical values $\hat{\alpha}_1$ and $\hat{\alpha}_2$ based on the estimator (2.9) in a DU model.

Theorem 2.8
Let $\alpha \in (0,1)$ and $k \in I_n$. By setting $n_1 = n - n_0$, the FWER of a BPI test with the threshold $\hat{\alpha}_1$ based on (2.9) *in a DU(n, n_0) model is given by*

$$\mathit{FWER}_{n,n_0} = 1 - \left(1 - \frac{\alpha}{n-k+\kappa}\right)^{n_0} \quad \mathit{for}\ \ k \leq n_1 \tag{2.15}$$

and

$$\mathit{FWER}_{n,n_0} = 1 - \left(1 - \frac{\alpha}{n-k+\kappa+\alpha}\right)^{n-k+1} \quad \mathit{for}\ \ k > n_1. \tag{2.16}$$

Moreover, the FWER of a BPI test with $\hat{\alpha}_2$ based on (2.9) *is given by*

$$\mathit{FWER}_{n,n_0} = 1 - (1-\alpha)^{n_0/(n-k+\kappa)} \ \ \mathit{for}\ \ k \leq n_1 \tag{2.17}$$

and for $k > n_1$ we get

$$\begin{aligned} \mathit{FWER}_{n,n_0} \ &= \ 1 - \frac{n_0!}{(k-n_1-1)!(n-k)!} \\ &\times \int_{t^*}^{1} \left(t - 1 + (1-\alpha)^{(1-t)/(n-k+\kappa)}\right)^{k-n_1-1} (1-t)^{n-k} dt, \end{aligned} \tag{2.18}$$

where

$$t^* = 1 + \frac{n-k+\kappa}{\ln(1-\alpha)} \mathit{LW}\left(\frac{\ln(1-\alpha)}{-n+k-\kappa}\right) \tag{2.19}$$

and LW denotes the Lambert W function, which is the inverse function of $f(x) = xe^x$.

Proof: At first, we consider the case $k \leq n_1$, which implies $p_{k:n} = 0$ almost surely. Then the estimator (2.9) is equal to $n - k + \kappa$ and the critical values $\hat{\alpha}_i$, $i = 1, 2$, are $\alpha/(n - k + \kappa)$ and $1 - (1 - \alpha)^{1/(n-k+\kappa)}$, respectively, that is, $\hat{\alpha}_i$, $i = 1, 2$, are almost surely constant. Hence, $\text{FWER}_{n,n_0} = 1 - (1 - \hat{\alpha}_i)^{n_0}$, $i = 1, 2$, yielding (2.15) and (2.17).

Now we investigate the case $k > n_1$, that is, $p_{k:n}$ corresponds to a true null hypothesis. It holds $\text{FWER}_{n,n_0} = 1 - \mathbb{P}_{n,n_0}(V_n = 0)$ and $\mathbb{P}_{n,n_0}(V_n = 0) = \mathbb{P}_{n,n_0}(\min_{j \in I_{n,0}} p_j > \hat{\alpha}_i)$. Then

$$\begin{aligned} \mathbb{P}_{n,n_0}(V_n = 0) \ &= \ \sum_{j \in I_{n,0}} \mathbb{P}_{n,n_0}\left(\min_{j \in I_{n,0}} p_j > \hat{\alpha}_i, p_{k:n} = p_j\right) \\ &= \ n_0 \mathbb{P}_{n,n_0}\left(\min_{j \in I_{n,0}} p_j > \hat{\alpha}_i, p_{k:n} = p_{i_0}\right), \end{aligned}$$

for some $i_0 \in I_{n,0}$. Obviously, $\{\min_{j \in I_{n,0}} p_j > \hat{\alpha}_i\} \subseteq \{p_{k:n} > \hat{\alpha}_i\}$. Thereby, the $\hat{\alpha}_i$s depend on $p_{k:n}$. If $p_{k:n} = t$ for some $t \in [0, 1]$, then

$$c_1(t) = \frac{\alpha(1-t)}{n-k+\kappa} \ \text{ and } \ c_2(t) = 1 - (1-\alpha)^{(1-t)/(n-k+\kappa)}$$

denote the realised thresholds under $\mathbb{P}_{n,n_0}$ according to $\hat{\alpha}_1$ and $\hat{\alpha}_2$, respectively. For $i = 1, 2$, the equality $t = c_i(t)$ has a unique solution t_i (say) in $[0, 1]$, where $t_1 = \alpha/(n - k + \kappa + \alpha)$ and

$t_2 = t^*$ with t^* given in (2.19). Altogether we get $\{p_{k:n} > \hat{\alpha}_i\} = \{p_{k:n} > t_i\}$. It follows that

$$\begin{aligned}
\mathbb{P}_{n,n_0}(V_n = 0) &= n_0 \int_{t_i}^1 \mathbb{P}_{n,n_0}\left(\min_{j \in I_{n,0}} p_j > c_i(t), p_{k:n} = p_{i_0} | p_{i_0} = t\right) dt \\
&= n_0 \int_{t_i}^1 \mathbb{P}_{n,n_0}\left(\min_{j \in I_{n,0} \setminus \{i_0\}} p_j > c_i(t), p_{i_0} > c_i(t), p_{k:n} = p_{i_0} | p_{i_0} = t\right) dt.
\end{aligned}$$

For $t > t_i$ we have $\{p_{i_0} = t\} \subseteq \{p_{i_0} > c_i(t)\}$. Moreover, under $\{p_{i_0} = t\}$ we get $\{p_{k:n} = p_{i_0}\} = \{\#\{j \in I_n \setminus \{i_0\} : p_j \leq t\} = k - 1\}$. Hence,

$$\begin{aligned}
\mathbb{P}_{n,n_0}(V_n = 0) &= n_0 \int_{t_i}^1 \mathbb{P}_{n,n_0}\left(\min_{j \in I_{n,0} \setminus \{i_0\}} p_j > c_i(t), \#\{j \in I_n \setminus \{i_0\} : p_j \leq t\} = k - 1\right) dt \\
&= n_0 \binom{n_0 - 1}{k - n_1 - 1} \int_{t_i}^1 (t - c_i(t))^{k - n_1 - 1} (1 - t)^{n-k} dt.
\end{aligned}$$

Note that the last formula immediately implies (2.18). For a BPI test with threshold $\hat{\alpha}_1$ we obtain that

$$t - c_1(t) = \frac{n - k + \kappa + \alpha}{n - k + \kappa}\left(t - \frac{\alpha}{n - k + \kappa + \alpha}\right) = \frac{n - k + \kappa + \alpha}{n - k + \kappa}(t - t_1)$$

and consequently

$$\begin{aligned}
\mathbb{P}_{n,n_0}(V_n = 0) &= n_0 \binom{n_0 - 1}{k - n_1 - 1} \left(\frac{n - k + \kappa + \alpha}{n - k + \kappa}\right)^{k - n_1 - 1} \\
&\quad \times \int_{t_1}^1 (t - t_1)^{k - n_1 - 1} (1 - t)^{n-k} dt.
\end{aligned}$$

By substituting $\tau = (t - t_1)/(1 - t_1)$ in the integral before, we get

$$\begin{aligned}
\mathbb{P}_{n,n_0}(V_n = 0) &= \frac{n_0!}{(k - n_1 - 1)!(n - k)!} \left(\frac{n - k + \kappa + \alpha}{n - k + \kappa}\right)^{k - n_1 - 1} \\
&\quad \times (1 - t_1)^{n_0} \int_0^1 \tau^{k - n_1 - 1} (1 - \tau)^{n-k} d\tau,
\end{aligned}$$

where the integral in the latter expression is the beta function $\mathcal{B}(k - n_1, n - k + 1)$, cf. Frampton [1986], p.57. Noting that for $x, y \in \mathbb{N}$

$$\mathcal{B}(x, y) = \frac{(x - 1)!(y - 1)!}{(x + y - 1)!},$$

we obtain

$$\mathbb{P}_{n,n_0}(V_n = 0) = \left(\frac{n - k + \kappa}{n - k + \kappa + \alpha}\right)^{n - k + 1},$$

which implies (2.16). ■

The next theorems provide conditions, under which a BPI test procedure with the considered thresholds controls the FWER.

Theorem 2.9
Let $\vartheta \in \Theta$ and assume (I1). Let $\alpha \in (0,1)$, $\lambda \in (0,1)$ and $\kappa \geq 1$ such that κ satisfies conditions (2.7) and (2.8), respectively. Then the BPI procedure with threshold $\hat{\alpha}_i$, $i = 1,2$, based on the estimator (2.6) *controls the FWER at level α.*

Proof: Let $n_0 = n_0(\vartheta)$. Lemma 2.2 yields that it suffices to check that FWER_{n,n_0} given in (2.14) does not exceed α, which is equivalent to the inequality

$$1-\alpha \leq (1-\lambda)^{n_0} \sum_{s=0}^{n_0} \binom{n_0}{s} (\lambda - c_i(s))^s (1-\lambda)^{-s},\ i=1,2. \tag{2.20}$$

Below, we write $c_i(s,\alpha)$ instead of $c_i(s)$, $i=1,2$ and define the functions

$$h_\lambda(\alpha) = \frac{1-\alpha}{(1-\lambda)^{n_0}} \tag{2.21}$$

and

$$g_{\lambda,i}(\alpha) = \sum_{s=0}^{n_0} \binom{n_0}{s} (\lambda - c_i(s,\alpha))^s (1-\lambda)^{-s},\ i=1,2. \tag{2.22}$$

Then (2.20) is equivalent to $h_\lambda(\alpha) \leq g_{\lambda,i}(\alpha)$, $i=1,2$. Obviously,

$$h_\lambda(0) = g_{\lambda,i}(0) = \frac{1}{(1-\lambda)^{n_0}}$$

and

$$h'_\lambda(0) = -\frac{1}{(1-\lambda)^{n_0}}.$$

Hence, (2.20) holds if $h'_\lambda(0) \leq g'_{\lambda,i}(0)$ and $g''_{\lambda,i}(\alpha) \geq 0$ for all $\alpha \in [0,1]$, $i=1,2$. We get

$$g'_{\lambda,i}(\alpha) = -\sum_{s=1}^{n_0} \binom{n_0}{s-1} (n_0-s+1)\left(\lambda - c_i(s,\alpha)\right)^{s-1} (1-\lambda)^{-s} c'_i(s,\alpha)$$

and

$$c'_1(s,\alpha) = \frac{1-\lambda}{n_0-s+\kappa},\quad c'_2(s,\alpha) = \frac{1-\lambda}{n_0-s+\kappa}(1-\alpha)^{\frac{1-\lambda}{n_0-s+\kappa}-1}.$$

Thus,

$$\begin{aligned}
g'_{\lambda,1}(\alpha) &= -\sum_{s=1}^{n_0} \binom{n_0}{s-1} \frac{n_0-s+1}{n_0-s+\kappa} \left(\frac{\lambda}{1-\lambda} - \frac{\alpha}{n_0-s+\kappa}\right)^{s-1}, \\
g'_{\lambda,2}(\alpha) &= -\sum_{s=1}^{n_0} \binom{n_0}{s-1} \frac{n_0-s+1}{n_0-s+\kappa} \left(\frac{(1-\alpha)^{\frac{1-\lambda}{n_0-s+\kappa}}}{1-\lambda} - 1\right)^{s-1} (1-\alpha)^{\frac{1-\lambda}{n_0-s+\kappa}-1}.
\end{aligned}$$

The assumptions (2.7) and (2.8) imply

$$\frac{\lambda}{1-\lambda} - \frac{\alpha}{n_0-s+\kappa} \geq 0 \ \text{ and } \ \frac{(1-\alpha)^{\frac{1-\lambda}{n_0-s+\kappa}}}{1-\lambda} - 1 \geq 0 \ \text{ for } \ s \in I_{n_0}.$$

Hence, $g'_{\lambda,i}(\alpha)$ is non-decreasing, that is, $g''_{\lambda,i}(\alpha) \geq 0$, $i = 1, 2$. Furthermore, the inequality $h'_\lambda(0) \leq g'_{\lambda,i}(0)$ is equivalent to

$$-\frac{1}{(1-\lambda)^{n_0}} \leq -\sum_{s=1}^{n_0} \binom{n_0}{s-1} \frac{n_0 - s + 1}{n_0 - s + \kappa} \left(\frac{\lambda}{1-\lambda}\right)^{s-1}. \tag{2.23}$$

Since

$$\frac{1}{(1-\lambda)^{n_0}} = \sum_{s=0}^{n_0} \binom{n_0}{s} \left(\frac{\lambda}{1-\lambda}\right)^s = \left(\frac{\lambda}{1-\lambda}\right)^{n_0} + \sum_{s=0}^{n_0-1} \binom{n_0}{s} \left(\frac{\lambda}{1-\lambda}\right)^s,$$

inequality (2.23) is equivalent to

$$\left(\frac{\lambda}{1-\lambda}\right)^{n_0} \geq \sum_{s=0}^{n_0-1} \binom{n_0}{s} \frac{n_0 - s}{n_0 - s - 1 + \kappa} \left(\frac{\lambda}{1-\lambda}\right)^s - \sum_{s=0}^{n_0-1} \binom{n_0}{s} \left(\frac{\lambda}{1-\lambda}\right)^s,$$

or

$$\left(\frac{\lambda}{1-\lambda}\right)^{n_0} \geq (1-\kappa) \sum_{s=0}^{n_0-1} \binom{n_0}{s} \frac{1}{n_0 - s - 1 + \kappa} \left(\frac{\lambda}{1-\lambda}\right)^s. \tag{2.24}$$

Obviously, the latter inequality is fulfilled for $\kappa \geq 1$ and therefore inequality (2.20) holds under the assumptions of Theorem 2.9, which finally yields that FWER is controlled at level α. ■

Remark 2.10
Note that in the case of a BPI procedure with $\hat{\alpha}_i$, $i = 1, 2$, based on the estimator (2.6), $\kappa = 1$ always fulfils conditions (2.7) and (2.8) if $\alpha \in (0, 1)$ and $\lambda \in [\alpha, 1)$. Violation of (2.7) or (2.8) can lead to an exceedance of the pre-specified FWER-level. For example, for $\alpha = 0.05$ and $\lambda = 0.06$ condition (2.7) implies $\kappa \leq 0.783$. By setting $\kappa = 0.1$ for the BPI test with (2.4) we get that (2.7) is not fulfilled and we obtain $\text{FWER}_{2,2} = \lambda^2 + 2(1-\lambda)^2\alpha/(1+\kappa) = 0.0839$ (note that (2.14) does not apply here). However, Guo [2009] showed that a BPI procedure with the critical value $\hat{\alpha}_1$ based on the estimator (2.6) with $\kappa = 1$ controls the FWER for all $\alpha \in (0, 1)$ and all $\lambda \in (0, 1)$, that is, condition (2.7) can be dispensed with. Thereby, this result was obtained by constructing an upper bound for the FWER. In contrast to that, our results are based on the exact formula (2.14) for the FWER in DU models.

Theorem 2.11
Let $\vartheta \in \Theta$ and assume (I1). Let $\alpha \in (0, 1)$ and $k \in I_n$. Then the BPI procedure with threshold $\hat{\alpha}_i$, $i = 1, 2$, based on the estimator (2.9) *controls the FWER at level α for all $k \leq n_1$ and $\kappa \geq 0$, where $n_1 = n - n_0$. Moreover, for $k > n_1$ the BPI procedure with threshold $\hat{\alpha}_1$ based on the estimator* (2.9) *controls the FWER for $\kappa \geq 1$.*

Proof: Lemma 2.2 yields that the FWER of a BPI test with $\hat{\alpha}_i$, $i = 1, 2$, based on (2.9) is maximal in a DU model so that FWER control follows if (2.15)-(2.18) are not greater than α. In case of $k \leq n_1$ the inequalities (2.15) and (2.17) in Theorem 2.8 immediately imply that the corresponding

FWER is not greater than α. Then we have to prove that (2.16) is not greater than α, which is equivalent to the inequality

$$\left(1-\frac{\alpha}{n-k+\kappa+\alpha}\right)^{n-k+1} \geq 1-\alpha. \tag{2.25}$$

Setting

$$h(\alpha)=1-\alpha \text{ and } g(\alpha)=\left(1-\frac{\alpha}{n-k+\kappa+\alpha}\right)^{n-k+1},$$

it suffices to check that $h(\alpha) \leq g(\alpha)$ for all $\alpha \in [0,1]$. Clearly, $h(0)=g(0)=1$, $h'(0)=-1$,

$$g'(0)=-\frac{n-k+1}{n-k+\kappa+\alpha}$$

and

$$g''(\alpha)=(n-k+2)\frac{(n-k+1)(n-k+\kappa)^{n-k+1}}{(n-k+\kappa+\alpha)^{n-k+3}} \geq 0, \quad \alpha \in [0,1].$$

For $\kappa \geq 1$ we get $h'(0) \leq g'(0)$ for all $\alpha \in [0,1]$, which implies (2.25). Therefore inequality (2.25) holds under the assumption of Theorem 2.11 and the FWER is controlled at level α. ■

Remark 2.12

We could not prove that a BPI test with the threshold $\hat{\alpha}_2$ based on (2.9) controls the FWER for $k > n_1$. But for fixed n, α and k, we can always find a $\kappa = \kappa(n, \alpha, k)$ such that the FWER, i.e. the expression in (2.18), is not greater than α. Moreover, we observed that $\kappa \equiv 1$ yields FWER control for all considered n-, α- and k-values.

Remark 2.13

Note that a smaller value of κ may result in a slightly more powerful BPI procedure. Hence, we can try to find a $\kappa < 1$ for fixed n, α and λ (or k resp.), i.e. $\kappa = \kappa(n, \alpha, \lambda)$ (or $\kappa = \kappa(n, \alpha, k)$ resp.), by checking that the FWER, i.e. the corresponding expression (2.14), (2.16) or (2.18), is not greater than α for all $n_0 \in I_n$. For illustration we consider BPI tests with the critical value $\hat{\alpha}_1$ based on (2.6). For $\alpha = 0.05$, $1 \leq n_0 \leq 200$, $\lambda = 0.5, 0.6, 0.7, 0.8$ the largest κ values are attained for $n_0^* = 7, 9, 13, 21$ and are given by $\kappa^* \approx 0.872, 0.867, 0.861, 0.857$. The left picture in Figure 2.2 (in Figure 2.3 resp.) suggests that a BPI test with threshold $\hat{\alpha}_1$ based on the estimator (2.6) (the estimator (2.9) resp.) and $\lambda = 0.5, 0.6, 0.7, 0.8$ ($k = n - 3n_0/4, n - n_0/2, n - n_0/4, n$ resp.) and corresponding κ^* controls the FWER for all n if (I1) is fulfilled. The picture on the right side in Figure 2.2 (in Figure 2.3 resp.) suggests that the best choice of κ for a BPI test with $\hat{\alpha}_2$ based on the estimator (2.6) (the estimator (2.9) resp.) converges to some limiting value that is less than or equal to 1 for $n_0 \to \infty$. We note that the κ values are not increasing if n_0 increases for a BPI test with $\hat{\alpha}_2$ based on (2.6); and κ increases for a BPI test with $\hat{\alpha}_2$ based on (2.9).

It seems that the apparently optimal κ^*-values are close to 1 such that the loss in power seems negligible by choosing $\kappa = 1$. In Sections 2.4 we will restrict our attention to BPI procedures with the threshold $\hat{\alpha}_1$ based on the estimator (2.6) with $\kappa = 1$.

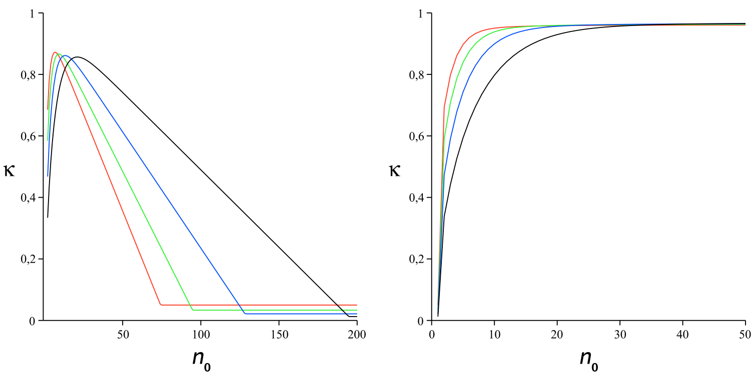

Figure 2.2: Values of κ such that $\text{FWER}_{n,n_0} = \alpha$ for a BPI test with threshold $\hat{\alpha}_1$ (left picture) and $\hat{\alpha}_2$ (right picture) based on the estimator (2.6) for $\alpha = 0.05$ and $\lambda = 0.5, 0.6, 0.7, 0.8$. The curves may be identified by noting that κ increases when λ increases in $n_0 = 50$ and decreases in $n_0 = 10$ in the left and right picture, respectively.

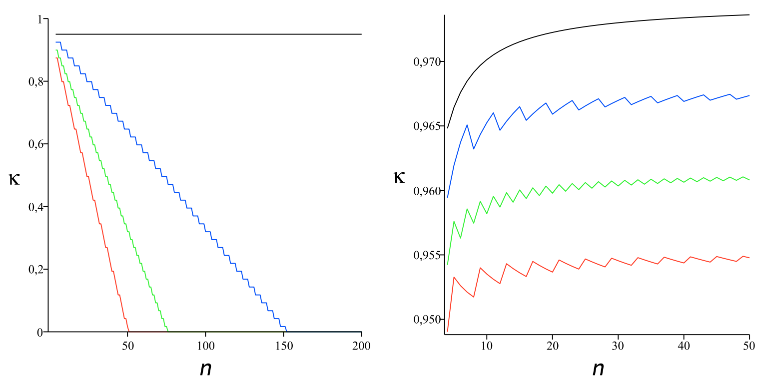

Figure 2.3: Values of κ such that $\text{FWER}_{n,n_0} \leq \alpha$ for a BPI test with threshold $\hat{\alpha}_1$ (left picture) and $\hat{\alpha}_2$ (right picture) based on the estimator (2.9) for $\alpha = 0.05$ and $k = \lceil n/4 \rceil, \lceil n/2 \rceil, \lceil 3n/4 \rceil, n$ and $n_0 \in I_n$. The curves may be identified by noting that κ increases when k increases in $n = 50$. In the case of BPI tests with $\hat{\alpha}_1$, for fixed n and k and the corresponding κ (left graph) we get $\text{FWER}_{n,n_0} = \alpha$ for all $n_0 \in \{n-k+1, \ldots, n\}$. In case of BPI tests with $\hat{\alpha}_2$, for fixed n and k and the corresponding κ (right graph) we obtain $\text{FWER}_{n,n} = \alpha$ and $\text{FWER}_{n,n_0^1} < \text{FWER}_{n,n_0^2}$ for all n_0^1 and n_0^2 such that $n-k+1 \leq n_0^1 < n_0^2 \leq n$.

2.2 Asymptotic behaviour of Bonferroni plug-in tests

The following theorem yields the asymptotic behaviour of the number of false rejections V_n in the least favourable DU configuration. That is in line with the asymptotic results in Finner and Roters [2002] for various traditional multiple-testing procedures; see Remark 2.18.

Theorem 2.14
Let $\alpha \in (0,1)$, $\lambda \in (0,1)$, $\kappa \in \mathbb{R}$ and set $\beta_1 = \alpha$, $\beta_2 = -\log(1-\alpha)$. Consider DU($n, n_0$) models with $n_0 = n_0(n) \to \infty$ as $n \to \infty$. Then, for $i = 1, 2$, it holds for a BPI test with threshold $\hat{\alpha}_i$ based on the estimator given in (2.6) *that*

$$\lim_{n\to\infty} \mathbb{P}_{n,n_0}(V_n = r) = \exp(-\beta_i)\frac{\beta_i^r}{r!} \quad for \quad r \in \mathbb{N} \cup \{0\}, \tag{2.26}$$

$$\lim_{n\to\infty} \mathbb{E}_{n,n_0} V_n = \beta_i. \tag{2.27}$$

Moreover, let $k = k(n) \in I_n$ satisfy

$$\liminf_{n\to\infty} \frac{k - n_1}{n_0} \geq 0 \;\; and \;\; \limsup_{n\to\infty} \frac{k - n_1}{n_0} < 1, \tag{2.28}$$

where $n_1 = n_1(n) = n - n_0$. Then (2.26) *and* (2.27) *hold also for a BPI test with thresholds $\hat{\alpha}_i$, $i = 1, 2$, based on* (2.9) *with given values of k.*

Proof: First we consider the case of a BPI test with $\hat{\alpha}_1 = \alpha/\hat{n}_0$. We obtain for $\epsilon > 0$, $r \in I_{n_0} \cup \{0\}$ and all $n \in \mathbb{N}$ that

$$\begin{aligned}
\mathbb{P}_{n,n_0}(V_n \leq r) &\leq \mathbb{P}_{n,n_0}\left(\left\{\#\left\{i \in I_{n,0}:\; p_i \leq \frac{\alpha}{\hat{n}_0}\right\} \leq r\right\} \cap \left\{\frac{\hat{n}_0}{n_0} < 1 + \epsilon\right\}\right) \\
&\quad + \mathbb{P}_{n,n_0}\left(\frac{\hat{n}_0}{n_0} \geq 1 + \epsilon\right) \\
&\leq \mathbb{P}_{n,n_0}\left(\left\{\#\left\{i \in I_{n,0}:\; p_i \leq \frac{\alpha}{n_0(1+\epsilon)}\right\} \leq r\right\} \cap \left\{\frac{\hat{n}_0}{n_0} < 1 + \epsilon\right\}\right) \\
&\quad + \mathbb{P}_{n,n_0}\left(\frac{\hat{n}_0}{n_0} \geq 1 + \epsilon\right) \\
&\leq \mathbb{P}_{n,n_0}\left(\#\left\{i \in I_{n,0}:\; p_i \leq \frac{\alpha}{n_0(1+\epsilon)}\right\} \leq r\right) + \mathbb{P}_{n,n_0}\left(\frac{\hat{n}_0}{n_0} \geq 1 + \epsilon\right) \\
&= \sum_{s=0}^{r} \binom{n_0}{s} \left(\frac{\alpha}{n_0(1+\epsilon)}\right)^s \left(1 - \frac{\alpha}{n_0(1+\epsilon)}\right)^{n_0 - s} + \mathbb{P}_{n,n_0}\left(\frac{\hat{n}_0}{n_0} \geq 1 + \epsilon\right) \\
&= G\left(r \,\middle|\, n_0, \frac{\alpha}{n_0(1+\epsilon)}\right) + \mathbb{P}_{n,n_0}\left(\frac{\hat{n}_0}{n_0} \geq 1 + \epsilon\right),
\end{aligned}$$

where $G(\cdot|m,p)$ denotes the distribution function of the binomial distribution $B(m,p)$. Similarly we get

$$
\begin{aligned}
\mathbb{P}_{n,n_0}(V_n \leq r) &\geq \mathbb{P}_{n,n_0}\left(\left\{\#\left\{i \in I_{n,0}:\ p_i \leq \frac{\alpha}{\hat{n}_0}\right\} \leq r\right\} \cap \left\{\frac{\hat{n}_0}{n_0} > 1-\epsilon\right\}\right) \\
&\geq \mathbb{P}_{n,n_0}\left(\left\{\#\left\{i \in I_{n,0}:\ p_i \leq \frac{\alpha}{n_0(1-\epsilon)}\right\} \leq r\right\} \cap \left\{\frac{\hat{n}_0}{n_0} > 1-\epsilon\right\}\right) \\
&\geq \mathbb{P}_{n,n_0}\left(\#\left\{i \in I_{n,0}:\ p_i \leq \frac{\alpha}{n_0(1-\epsilon)}\right\} \leq r\right) - \mathbb{P}_{n,n_0}\left(\frac{\hat{n}_0}{n_0} \leq 1-\epsilon\right) \\
&= \sum_{s=0}^{r} \binom{n_0}{s} \left(\frac{\alpha}{n_0(1-\epsilon)}\right)^s \left(1 - \frac{\alpha}{n_0(1-\epsilon)}\right)^{n_0-s} - \mathbb{P}_{n,n_0}\left(\frac{\hat{n}_0}{n_0} \leq 1-\epsilon\right) \\
&= G\left(r \,\middle|\, n_0, \frac{\alpha}{n_0(1-\epsilon)}\right) - \mathbb{P}_{n,n_0}\left(\frac{\hat{n}_0}{n_0} \leq 1-\epsilon\right).
\end{aligned}
$$

Moreover, since $\mathbb{E}_{n,n_0} V_n = \sum_{r=1}^{n_0} \mathbb{P}_{n,n_0}(V_n \geq r)$, the inequalities derived before imply

$$
\frac{\alpha}{1+\epsilon} - n_0 \mathbb{P}_{n,n_0}\left(\frac{\hat{n}_0}{n_0} \geq 1+\epsilon\right) \leq \mathbb{E}_{n,n_0} V_n \leq \frac{\alpha}{1-\epsilon} + n_0 \mathbb{P}_{n,n_0}\left(\frac{\hat{n}_0}{n_0} \leq 1-\epsilon\right).
$$

Therefore, if the following condition

$$
n_0 \mathbb{P}_{n,n_0}\left(\left|\frac{\hat{n}_0}{n_0} - 1\right| \geq \epsilon\right) \to 0 \ \text{ for } \ n \to \infty \tag{2.29}
$$

is fulfilled, then (2.26) and (2.27) apply by choosing $\epsilon = \epsilon_n$ such that $\epsilon_n \downarrow 0$ for $n \to \infty$.

Analogously, it follows for a BPI test with $\hat{\alpha}_2 = 1 - (1-\alpha)^{1/\hat{n}_0}$ that

$$
\mathbb{P}_{n,n_0}(V_n \leq r) \leq G\left(r \,\middle|\, n_0, 1 - (1-\alpha)^{1/(n_0(1+\epsilon))}\right) + \mathbb{P}_{n,n_0}\left(\frac{\hat{n}_0}{n_0} \geq 1+\epsilon\right)
$$

and

$$
\mathbb{P}_{n,n_0}(V_n \leq r) \geq G\left(r \,\middle|\, n_0, 1 - (1-\alpha)^{1/(n_0(1-\epsilon))}\right) - \mathbb{P}_{n,n_0}\left(\frac{\hat{n}_0}{n_0} \leq 1-\epsilon\right).
$$

Since

$$
n_0\left(1 - (1-\alpha)^{1/(n_0(1\pm\epsilon))}\right) \to -\frac{\log(1-\alpha)}{1\pm\epsilon} \ \text{ for } \ n_0 \to \infty,
$$

the distribution of V_n converges to the desired Poisson distribution if condition (2.29) applies.

For proving (2.27) it suffices to show that the estimators given in (2.6) and (2.9) fulfil (2.29). The next lemma yields this result. ■

Lemma 2.15

Let $\hat{n}_0$ be an estimator for the number n_0 of true null hypotheses defined in (2.6) *or* (2.9) *with $\kappa \in \mathbb{R}$, $\lambda \in (0,1)$ or $k = k(n) \in I_n$ that satisfies* (2.28)*, respectively. Then*

$$
\forall\, \epsilon > 0:\ \exists\, C_1, C_2 > 0:\ \forall\, n \in \mathbb{N}:\ \mathbb{P}_{n,n_0}\left(\left|\frac{\hat{n}_0}{n_0} - 1\right| \geq \epsilon\right) \leq C_1 e^{-n_0 C_2}. \tag{2.30}
$$

Proof: First we consider the estimator given in (2.6). Noting that

$$\frac{\hat{n}_0}{n_0} = \frac{1 - \hat{F}_{n,0}(\lambda) + \kappa/n_0}{1 - \lambda},$$

we obtain

$$\left\{ \frac{\hat{n}_0}{n_0} \leq 1 - \epsilon \right\} = \left\{ \hat{F}_{n,0}(\lambda) - \lambda \geq \frac{\kappa}{n_0} + \epsilon(1 - \lambda) \right\}$$

and

$$\left\{ \frac{\hat{n}_0}{n_0} \geq 1 + \epsilon \right\} = \left\{ \lambda - \hat{F}_{n,0}(\lambda) \geq -\frac{\kappa}{n_0} + \epsilon(1 - \lambda) \right\}.$$

For fixed $\epsilon > 0$ and $\kappa > 0$ there exists an $N_{\epsilon,\kappa} \in \mathbb{N}$ such that for all $n_0 \geq N_{\epsilon,\kappa}$ we get $\epsilon(1-\lambda) \pm \kappa/n_0 \geq \epsilon(1-\lambda)/2$. Altogether this implies

$$\left\{ \left| \frac{\hat{n}_0}{n_0} - 1 \right| \geq \epsilon \right\} \subseteq \left\{ \left| \hat{F}_{n,0}(\lambda) - \lambda \right| \geq \frac{\epsilon(1-\lambda)}{2} \right\} \text{ for } n_0 \geq N_{\epsilon,\kappa}.$$

Hence, for $n_0 \geq N_{\epsilon,\kappa}$ we get

$$\begin{aligned} \mathbb{P}_{n,n_0}\left(\left| \frac{\hat{n}_0}{n_0} - 1 \right| \geq \epsilon \right) &\leq \mathbb{P}_{n,n_0}\left(\left| \hat{F}_{n,0}(\lambda) - \lambda \right| \geq \frac{\epsilon(1-\lambda)}{2} \right) \\ &\leq \mathbb{P}_{n,n_0}\left(\sup_{x \in [0,1]} \left| \hat{F}_{n,0}(x) - x \right| \geq \frac{\epsilon(1-\lambda)}{2} \right) \\ &\leq 2\exp\left(-\frac{n_0 \epsilon^2 (1-\lambda)^2}{2} \right), \end{aligned}$$

where the latter inequality follows by applying the Dvoretzky-Kiefer-Wolfowitz (DKW) inequality, cf. Theorem A.10.

Now we show that the estimator given in (2.9) fulfils (2.30). We divide the proof into two parts: (i) $n_1 \geq k$ and (ii) $k = n_1 + s$ for $s \in I_{n_0}$.

(i) Since $p_{k:n} = 0$ in DU models, we get

$$\frac{\hat{n}_0}{n_0} = 1 + \frac{n_1 - k}{n_0} + \frac{\kappa}{n_0} \text{ almost surely.}$$

The first expression in (2.28) implies $\lim_{n\to\infty}(k - n_1)/n_0 = 0$. Hence, $\hat{n}_0/n_0 = 1 + o(1)$ almost surely and consequently we obtain $\mathbb{P}_{n,n_0}(|\hat{n}_0/n_0 - 1| \geq \epsilon) = 0$ for fixed $\epsilon > 0$, $\kappa \in \mathbb{R}$ and sufficiently large n-values. Then (2.30) is trivially fulfilled.

(ii) W.l.o.g. let $s < n_0$ and $\lim_{n\to\infty} s/n_0 = \eta \in [0,1)$, because the second property in (2.28) applies. Note that

$$\frac{\hat{n}_0}{n_0} = \frac{1 - s/n_0 + \kappa/n_0}{1 - p_{s:n_0}} = \frac{1 - s/n_0 + \kappa/n_0}{1 - \hat{F}_{n,0}^{-1}(s/n_0)},$$

where $p_{s:n_0}$ is the sth smallest p-value corresponding to true null hypotheses, $\hat{F}_{n,0}$ is the ecdf of null p-values and $\hat{F}_{n,0}^{-1}(u) = \inf\{t \in [0,1] : \hat{F}_{n,0}(t) \geq u\}$. Then

$$\left\{ \frac{\hat{n}_0}{n_0} \leq 1 - \epsilon \right\} = \left\{ \hat{F}_{n,0}^{-1}(s/n_0) \leq \frac{s/n_0 - \epsilon - \kappa/n_0}{1 - \epsilon} \right\}.$$

Since $\hat{F}_{n,0}^{-1}(y) \le x$ if and only if $y \le \hat{F}_{n,0}(x)$ for $x \in \mathbb{R}$ and $y \in [0,1]$ (cf. Witting [1985], p. 20), we get

$$\left\{\frac{\hat{n}_0}{n_0} \le 1-\epsilon\right\} = \left\{\hat{F}_{n,0}\left(\frac{s/n_0 - \epsilon - \kappa/n_0}{1-\epsilon}\right) \ge \frac{s}{n_0}\right\}$$

and setting $y = (s/n_0 - \epsilon - \kappa/n_0)/(1-\epsilon)$ we obtain

$$\left\{\frac{\hat{n}_0}{n_0} \le 1-\epsilon\right\} = \left\{\hat{F}_{n,0}\left(y\right) - y \ge \frac{\epsilon(1-s/n_0) + \kappa/n_0}{1-\epsilon}\right\}.$$

Thus,

$$\left\{\frac{\hat{n}_0}{n_0} \le 1-\epsilon\right\} \subseteq \left\{\sup_{x\in[0,1]} (\hat{F}_{n,0}\left(x\right) - x) \ge \frac{\epsilon(1-s/n_0) + \kappa/n_0}{1-\epsilon}\right\}.$$

Obviously, for fixed $\epsilon > 0$ and $\kappa \in \mathbb{R}$ there exists some $N_{\epsilon,\kappa} \in \mathbb{N}$ such that for all $n_0 \ge N_{\epsilon,\kappa}$ it holds

$$\left\{\frac{\hat{n}_0}{n_0} \le 1-\epsilon\right\} \subseteq \left\{\sup_{x\in[0,1]} |\hat{F}_{n,0}\left(x\right) - x| \ge \frac{\epsilon(1-\eta)}{2(1-\epsilon)}\right\}.$$

The latter relation together with the DKW inequality yields

$$\mathbb{P}_{n,n_0}\left(\frac{\hat{n}_0}{n_0} \le 1-\epsilon\right) \le 2\exp\left(-n_0\frac{\epsilon^2(1-\eta)^2}{2(1-\epsilon)^2}\right). \tag{2.31}$$

Similarly we obtain

$$\left\{\frac{\hat{n}_0}{n_0} \ge 1+\epsilon\right\} = \left\{\hat{F}_{n,0}^{-1}(s/n_0) \ge \frac{s/n_0 + \epsilon - \kappa/n_0}{1+\epsilon}\right\}.$$

Noting that the inverse ecdf $\hat{F}_{n,0}^{-1}$ is left continuous, we get

$$\left\{\frac{\hat{n}_0}{n_0} \ge 1+\epsilon\right\} \subseteq \left\{\hat{F}_{n,0}^{-1}(s/n_0 + 0) \ge \frac{s/n_0 + \epsilon - \kappa/n_0}{1+\epsilon}\right\}.$$

Moreover, since $s/n_0 \in (0,1)$ and $x \le \hat{F}_{n,0}^{-1}(y+0)$ if and only if $\hat{F}_{n,0}(x-0) \le y$ for $x \in \mathbb{R}$ and $y \in (0,1)$ (cf. Witting [1985], p.20), it follows for a fixed $\epsilon > 0$ and sufficiently large n that

$$\left\{\frac{\hat{n}_0}{n_0} \ge 1+\epsilon\right\} \subseteq \left\{\hat{F}_{n,0}\left(\frac{s/n_0 + \epsilon - \kappa/n_0}{1+\epsilon} - 0\right) \le \frac{s}{n_0}\right\}.$$

Note that $\hat{F}_{n,0}(x) \le \hat{F}_{n,0}(x-0) + 1/n_0$ almost surely for all $x \in (0,1)$. Hence,

$$\left\{\frac{\hat{n}_0}{n_0} \ge 1+\epsilon\right\} \subseteq \left\{\hat{F}_{n,0}\left(\frac{s/n_0 + \epsilon - \kappa/n_0}{1+\epsilon}\right) \le \frac{s+1}{n_0}\right\}.$$

Setting $y = (s/n_0 + \epsilon - \kappa/n_0)/(1+\epsilon)$ we obtain

$$\left\{\frac{\hat{n}_0}{n_0} \ge 1+\epsilon\right\} \subseteq \left\{y - \hat{F}_{n,0}(y) \ge \frac{\epsilon(1-s/n_0)}{1+\epsilon} - \frac{\kappa+1+\epsilon}{n_0(1+\epsilon)}\right\}$$

and herewith

$$\left\{\frac{\hat{n}_0}{n_0} \ge 1+\epsilon\right\} \subseteq \left\{\sup_{x\in[0,1]} |x - \hat{F}_{n,0}(x)| \ge \frac{\epsilon(1-s/n_0)}{1+\epsilon} - \frac{\kappa+1+\epsilon}{n_0(1+\epsilon)}\right\}.$$

For fixed $\epsilon > 0$ and $\kappa \in \mathbb{R}$ there exists some $N_{\epsilon,\kappa} \in \mathbb{N}$ such that for all $n_0 \geq N_{\epsilon,\kappa}$ it holds

$$\left\{\frac{\hat{n}_0}{n_0} \geq 1+\epsilon\right\} \subseteq \left\{\sup_{x\in[0,1]} |x - \hat{F}_{n,0}(x)| \geq \frac{\epsilon(1-\eta)}{2(1+\epsilon)}\right\}.$$

Then the DKW inequality implies

$$\mathbb{P}_{n,n_0}\left(\frac{\hat{n}_0}{n_0} \geq 1+\epsilon\right) \leq 2\exp\left(-n_0\frac{\epsilon^2(1-\eta)^2}{2(1+\epsilon)^2}\right). \tag{2.32}$$

Conditions (2.31) and (2.32) yield (2.30). ■

Remark 2.16
For estimators given in (2.6), the choice of $\kappa = 0$ may be preferred, because $\kappa = 0$ leads to unbiased estimators of n_0. For estimators given in (2.9), $\kappa = 0$ also leads to unbiased estimators of n_0 if (2.28) is fulfilled. The first condition in (2.28) means that the kth smallest p-value corresponds asymptotically to a true null hypothesis (i.e. $k > n_1$ and consequently $\hat{n}_0/n_0 \to 1$, $n \to \infty$, almost surely if $n_0 \to \infty$) or that $p_{k:n}$ corresponds asymptotically to a false hypothesis (i.e. $k \leq n_1$ and consequently $\hat{n}_0 = n - k + \kappa \geq n_0 + \kappa$) but $\hat{n}_0$ is not too large. In general, $\hat{n}_0/n_0$ may be considerably larger than n_0 if $k < n_1$. If the proportion of true null hypotheses is asymptotically larger than 0, then the second condition in (2.28) can be replaced by $\limsup_{n\to\infty} k/n < 1$.

Remark 2.17
If the alternative distributions are not Dirac, estimators for the number of true null hypotheses become stochastically larger. Hence, the critical values $\hat{\alpha}_1$ and $\hat{\alpha}_2$ become stochastically smaller. It follows that V_n becomes stochastically smaller than under a DU distribution. For estimators given in (2.9), parameters $k = k(n)$ fulfilling $\limsup_{n\to\infty}(k - n_1) < 0$ may lead to an overestimation of n_0 and consequently V_n becomes stochastically smaller than in DU models.

Remark 2.18
In Finner and Roters [2002] the distribution of V_n and its limiting distribution for iid uniformly distributed p-values were computed, assuming that all hypotheses are true, especially for traditional single-parameter and certain stepwise procedures. Their limiting results for single-parameter procedures (without plug-in estimate) coincide with Theorem 2.14.

2.3 Step-down plug-in procedures

It this section we consider the possibility of a **step-down plug-in (SDPI)** procedure related to the Bonferroni-Holm test. Let $\vartheta \in \Theta$ be given and suppose that the assumption (I1) is satisfied. One possibility to define critical values for an SDPI procedure corresponding to the Bonferroni test is given by

$$\hat{\alpha}^{(1)}_{i:n} = \max\left(\frac{\alpha}{\hat{n}_0}, \frac{\alpha}{n-i+1}\right),\ i \in I_n. \tag{2.33}$$

Analogously, critical values for an SDPI test corresponding to the Šidàk procedure are given by

$$\hat{\alpha}_{i:n}^{(2)} = \max\left(1-(1-\alpha)^{1/\hat{n}_0}, 1-(1-\alpha)^{1/(n-i+1)}\right),\ i \in I_n. \tag{2.34}$$

An SDPI procedure rejects all H_i with $p_i \leq \hat{\alpha}_{m:n}^{(i)}$, $i = 1, 2$, where

$$m = \max\{j \in I_n : p_{s:n} \leq \hat{\alpha}_{s:n}^{(i)} \text{ for all } s \leq j\}.$$

Remark 2.19

As in the case of BPI procedures, the probability of at least one false rejection for these SDPI procedures is largest if p-values corresponding to true null hypotheses are iid uniformly distributed on $[0, 1]$ and p-values under alternatives follow a Dirac distribution with point mass 1 at 0, that is, DU(n, n_0) models are LFCs for the FWER and hence $\text{FWER}_\vartheta \leq \text{FWER}_{n,n_0}$.

The next theorems give formulas for the calculation of the FWER in DU models.

Theorem 2.20

Let $\alpha \in (0, 1)$, $\lambda \in (0, 1)$ and let κ satisfy conditions (2.7) and (2.8)*, respectively. Then the FWER of the SDPI procedure with thresholds (2.33) based on the estimator* (2.6) *in a DU(n, n_0) model is given by*

$$\begin{aligned} FWER_{n,n_0} &= 1 - \sum_{s=0}^{\min(\lfloor \lambda n_0 + \kappa \rfloor, n_0)} \binom{n_0}{s} (1-\lambda)^{n_0-s} \left(\lambda - \frac{\alpha}{n_0}\right)^s \\ &\quad - \sum_{s=\lfloor \lambda n_0+\kappa \rfloor+1}^{n_0} \binom{n_0}{s} (1-\lambda)^{n_0-s} \left(\lambda - \frac{\alpha(1-\lambda)}{n_0 - s + \kappa}\right)^s \end{aligned}$$

and the FWER of the SDPI test with (2.34) based on (2.6) *in a DU(n, n_0) model is given by*

$$\begin{aligned} FWER_{n,n_0} &= 1 - \sum_{s=0}^{\min(\lfloor \lambda n_0 + \kappa \rfloor, n_0)} \binom{n_0}{s} (1-\lambda)^{n_0-s} \left(\lambda - 1 + (1-\alpha)^{1/n_0}\right)^s \\ &\quad - \sum_{\lfloor \lambda n_0+\kappa \rfloor+1}^{n_0} \binom{n_0}{s} (1-\lambda)^{n_0-s} \left(\lambda - 1 + (1-\alpha)^{(1-\lambda)/(n_0-s+\kappa)}\right)^s, \end{aligned}$$

where $\lfloor x \rfloor$ denotes the smallest integer greater than or equal to x.

Proof: Let $n_1 = n - n_0$. An SDPI procedure implies that the smallest p-value corresponding to true null hypotheses should be compared with the critical value $\hat{\alpha}_{n_1+1:n}^{(i)}$ in DU models. Hence, the event $\{V_n = 0\}$ is equivalent to the event $\{\min_{i \in I_{n,0}} p_i > \hat{\alpha}_{n_1+1:n}^{(i)}\}$. If $V_n(\lambda) = s$ for $s \in I_{n_0} \cup \{0\}$, then

$$c_1(s) = \max\left(\frac{\alpha(1-\lambda)}{n_0 - s + \kappa}, \frac{\alpha}{n_0}\right) \text{ and } c_2(s) = \max(1-(1-\alpha)^{(1-\lambda)/(n_0-s+\kappa)}, 1-(1-\alpha)^{1/n_0})$$

denote the realised critical values $\hat{\alpha}^{(i)}_{n_1+1:n}$, $i = 1, 2$, under $\mathbb{P}_{n,n_0}$. It follows

$$
\begin{aligned}
\mathbb{P}_{n,n_0}(V_n = 0) &= \sum_{s=0}^{n_0} \mathbb{P}_{n,n_0}\left(\min_{j \in I_{n,0}} p_j > c_i(s), V_n(\lambda) = s\right) \\
&= \sum_{s=0}^{n_0} \mathbb{P}_{n,n_0}\left(\#\{j \in I_{n,0} : p_j \in (c_i(s), \lambda]\} = s, \#\{j \in I_{n,0} : p_j > \lambda\} = n_0 - s\right) \\
&= \sum_{s=0}^{n_0} \binom{n_0}{s} (1-\lambda)^{n_0 - s} (\lambda - c_i(s))^s.
\end{aligned}
$$

By noting that $c_1(s) = \alpha(1-\lambda)/(n_0 - s + \kappa)$ and $c_2(s) = 1 - (1-\alpha)^{(1-\lambda)/(n_0 - s + \kappa)}$ for $s > \lambda n_0 + \kappa$, i.e. $s \geq \lfloor \lambda n_0 + \kappa \rfloor + 1$, and $c_1(s) = \alpha/n_0$, $c_2(s) = 1 - (1-\alpha)^{1/n_0}$ otherwise, we obtain the desired formulas for the FWER. ■

Theorem 2.21

Let $\alpha \in (0,1)$, $k \in I_n$ and $\kappa \geq 0$. Setting $n_1 = n - n_0$, $t' = (k - n_1 - \kappa)/n_0$, $t_1 = \alpha/(n - k + \kappa + \alpha)$ and $t_2 = t^*$ with t^* given in (2.19)*, the FWER of the SDPI procedure with thresholds (2.33) based on the estimator* (2.9) *in a DU(n, n_0) model is given by*

$$
\mathit{FWER}_{n,n_0} = 1 - \left(1 - \frac{\alpha}{n_0}\right)^{n_0} \leq \alpha \ \ \mathit{for} \ \ k \leq n_1 \tag{2.35}
$$

and

$$
\begin{aligned}
\mathit{FWER}_{n,n_0} = n_0 \binom{n_0 - 1}{k - n_1 - 1} &\left(\int_{t_1}^{\max(t_1, t')} \left(t - \frac{\alpha(1-t)}{n-k+\kappa}\right)^{k - n_1 - 1} (1-t)^{n-k} dt \right. \\
&\left. + \int_{\max(t_1,t')}^{1} \left(t - \frac{\alpha}{n_0}\right)^{k-n_1-1} (1-t)^{n-k} dt \right) \ \ \mathit{for} \ \ k > n_1.
\end{aligned} \tag{2.36}
$$

Moreover, the FWER of the SDPI test with (2.34) based on (2.9) *in a DU(n, n_0) model is given by*

$$
\mathit{FWER}_{n,n_0} = \alpha \ \ \mathit{for} \ \ k \leq n_1, \tag{2.37}
$$

and for $k > n_1$ we obtain

$$
\begin{aligned}
\mathit{FWER}_{n,n_0} = n_0 \binom{n_0 - 1}{k - n_1 - 1} &\left(\int_{t_2}^{\max(t_2, t')} \left(t - 1 + (1-\alpha)^{(1-t)/(n-k+\kappa)}\right)^{k-n_1-1} \right. \\
&\left. \times (1-t)^{n-k} dt + \int_{\max(t_2,t')}^{1} \left(t - 1 + (1-\alpha)^{1/n_0}\right)^{k-n_1-1} (1-t)^{n-k} dt \right).
\end{aligned} \tag{2.38}
$$

.

Proof: It holds $\mathrm{FWER}_{n,n_0} = 1 - \mathbb{P}_{n,n_0}(V_n = 0)$ and $\mathbb{P}_{n,n_0}(V_n = 0) = \mathbb{P}_{n,n_0}(\min_{j \in I_{n,0}} p_j \geq \hat{\alpha}^{(i)}_{n_1+1:n})$, $i = 1, 2$. First we investigate the case $k \leq n_1$, which implies $p_{k:n} = 0$ almost surely. Then the estimator $\hat{n}_0$ given in (2.9) equals $n - k + \kappa$ almost surely and consequently we get

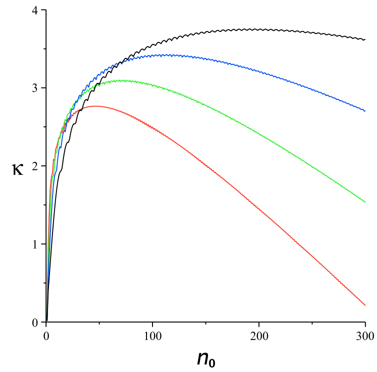

Figure 2.4: Values of κ such that $\text{FWER}_{n,n_0} = \alpha$ for the SDPI procedure with thresholds given in (2.33) based on the estimator (2.6) for $\alpha = 0.05$ and $\lambda = 0.5, 0.6, 0.7, 0.8$. The curves may be identified by noting that κ increases when λ increases in $n_0 = 300$.

$\hat{n}_0 \geq n_0$ since $\kappa \geq 0$. Hence, $\hat{\alpha}^{(1)}_{n_1+1:n} = \alpha/n_0$ and $\hat{\alpha}^{(2)}_{n_1+1:n} = 1-(1-\alpha)^{1/n_0}$. Moreover, SDPI procedures with $k \leq n_1$ and $\kappa \geq 0$ coincide with the corresponding classical SD tests. It follows that $\text{FWER}_{n,n_0} = 1-(1-\hat{\alpha}^{(i)}_{n_1+1:n})^{n_0}$ which yields (2.35) and (2.37).

Now we consider the case $k > n_1$, which implies that $p_{k:n}$ corresponds to a true null hypothesis. If $p_{k:n} = t$ for some $t \in [0,1]$, then

$$c_1(t) = \max\left(\frac{\alpha(1-t)}{n-k+\kappa}, \frac{\alpha}{n_0}\right) \quad \text{and} \quad c_2(t) = \max\left(1-(1-\alpha)^{(1-t)/(n-k+\kappa)}, 1-(1-\alpha)^{1/n_0}\right)$$

denote the realised thresholds under $\mathbb{P}_{n,n_0}$ according to $\hat{\alpha}^{(1)}_{n_1+1:n}$ and $\hat{\alpha}^{(2)}_{n_1+1:n}$, respectively. Similar as in the proof of Theorem 2.8 we get

$$\mathbb{P}_{n,n_0}(V_n = 0) = n_0 \binom{n_0-1}{k-n_1-1} \int_{t_i}^{1} (t-c_i(t))^{k-n_1-1}(1-t)^{n-k} dt.$$

By noting that $c_1(t) = \alpha(1-t)/(n-k+\kappa)$ and $c_2(t) = 1-(1-\alpha)^{(1-t)/(n-k+\kappa)}$ for $t \leq t' = (k-n_1-\kappa)/n_0$ and $c_1(t) = \alpha/n_0$, $c_2(t) = 1-(1-\alpha)^{1/n_0}$ otherwise, we obtain (2.36) and (2.38).

■

Remark 2.22

In contrast to BPI tests, a SDPI procedure does not always control the FWER if $\kappa = 1$. There are $\lambda \in (0,1)$ and $n_0 \in \mathbb{N}$ such that FWER_{n,n_0} exceeds the level α. For example, $\alpha = 0.05$, $\lambda = 0.5$, $\kappa = 1$ and $n_0 = 3$ yield $\text{FWER}_{3,3} \approx 0.055$ for the SDPI test with critical values given in (2.33) based on the estimator (2.6). But for fixed λ, α and n it is possible to calculate the minimum κ such that

$$\max_{1 \leq n_0 \leq n} \text{FWER}_{n,n_0} = \alpha.$$

Then the SDPI procedure with this κ controls the FWER. On the other hand, the case $\kappa > 1$ may result in less rejections as obtained with the corresponding BPI procedure with $\kappa = 1$. For example, for $\alpha = 0.05$, $1 \leq n_0 \leq 300$, $\lambda = 0.5, 0.6, 0.7, 0.8$ the largest κ-values for the SDPI tests with (2.33) based on (2.6) are attained for $n_0^* = 30, 69, 113, 191$ and are given by $\kappa^* \approx 2.765$, $3.100, 3.427, 3.754$. Figure 2.4 suggests that an SDPI test with thresholds defined in the expression (2.33) based on the estimator (2.6) and one of these λ's and corresponding κ^* control the FWER for all n if (D1) and (I1) apply.

Even though the SDPI procedure with $\kappa = 1$ does not always control the FWER, the FWER is controlled asymptotically if $\hat{n}_0/n_0 \to 1$ for $n_0 \to \infty$ almost surely. The next theorem gives this result.

Theorem 2.23
Let $\alpha \in (0,1)$, $\kappa \in \mathbb{R}$, $\lambda \in (0,1)$ be fixed and/or $k = k(n)$ satisfy (2.28). *Consider DU models with $n_0(n) \to \infty$ as $n \to \infty$. Then the limiting FWER of an SDPI procedure with thresholds* (2.33) *based on* (2.6) *or* (2.9) *is given by*

$$\lim_{n\to\infty} FWER_{n,n_0} = 1 - \exp(-\alpha) < \alpha,$$

and the limiting FWER of an SDPI test with (2.34) *based on* (2.6) *or* (2.9) *is given by*

$$\lim_{n\to\infty} FWER_{n,n_0} = \alpha.$$

Proof: Note that $\hat{\alpha}^{(1)}_{n_1+1:n} \geq \alpha/n_0$ for an SDPI test with thresholds given in (2.33) and $\hat{\alpha}^{(2)}_{n_1+1:n} \geq 1-(1-\alpha)^{1/n_0}$ for an SDPI test with thresholds given in (2.34). Hence,

$$\mathbb{P}_{n,n_0}(V_n = 0) = \mathbb{P}_{n,n_0}\left(\min_{j\in I_{n,0}} p_j \geq \hat{\alpha}^{(1)}_{n_1+1:n}\right) \leq \mathbb{P}_{n,n_0}\left(\min_{j\in I_{n,0}} p_j \geq \alpha/n_0\right) = \left(1-\frac{\alpha}{n_0}\right)^{n_0}$$

for an SDPI test with thresholds given in (2.33) and

$$\mathbb{P}_{n,n_0}(V_n = 0) = \mathbb{P}_{n,n_0}\left(\min_{j\in I_{n,0}} p_j \geq \hat{\alpha}^{(2)}_{n_1+1:n}\right) \leq \mathbb{P}_{n,n_0}\left(\min_{j\in I_{n,0}} p_j \geq 1-(1-\alpha)^{1/n_0}\right) = 1-\alpha$$

for an SDPI test with thresholds given in (2.34). As in Theorem 2.14 we get

$$\begin{aligned}\mathbb{P}_{n,n_0}(V_n = 0) &\geq \mathbb{P}_{n,n_0}\left(\{V_n = 0\} \cap \left\{\frac{\hat{n}_0}{n_0} \in [1-\epsilon, 1+\epsilon]\right\}\right)\\ &= \mathbb{P}_{n,n_0}\left(\left\{\min_{j\in I_{n,0}} p_j \geq \hat{\alpha}^{i}_{n_1+1:n}\right\} \cap \left\{\frac{\hat{n}_0}{n_0} \in [1-\epsilon, 1-\epsilon]\right\}\right).\end{aligned}$$

If $n_0 \leq \hat{n}_0 \leq (1+\epsilon)n_0$, then $\hat{\alpha}^{(1)}_{n_1+1:n} = \alpha/n_0$ and $\hat{\alpha}^{(2)}_{n_1+1:n} = 1-(1-\alpha)^{1/n_0}$. Moreover, if $(1-\epsilon)n_0 \leq \hat{n}_0 < n_0$, then $\alpha/n_0 < \hat{\alpha}^{(1)}_{n_1+1:n} \leq \alpha/(n_0(1-\epsilon))$ and $1-(1-\alpha)^{1/n_0} < \hat{\alpha}^{(2)}_{n_1+1:n} \leq 1-(1-\alpha)^{1/(n_0(1-\epsilon))}$. By noting that $\hat{\alpha}^{(1)}_{n_1+1:n} \leq \alpha/(n_0(1-\epsilon))$ and $\hat{\alpha}^{(2)}_{n_1+1:n} \leq 1-(1-\alpha)^{1/(n_0(1-\epsilon))}$

for all $\hat{n}_0$-values, we get for all $n_0 \geq N_{\epsilon,\delta}$ that

$$\begin{aligned}\mathbb{P}_{n,n_0}(V_n = 0) &\geq \mathbb{P}_{n,n_0}\left(\left\{\min_{j\in I_{n,0}} p_j \geq \frac{\alpha}{(1-\epsilon)n_0}\right\} \cap \left\{\frac{\hat{n}_0}{n_0} \in [1-\epsilon, 1+\epsilon]\right\}\right)\\ &\geq \mathbb{P}_{n,n_0}\left(\min_{j\in I_{n,0}} p_j \geq \frac{\alpha}{(1-\epsilon)n_0}\right) - \mathbb{P}_{n,n_0}\left(\left|\frac{\hat{n}_0}{n_0} - 1\right| \geq \epsilon\right)\\ &= \left(1 - \frac{\alpha}{(1-\epsilon)n_0}\right)^{n_0} - \mathbb{P}_{n,n_0}\left(\left|\frac{\hat{n}_0}{n_0} - 1\right| \geq \epsilon\right)\end{aligned}$$

for an SDPI test with thresholds given in (2.33) and

$$\begin{aligned}\mathbb{P}_{n,n_0}(V_n = 0) &\geq \mathbb{P}_{n,n_0}\left(\left\{\min_{j\in I_{n,0}} p_j \geq 1-(1-\alpha)^{1/(n_0(1-\epsilon))}\right\} \cap \left\{\frac{\hat{n}_0}{n_0} \in [1-\epsilon, 1+\epsilon]\right\}\right)\\ &\geq \mathbb{P}_{n,n_0}\left(\min_{j\in I_{n,0}} p_j \geq 1-(1-\alpha)^{1/(n_0(1-\epsilon))}\right) - \mathbb{P}_{n,n_0}\left(\left|\frac{\hat{n}_0}{n_0} - 1\right| \geq \epsilon\right)\\ &= (1-\alpha)^{1/(1-\epsilon)} - \mathbb{P}_{n,n_0}\left(\left|\frac{\hat{n}_0}{n_0} - 1\right| \geq \epsilon\right)\end{aligned}$$

for an SDPI test with thresholds given in (2.34). Lemma 2.15 implies $\mathbb{P}_{n,n_0}(|\hat{n}_0/n_0 - 1| \geq \epsilon) = o(1)$. Since $\text{FWER}_{n,n_0} = 1 - \mathbb{P}_{n,n_0}(V_n = 0)$, for each $\epsilon > 0$ it follows for SDPI tests with thresholds given in (2.33) and (2.34) that

$$1 - \exp(-\alpha) \leq \lim_{n_0\to\infty} \text{FWER}_{n,n_0} \leq 1 - \exp(-\alpha/(1-\epsilon)) + o(1)$$

and

$$\alpha \leq \lim_{n_0\to\infty} \text{FWER}_{n,n_0} \leq 1 - (1-\alpha)^{1/(1-\epsilon)} + o(1),$$

respectively. Setting $\epsilon = \epsilon_n \downarrow 0$ completes the proof. ■

2.4 Power investigation

In this section we compare the power of BPI tests with the power of corresponding classical tests with fixed threshold. We restrict our attention to a simple normal model with fixed alternatives and fixed number of true null hypotheses n_0. Let X_{ij}, $i \in I_n$, $j \in I_m$, be independent normally distributed random variables with unknown mean ϑ_i and known variance $\sigma^2 > 0$. We consider the following multiple-testing problem

$$H_i : \vartheta_i \leq 0 \text{ versus } K_i : \vartheta_i > 0,\ i \in I_n.$$

The associated test statistics and p-values are given by $T_i = \sum_{j=1}^m X_{ij}/(\sigma\sqrt{m})$, $i \in I_n$, and $p_i = p_i(t_i) = 1 - \Phi(t_i)$, $i \in I_n$, respectively, where t_i denotes the realisation of T_i. Let $\vartheta = (\vartheta_1, \ldots, \vartheta_n)$ and φ_i denote the test for H_i. The power of a single test φ_i in terms of ϑ is defined as

$$\beta_i(\vartheta) = \mathbb{P}_\vartheta(\varphi_i = 1),\ i \in I_{n,1}.$$

Alternatively, we may consider the expected proportion of correct rejections as power, that is,

$$\beta(\vartheta) = \frac{1}{n-n_0} \sum_{i \in I_{n,1}} \mathbb{P}_\vartheta(\varphi_i = 1) = \frac{1}{n-n_0} \sum_{i \in I_{n,1}} \beta_i(\vartheta). \tag{2.39}$$

Obviously, the power of a multiple test depends on the distribution of the p-values under alternatives. We consider the case, where means corresponding to alternatives all have the same value, i.e. $\vartheta_i = \mu$ for $i \in I_{n,1}$ for some fixed $\mu > 0$. For simplicity, we confine ourselves to considering a BPI test with $\kappa = 1$ and critical value $\alpha/\hat{n}_0$ given in (2.4) based on $\hat{n}_0$ given in (2.6). Under the assumptions stated before, the power of single tests φ_i, $i \in I_{n,1}$, and the overall power defined in equation (2.39) only depend on $\delta = \mu\sqrt{m}/\sigma$ and coincide, i.e. $\beta_i(\delta) = \beta(\delta)$. For $i \in I_{n,1}$, a straightforward calculation yields

$$\begin{aligned}
\beta(\delta) &= \mathbb{P}_\vartheta\left(p_i \le \alpha/\hat{n}_0\right) \\
&= \sum_{s=1}^{n} \mathbb{P}_\vartheta\left(p_i \le \frac{\alpha}{\hat{n}_0} \mid \#\{k \in I_n \setminus \{i\} : p_k \le \lambda\} = s-1\right) \\
&\quad \times \mathbb{P}_\vartheta\left(\#\{k \in I_n \setminus \{i\} : p_k \le \lambda\} = s-1\right) \\
&= \sum_{s=1}^{n} \mathbb{P}_\vartheta\left(p_i \le \frac{\alpha}{\hat{n}_0}, \mid \#\{k \in I_n \setminus \{i\} : p_k \le \lambda\} = s-1\right) \\
&\quad \times \sum_{j=0}^{s-1} \mathbb{P}_\vartheta\left(\#\{k \in I_n \setminus \{i\} : p_k \le \lambda\} = s-1,\ \#\{k \in I_{n,0} : p_k \le \lambda\} = j\right) \\
&= \sum_{s=1}^{n} \Phi\left(\delta - u_{\alpha(1-\lambda)/(n-s+1)}\right) \sum_{j=0}^{s-1} \binom{n_0}{j} \binom{n-n_0-1}{s-1-j} \lambda^j (1-\lambda)^{n_0-j} \\
&\quad \times \left(\Phi(\delta - u_\lambda)\right)^{s-1-j} \left(1 - \Phi(\delta - u_\lambda)\right)^{n-n_0-s+j},
\end{aligned}$$

where u_z denotes the $(1-z)$-quantile of a normal distribution. Note that the power of the Bonferroni test and the power of the OB test are given by

$$\beta(\delta) = \Phi\left(\delta - u_{\alpha/n}\right) \quad \text{and} \quad \beta(\delta) = \Phi\left(\delta - u_{\alpha/n_0}\right),$$

respectively. Clearly, if $\hat{n}_0 < n$, then a BPI test rejects at least as many hypotheses as the classical Bonferroni procedure. Thereby, additional rejections appear if there are $i \in I_n$ such that $p_i \in (\alpha/n, \alpha/\hat{n}_0]$. On the other hand, the power of the OB test seems to be an upper bound of the power of a BPI test. For $n = 50$, $n_0 = 10, 30$, $\alpha = 0.05$ and $\lambda = 0.5$ we compare the power of the BPI test with the threshold $\hat{\alpha}_1$ based on (2.6), the classical Bonferroni test and the OB test in terms of $\delta = \mu\sqrt{k}/\sigma \in [0, 6]$. Figure 2.5 shows that the power of the OB test (full curve) and the BPI (asterisks) test differs only slightly. Clearly, in the model considered here the BPI test is more powerful than the classical Bonferroni test. Although the gain in power for $n_0 = 30$ is not as large as we might wish, the gain in power for $n_0 = 10$ is remarkable. For example, for $\delta = 2, 3, 4$ we obtain $\beta(\delta) = 0.138, 0.464, 0.819$ for the Bonferroni test, $\beta(\delta) = 0.252, 0.645, 0.915$ if $n_0 = 10$

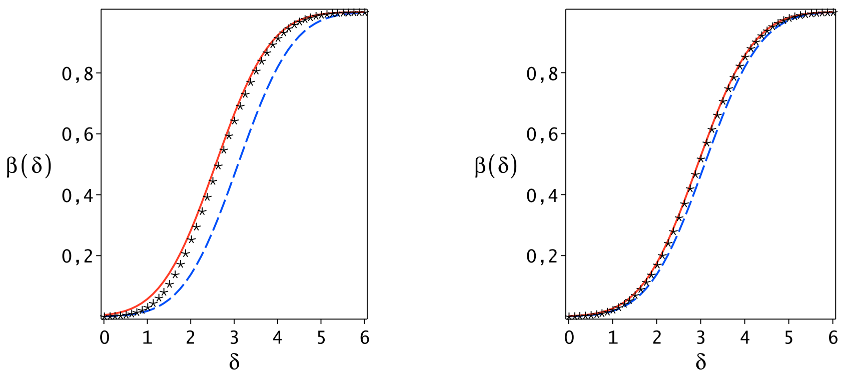

Figure 2.5: Power $\beta(\delta)$ in terms of $\delta = \sqrt{m}\mu/\sigma$ (Bonferroni: dashed line; BPI: asterisks; OB: full curve) for $n = 50$, $\lambda = 0.5$, $n_0 = 10$ (left picture) and $n_0 = 30$ (right picture).

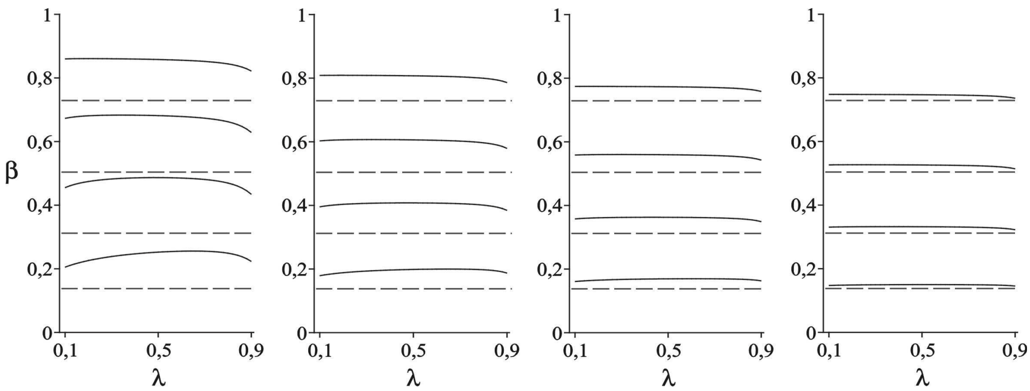

Figure 2.6: Power $\beta \equiv \beta(\delta)$ of the BPI procedure (full curves) for $\delta = \sqrt{m}\mu/\sigma = 2.0,\ 2.6,\ 3.1,\ 3.7$ (from bottom to the top) in terms of λ for $n = 50$ and $n_0 = 10, 20, 30, 40$ (from left to right picture). The power of the Bonferroni test (dashed line) always lies below the corresponding power of the BPI procedure.

and $\beta(\delta) = 0.169, 0.520, 0.853$ if $n_0 = 30$ for the BPI test. In Section 4.7 in Chapter 4 we consider a simulation study, which shows that there are different distributions for which the gain in power is large. In any case, we have to keep in mind that control of the FWER is a very strict criterion and that even critical values of the OB test remain small compared to α as long as the number of true null hypotheses n_0 is not very small.

In conclusion, we look at the dependence of the power of the BPI procedure on the tuning parameter λ in this specific model with $n = 50$ and $\alpha = 0.05$ for $n_0 = 10, 20, 30, 40$ and $\delta = \sqrt{m}\mu/\sigma = 2.0,\ 2.6,\ 3.1,\ 3.7$. Figure 2.6 shows that differences in the power of the BPI procedure (full curve) for various λ-values are small if n_0 and/or δ is large. Moreover, the power decreases in all cases if λ approaches 1. It seems that a λ of around 0.5 is a good compromise. Note that in all the cases that are considered here the power of the BPI procedure is always greater than the power of the Bonferroni procedure (dashed line). Figure 2.6 indicates again that the power gain becomes more apparent for smaller values of n_0.

We conclude this section with a simulated example, where we compare the number of hypotheses which are rejected with the test procedures considered before.

Example 2.24
In the multiple-testing problem given at the beginning of this section we set $\vartheta_i = 0$ for all $i \in I_{n,0}$ and $\vartheta_i = \mu$ for $i \in I_{n,1}$, where μ denotes a random variable following a uniform distribution on $[0, 3]$. Let $n = 40$, $n_0 = 18$, $\alpha = 0.05$ and $\lambda = 0.5$. The BPI test with the threshold $\hat{\alpha}_1$ based on (2.6) and $\kappa = 1$ yielded $\hat{n}_0 = 28$ and for the SDPI test with the optimal $\kappa = 2.76$ we obtained $\hat{n}_0 = 31.52$. The OB test rejected 5 hypotheses, the BPI and SDPI tests rejected 4 each and the Bonferroni test rejected only 2 hypotheses. Thereby, the smallest critical values of the SDPI test were a little smaller than the threshold of the BPI procedure.

2.5 Conclusions

In this chapter, we have proved that a Bonferroni-type procedure based on a suitable plug-in estimate for the number n_0 of true null hypotheses controls the FWER under several distributional assumptions. Typically, the power of a plug-in test is larger than the power of the corresponding classical test and smaller than the power of the associated oracle procedure. The latter implies that we may have a gain in power by a BPI procedure if the corresponding oracle procedure has more power than the classical test.

Note that a plug-in procedure can be more conservative than the corresponding classical test. In fact, $\hat{n}_0$ can be larger than n and consequently the threshold $\hat{\alpha}$ of a plug-in test can be smaller than the threshold of the classical test. This is more likely to occur when n_0 is close to n. Therefore, we do not recommend a BPI procedure if there is prior knowledge that the proportion of true null hypotheses is large. However, if this proportion is not too large, BPI tests are more attractive than classical tests.

Furthermore, we have shown that corresponding SD procedures can be adjusted so that their

FWER is controlled at pre-specified level α. Unfortunately, we cannot recommend this method for small n-values, because our simulations have shown that the power of SDPI tests seems to be smaller than the power of BPI procedures. The reason for this is that κ utilised in SDPI tests need to be larger that in BPI tests. This implies that the smallest critical values of a SDPI procedure are typically smaller than a BPI threshold.

The tuning parameter λ appearing in the estimators (2.6) has to be chosen independently of the data, and the results presented in this chapter for a BPI procedure based on (2.6) heavily depend on this assumption. Note that the estimator (2.9) is a data-dependent version of (2.6) with $\lambda = p_{k:n}$. Some investigations concerning the case of a data-dependent λ can be found in Storey *et al.* [2004]. Obviously, to obtain a meaningful estimate for the number of true null hypotheses n_0, the number of p-values greater than λ should be large enough. In Section 2.4 we speculated that $\lambda \approx 0.5$ may be a good compromise. A further indication for this choice may be that rejection of hypotheses with p-values greater than 0.5 is typically disliked. In any case, it seems there is no uniform best choice for the parameter λ.

A further issue is the choice of k for the estimator $\hat{n}_0$ given in (2.9). Moreover, for $k \leq n - n_0$ the estimator (2.9) can be considerably larger than n_0 so that we prefer to recommend a BPI procedure based on the estimator (2.6).

In contrast to λ and/or k, the choice of κ does not seem to be problematic. It has been proved that in case of independent null p-values $\kappa \equiv 1$ always implies FWER control for a BPI test with critical value (2.4) and/or (2.5) based on (2.6) and α not greater than λ or for a BPI procedure with critical value (2.4) based on (2.9). Thereby, optimal κ-values are only slightly smaller than 1 such that the power of a BPI test with $\kappa = 1$ is almost the same as one of the BPI test with an optimal κ. Note that a BPI test with $\hat{\alpha}_1$ based on (2.6) and $\kappa = 1$ controls the FWER for all $\alpha \in (0, 1)$ and $\lambda \in (0, 1)$ (i.e. α and λ such that $\lambda < \alpha$ are allowed), cf. Guo [2009].

In conclusion, we mention again that if the number of hypotheses n is very large, then the power of any multiple-test procedure controlling the FWER often tends to 0 so that the advantage of a plug-in procedure becomes negligible. For such multiple-testing problems the false discovery rate (FDR) is an attractive alternative error rate criterion. In Chapter 3 we introduce various methods for constructing multiple tests controlling the FDR. Moreover, in Chapter 4 we investigate the FWER of BPI tests in the case of dependent p-values.

Chapter 3

FDR controlling multiple tests related to the asymptotically optimal rejection curve

As mentioned in Chapter 1, application of the FDR criterion allows for more type I errors on the average than application of the FWER criteria, but bounds the proportion of false rejections. Therefore, the usage of the FDR criterion can lead to more rejections. Benjamini and Hochberg [1995] proposed the linear step-up (LSU) procedure, which controls the FDR under several assumptions, cf. Chapter 1. Thereby, the pre-specified α-level is exhausted only if all hypotheses are true while the actual FDR is distinctively smaller than α if the proportion of true null hypotheses is small. Various approaches are available which improve the LSU procedure with respect to the power. For example, Storey *et al.* [2004] suggested plug-in LSU tests which use a plug-in estimate for the number of true null hypotheses n_0, cf. Chapter 2. Another approach can be found in Finner *et al.* [2009]. They constructed a non-linear asymptotically optimal rejection curve (AORC) such that for extreme parameter configurations SUD procedures based on this curve control the FDR at least asymptotically. For a fixed $\alpha \in (0, 1)$, the AORC is defined in (1.2) and the corresponding critical values are given by

$$\alpha_{i:n} = f_\alpha^{-1}(i/n) = \frac{i\alpha}{n - i(1-\alpha)},\ i \in I_n. \tag{3.1}$$

Note that $\alpha_{n:n} = 1$ for all $\alpha \in (0, 1)$ implies that an SU test procedure based on (3.1) always rejects all hypotheses. Moreover, Finner *et al.* [2009] showed that SUD procedures based on the AORC critical values typically do not control the FDR for a finite number of all hypotheses. It follows that the critical values (3.1) have to be adjusted in order to obtain finite FDR control for an SUD test. Finner *et al.* [2009] proposed SUD procedures with slightly adjusted AORC critical values (replace n by $n + \beta_n$ in the denominator of the AORC critical values for a suitable β_n). Gavrilov *et al.* [2009] proved that SD tests with $\beta_n = 1$ control the FDR under the usual independence assumptions. Clearly, an SUD procedure rejects at least as many null hypotheses as the SD test with the same set of critical values and the corresponding SU test is the most powerful. Hence, the construction of AORC related multiple tests controlling the FDR for fixed n-values, and exhausting the pre-specified FDR level as sharply as possible, remains an open problem.

In this chapter we focus on exact control of the FDR for step-up-down (SUD) test procedures related to the asymptotically optimal rejection curve (AORC). In Section 3.1 we introduce the class of SUD tests, which includes SU and SD procedures, and derive explicit formulas for upper bounds of their FDR. In the case of SU tests we obtain that upper bounds for the FDR are the FDR-values in DU models. We show under several assumptions that upper bounds and FDRs of SUD tests in DU models coincide asymptotically. Moreover, we prove that FDR control of an SU test implies FDR control of all SUD tests with the same set of critical values. We provide conditions under which FDR control of an SUD test follows from FDR control of the corresponding SD test. In Section 3.2 we provide a recursive scheme for the computation of critical values leading to the pre-specified FDR-values. We also consider a possibility to compute a feasible set of critical values such that the corresponding FDR-values coincide with the pre-specified FDR-values for larger numbers of true hypotheses. In Section 3.3 we introduce alternative FDR bounding curves and show their connection to rejection curves. We give some examples of FDR bounding curves and discuss the solvability of the corresponding recursive schemes. Section 3.4 deals with various methods based on the AORC. We show how critical values corresponding to the AORC or to a modified AORC can be adjusted in order to obtain finite FDR control. For single-parameter adjustment methods we investigate the behaviour of the adjusting parameters for SUD test procedures. We also consider an adjustment method, which modifies critical values $\alpha_{i:n}$ depending on $i \in I_n$ and discuss a possibility of exact solving. In Section 3.5 we introduce an approach for the computation of critical values yielding finite FDR control which is based on the fixed point theorem. This iterative method combined with a β-adjustment yields a good (and may be the best) set of critical values. Finally, in Section 3.6 we discuss advantages and disadvantages of each method.

3.1 SUD tests and upper FDR bounds

Throughout this chapter, we consider a multiple-testing problem described in Notation 1.1. Moreover, we make the general assumptions that the conditions (I1) and (I2) are fulfilled, that is, p_i, $i \in I_{n,0}(\vartheta)$, are independent and that $(p_i : i \in I_{n,0})$ and $(p_i : i \in I_n \setminus I_{n,0})$ are independent random vectors. Suppose that $\varphi = (\varphi_i : i \in I_n)$ is defined in terms of critical values (1.1) such that the corresponding continuous critical value function ρ fulfils the condition (A1), which implies in particular that ρ is strictly increasing. Below, we call critical values (1.1) fulfilling (A1) **feasible**. As before, a rejection curve associated with ρ is defined by $r = \rho^{-1}$. Note that r and ρ may depend on the number of hypotheses n but do not depend on n in asymptotic considerations. Moreover, we define $q(t) = \rho(t)/t$ for $t \in (0, 1]$ and $q(0) = \lim_{t \to 0} \rho(t)/t$. Thereby, $0 \leq q(0) \leq 1$ if condition (A1) applies. It holds $q(1) \leq 1$.

First we give a formal definition of SUD test procedures.

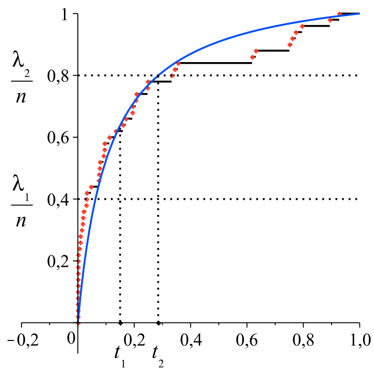

Figure 3.1: The ecdf of $n = 50$ p-values, where $n_0 = 15$ p-values correspond to true null hypotheses, and the AORC with $\alpha = 0.1$. An SUD test based on the AORC with $\lambda_1 = 40$ ($\lambda_2 = 80$) rejects hypotheses with p-values which are not greater than t_i.

Definition 3.1
For $\lambda \in I_n$ an SUD(λ) procedure $\varphi^\lambda = (\varphi_1, \ldots, \varphi_n)$ of order λ is defined as follows. If $p_{\lambda:n} \leq \alpha_{\lambda:n}$, set $m_n = \max\{j \in \{\lambda, \ldots, n\} : p_{i:n} \leq \alpha_{i:n}$ for all $i \in \{\lambda, \ldots, j\}\}$, whereas for $p_{\lambda:n} > \alpha_{\lambda:n}$, put $m_n = \sup\{j \in \{1, \ldots, \lambda - 1\} : p_{j:n} \leq \alpha_{j:n}\}$ $(\sup \emptyset = -\infty)$. Define $\varphi_i = 1$ if $p_i \leq \alpha_{m_n:n}$ and $\varphi_i = 0$ otherwise $(\alpha_{-\infty:n} = -\infty)$. Thereby, $\lambda = 1$ yields an SD procedure and $\lambda = n$ yields an SU procedure.

An SUD test φ^λ can be defined in terms of a random threshold t^* depending on the data, that is, $\varphi_i = 1$ if and only if $p_i \leq t^*$, cf. Section 1.2 in Chapter 1. If $p_{\lambda:n} \leq \alpha_{\lambda:n}$, then t^* is the smallest crossing point between r and the ecdf $\hat{F}_n$ such that $p_{\lambda:n} \leq t^*$. If $p_{\lambda:n} > \alpha_{\lambda:n}$, then t^* is the largest crossing point between r and the ecdf $\hat{F}_n$ such that $t^* < p_{\lambda:n}$. Figure 3.1 shows the ecdf of $n = 50$ p-values ($n_0 = 15$) and the AORC defined in (1.2) with $\alpha = 0.1$. An SUD test based on the AORC with $\lambda_1 = 40$ ($\lambda_2 = 80$) rejects null hypotheses with p-values being not greater than t_i. Clearly, an SUD(λ_2) test rejects at least as many hypotheses as an SUD(λ_1) procedure if $\lambda_1 \leq \lambda_2$. Therefore, an SU test rejects the most hypotheses.

As mentioned before, although an SUD procedure φ^λ based on the AORC (i.e. with critical values (3.1)) with $\lambda \in I_{n-1}$ does not necessarily reject all hypotheses, the FDR is not controlled for a fixed $n \in \mathbb{N}$. Benjamini and Yekutieli [2001] showed that in the case of SU tests, feasible critical values imply that the FDR becomes larger if p-values decrease stochastically, cf. Theorem 1.2. This implies that DU models are least favourable parameter configurations (LFC) for the FDR of an SU test if p-values corresponding to true null hypotheses are iid uniformly distributed on $[0, 1]$, cf. Chapter 1. Thereby, $\mathbb{P}_{n,n_0}$ denotes the probability measure in the DU(n, n_0) model, that is, all p-values p_i, $i \in I_{n,0}$, are iid uniformly distributed on $[0, 1]$, and all p_i, $i \in I_{n,1}$, follow a Dirac distribution with point mass 1 at 0. Unfortunately, for an SUD procedure with $\lambda \in I_{n-1}$ it is not known whether DU configurations are least favourable. It is also not known whether the FDR of an SU test is maximum in DU models if p-values corresponding to true null hypotheses are stochastically larger than a uniform variate. However, Finner *et al.* [2009] showed that DU configurations yield an upper bound for the FDR of SUD procedures. Moreover, this upper bound for an SU procedure is sharp if the corresponding DU configuration belongs to the model. For the computation of upper bounds $b(n, n_0)$ we need formulas for the probability mass function (pmf) of V_n under DU configurations. They can be obtained in terms of the joint cumulative distribution function (cdf) of order statistics.

Let $0 \leq c_{1:n} \leq \cdots \leq c_{n:n} \leq 1$ and $n \in \mathbb{N}$ be given. Then a general recursive formula for the joint cdf F_n^k of the order statistics $U_{1:n}, \ldots, U_{n-k:n}$, $0 \leq k \leq n$, of n iid uniformly distributed random variables U_i is given by

$$F_n^k(c_{1:n}, \ldots, c_{n-k:n}) = 1 - \sum_{j=0}^{n-k-1} \binom{n}{j} F_j(c_{1:n}, \ldots, c_{j:n})(1 - c_{j+1:n})^{n-j}, \tag{3.2}$$

with $F_n^0 = F_n$ and $F_0^0 \equiv F_n^n \equiv 1$. We apply formula (3.2) (for $k = 0$), which is essentially Bolshev's recursion, cf. Shorack and Wellner [1986], pp. 366-367, for the calculation of the pmf of V_n for an SUD procedure of order λ under DU configurations. The next lemma yields this result.

Lemma 3.2

For the pmf of V_n of an SUD(λ) procedure based on critical values $0 \leq \alpha_{1:n} \leq \ldots \leq \alpha_{n:n} \leq 1$ under a DU configuration with n_0 true null hypotheses and $n_1 = n - n_0$ false hypotheses, we obtain that $\mathbb{P}_{n,n_0}(V_n = j)/\binom{n_0}{j}$ is equal to

$$\begin{cases} F_j(\alpha_{n_1+1:n}, \ldots, \alpha_{n_1+j:n})\bar{\alpha}_{n_1+j+1:n}^{n_0-j}, & \text{if } \lambda \leq n_1, \\ F_{n_0-j}(\underbrace{\bar{\alpha}_{\lambda:n}, \ldots, \bar{\alpha}_{\lambda:n}}_{n-\lambda+1}, \bar{\alpha}_{\lambda-1:n}, \ldots, \bar{\alpha}_{n_1+j+1:n})\alpha_{n_1+j:n}^{j}, & \text{if } \lambda > n_1 \wedge j < \lambda - n_1, \\ F_j(\underbrace{\alpha_{\lambda:n}, \ldots, \alpha_{\lambda:n}}_{\lambda-n_1}, \alpha_{\lambda+1:n}, \ldots, \alpha_{n_1+j:n})\bar{\alpha}_{n_1+j+1:n}^{n_0-j}, & \text{if } \lambda > n_1 \wedge j \geq \lambda - n_1, \end{cases}$$

for $j = 0, \ldots, n$, where $\bar{\alpha}_{j:n} = 1 - \alpha_{j:n}$, $j \in I_n$.

Proof: For notational convenience, we denote the p-values corresponding to true null hypotheses by p^0-values. The vector of ordered p-values $(p_{1:n}, \ldots, p_{n:n})$ is almost surely of the form

$$(\underbrace{p_{1:n} = 0 = \ldots = 0 = p_{n_1:n}}_{n_1}, p_{n_1+1:n} = p^0_{1:n_0}, \ldots, p^0_{n_0:n_0} = p_{n:n}).$$

Case 1. Let $\lambda \leq n_1$. In this case, we necessarily fall into the SD branch of the test procedure, because at least the first λ components of the vector of ordered p-values are 0 such that $p_{\lambda:n} \leq \alpha_{\lambda:n}$ is true with probability 1. Consequently, the event $\{V_n = j\}$ can be expressed as

$$\{V_n = j\} = \{p_{n_1+1:n} \leq \alpha_{n_1+1:n}, \ldots, p_{n_1+j:n} \leq \alpha_{n_1+j:n}\} \cap \{p_{n_1+j+1:n} > \alpha_{n_1+j+1:n}\}.$$

Since the second event implies that all ordered p-values with ordered indices $n_1 + j + 1$ or greater are larger than $\alpha_{n_1+j+1:n}$, the event means that $(n_0 - j)$ p^0-values are greater than $\alpha_{n_1+j+1:n}$.

Since we have $\binom{n_0}{j}$ possibilities to choose these p^0-values and all p-values are assumed to be independent, we immediately obtain the result.

Case 2. Let $\lambda > n_1$ and $j < \lambda - n_1$. In order to reach this case, we must have $p_{\lambda:n} > \alpha_{\lambda:n}$ and fall into the SU branch of the procedure. Consequently, we can write

$$\{V_n = j\} = \{p_{\lambda:n} > \alpha_{\lambda:n}, p_{\lambda-1:n} > \alpha_{\lambda-1:n}, \ldots, p_{n_1+j+1:n} > \alpha_{n_1+j+1:n}\} \cap \{p_{n_1+j:n} \leq \alpha_{n_1+j:n}\}.$$

Since $p_{n_1+j:n} \leq \alpha_{n_1+j:n}$ automatically implies that $p^0_{k:n_0} \leq \alpha_{n_1+j:n}$ for all $k = 1, \ldots, j$, we can again choose j of the n_0 p^0-values to fulfil this relationship.

Case 3. Let $\lambda > n_1$ and $j \geq \lambda - n_1$. In this third case, we fall into the SD branch of the procedure, resulting in

$$\{V_n = j\} = \{p_{\lambda:n} \leq \alpha_{\lambda:n}, p_{\lambda+1:n} \leq \alpha_{\lambda+1:n}, \ldots, p_{n_1+j:n} \leq \alpha_{n_1+j:n}\} \cap \{p_{n_1+j+1:n} > \alpha_{n_1+j+1:n}\}.$$

The assertion then follows in analogy to the SD considerations under case 1. ■

Corollary 3.3

For an SU procedure with critical values $0 \leq \alpha_{1:n} \leq \cdots \leq \alpha_{n:n} \leq 1$ in a DU(n, n_0) model we obtain

$$\mathbb{P}_{n,n_0}(V_n = j) = \binom{n_0}{j} F_{n_0-j}(1 - \alpha_{n:n}, \ldots, 1 - \alpha_{n-n_0+j+1:n}) \alpha^j_{n-n_0+j:n}, \tag{3.3}$$

cf. Lemma 3.2 in Finner and Roters [2002].

This result is immediate if we consider the case $\lambda = n$ in Lemma 3.2. Alternatively, the pmf of V_n in this case can be calculated by the recursive formula (3.9) below.

Now we present an upper bound for the FDR of an SUD(λ) procedure φ_n which was introduced in Finner *et al.* [2009]. The following result corresponds to the slightly more general Theorem 4.3 in Finner *et al.* [2009]. In what follows, $\mathbb{P}_{\vartheta^i}$ refers to the situation where $(p_j : j \in I_n \backslash \{i\})$ has the same distribution under ϑ^i as under ϑ except that we put $p_i \equiv 0$ under ϑ^i.

Theorem 3.4

Let $\vartheta \in \Theta$ be such that $n_0 \in \mathbb{N}$ hypotheses are true and the remaining ones are false. Let $i \in I_{n,0}$. Then, for an SUD(λ) test with $\lambda \in I_n$ based on a rejection curve ρ it holds under (I1), (I2) and (A1) by setting $q(t) = \rho(t)/t$ that

$$\begin{aligned} FDR_\vartheta(\varphi^\lambda) &\leq \frac{n_0}{n} \sum_{j=1}^{n} q(j/n) \mathbb{P}_{\vartheta^i}(R_n/n = j/n) = \frac{n_0}{n} \mathbb{E}_{\vartheta^i} q(R_n/n) & (3.4) \\ &\leq \frac{n_0}{n} \mathbb{E}_{n,n_0-1} q(R_n/n), & (3.5) \end{aligned}$$

with equality in (3.4) for an SU test (i.e. for $\lambda = n$) if (D1) is fulfilled.

Remark 3.5

In Theorem 4.3 in Finner *et al.* [2009] p-values corresponding to true null hypotheses have to be uniformly distributed on $[0,1]$, i.e. (D1) has to be fulfilled. However, it can be easily seen that for p_i, $i \in I_{n,0}$, being stochastically larger than a uniform variate, the expression in (3.4) is also an upper bound for the FDR, cf. proof of Theorem 4.1 (inequality (4.4) in particular) and proof of Theorem 4.3 in Finner *et al.* [2009].

Note that (3.5) is a ϑ-free upper bound for the FDR for all $n_0 \in I_n$. Setting

$$b(n, n_0|\lambda) = \frac{n_0}{n}\mathbb{E}_{n,n_0-1}\left[q(R_n/n)\right],\ n_0 \in I_n, \tag{3.6}$$

and $b_n^* = \max_{1\le n_0\le n} b(n,n_0)$, we obtain $\sup_{\vartheta\in\Theta} \mathrm{FDR}_\vartheta(\varphi) \le b_n^*$.

An explicit representation of the upper FDR bound $b(n, n_0|\lambda)$ for SUD(λ) tests is given in the next theorem.

Theorem 3.6
For an SUD(λ) procedure with $\lambda \in I_n$ and critical values (1.1) *satisfying (A1), it holds*

$$b(n, n_0|\lambda) = n_0 \sum_{j=1}^{n_0} \frac{\alpha_{n_1+j:n}}{n_1+j}\mathbb{P}_{n,n_0-1}(V_n = j-1), \tag{3.7}$$

where $n_1 = n - n_0$. For an SU test, that is $\lambda = n$, $b(n, n_0|n)$ can alternatively be calculated by

$$b(n, n_0|n) = \sum_{j=1}^{n_0} \frac{j}{n_1+j}\mathbb{P}_{n,n_0}(V_n = j) = FDR_{n,n_0}(\varphi^n) \tag{3.8}$$

and it even holds equality in every summand in (3.7) and (3.8), i.e.

$$\mathbb{P}_{n,n_0}(V_n = j) = \frac{n_0}{j}\alpha_{n_1+j:n}\mathbb{P}_{n,n_0-1}(V_n = j-1)\ \ for\ \ j \in I_{n_0}. \tag{3.9}$$

Proof: In order to prove (3.7), we keep in mind that the expectation in (3.6) refers to a DU configuration with (n_0-1) true null hypotheses and (n_1+1) false hypotheses and since $p_j \sim \varepsilon_0$ for all $j \in I_{n,1}$, we get $R_n = V_n + (n_1+1)$ $\mathbb{P}_{n,n_0-1}$-almost surely. A straightforward calculation now yields

$$\begin{aligned}
\frac{n_0}{n}\mathbb{E}_{n,n_0-1}\left[q\left(\frac{R_n}{n}\right)\right] &= \frac{n_0}{n}\mathbb{E}_{n,n_0-1}\left[\frac{\rho(R_n/n)}{R_n/n}\right] = n_0\mathbb{E}_{n,n_0-1}\left[\frac{\alpha_{R_n:n}}{R_n}\right] \\
&= n_0\mathbb{E}_{n,n_0-1}\left[\frac{\alpha_{V_n+n_1+1:n}}{V_n+n_1+1}\right] \\
&= n_0\sum_{k=0}^{n_0-1}\frac{\alpha_{k+n_1+1:n}}{k+n_1+1}\mathbb{P}_{n,n_0-1}(V_n = k) \\
&= n_0\sum_{j=1}^{n_0}\frac{\alpha_{n_1+j:n}}{n_1+j}\mathbb{P}_{n,n_0-1}(V_n = j-1),
\end{aligned}$$

which is formula (3.7). Equality (3.9) and consequently the left-hand side equality of (3.8) are immediate consequences of the representation of the pmf of V_n for an SU test φ^n given in Corollary 3.3. The right-hand side equality follows with Theorem 3.4. ■

A natural question concerns the quality of the upper bounds (3.6). The next lemma shows that (3.6) and the FDR often coincide asymptotically in DU models.

Lemma 3.7
Let φ_n be an SUD(λ_n) test based on some rejection curve r with $\rho = r^{-1}$ satisfying (A1) and $\lambda_n/n \to \kappa \in [0,1]$. Consider a sequence of DU(n,n_0) models with $n_0(n)/n \to \zeta \in [0,1]$ and suppose that R_n/n converges to some fixed value at least in probability. Then the bound given in (3.6) converges to the limiting FDR under DU(n,n_0), that is,

$$\lim_{n\to\infty} b(n,n_0(n)) = \lim_{n\to\infty} FDR_{n,n_0(n)} \tag{3.10}$$

for all $\zeta \in [0,1]$ if $\kappa \in (0,1]$ and for all $\zeta \in [0,1)$ if $\kappa = 0$ (which includes SD procedures).

Proof: Let $t_n^* \in [0,1]$ be the crossing point between r and the ecdf of p-values $\hat{F}_n$ such that $r(t_n^*) = \hat{F}_n(t_n^*) = R_n/n$, that is, φ_n rejects hypotheses with p-values not greater than t_n^*. Note that the existence of t_n^* is guaranteed by the structure of SUD test procedures. From the convergence of R_n/n we get that there exists a $t^* \in [0,1]$ such that $t_n^* \to t^*$, $n \to \infty$, in probability. Then φ_n rejects asymptotically all hypotheses with p-values not greater than t^*. For $t^* > 0$ this implies that $\mathrm{FDR}_{n,n_0} \to \zeta t^*/(1-\zeta+\zeta t^*)$, $n \to \infty$. Moreover, if $t^* > 0$ we obtain $b(n,n_0) \to \zeta t^*/r(t^*) = \zeta t^*/F_\infty(t^*|\zeta)$, where $F_\infty(t^*|\zeta) = \lim_{n\to\infty} \hat{F}_n(t^*) = 1-\zeta+\zeta t^*$. Hence, for $\zeta < 1$ (i.e. $t^* > 0$) we get equation (3.10) for all $\kappa \in [0,1]$.

Now we consider the case of $t^* = 0$ (i.e. $\zeta = 1$) and $\kappa > 0$. Theorem 4.3 in Finner *et al.* [2009] yields

$$\mathrm{FDR}_{n,n_0} = n_0 \sum_{j=1}^{n} \frac{\alpha_{j:n}}{j} \mathbb{P}_{n,n_0}(R_n = j | p_{i_0} \le \alpha_{j:n}),$$

where $i_0 \in I_{n,0}$ and $\alpha_{j:n} = \rho(j/n)$, $j \in I_n$. Then setting

$$C_{1,n} = n_0 \sum_{j=1}^{\lambda_n} \frac{\alpha_{j:n}}{j} \mathbb{P}_{n,n_0}(R_n = j | p_{i_0} \le \alpha_{j:n})$$

and

$$C_{2,n} = n_0 \sum_{j=\lambda_n+1}^{n} \frac{\alpha_{j:n}}{j} \mathbb{P}_{n,n_0}(R_n = j | p_{i_0} \le \alpha_{j:n}),$$

we obtain $\mathrm{FDR}_{n,n_0} = C_{1,n} + C_{2,n}$. The statement (4.2) in Finner *et al.* [2009] yields

$$\mathbb{P}_{n,n_0}(R_n = j | p_{i_0} \le \alpha_{j:n}) = \mathbb{P}_{n,n_0-1}(R_n = j) \ \text{ for } \ j \le \lambda_n$$

and consequently

$$C_{1,n} = n_0 \sum_{j=1}^{\lambda_n} \frac{\alpha_{j:n}}{j} \mathbb{P}_{n,n_0-1}(R_n = j).$$

Since $\mathbb{P}_{n,n_0-1}(R_n = j + n_1) = \mathbb{P}_{n,n_0-1}(V_n = j-1)$ for $j \in I_{n_0}$, the representation (3.7) of the upper bound in Theorem 3.6 implies $b(n,n_0|\lambda_n) = C_{1,n} + C_{3,n}$, where

$$C_{3,n} = n_0 \sum_{j=\lambda_n+1}^{n} \frac{\alpha_{j:n}}{j} \mathbb{P}_{n,n_0-1}(R_n = j).$$

We will show that $C_{2,n}$ and $C_{3,n}$ converge to 0 if n increases, which implies $\lim_{n\to\infty} \mathrm{FDR}_{n,n_0} = \lim_{n\to\infty} C_{1,n} = \lim_{n\to\infty} b(n,n_0)$. Since (A1) applies and $\alpha_{n:n} \le 1$, we get $\alpha_{j:n}/j \le 1/n$ for all $j \in I_n$, hence

$$C_{2,n} \le C_{3,n} \le \frac{n_0}{n}\mathbb{P}_{n,n_0-1}(R_n > \lambda_n). \tag{3.11}$$

Note that $t^* = 0$ is equivalent to $R_n/n \to 0$, $n \to \infty$, in probability in DU(n, n_0) models. It follows that there are no crossing points between the limiting ecdf of p-values (i.e. $F_\infty(t) = t$) and the rejection curve r in $(0, \kappa]$. Hence, $R_n/n \to 0$, $n \to \infty$, in probability in DU($n, n_0 - 1$) models, too. For $\lambda_n/n \to \kappa > 0$, we obtain

$$\forall\, \epsilon > 0 : \ \exists\, N_\epsilon \in \mathbb{N} : \ \forall\, n \ge N_\epsilon : \ \mathbb{P}_{n,n_0-1}(R_n > \lambda_n) \le \epsilon.$$

Thus, (3.11) yields that for all $\epsilon > 0$ there exists an $N_\epsilon \in \mathbb{N}$ such that for all $n \ge N_\epsilon$ we get $C_{2,n} \le C_{3,n} \le \epsilon$, which completes the proof. ■

Remark 3.8

Note that for $\kappa = 0$ and $\zeta = 1$ the bound and the FDR may not be equal in the limit. For example, for $n_0 = n$ the FDR of an SD test based on f_α equals $1 - (1 - \alpha_{1:n})^n$ which converges to $1 - \exp(-\alpha) < \alpha = \lim_{n\to\infty} b(n, n|1)$.

As shown before, upper bounds given in (3.7) (which are equal to (3.6)) for an SD test are not sharp such that they can be improved. One possibility can be derived from results that are implicitly contained in Gavrilov *et al.* [2009], cf. proof of Theorem 1.1, p. 623, the second line in formula (3.5). The next corollary yields this result. Below, $\mathbb{P}_{\vartheta-i}$ denotes a probability measure for which p-values have almost the same distribution under $\mathbb{P}_\vartheta$ as under $\mathbb{P}_{\vartheta-i}$, the only difference being that $p_i \equiv 1$ under $\mathbb{P}_{\vartheta-i}$.

Corollary 3.9

Let $\vartheta \in \Theta$ such that $n_0 \in \mathbb{N}$ hypotheses are true and the remaining ones are false. Let $i_0 \in I_{n,0}$. If (I1) and (I2) are fulfilled for an SD test φ^1 with critical values (1.1)*, then*

$$\begin{aligned} \mathit{FDR}_\vartheta(\varphi^1) &\le \frac{n_0}{n}\sum_{j=1}^{n} q(j/n)\mathbb{P}_{\vartheta-i_0}(R_n/n = (j-1)/n) \\ &= \frac{n_0}{n}\mathbb{E}_{\vartheta-i_0} q\left(\frac{R_n+1}{n}\right). \end{aligned} \tag{3.12}$$

Moreover, the upper bound (3.12) attains its maximum in DU configurations, i.e.

$$\frac{n_0}{n}\mathbb{E}_{\vartheta-i_0} q\left(\frac{R_n+1}{n}\right) \le b^{SD}(n, n_0) \ \ \textit{(say)},$$

where

$$b^{SD}(n,n_0) = n_0 \sum_{j=1}^{n_0} \frac{\alpha_{n_1+j:n}}{n_1+j}\mathbb{P}_{n-1,n_0-1}(V_{n-1} = j-1|\alpha_{1:n},\ldots,\alpha_{n-1:n}). \tag{3.13}$$

It is well-known that an SU test rejects at least as many null hypotheses as an SD test with the same set of critical values. But a question of general interest is whether FDR control of an SU procedure implies FDR control of the corresponding SUD procedures. An investigation concerning this problem can be found in Blanchard and Roquain [2008]. They gave a specific dependency condition, under which FDR control of an SD test follows from FDR control of the corresponding SU procedure. Note that the dependence condition given in Blanchard and Roquain [2008] results in very restrictive conditions on the critical values.

The next theorem yields the desired result for SUD tests requiring more restrictive distributional assumptions but only the simple monotonicity property (A1) on the critical values.

Theorem 3.10
Consider an SU test φ^n and an SUD(λ) test φ^λ with the same set of critical values $0 \leq \alpha_{1:n} \leq \ldots \leq \alpha_{n:n} \leq 1$ and $\lambda \in I_{n-1}$. Then, under assumptions (D1),(I1),(I2) and (A1) it holds

$$FDR_\vartheta(\varphi^\lambda) \leq FDR_\vartheta(\varphi^n) \text{ for all } \vartheta \in \Theta. \tag{3.14}$$

Hence, if the FDR is controlled by the SU test φ^n, then the SUD(λ) test φ^λ also controls the FDR. Moreover, the bounds $b(n, n_0|\lambda)$ defined in (3.6) are non-decreasing in $\lambda \in I_n$ ((D1) is not required for this).

Proof: Set $R_n^\lambda = R_n$ for an SUD(λ) test. An SUD(λ_2) test rejects at least as many hypotheses as an SUD(λ_1) test for any $1 \leq \lambda_1 \leq \lambda_2 \leq n$, which implies that $R_n^{\lambda_1}$ is stochastically not greater than $R_n^{\lambda_2}$. Under (A1) we obtain that $\rho(R_n^\lambda/n)/(R_n^\lambda/n)$ is stochastically non-decreasing in λ, hence the bounds $b(n, n_0|\lambda)$ defined in (3.6) and $\mathbb{E}_{\vartheta^i} q(R_n^\lambda/n)$ are non-decreasing in λ. Since (D1) is fulfilled, we get together with Theorem 3.4 that

$$\mathrm{FDR}_\vartheta(\varphi^\lambda) \leq \frac{n_0}{n}\mathbb{E}_{\vartheta^i} q(R_n^\lambda/n) \leq \frac{n_0}{n}\mathbb{E}_{\vartheta^i} q(R_n^n/n) = \mathrm{FDR}_\vartheta(\varphi^n).$$

■

By means of Theorem 3.10 we have an alternative method of obtaining FDR controlling SUD procedures. Once we have an SU procedure with critical values (1.1) controlling the FDR, all corresponding SUD procedure with the same set of critical values control the FDR, too. Unfortunately, for $\lambda < n$ the calculation time for the pmf of V_n via the formula in Lemma 3.2 increases rapidly if n increases. For an SU test (i.e. $\lambda = n$) all computations are much easier and faster due to the efficient recursive formula (3.9). In any case, as long as we are able to compute the pmf of V_n for an SUD(λ) procedure with fixed critical values, we can easily compute the bounds for the FDR given in Theorem 3.6.

Note that Theorem 3.10 also implies that an FDR controlling SUD test can be based on larger critical values than an SU procedure which controls the FDR. On the other hand, for fixed critical values an SUD(λ_1) test rejects at least as many hypotheses as an SUD(λ_2) test if λ_1 is larger than λ_2. Hence, there is a trade-off between the conservativity of critical values and the conservativity of the test structure, quantified by the parameter λ of the SUD test.

The following lemma is a partial reverse of Theorem 3.10 and shows that FDR control of an SD test sometimes implies FDR control of the corresponding SUD(λ) test for certain values of λ.

Lemma 3.11
Let φ^λ with $\lambda \in I_n$ denote SUD(λ) tests with fixed critical values satisfying (A1) such that $b(n, n_0|1) \leq \alpha$ for all $n_0 \in I_n$, that is, the SD test controls the FDR at level α. Define

$$n_0^* = \min\{k \in I_n : \mathit{FDR}_{n,n_0}(\varphi^n) \leq \alpha \ \text{ for all } \ n_0 = k+1, \dots, n\} \tag{3.15}$$

with the convention $\min \emptyset = \infty$. If $n_0^* \leq n$, then $\mathit{FDR}_{n,n_0}(\varphi^\lambda) \leq \alpha$ for all $n_0 \in I_n$ and all $\lambda \leq n - n_0^* + 1$, that is, an SUD($\lambda$) test controls the FDR at level α if $\lambda \leq n - n_0^* + 1$.

Proof: Suppose that $n_0^* \leq n$. Theorem (3.10) yields that $\mathrm{FDR}_{n,n_0}(\varphi^\lambda) \leq \alpha$ for $n_0 = n_0^* + 1, \dots, n$ and $\lambda \in I_n$. A look at Lemma 3.2 and formula (3.7) immediately yields for $\lambda \in I_n$ that

$$b(n, n_0|\lambda) = b(n, n_0|1) \ \text{ for all } \ n_0 \leq n - \lambda + 1.$$

Hence, for $\lambda \leq n - n_0^* + 1$ we obtain

$$\mathrm{FDR}_{n,n_0}(\varphi^\lambda) \leq b(n, n_0|\lambda) = b(n, n_0|1) \leq \alpha \ \text{ for all } \ n_0 \leq n_0^*$$

which completes the proof. ∎

If it is known that an SD test controls the FDR for some fixed critical values, then we can try to find some $n_0^* \in I_n$, which ensure the conditions in Lemma 3.11. Note that in this case it is only necessary to check whether the corresponding SU test with the same critical values controls the FDR for larger numbers of true null hypotheses. Thereby, an SU test requires less computation time than an SUD procedure.

3.2 General computational issues

The formulas derived in Section 2.1 imply that it suffices to check FDR control of an SUD procedure at level $\alpha \in (0, 1)$ for all DU configurations. Since each SUD(λ) procedure with $\lambda \in I_n$ rejects all $n - n_0$ false hypotheses with probability 1 under DU(n, n_0) configurations, we only have to prove that the FDR is less than or equal to $g^*(n_0/n)$ in this case, where the function g^* is defined by $g^*(\zeta) = \min\{\alpha, \zeta\}$ for $\zeta \in [0, 1]$ and plays an important role below. It follows that

$$b(n, n_0|\lambda) \leq g^*(n_0/n) \ \text{ for all } \ n_0 \in I_n, \tag{3.16}$$

yields that the SUD test φ^λ controls the FDR at level α. Clearly, our objective is to exhaust the FDR level given by the function g^* for SUD procedures.

For a start, suppose for a moment that for each $n_0 \in I_n$ the FDR under a DU(n, n_0) configuration should be bounded by $g(n_0/n)$ for an arbitrary but fixed function $g : [0, 1] \to [0, 1]$. To achieve this, we require with respect to (3.7) that

$$n_0 \sum_{j=1}^{n_0} \frac{\alpha_{n_1+j:n}}{n_1 + j} \mathbb{P}_{n,n_0-1}(V_n = j - 1) = g(n_0/n) \ \text{ for all } \ n_0 \in I_n. \tag{3.17}$$

For $n_0 = 1$ this results in

$$\alpha_{n:n} = ng(1/n). \tag{3.18}$$

Setting

$$h_{n_0}(\alpha_{n-n_0+2:n}, \ldots, \alpha_{n:n}) =$$
$$\frac{n - n_0 + 1}{n_0 \mathbb{P}_{n,n_0-1}(V_n = 0)} \left[g(n_0/n) - n_0 \sum_{j=2}^{n_0} \frac{\alpha_{n_1+j:n}}{n_1 + j} \mathbb{P}_{n,n_0-1}(V_n = j - 1) \right],$$

we obtain

$$\alpha_{n-n_0+1:n} = h_{n_0}(\alpha_{n-n_0+2:n}, \ldots, \alpha_{n:n}) \tag{3.19}$$

for $2 \leq n_0 \leq n$, i.e., we get a recursive scheme for the determination of critical values. As a matter of course, we have to check whether the resulting solution is feasible.

Unfortunately, for $g \equiv g^*$ this recursive scheme only leads to feasible critical values for very small values of n. For example, for $\alpha = 0.05$ and SU tests, we only get feasible solutions for $n \leq 6$, cf. Kwong and Wong [2002].

However, a question of more general interest is to find functions g such that condition (3.17) leads to feasible critical values for all $n \in \mathbb{N}$. There exists at least one such function, that is $g(\zeta) = \zeta\alpha$, $\zeta \in [0, 1]$, which corresponds to the LSU procedure introduced in Benjamini and Hochberg [1995]. Further candidates will be presented in Section 3.3.

In order to exhaust the FDR-level and to find feasible critical values close to AORC-based critical values, we can try to relax (3.18) and (3.19) as follows. In a first step one may choose $m \in I_{n-1}$ starting values $\alpha_{n-i+1:n} \leq \cdots \leq \alpha_{n:n}$, $i \in I_m$, satisfying all constraints required for a feasible solution and

$$b(n, i|\lambda) \leq g^*(i/n) \ \text{ for } \ i = 1, \ldots, m, \tag{3.20}$$

where some of the inequalities may be strict. In a second step one can try to examine whether recursive computation of the remaining critical values via (3.19) leads to a feasible solution with

$$b(n, i|\lambda) = g^*(i/n) \ \text{ for } \ i = m + 1, \ldots, n. \tag{3.21}$$

Although this proposal sounds attractive, it turns out to be a balancing act and extremely sensitive with respect to the initial critical values, which will be shown in Section 3.4. Our experience is that one needs to be lucky to find a feasible solution with this method for larger values of n. The main reason for the sensitivity of this method seems to be that the new critical value to be calculated via (3.19) is the smallest critical value in the support of the distribution of V_n and typically has very small impact on the actual FDR. Figure 3.2 shows the AORC (red curve) and the cdf of p-values (black line) in the DU(n, n_0) model. The crossing point t_ζ (say), which specifies the FDR for an SUD test, is typically greater (and asymptotically strictly greater) than the smallest critical value $\alpha_{n:n-n_0+1}$, such that it is not possible to obtain $b(n, n_0) = \alpha$ by adjusting $\alpha_{n:n-n_0+1}$.

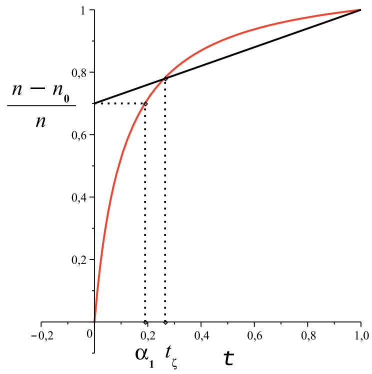

Figure 3.2: AORC (red curve) and the cdf of p-values (black line) in the DU(n, n_0) model. The crossing point t_ζ (say), which specifies the FDR, is typically greater than the smallest critical value $\alpha_{n:n-n_0+1}$ (denoted by α_1 in the figure).

3.3 Alternative FDR curves and exact solving

In this section we investigate the question whether there exist further functions $g : [0,1] \to [0,\alpha]$, $\alpha \in (0,1)$, such that the FDR of an SU test procedure φ^n under a DU(n, n_0) configuration fulfils the following equalities

$$\mathrm{FDR}_{n,n_0}(\varphi^n) = g(n_0/n) \text{ for all } n_0 \in I_n \tag{3.22}$$

for a fixed $n \in \mathbb{N}$ or probably for all $n \in \mathbb{N}$. We call any function g an **FDR bounding curve** if it satisfies the natural restrictions $g(0) = 0$ and $0 < g(\zeta) \leq \min\{\zeta, \alpha\}$ for all $\zeta \in (0,1]$ and some $\alpha \in (0,1)$. As noted in Section 3, $g(\zeta) = \alpha\zeta$ leads to the LSU procedure while $g^*(\zeta) = \min\{\alpha, \zeta\}$ does not work for the most $n \in \mathbb{N}$. At present, $g(\zeta) = \alpha\zeta$ is the only known type of an FDR bounding function which solves (3.22).

For SUD(λ) tests (3.22) may be replaced by $b(n, n_0|\lambda) = g(n_0/n)$ for all $n_0 \in I_n$. Among others, we investigate conditions such that

$$\lim_{n\to\infty} \mathrm{FDR}_{n,n_0}(\varphi^n) = g(\zeta) \text{ or } \lim_{n\to\infty} \mathrm{FDR}_{n,n_0}(\varphi^\lambda) = \lim_{n\to\infty} b(n, n_0|\lambda) = g(\zeta)$$

holds for all ζ if $n_0/n \to \zeta$.

Similarly as in Finner *et al.* [2009], we can try to find the asymptotic rejection curve r and the asymptotic critical value curve ρ associated with an FDR bounding curve g. Since ρ should satisfy (A1), this imposes further conditions on g as will be seen below. Assume for a moment that $\lim_{n\to\infty} n_0/n = \zeta \in (0,1)$. Then, for a fixed threshold t, the asymptotic FDR with respect to DU configurations is given by

$$\mathrm{FDR}_\zeta(t) = \frac{t\zeta}{(1-\zeta) + t\zeta}. \tag{3.23}$$

Solving $\mathrm{FDR}_\zeta(t) = g(\zeta)$ for t leads to

$$t_\zeta = \frac{g(\zeta)(1-\zeta)}{\zeta(1-g(\zeta))}. \tag{3.24}$$

Note that the threshold for the p-values is determined by the asymptotic crossing point between the rejection curve r and the asymptotic ecdf $F_\infty(t|\zeta) = \zeta t + (1-\zeta)$ of p-values with respect to DU configurations, cf. Chapter 1. This results in an implicit definition of the asymptotic rejection curve r given by $r(t_\zeta) = F_\infty(t_\zeta|\zeta)$, or equivalently,

$$r\left(\frac{g(\zeta)(1-\zeta)}{\zeta(1-g(\zeta))}\right) = \frac{1-\zeta}{1-g(\zeta)},\ \zeta \in (0,1). \tag{3.25}$$

Analogously, the asymptotic critical value function $\rho \equiv \rho(\cdot|\eta) = r^{-1}$ is implicitly defined by

$$\rho\left(\frac{1-\zeta}{1-g(\zeta)}\right) = \frac{g(\zeta)(1-\zeta)}{\zeta(1-g(\zeta))},\ \zeta \in (0,1). \tag{3.26}$$

The following lemma shows that r and ρ are well defined for suitable FDR bounding curves g.

Lemma 3.12

Let $g : [0,1] \to [0,\alpha]$, $\alpha \in (0,1)$, be a continuous FDR bounding curve such that $g(\zeta)/\zeta$ is non-increasing in $\zeta \in (0,1]$ and $b = \lim_{\zeta\to 0} g(\zeta)/\zeta \in (0,1]$. Then $r : [0,b] \to [0,1]$ and $\rho : [0,1] \to [0,b]$ are well defined via (3.25) and (3.26), respectively, and by setting $r(0) = \rho(0) = 0$ and $r(b) = 1$, $\rho(1) = b$. Moreover, ρ fulfils condition (A1).

Proof: Let $\underline{\zeta} = \sup\{\zeta \in [0,1] : g(\zeta) = \zeta\}$. Then $g(\zeta) = \zeta$ for $\zeta \in [0,\underline{\zeta}]$ and $g(\zeta) < \zeta$ for $\zeta \in (\underline{\zeta},1]$. Moreover, if there exists a $\zeta \in (0,\underline{\zeta})$, then $b = 1$ and (3.25) yields $r(1) = 1$ and (3.26) yields $\rho(1) = 1$. Setting $g_1(\zeta) = (1-\zeta)/(1-g(\zeta))$, $\zeta \in [0,1]$, $g_2(\zeta) = g(\zeta)/\zeta$, $\zeta \in (\underline{\zeta},1]$ and $g_2(\zeta) = b$ for $\zeta \in [0,\underline{\zeta}]$, (3.26) can be written as

$$\frac{\rho(g_1(\zeta))}{g_1(\zeta)} = g_2(\zeta).$$

Since g_2 is non-increasing and g_1 is strictly decreasing on $[\underline{\zeta},1]$, we obtain that $r : [0,b] \to [0,1]$ and $\rho : [0,1] \to [0,b]$ are well defined and ρ fulfils condition (A1). From $g_1(0) = 1$, $g_1(1) = 0$ and $g_2(0) = b$ we obtain the remainder. ■

We note that if ζ_i denotes the solution of $(1-\zeta)/(1-g(\zeta)) = i/n$ with respect to ζ, the asymptotic critical values can be computed by

$$\alpha_{i:n} = \rho(i/n) = \begin{cases} \dfrac{g(\zeta_i)(1-\zeta_i)}{\zeta_i(1-g(\zeta_i))} & ,\ i \in I_{n-1}, \\[2ex] b & ,\ i = n. \end{cases}$$

Typically, for a given bounding function g we can determine ζ_i-values for $i \in I_{n-1}$ (and hence, critical values) only numerically. But in the next example we give a bounding function g for which the corresponding critical value function ρ can be outlined analytically.

Example 3.13

The FDR bounding function

$$g(\zeta) = \frac{\alpha\zeta}{\zeta + \alpha(1-\zeta)}$$

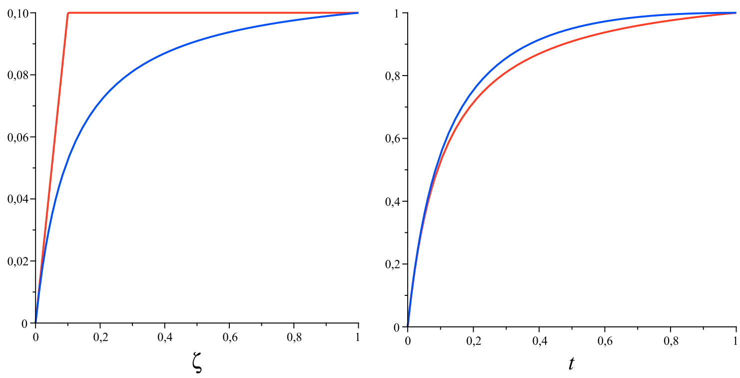

Figure 3.3: FDR bounding functions g^* (upper curve) and g (lower curve) given in Example 3.13 with $\alpha = 0.1$ (picture on the left) and the corresponding rejection curves f_α (lower curve) and r (upper curve) (picture on the right).

leads to

$$t_\zeta = \frac{\alpha(1-\zeta)}{\alpha+\zeta-2\alpha\zeta} \quad \text{and} \quad \zeta(t) = \frac{\alpha(1-t)}{t+\alpha-2\alpha t}.$$

Then the rejection curve related to g is given by

$$r(t) = \frac{t(1-t\alpha)}{t+\alpha-2\alpha t},\ t \in [0,1]$$

and the corresponding critical value function is given by

$$\rho(t) = \frac{2t\alpha - t + 1 - \sqrt{4t^2\alpha^2 - 4t^2\alpha + 4t\alpha + t^2 - 2t + 1 - 4\alpha^2 t}}{2\alpha}.$$

Figure 3.13 shows the FDR bounding functions g and g^* with $\alpha = 0.1$ on the left as well as the rejection curve r and the AORC f_α on the right-hand side of this figure.

The next theorem shows that in DU models the asymptotic FDR of an SUD test based on the rejection curve defined in (3.25) equals the given FDR bounding curve.

Theorem 3.14
Let g be an FDR bounding curve with the same properties as in Lemma 3.12. Consider SUD(λ_n) tests φ_n based on r defined in (3.25) *with $\lambda_n/n \to \kappa$. Then we obtain for the limiting FDR in DU(n, n_0) models with $n_0/n \to \zeta$ that*

$$\lim_{n\to\infty} FDR_{n,n_0} = g(\zeta)$$

for (i) $\kappa \in (0,1]$ and $\zeta \in [0,1]$ if $b < 1$, (ii) $\kappa \in (0,1)$ and $\zeta \in [0,1]$ if $b = 1$ and (iii) $\kappa = 0$ and $\zeta \in [0,1)$.

Proof: Let g_1 and g_2 be defined as in the proof of Lemma 3.12. Setting $t_\zeta = g_1(\zeta)g_2(\zeta)$ we obtain that t_ζ as a function of ζ is continuous for $\zeta \in [0,1]$ and strictly decreasing for $\zeta \in [\underline{\zeta}, 1]$ with

$t_\zeta = b$ for $\zeta \leq \underline{\zeta}$ and $t_\zeta = 0$ for $\zeta = 1$. For each $\zeta \in [0,1]$ it will be shown that $r(t) = F_\infty(t|\zeta)$ has at least one solution and at most two solutions in $[0,b]$. Note, that from (3.25) we obtain $r(t_\zeta) = F_\infty(t_\zeta|\zeta)$, which implies that there exists at least one solution, namely t_ζ. Now suppose there exists a further solution $t' \neq t_\zeta$. The strict monotonicity of t_ζ in $\zeta \in [\underline{\zeta}, 1]$ yields that there exists a $\zeta' \in [\underline{\zeta}, 1]$ such that $t' = t_{\zeta'}$. Altogether we get $r(t_{\zeta'}) = F_\infty(t_{\zeta'}|\zeta') = F_\infty(t'|\zeta)$, hence $\zeta = \zeta'$ or $t' = 1$ which implies the existence of at most two solutions, namely $t_\zeta < 1$ and 1 or only t_ζ. Finally, we get $R_n/n \to F_\infty(t_\zeta|\zeta) = r(t_\zeta) = g_1(\zeta)$ and $\rho(R_n/n) \to \rho(r(t_\zeta)) = t_\zeta = g_1(\zeta)g_2(\zeta)$, hence $b(n,n_0) \to g(\zeta)$ for $\zeta \in [0,1]$ with formula (3.6). Lemma 3.7 completes the proof. ∎

In order to complete the picture concerning the relationship between asymptotic rejection curves, asymptotic critical value curves and asymptotic FDR bounding curves, we consider the case where we start with an asymptotic rejection curve r.

Remark 3.15
Let $r : [0,b] \to [0,1]$ be continuous with $b \in (0,1]$ and $r(b) = 1$ and suppose there exists a $\zeta_0 \in [0,1)$ such that for each $\zeta \in (\zeta_0, 1]$ there exists a unique crossing point $t(\zeta)$ between $F_\infty(\cdot|\zeta)$ and r on $[0,b]$ if $b < 1$ or on $[0,1)$ if $b = 1$ while the unique crossing point $t(\zeta)$ on $[0,1]$ is b for $\zeta \in [0,\zeta_0]$. Moreover, suppose that $r(t)/t$ is non-increasing in $t \in (0,b]$. Consider a sequence of DU(n,n_0) models and a sequence of SUD(λ_n) tests based on r such that $R_n/n \to r(t(\zeta))$ as $n_0(n)/n \to \zeta$ for all $\zeta \in [0,1]$. Then the asymptotic FDR bounding curve on $[0,1)$ is given by

$$g(\zeta) = \frac{\zeta t(\zeta)}{1 - \zeta + \zeta t(\zeta)}$$

and $g(\zeta)/\zeta$ is non-increasing in $\zeta \in (0,1)$ with $\lim_{\zeta \to 0} g(\zeta)/\zeta = b$. Moreover, with $\rho = r^{-1}$ and $\rho(1 - \zeta + \zeta t(\zeta)) = t(\zeta)$ we get

$$\lim_{\zeta \to 1} g(\zeta) = \lim_{\zeta \to 1} \frac{\zeta t(\zeta)}{1 - \zeta + \zeta t(\zeta)} = \lim_{\zeta \to 1} \zeta \frac{\rho(1 - \zeta + \zeta t(\zeta))}{1 - \zeta + \zeta t(\zeta)} = \lim_{t \to 0} \frac{\rho(t)}{t} = q(0),$$

which is in line with the asymptotic results in Finner *et al.* [2009] for SUD procedures, where it is shown that under suitable assumptions the asymptotic FDR for $n \to \infty$ and $\zeta \to 1$ (or $\zeta = 1$) is $q(0)$.

Example 3.16
A class of FDR bounding functions g for which the system of equations given by the recursive scheme (3.18) and (3.19) can be solved at least for a broad range of n-values is given as follows. These functions depend on two further parameters γ, η with $1 \leq \eta \leq \gamma/\alpha$, $\alpha \leq \gamma \leq 1$, and are defined by

$$g(\zeta|\gamma,\eta) = \begin{cases} \alpha(1 - (1 - \zeta/\gamma)^\eta), & , 0 \leq \zeta < \gamma, \\ \alpha, & , \gamma \leq \zeta \leq 1. \end{cases}$$

We first note that $g(\zeta|1,1) = \alpha\zeta$, $g(\zeta|\alpha,1) = g^*(\zeta)$ and $g(\zeta|\gamma,\eta) \leq g^*(\zeta)$ for all $\zeta \in [0,1]$. Moreover, $g(\cdot|\gamma,\gamma/\alpha)$ and g^* have the same slope in $\zeta = 0$, $g(\zeta|\gamma,\eta)$ is non-decreasing in η and

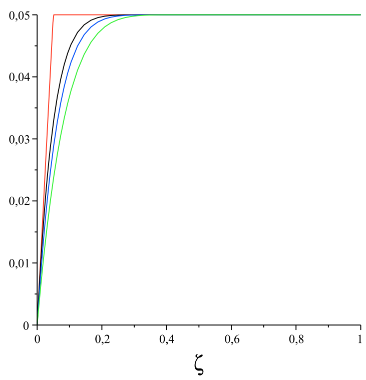

Figure 3.4: FDR bounding curves $g^*(\zeta)$ and $g(\zeta|\gamma,\eta)$ in Example 3.16 with $\gamma = 0.5$, $\eta = 6, 8, 10$ (from bottom to top in $\zeta = 0.1$) and $\alpha = 0.05$.

in γ for $\zeta \in [0,1]$. Note that $g(\cdot|\gamma,\eta)$ gets closer to g^* if η increases and/or γ decreases. Figure 3.4 displays the situation for $\alpha = 0.05$ and $\gamma = 1/2$. In this example, for $\eta = 6, 8, 10$, $g(\zeta|\gamma,\eta)$ and $g^*(\zeta)$ are equal for $\zeta \in [0.5, 1]$ and nearly coincide for $\zeta \in [0.3, 0.5)$. Whether (3.22) can be solved heavily depends on α and the choice of γ and η. It seems that a smaller α increases the chance to solve (3.22) for larger values of n. For example, for $\alpha = 0.01$ and $\gamma = 0.1$ there always exists an η at least for $n \leq 500$ such that (3.22) is solvable, whereas for $\alpha = 0.05$, $n = 7$ and $\alpha = 0.1$, $n = 4$ we could not find any solution. For $\gamma = 0.5$ we can find suitable η's for $\alpha = 0.01, 0.05$ and $n \leq 500$ (probably also for much larger n-values), as well as for $\alpha = 0.1$ and $n \leq 341$, but not for $n = 342$. Moreover, for $\gamma = 1$, $\alpha = 0.01, 0.05, 0.1$ we can find suitable η's at least for $n \leq 500$.

In the case $\alpha = 0.1$ we fail to find feasible critical values for larger n. The reason for this is that the parameter η is bounded by γ/α which decreases if α increases. This results in a worse approximation of g^* for smaller values of ζ. Thereby, we observed that for arbitrary but fixed α and γ a suitable parameter η, i.e. an η such that the recursive scheme (3.18) and (3.19) can be solved, increases if n increases. It seems that the larger the value of n, the better g^* has to be approximated by an FDR bounding curve.

An idea how g^* can be approximated in a smooth way is as follows. For a given function $G : [0,1] \to [0,\alpha]$ we can apply a linear transformation, such that the corresponding transformed function $g : [0,1] \to [0,\alpha]$ fulfils the condition $g(\zeta) \leq \zeta$. For example, Figure 3.5 shows the function $G(\zeta) = \alpha(1 - e^{\zeta\eta})$ and the transformed function g that lies below g^*. Thereby, this considered linear transformation maps the vector $(1,0)$ to itself and the vector $(0,1)$ to the vector $(1,1)$.

Now we give a formal definition of a general class of functions g which allow to approximate g^* in a smooth way.

Let $E = [\underline{\eta}, \infty)$ or $E = (\underline{\eta}, \infty)$ for some $\underline{\eta} \in \mathbb{R}$ and let $G_\eta : [0,1] \to [0,\alpha]$, $\eta \in E$, be continuous and non-decreasing functions such that $G_\eta(x)/x$ is non-increasing in $x \in [0,1]$ with

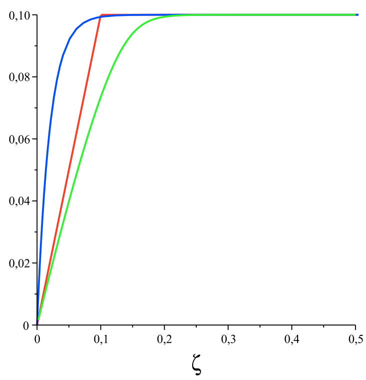

Figure 3.5: FDR bounding curve g (the lowest curve in $\zeta = 0.1$) obtained by the linear transformation of the function $G(\zeta) = \alpha(1 - e^{\zeta\eta})$ (the highest curve in $\zeta = 0.4$) with $\alpha = 0.1$ and $\eta = 50$.

$\lim_{x \downarrow 0} G_\eta(x)/x = b_\eta \in (0, \infty)$, $G_\eta(0) = 0$ for all $\eta \in E$ and $\lim_{\eta \to \infty} G_\eta(x) = \alpha$ for all $x \in (0, 1]$. Moreover, G_η is assumed to satisfy either
(G1) $\exists\, \gamma \in (0, 1 - \alpha)$ such that $G_\eta(\gamma) = \alpha$ for all $\eta \in E$ and $G_\eta(x)$ is strictly increasing in $\eta \in E$ for all $x \in (0, \gamma)$;
or
(G2) $G_\eta(x)$ is strictly increasing in $\eta \in E$ for all $x \in (0, 1]$.
In case of (G2) we formally set $\gamma = 1$. We denote the set of all these $(G_\eta)_{\eta \in E}$ by $\mathcal{G}$. Now define h_η by

$$h_\eta(x) = x + G_\eta(x)$$

and $g(\cdot|\eta) : [0, 1] \to [0, \alpha]$ by

$$g(\zeta|\eta) = G_\eta(h_\eta^{-1}(\zeta)),\ \zeta \in [0, 1]. \tag{3.27}$$

A little analysis yields that

$$\begin{aligned} g(\zeta|\eta) &\leq g^*(\zeta)\ \ \forall\, \eta \in E \ \text{ and } \ \forall\, \zeta \in [0, 1], \\ g(\zeta|\eta) &< g^*(\zeta)\ \ \forall\, \eta \in E \ \text{ and } \ \forall\, \zeta \in (0, \min\{\gamma + \alpha, 1\}), \\ \lim_{\eta \to \infty} g(\zeta|\eta) &= g^*(\zeta)\ \ \forall\, \zeta \in [0, 1], \\ \lim_{\zeta \to 0} g(\zeta|\eta)/\zeta &= b_\eta/(1 + b_\eta)\ \ \forall\, \eta \in E. \end{aligned}$$

If (G1) applies, we obtain $g(\zeta|\eta) = \alpha$ for $\zeta \in [\alpha + \gamma, 1]$.

Lemma 3.17
Let $(G_\eta)_{\eta \in E} \in \mathcal{G}$ and let $g(\cdot|\eta)$ be defined by (3.27). Then the asymptotic rejection curve $r \equiv r(\cdot|\eta)$ defined via (3.25) is strictly increasing on $[0, b_\eta/(1 + b_\eta)]$ with

$$\lim_{\eta \to \infty} r(t|\eta) = f_\alpha(t)\ \forall t \in [0, 1].$$

If (G1) applies, i.e. $\gamma + \alpha < 1$, then

$$r(t|\eta) = f_\alpha(t) \; \forall t \in [0, t_\gamma],$$

where $t_\gamma = \{\alpha(1-\alpha-\gamma)\}/\{(1-\alpha)(\gamma+\alpha)\}$. The asymptotic critical value function $\rho \equiv \rho(\cdot|\eta)$ defined via (3.26) satisfies the monotonicity condition (A1).

Proof: For $\min\{\gamma+\alpha, 1\} < \zeta \leq 1$, the asymptotic rejection curve r implicitly defined by (3.25) coincides with the AORC which has all desired properties. Therefore, it suffices to show the assertions of the lemma for $0 \leq \zeta \leq \min\{\gamma+\alpha, 1\}$. In view of Lemma 3.12 we have to show that $g(\zeta|\eta)/\zeta$ is continuous (which is trivial) and non-increasing in ζ. Substituting $\zeta = h_\eta(y)$ in $g(\zeta|\eta)/\zeta = G_\eta(h_\eta^{-1}(\zeta))/\zeta$, we see that $g(\zeta|\eta)/\zeta$ is non-increasing if

$$\frac{G_\eta(y)/y}{G_\eta(y)/y + 1}$$

is non-increasing which is implied by the assumptions. ■

Clearly, there are uncountable choices of G_η to approach g^* in a smooth way. For example, we can choose $G_\eta = \alpha H_\eta I_{[0,1]}$ for a suitable family of cdfs H_η on $[0, \infty)$ such that G_η has the desired properties, see the following example.

Example 3.18 (Families of probability distributions for generating FDR bounding curves)
Let $\alpha \in (0, 1)$.
(a) (Beta distributions.) Let $E = [1, \infty)$ and consider the family of beta distributions with cdfs $H_\eta(u) = (1-(1-u)^\eta)I_{[0,1]}(u) + I_{(1,\infty)}(u)$ for $\eta \in E$. Setting $G_\eta = \alpha H_\eta$ and $x = u\gamma$ for some $\gamma \in (0, 1-\alpha]$ this leads to (compare with Example 3.16)

$$G_\eta(x) = \alpha(1-(1-x/\gamma)^\eta)I_{[0,\gamma)}(x) + \alpha I_{[\gamma,1]}(x), \;\; \eta \in E.$$

Then $(G_\eta)_{\eta\in E} \in \mathcal{G}$, hence Lemma 3.17 applies. For convenience, we denote the resulting FDR bounding curves by $g(\cdot|\eta,\gamma)$. Note that $g(\cdot|\eta,\gamma)$ is non-increasing in $\gamma \in (0, 1-\alpha]$ for $\zeta \in [0, 1]$. Moreover, $g(\zeta|1, 1-\alpha) = \alpha\zeta$ which is the FDR bounding curve of the LSU procedure.
(b) (Exponential distributions.) Let $E = (0, \infty)$ and consider the family of exponential distributions with parameter $\eta \in E$ and cdf H_η (say) and define again $G_\eta = \alpha H_\eta$. Then we have

$$G_\eta(x) = \alpha(1-\exp(-\eta x))I_{[0,1]}(x), \;\; \eta \in E,$$

and $(G_\eta)_{\eta\in E} \in \mathcal{G}$ with $\gamma = 1$, hence Lemma 3.17 applies again. The resulting FDR bounding curves are denoted by $g(\cdot|\eta)$.

It seems that one can choose FDR bounding curves of the type introduced in Example 3.18 being closer to g^* and allowing for exact solving of (3.18) and (3.19) for larger values of n than the ones in Example 3.16. For suitable choices of η and γ in (a) and (b) in Example 3.18 we obtain approximately identical FDR curves and critical value functions (rejection curves). Moreover, for $\alpha = 0.01, 0.05$ and $\gamma = 0.5$ in Example 3.18(a), we can find suitable ηs for $n \leq 500$ (and probably for much larger values of n) such that (3.22) is solvable for both examples. For instance, if $\alpha = 0.05$, then for $\eta = 16$, $\gamma = 0.5$ in Example 3.18(a) and $\eta = 35$ in Example 3.18(b) there are feasible critical values with (3.22) for at least $n \leq 500$. As noted before, the case of larger α-values is problematic. At least for $\alpha = 0.1$, (3.22) can be solved for both examples for larger values of n than in Example 3.16, i.e., for at least $n \leq 700$ we find an η such that (3.22) is solvable. All in all this approach (as long as it works) yields an attractive possibility to obtain a feasible set of critical values which should not differ too much from the AORC based critical values (3.1). Anyhow, it remains completely unclear whether for each n there exists an η such that (3.22) can be solved.

Of course, for SUD procedures it is also possible to apply the recursive scheme (3.18) and (3.19) such that the upper bound is equal to one of the FDR bounding curves considered in Examples 3.16 and 3.18. But, as mentioned before, computations for SUD tests can take a long time.

3.4 AORC adjustments

In this section we present different adjustment methods related to the AORC or to a modified AORC, such that the FDR is controlled for a finite number of hypotheses. We consider single-parameter and multiple-parameter adjustment methods. In the case of single-parameter adjustments we investigate the behaviour of the adjusting parameter β_n for various SUD test procedures. We show that exact solving (i.e. the most FDR-values should be α) seems to be possible only if the number of all hypotheses n is very small. On the other hand, β-adjustment methods yield a good approximation of the α level even for n-values being not too large. Moreover, it is mostly easy to implement critical values corresponding to a single-parameter adjustment approach. Thereby, critical values corresponding to theses tests depend on the number of all hypotheses, the pre-specified parameter α and an adjusting parameter β_n so that one has only to determine the corresponding adjusting parameter β_n. Since for a large number of all hypotheses computation complexity increases rapidly, AORC adjustments yield a good alternative for other multiple test procedures.

3.4.1 Single-parameter adjustment

One way to get a feasible set of critical values for an SUD(λ) procedure controlling the FDR is to adjust the AORC. For example, as already mentioned in Finner *et al.* [2009], we can try to find a suitable $\beta_n > 0$ such that the adjusted rejection curve

$$f_{\alpha,\beta_n}(t) = \left(1 + \frac{\beta_n}{n}\right) f_\alpha(t), \quad t \in \left[0, \frac{\alpha}{\alpha + \beta_n/n}\right],$$

with corresponding, always feasible critical values

$$\alpha_{i:n} = \frac{\frac{i}{n+\beta_n}\alpha}{1 - \frac{i}{n+\beta_n}(1-\alpha)} = \frac{i\alpha}{n + \beta_n - i(1-\alpha)}, \; i \in I_n, \tag{3.28}$$

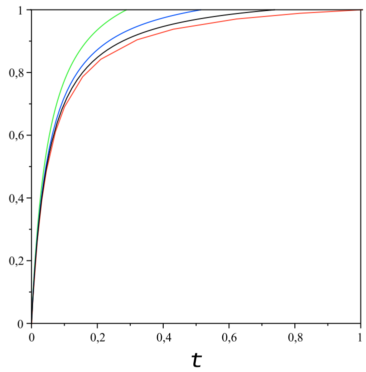

Figure 3.6: Rejection curves $f_{\alpha,\beta_n}(t)$ for SU tests with $n = 10, 30, 100$ and $f_\alpha(t)$ (from top to bottom in $t = 0.2$) with $\alpha = 0.05$, $\beta_{10} = 1.23$, $\beta_{30} = 1.41$ and $\beta_{100} = 1.76$.

yields FDR control by an SUD(λ) test at level α. Below, we say the parameter β_n is **optimal** if β_n is the minimum value, which yields FDR control. For example, for $\alpha = 0.05$ and $n = 100, 1000$ we obtain that $\beta_{100} = 1.76$, $\beta_{1000} = 3.07$ for an SU test and $\beta_{100} = 1.54$, $\beta_{1000} = 1.82$ for an SUD(λ_n) test with $\lambda_n = \lceil n/(1+\alpha)\rceil$ yielding strict FDR control. Note that this choice of λ_n yields $\alpha_{\lambda_n:n} \to 1/2$ for $n \to \infty$, because $f_{\alpha,\beta_n} \to f_\alpha$ for $n \to \infty$ (it will be proved later) and $\alpha_{\lambda_n:n} \approx f_\alpha^{-1}(\lambda_n/n) = \kappa$ for some $\kappa \in (0,1)$ (for example $\kappa = 1/2$) yield $\lambda_n = n\kappa/(\alpha + \kappa(1-\alpha))$.

For $\alpha = 0.05$, Figure 3.6 depicts the modified curves f_{α,β_n} for SU procedures for $n = 10, 30, 100$ together with f_α, where $\beta_{10} = 1.23$, $\beta_{30} = 1.41$, $\beta_{100} = 1.76$.

It follows from the monotonicity of the upper bounds $b(n, n_0|\lambda)$ in λ stated in Theorem 3.10 that for a fixed $n \in \mathbb{N}$ the value of the parameter β_n needed to ensure strict FDR control, increases with increasing parameter λ of an SUD procedure; i.e. larger values for λ lead to larger β_n-values. But for fixed critical values an SUD(λ_1) test rejects at least as many hypotheses as an SUD(λ_2) test with the same critical values if λ_1 is larger than λ_2. Lemma 3.11 shows that critical values ensuring FDR control for an SD test procedure yield FDR control for an SUD(λ) test for some smaller λs if the corresponding SU test controls the FDR for larger n_0-values.

We apply this result for β_n-adjusted critical values (3.28). Although an SU test with critical values (3.28) and β_n optimal for the corresponding SD test does not control the FDR for certain values of n_0, we observed in all our calculations that the pre-chosen α-level is exceeded only for a certain set of small n_0-values, that is, for each $n \in \mathbb{N}$ there seems to exists an $n_0^* \leq n$ defined by (3.15) such that Lemma 3.11 applies.

For example, for $\alpha = 0.05$ and $n = 100, 500, 1000, 2000$ the smallest β_n-values such that the SD test with (3.28) controls the FDR are given by $\beta_n = 1.34, 1.47, 1.53, 1.58$. Due to Lemma 3.11 this results in $n_0^* = 29, 134, 271, 565$. Hence, an SUD($\lambda_n$) test with appropriately chosen β_n and $\lambda_n \leq 72, 367, 730, 1436$ (or respectively $\lambda_n/n \leq 0.72, 0.734, 0.73, 0.7185$) controls the FDR. Figure 3.7 shows that an SUD(λ_n) test at level $\alpha = 0.1, 0.05, 0.01$ with $\lambda_n \approx 0.4n, 0.7n, 0.9n$

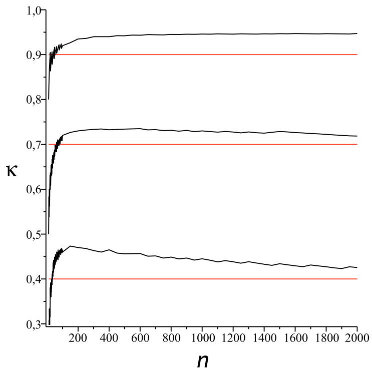

Figure 3.7: Maximal values of κ such that an SUD(λ_n) test with $\lambda_n/n \leq \kappa$ and optimal β_n with respect to an SD test controls the FDR at level $\alpha = 0.01, 0.05, 0.1$ (from top to bottom).

and β_n optimal for the corresponding SD test controls the FDR for larger n-values. Moreover, our simulation study for $\alpha = 0.05$ and larger n-values shows that the upper bound of an SUD(λ_n) test with $\lambda_n = 0.7n$ and $\beta_n \equiv 1$ exceeds the α-level only slightly and it seems that the upper bound decreases to α, i.e. for $n = 5000, 10000, 50000$ the maximum upper bound is about $0.05022, 0.05020, 0.05008$.

3.4.2 Adjustment of the modified AORC

Realised FDR values for SU tests based on (3.28) in DU(n, n_0) models with varying numbers n_0 of true null hypotheses have a maximum peak which is taken typically for smaller n_0-values (cf. Figure 3.9). It seems that if we diminish larger critical values (which correspond to FDRs with smaller n_0-values), then we can enlarge smaller ones, such that the corresponding FDR curve is flatter for most n_0-values. For some fixed $k \in I_n$, we therefore replace the larger critical values $\alpha_{i:n}$, $i \geq k$, in (3.28) such that they just fulfil the monotonicity condition (A1) with equality. This corresponds to the adjustment SU procedures proposed in Example 3.2 in Finner *et al.* [2009]. For example, we can choose k appropriately equal to $n(1-\alpha)$ or $n(1-2\alpha)$. Then, we search for a suitable constant $\beta_n^* > 0$ such that the critical values

$$\alpha_{i:n} = \begin{cases} \frac{i\alpha}{n+\beta_n^*-i(1-\alpha)}, & 1 \leq i \leq k-1, \\ i\alpha_{i-1:n}/(i-1), & k \leq i \leq n \end{cases} \tag{3.29}$$

yield FDR control at level α. The underlying rejection curve is given by

$$f_{\alpha,\beta_n^*}(t) = \begin{cases} f_{\alpha,\beta_n^*}(t), & 0 \leq t \leq t^*, \\ t\frac{f_{\alpha,\beta_n^*}(t^*)}{t^*}, & t^* < t \leq 1 \end{cases}$$

with $t^* = f^{-1}_{\alpha,\beta_n^*}((k-1)/n)$, cf. Figure 3.8. Note that $f_{\alpha,\beta_n^*}(t)$ for $t \in (t^*, 1]$ corresponds a Simes line with level $t^*/f_{\alpha,\beta_n^*}(t^*)$.

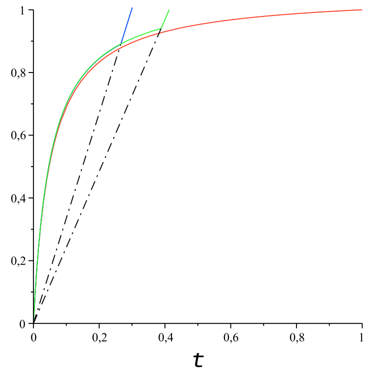

Figure 3.8: The AORC is red, $f_{\alpha,\beta_n^*}(t)$ for $n = 100$, $k = 90$ and $\beta_n^* = 1.3$ is blue and $f_{\alpha,\beta_n^*}(t)$ for $n = 100$, $k = 95$ and $\beta_n^* = 1.41$ is green.

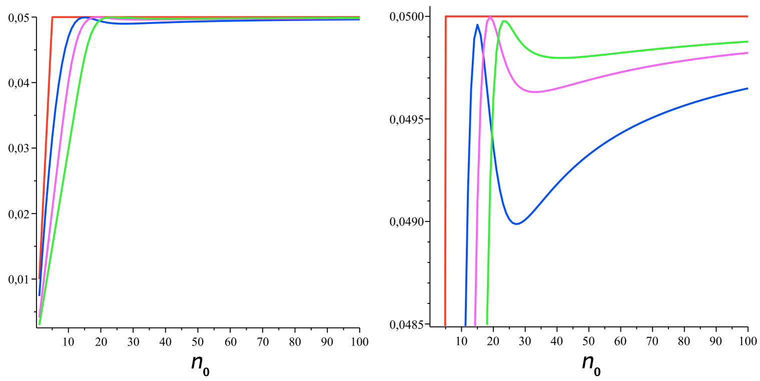

Figure 3.9: FDR-values for SU tests with $\alpha = 0.05$ and $n = 100$ in DU models based on simultaneously β_n-adjusted critical values and critical values (3.29) with $\beta_n^* = 1.41$, $k = 95$ and $\beta_n^* = 1.3$, $k = 90$ (left graph: from top to bottom in $n_0 = 10$, right graph: from bottom to top in $n_0 = 50$), the right graph is zoomed.

The critical values (3.29) are always feasible and the optimal β_n^* (i.e. the minimum valid β_n^*) is smaller than the corresponding β_n of the simultaneous adjustment method with critical values (3.28). For example, for $\alpha = 0.05$, $n = 100, 1000$ and $k = 90, 900$ we obtain $\beta_n^* = 1.3, 1.42$ for an SU test. This results in flatter FDR curves, which are closer to α for larger n_0-values than the corresponding FDR curves of the simultaneous β_n-adjustment. For $n = 100$ and $\alpha = 0.05$, Figure 3.9 shows realised FDR values for the simultaneous β_n-adjustment method with $\beta_{100} = 1.76$ and for β_n^*-adjustment methods with $k = 95$, $\beta_n^* = 1.41$, and, $k = 90$, $\beta_n^* = 1.3$.

3.4.3 Behaviour of the adjusting parameters

Now we consider the behaviour of the parameters β_n and β_n^* for different SUD(λ) test procedures. Note that for an SUD(λ_n)-test the smallest parameter β_n yielding strict FDR control is

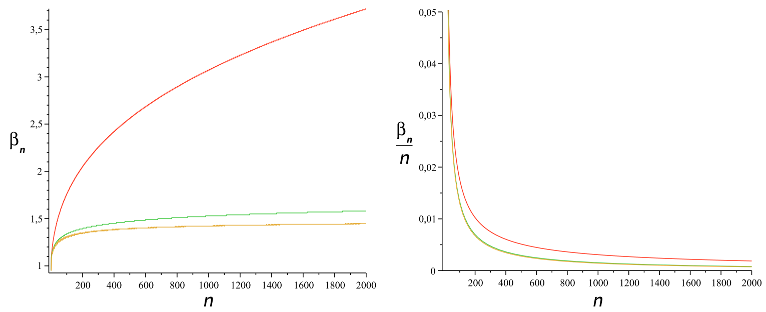

Figure 3.10: Optimal parameters β_n resp. β_n^* (left graph) and β_n/n resp. β_n^*/n (right graph) for SU- and SD-tests based on f_{α,β_n} and SU test based on (3.29). The curves can be distinguished by noticing that for all n-values, the optimal β_n for an SU test is larger than for an SD test and the parameter β_n for SD is in turn larger than β_n^* for SU. In the right graph, the curves corresponding to SD with β_n and to SU with β_n^* are nearly identical.

non-decreasing in λ_n, such that the marginal cases (i.e. SU and SD tests) are a matter of particular interest. For $1 \leq n \leq 2000$, the left graph in Figure 3.10 shows the minimum values for β_n (or β_n^*, respectively) which have to be used to ensure strict FDR control for SU and SD procedures based on critical values in (3.28) and SU tests with critical values (3.29) and $k = \lceil n(1-2\alpha) \rceil$ for $\alpha = 0.05$ and varying n. In the right graph of Figure 3.10, the corresponding factors β_n/n (or β_n^*/n, respectively) are displayed. Thereby, critical values (3.28) for SD-tests and critical values (3.29) for SU-tests are nearly identical (lower curves).

A complete characterisation of the asymptotic behaviour of the parameter β_n resp. β_n^* remains an interesting open question for the considered adjustment procedures. It is not entirely clear for SUD tests whether β_n and β_n^* are bounded or diverge for $n \to \infty$. Note that $\beta_n^* \leq \beta_n$ for each SUD procedures with a fixed λ_n.

Remark 3.19

In Benjamini *et al.* [2005] (Remark to Definition 7), an SD procedure with the universal adjustment constant $\beta_n \equiv 1$ was proposed. FDR control in case of independent p-values for this SD procedure was proven in Gavrilov *et al.* [2009] making use of special structural properties of SD tests. More precisely, $\beta_n \geq 1$ yields control of the upper bound (3.13) which implies FDR control. We note that the bound $b(n, n_0|1)$ given in (3.7) for the SD procedure with AORC-based adjusted critical values for $\beta_n \equiv 1$ can exceed α. For example, setting $\alpha = 0.05$, $n = 50$ and $n_0 = 10$ leads to $b(n, n_0|1) = 0.052$.

The next lemma yields a similar result as in Gavrilov *et al.* [2009] concerning the behaviour of the adjustment parameter β_n yielding control of the upper bound (3.7) for an SD test.

Lemma 3.20

For an SD test with critical values (3.28) and $\beta_n \geq 2$ the upper bound (3.7) is not greater than the *pre-specified level α.*

Proof: Let $p_1, \ldots, p_i$ be p-values that are iid uniformly distributed on $[0,1]$ and let $p_{1:i}, \ldots, p_{i:i}$ be the corresponding order statistics. For simplicity we write $\alpha_j \equiv \alpha_{j:n}$, $j \in I_n$. Then the upper bound (3.7) for an SD test can be rewritten as

$$\begin{aligned} b(n, n_0|1) &= n_0 \sum_{j=1}^{n_0-1} \frac{\alpha_{n_1+j}}{n_1+j} \mathbb{P}(p_{1:n_0-1} \le \alpha_{n_1+2}, \ldots, p_{j-1:n_0-1} \le \alpha_{n_1+j}, p_{j:n_0-1} > \alpha_{n_1+j+1}) \\ &\quad + n_0 \frac{\alpha_n}{n} \mathbb{P}(p_{1:n_0-1} \le \alpha_{n_1+2}, \ldots, p_{n_0-1:n_0-1} \le \alpha_n). \end{aligned}$$

Noting that

$$\frac{\alpha_{n_1+j}}{n_1+j} = \alpha \frac{1-\alpha_{n_1+j}}{n_0 - j + \beta_n} \quad \text{for } j \in I_{n_0},$$

the monotonicity property (A1) yields

$$\frac{\alpha_{n_1+j}}{n_1+j} \le \alpha \frac{1-\alpha_{n_1+j+1}}{n_0 - j - 1 + \beta_n} \quad \text{for } j \in I_{n_0-1}.$$

Therefore, for $\beta_n \ge 2$ we obtain

$$\frac{\alpha_{n_1+j}}{n_1+j} \le \alpha \frac{1-\alpha_{n_1+j+1}}{n_0 - j + 1} \quad \text{for } j \in I_{n_0-1}.$$

Furthermore,

$$\frac{\alpha_n}{n} = \alpha \frac{1-\alpha_n}{\beta_n} < \alpha \frac{1-\alpha_n}{1}.$$

Hence, we get

$$\begin{aligned} b(n, n_0|1) &< \alpha n_0 \sum_{j=1}^{n_0-1} \frac{1-\alpha_{n_1+j+1}}{n_0-j+1} \mathbb{P}(p_{1:n_0-1} \le \alpha_{n_1+2}, \ldots, p_{j-1:n_0-1} \le \alpha_{n_1+j}, \\ &\quad p_{j:n_0-1} > \alpha_{n_1+j+1}) \\ &\quad + \alpha n_0 \frac{1-\alpha_n}{1} \mathbb{P}(p_{1:n_0-1} \le \alpha_{n_1+2}, \ldots, p_{n_0-1:n_0-1} \le \alpha_n) \\ &= \alpha \sum_{j=1}^{n_0-1} \frac{n_0}{n_0-j+1} \mathbb{P}(p_{1:n_0-1} \le \alpha_{n_1+2}, \ldots, p_{j-1:n_0-1} \le \alpha_{n_1+j}, \\ &\quad p_{j:n_0-1} > \alpha_{n_1+j+1}, p_{n_0} > \alpha_{n_1+j+1}) \\ &\quad + \alpha n_0 \mathbb{P}(p_{1:n_0-1} \le \alpha_{n_1+2}, \ldots, p_{n_0-1:n_0-1} \le \alpha_n, p_{n_0} > \alpha_n) \\ &= A \text{ (say).} \end{aligned}$$

Setting $\alpha_{n+1} = \alpha_n$, we obtain

$$A = \alpha \sum_{j=1}^{n_0} \frac{n_0}{n_0-j+1} \mathbb{P}(p_{1:n_0-1} \le \alpha_{n_1+2}, \ldots, p_{j-1:n_0-1} \le \alpha_{n_1+j}, p_{j:n_0-1} > \alpha_{n_1+j+1}, p_{n_0} > \alpha_{n_1+j+1}).$$

Obviously, for each $j \in I_{n_0}$ the event

$$\{p_{1:n_0-1} \leq \alpha_{n_1+2}, \ldots, p_{j-1:n_0-1} \leq \alpha_{n_1+j}, p_{j:n_0-1} > \alpha_{n_1+j+1}, p_{n_0} > \alpha_{n_1+j+1}\}$$

can be expressed as

$$\{p_{1:n_0} \leq \alpha_{n_1+2}, \ldots, p_{j-1:n_0} \leq \alpha_{n_1+j}, p_{j:n_0} > \alpha_{n_1+j+1}, p_{n_0} \geq p_{j:n_0}\}.$$

The latter can be expressed in terms of the number $\tilde{V}_{n_0}$ of false rejections which corresponds to an SD test with critical values $(\alpha_{n_1+2}, \ldots, \alpha_n, \alpha_{n+1})$, that is,

$$\{\tilde{V}_{n_0} = j-1, p_{n_0} \geq p_{j:n_0}\}.$$

Hence,

$$\begin{aligned}
A &= \alpha \sum_{j=1}^{n_0} \frac{n_0}{n_0-j+1} \mathbb{P}_{n_0,n_0}(\tilde{V}_{n_0} = j-1, p_{n_0} \geq p_{j:n_0}) \\
&= \alpha \sum_{j=1}^{n_0} \frac{1}{n_0-j+1} \sum_{i=1}^{n_0} \mathbb{P}_{n_0,n_0}(\tilde{V}_{n_0} = j-1, p_i \geq p_{j:n_0}) \\
&= \alpha \sum_{j=1}^{n_0} \frac{1}{n_0-j+1} \sum_{i \in I_{n_0}: p_i \geq p_{j:n_0}} \mathbb{P}_{n_0,n_0}(\tilde{V}_{n_0} = j-1) \\
&= \alpha \sum_{j=1}^{n_0} \mathbb{P}_{n_0,n_0}(\tilde{V}_{n_0} = j-1).
\end{aligned}$$

With this we obtain

$$b(n, n_0|1) < A = \alpha(1 - \mathbb{P}_{n_0,n_0}(\tilde{V}_{n_0} = n_0)) < \alpha,$$

which completes the proof.

■

In order to characterise the behaviour β_n for an SU test we first prove two lemmas. The next lemma yields some results for an SUD test with a unique crossing point between the limiting rejection curve and the asymptotic distribution function.

Lemma 3.21

Consider DU(n, n_0) models with $n_0/n \to \zeta$, $\zeta \in [0,1]$. Let φ_n, $n \in \mathbb{N}$, be SUD(λ_n) tests based on rejection curves $r_n : [0, b_n] \to [0,1]$ with $b_n \in (0,1)$, and suppose there exists a rejection curve $r : [0,b] \to [0,1]$ with $b \in (0,1)$ such that $\lim_{n\to\infty} r_n(t) = r(t)$ for $t \in [0,b]$ and $\lim_{n\to\infty} b_n = b$. Suppose there exists a unique crossing point $t^*_\zeta \in [0,b]$ between r and $F_\infty(t) = 1-\zeta+t\zeta$. Then

$$t_n \to t^*_\zeta \text{ and } R_n/n \to r(t^*_\zeta) \text{ for } n \to \infty \text{ almost surely },$$

where t_n is the crossing point between r_n and $\hat{F}_n$ determined by the SUD(λ_n) test.

Moreover, if $\rho_n = r_n^{-1}$, $n \in \mathbb{N}$, and $\rho = r^{-1}$ fulfil (A1), then the asymptotic upper bound for the FDR is given by

$$\lim_{n\to\infty} b(n, n_0) = \zeta q(r(t^*_\zeta)),$$

where $q(t) = \rho(t)/t$ for $t \in (0, 1]$ and $q(0) = \lim_{t\to 0} \rho(t)/t$.

Proof: For a fixed but arbitrary $\epsilon > 0$ define $A_\epsilon = \{t \in [0, 1] : |t - t^*_\zeta| > \epsilon\}$. In order to prove almost sure convergence of t_n we will show that the probability of $\{t_n \notin A_\epsilon\}$ increases to 1 for $n \to \infty$.

For a given $\epsilon > 0$ there always exists a $\delta_\epsilon > 0$ such that

$$\forall\, t \in [0, b] \cap A_\epsilon :\ |F_\infty(t) - r(t)| > \delta_\epsilon, \tag{3.30}$$

which follows from the assumption that r and F_∞ have only one crossing point in $[0, b]$. It is easy to show that assumptions of this lemma imply uniform convergence of r_n in $[0, b]$, that is,

$$\exists\, N_1 \in \mathbb{N} :\ \forall\, n \geq N_1 :\ \forall\, t \in [0, \min(b, b_n)] :\ |r_n(t) - r(t)| < \delta_\epsilon/3. \tag{3.31}$$

Then the triangle inequality applying to (3.30) and (3.31) yields

$$\forall\, n \geq N_1 :\ \forall\, t \in [0, \min(b, b_n)] \cap A_\epsilon :\ |F_\infty(t) - r_n(t)| > 2\delta_\epsilon/3. \tag{3.32}$$

Obviously, if $b_n \leq b$, then $t_n \in [0, \min(b, b_n)]$. If $b_n > b$ it holds for ϵ being small enough that

$$\exists\, N_2 \in \mathbb{N} :\ \forall\, n \geq N_2 :\ [b, b_n] \subseteq A_\epsilon \text{ if } b \neq t^*_\zeta \text{ and } [b, b_n] \subseteq A^c_\epsilon \text{ else.}$$

Therefore, (3.31) and monotonicity of r_n yield

$$\forall\, n \geq \max(N_1, N_2) :\ \forall\, t \in [b, b_n] \cap A_\epsilon :\ r_n(t) \geq r_n(b) \geq r(b) - \delta_\epsilon/3. \tag{3.33}$$

The convergence of $b_n \to b$ for $n \to \infty$ implies

$$\exists\, N_3 \in \mathbb{N} :\ \forall\, n \geq N_3 :\ |F_\infty(b_n) - F_\infty(b)| < \delta_\epsilon/3$$

and hence,

$$\forall\, n \geq N_3 :\ \forall\, t \in [b, b_n] :\ F_\infty(t) \leq F_\infty(b) + \delta_\epsilon/3. \tag{3.34}$$

Setting $N_4 = \max(N_1, N_2, N_3)$, (3.33), (3.34) and (3.30) imply

$$\forall\, n \geq N_4 :\ \forall\, t \in [b, b_n] \cap A_\epsilon :\ |F_\infty(t) - r_n(t)| > \delta_\epsilon/3.$$

The latter together with (3.32) yields

$$\forall\, n \geq N_4 :\ \forall\, t \in [0, b_n] \cap A_\epsilon :\ |F_\infty(t) - r_n(t)| > \delta_\epsilon/3. \tag{3.35}$$

Moreover, the Glivenko-Cantelli Lemma guarantees that $\sup_{t\in[0,1]} |\hat{F}_n(t, \omega) - F_\infty(t)| \to 0$ for $n \to \infty$ almost surely, i.e.

$$\forall\, \epsilon_1 > 0 :\ \exists\, N_5 \in \mathbb{N} :\ \mathbb{P}_{n,n_0}\left(\bigcap_{n \geq N_5} \{ \sup_{t\in[0,1]} |\hat{F}_n(t, \omega) - F_\infty(t)| < \delta_\epsilon/3\} \right) > 1 - \epsilon_1.$$

Applying the triangle inequality to (3.35) and $|\hat{F}_n(t,\omega) - F_\infty(t)| < \delta_\epsilon/3$ in the last expression, we obtain

$$\mathbb{P}_{n,n_0}\left(\bigcap_{n\geq N_5}\{\sup_{t\in[0,b_n]\cap A_\epsilon}|\hat{F}_n(t,\omega)-r_n(t)|>0\}\right)>1-\epsilon_1.$$

Altogether we get

$$\forall\,\epsilon>0:\ \forall\,\epsilon_1>0:\ \exists\,N_5\in\mathbb{N}:\ \mathbb{P}_{n,n_0}\left(\bigcap_{n\geq N_5}\{|t_\zeta^*-t_n|<\epsilon\}\right)>1-\epsilon_1,$$

which implies $t_n \to t_\zeta^*$ for $n \to \infty$ almost surely.

Now, continuity of r and convergence of r_n in $[0,\min(b,b_n)]$ imply $R_n/n = r_n(t_n) \to r(t_\zeta^*)$ for $n \to \infty$ almost surely.

In the case ρ_n, $n \in \mathbb{N}$, and ρ fulfil (A1), we get $\rho_n(t) \to \rho(t)$ and consequently $q_n(t) \to q(t)$ for all $t \in (0,\min(b,b_n)]$ if $n \to \infty$. Due to the definition of $q_n(0)$, $q(0)$ and continuity of q_n, q we get $q_n(0) \to q(0)$. Since $0 \leq q_n(t), q(t) \leq 1$ for $t \in [0,1]$ and q_n, q are continuous and non-decreasing, we obtain that $q_n \to q$ for $n \to \infty$ uniformly in $[0,1]$. Dominated convergence together with (3.6) yields $b(n,n_0) \to \zeta q(r(t_\zeta^*))$ almost surely, where $b(n,n_0)$ is the upper bound of the FDR of φ_n.

■

Now it will be shown that for an SUD procedure smaller critical values imply smaller upper bound for the FDR.

Lemma 3.22

Let r_1 and r_2 be rejection curves satisfying (A1) and suppose $r_1(t) \geq r_2(t)$, $t \in [0,1]$. Then,

$$b_1(n,n_0) \leq b_2(n,n_0),$$

where $b_i(n,n_0)$ is the upper bound (3.7) for an SUD test based on r_i for $i = 1,2$.

Proof: It holds $\rho_1(t) \leq \rho_2(t)$ and $q_1(t) \leq q_2(t)$ for $t \in [0,1]$, where $\rho_i = r_i^{-1}$, $q_i(t) = \rho_i(t)/t$ and $q_i(0) = \lim_{t\to 0}\rho_i(t)/t$ for $i = 1,2$. Moreover, for each realisation of p-values the test based on r_2 rejects at least as many hypotheses as the test based on r_1. Then formula (3.6) of the upper bound for an FDR of SUD test yields the result. ■

The next lemma shows that for an SU test based on (3.28) β_n/n is bounded.

Lemma 3.23

For an SU test with critical values (3.28) the optimal β_n is not greater than $(1-\alpha)n$.

Proof: It is known that $\text{FDR}_{n,n_0} \leq \alpha$ for an SU test based on the Simes rejection curve t/α. From Lemma 3.22 we obtain that an SU test based on $f_{\alpha,\beta}$ with some $\beta > 0$ controls the FDR if $f_{\alpha,\beta}(t) \geq t/\alpha$. Note that this condition is sufficient but not necessary. We obtain $f_{\alpha,\beta}(t) \geq t/\alpha$ if and only if $\beta/n \geq t(1-\alpha)/\alpha$ for $t \in (0,\alpha]$. Hence, choosing $\beta \geq (1-\alpha)n$ yields FDR control for $n \in \mathbb{N}$, i.e. the optimal value of β_n has to be less than or equal to $(1-\alpha)n$. ■

The lemmas considered before allow the following important result concerning the asymptotic behaviour of the optimal β_n.

Lemma 3.24

Let $\beta_n > 0$ be the minimum value which ensures FDR control of an SU test φ^n with critical values (3.28) in all possible DU(n, n_0) configurations, i.e. the FDR is not greater than α for all n_0-values and is equal to α for at least one $n_0 \in I_n$. Then it holds $\lim_{n\to\infty} \beta_n/n = 0$ and consequently $\lim_{n\to\infty} f_{\alpha,\beta_n} = f_\alpha$.

Proof: W.l.o.g. we assume $\lim_{n\to\infty} \beta_n/n = c$ for some $c \in (0, 1-\alpha]$, cf. Lemma 3.23 (otherwise we can take a subsequence with the desired property). Then $\lim_{n\to\infty} f_{\alpha,\beta_n}(t) = (1+c)f_\alpha(t)$. Lemma 3.21 yields that for $n_0/n \to \zeta \in [0,1]$ the asymptotic FDR based on $(1+c)f_\alpha$ is given by

$$\lim_{n\to\infty} \mathrm{FDR}_{n,n_0} = \zeta \frac{\alpha}{1 + c - (1+c)f_\alpha(t^*_\zeta)(1-\alpha)},$$

where t^*_ζ is the unique crossing point between $(1+c)f_\alpha(t)$ and $F_\infty(t) = 1 - \zeta + \zeta t$. Obviously, it holds that $t^*_\zeta \le t_\zeta = \alpha(1-\zeta)/\zeta/(1-\alpha)$ with t_ζ being the smallest crossing point between f_α and F_∞. Then

$$\lim_{n\to\infty} \mathrm{FDR}_{n,n_0} \le \alpha/(1+c) < \alpha \text{ for all } \zeta \in [0,1]. \tag{3.36}$$

On the other hand, β_n is defined such that the FDR equals α for at least one $n^*_0 = n^*_0(n) \in I_n$ in a DU model. Hence, there exists a sequence $\{n_k\} \in \mathbb{N}$ with $n^*_0(n_k)/n_k \to \zeta$ for some $\zeta \in [\alpha, 1]$ and $\lim_{k\to\infty} \mathrm{FDR}_{n_k, n^*_0(n_k)} = \alpha$ (note that for any $n, n_0 \in \mathbb{N}$ $FDR_{n,n_0} = \alpha$ is possible only for $\zeta \ge \alpha$), which contradicts (3.36).

■

The last lemma implies that the parameters β_n and β^*_n may increase but not faster than $o(n)$ for SUD test procedures with critical values (3.28) or (3.29), respectively. It is not clear, whether β_n and β^*_n diverge or are bounded for SUD tests. For SU tests with critical values (3.28) the next lemma gives a partial answer.

Lemma 3.25

Let β_n be the smallest parameter yielding FDR control for an SU test procedure based on (3.28) in a DU model. Then, we obtain $\beta_n \to \infty$ for $n \to \infty$.

Proof: Suppose $\beta_n \le \beta$ for all $n \in \mathbb{N}$ and some $\beta \in (0, \infty)$ and let $n_0/n \to \zeta \in (\alpha, 1]$. It will be shown that $\lim_{n\to\infty} \mathrm{FDR}_{n,n_0} > \alpha$, which contradicts the definition of β_n.

First note that $F_\infty(t) = f_\alpha(t)$ if and only if $t \in \{t_\zeta, 1\}$. Let $t^* \in (t_\zeta, 1)$. Then the FDR based on f_{α,β_n} can be represented as

$$\mathrm{FDR}_{n,n_0} = \mathbb{E}_{n,n_0}\left[\frac{V_n}{R_n} | t_n \le t^*\right] \mathbb{P}_{n,n_0}(t_n \le t^*) + \mathbb{E}_{n,n_0}\left[\frac{V_n}{R_n} | t_n > t^*\right] \mathbb{P}_{n,n_0}(t_n > t^*),$$

where t_n is the largest crossing point between $\hat{F}_n$ and f_{α,β_n}. Similarly as in Lemma 3.21 it can be proved that conditionally on $\{t_n \le t^*\}$ we obtain $t_n \to t_\zeta$ for $n \to \infty$ almost surely and under the condition $\{t_n > t^*\}$ we get $t_n \to 1$ for $n \to \infty$ almost surely. Moreover, $\{t_n \leq t^*\}$ implies that

$$\frac{V_n}{R_n} = \frac{V_n/n}{R_n/n} \to \frac{\zeta t_\zeta}{1 - \zeta + \zeta t_\zeta} = \alpha \text{ for } n \to \infty \text{ almost surely}$$

and conditionally on $\{t_n > t^*\}$ we obtain $V_n/R_n \to \zeta$ for $n \to \infty$ almost surely. Then

$$\lim_{n\to\infty} \mathrm{FDR}_{n,n_0} = \alpha \lim_{n\to\infty} \mathbb{P}_{n,n_0}(t_n \leq t^*) + \zeta \lim_{n\to\infty} \mathbb{P}_{n,n_0}(t_n > t^*).$$

We now show that $\lim_{n\to\infty} \mathbb{P}_{n,n_0}(t_n > t^*) > 0$, which implies $\lim_{n\to\infty} \mathrm{FDR}_{n,n_0} > \alpha$. For the largest critical value we have

$$\alpha_{n:n} = \frac{\alpha}{\beta_n/n + \alpha} \geq \frac{\alpha}{\beta/n + \alpha},$$

which implies

$$\mathbb{P}_{n,n_0}(R_n = n) = \mathbb{P}_{n,n_0}(V_n = n_0) = (\alpha_{n:n})^{n_0} \geq \left(\frac{\alpha}{\beta/n + \alpha}\right)^{n_0},$$

hence

$$\lim_{n\to\infty} \mathbb{P}_{n,n_0}(R_n = n) \geq \lim_{n\to\infty} \left[\frac{1}{\left(\frac{\beta/\alpha}{n} + 1\right)^n}\right]^{\zeta_n} = e^{-\frac{\beta}{\alpha}\zeta}.$$

From $\{R_n = n\} \equiv \{t_n = \alpha_{n:n}\}$ and $\lim_{n\to\infty} \alpha_{n:n} = 1$ we obtain $\{R_n = n\} \subseteq \{t_n > t^*\}$ for all larger n-values. Consequently, we get $\lim_{n\to\infty} \mathbb{P}_{n,n_0}(t_n > t^*) \geq e^{-\frac{\beta}{\alpha}\zeta} > 0$ and $\lim_{n\to\infty} \mathrm{FDR}_{n,n_0} > \alpha$, which is the desired contradiction. ■

For SUD(λ_n) tests based on (3.29) with $\lambda_n \in I_n$ and $k/n \to b > 0$ there are some indications that β_n^* may be bounded. For SUD(λ_n) tests based on (3.28) with $\lambda_n/n \to \kappa \in [0, 1)$ it also seems possible that β_n is bounded. But until now there is no proof for these statements.

3.4.4 Multiple-parameter adjustment

Now we consider another possibility to modify the AORC. As mentioned before, it seems that for an optimal β_n larger critical values in (3.28) are too large (i.e. the FDR takes its maximum typically for smaller n_0-values) and smaller critical values are too small (i.e. the FDR for larger n_0-values is typically smaller than α), cf. Figure 3.9. Therefore, we replace β_n by

$$\beta_{n,i} = \beta_{1,n} + \beta_{2,n}\left(\frac{i}{n}\right)^{\beta_{3,n}}$$

for suitable values $\beta_{1,n}, \beta_{2,n}$ and $\beta_{3,n}$, which influence each index i individually. This results in critical values of the form

$$\alpha_{i:n} = \frac{i\alpha}{n + \beta_{n,i} - i(1-\alpha)},\ i \in I_n, \tag{3.37}$$

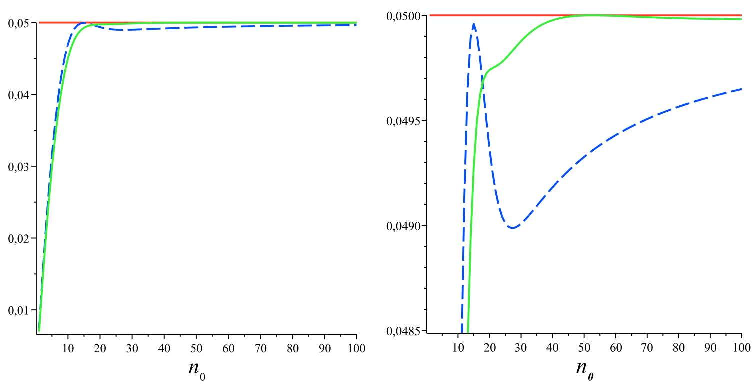

Figure 3.11: FDR-values for SU tests with $n = 100$ under DU configurations based on simultaneously β-adjustment with $\beta_n = 1.76$ (dashed line) and individually $\beta_{n,i}$-adjustment with $\beta_{1,n} = 1.09$, $\beta_{2,n} = 1.07$ and $\beta_{3,n} = 6$ (solid line), the right graph is zoomed.

which are always feasible for $\beta_{3,n} > 1$ and $\beta_{2,n} \leq n(1-\alpha)/\beta_{3,n}$. Here the term $\beta_{2,n}\,(i/n)^{\beta_{3,n}}$ leads to a leveling out of the FDR curve. As a criterion for a search algorithm for a suitable triple $(\beta_{1,n}, \beta_{2,n}, \beta_{3,n})$ such that $\mathrm{FDR}_{n,n_0} \leq \alpha$ for all $n_0 \in I_n$ we can try to minimize

$$t(\beta_{1,n}, \beta_{2,n}, \beta_{3,n}) = \sum_{n_0=\lceil n(1-\pi)\rceil}^{n} (\alpha - \mathrm{FDR}_{n,n_0})$$

for some $\pi \in (0, 1]$. For example, we can choose $\pi = 3/4n$. Obviously, the three-dimensional parameter space has to be restricted in order to make computations possible in reasonable time. For example, $\beta_{3,n}$ may be restricted to the integers $\{4, \ldots, 8\}$ or fixed in advance. For instance, $\beta_3 = 6$ can be taken for $n \leq 500$ and $\alpha = 0.05$. Thereby the search algorithm for the remaining parameters becomes faster. As an example, Figure 3.11 displays the resulting FDR-values under DU configurations for $n = 100$ and $\beta_{1,n} = 1.09$, $\beta_{2,n} = 1.07$, $\beta_{3,n} = 6$ in comparison with those originating from the simultaneously β-adjustment SU procedure with $\beta_n = 1.76$. The picture shows that the FDR is closer to α for $n_0 \geq 18$ with the refined adjustment.

3.4.5 Exact solving

As outlined at the end of Section 3.2, it is also possible to modify only m largest critical values for some $m \in I_{n-1}$ with an adjustment method such that (3.20) is fulfilled and to try to apply the recursive scheme (3.19) for $n_0 > m$ with $g \equiv g^*$ in order to determine the remaining critical values.

Let us consider an SU test with critical values (3.28). For example, SU procedures for $n = 90$ with $\beta_{90} = 1.718$ and for $n = 100$ with $\beta_{100} = 1.76$ control the FDR at level $\alpha = 0.05$. For $n = 90$ we obtained that one can choose the $m = 13$ largest β_{90}-adjusted critical values as starting values and apply the recursive scheme (3.19) with $g \equiv g^*$ for the remaining ones, so that $g(n_0/90) = 0.05$ for $n_0 = 14, \ldots, 100$. But for each $m \in \{14, \ldots, 87\}$, the recursive scheme

(3.19) with $n_0 \geq m + 1$ does not yield feasible critical values, whereas for $m \geq 88$ it does. Unfortunately, for $n = 100$, one has to choose at least the $m = 96$ largest critical values in order to apply (3.19) with $g(n_0/100) = 0.05$, $n_0 = m + 1, \ldots, 100$ successfully. This may illustrate the sensitivity of the recursive scheme. The other adjustment methods yield similar results so that it is not necessary to pursue them.

3.5 Iterative method

In this section we consider an iterative method for calculating of critical values for an SU test. Let $\alpha_{1:n}, \ldots, \alpha_{n:n}$ be feasible critical values, fulfilling that $\text{FDR}_{n,n_0}(\varphi^n) \approx \alpha$ for all $n_0 \geq k$ for some integer $k \equiv k(n, \alpha) \geq n\alpha$, where φ^n is an SU test based on these critical values. For example, in case of $\alpha = 0.05$ and $n = 100$, the right graph in Figure 3.9 suggests that β_n-adjusted AORC-based critical values fulfil this requirement for $k = 15$ and can therefore be taken as initial values. Now, we can try to iteratively modify certain critical values in order to reduce the corresponding distances $|\alpha - \text{FDR}_{n,n_0}(\varphi^n)|$ even further.

As showed in Section 3.2, the smallest critical value usually does not influence the FDR. Therefore, we have to identify which critical values have the most impact on $\text{FDR}_{n,n_0}(\varphi^n)$ for a given value of n_0. We recall that, at least for $\zeta < 1$, the FDP, i.e. the ratio V_n/R_n, convergences to the asymptotic FDR in DU models with $n_0/n \to \zeta$. Since p-values under alternatives follow a Dirac distribution in DU models, we conclude that

$$\text{FDP} = \frac{V_n}{R_n} = \frac{V_n}{V_n + n_1} \approx \alpha.$$

This leads to $V_n \approx n_1\alpha/(1 - \alpha)$. Then we obtain that about $R_n = V_n + n_1 \approx n_1/(1 - \alpha)$ hypotheses are rejected by an α-exhausting SU test in DU models. Therefore, the critical values $\alpha_{i:n}$ with i close to $n_1/(1 - \alpha)$ are crucial for $\text{FDR}_{n,n_0}(\varphi^n)$ and accordingly for a given $i \in I_n$ the critical value $\alpha_{i:n}$ has the most impact on $\text{FDR}_{n,n_0}(\varphi^n)$ with $n_0 \approx n - i(1 - \alpha)$. In order to modify $\text{FDR}_{n,n_0}(\varphi^n)$-values for all $n_0 = k, \ldots, n$, we have to modify critical values with indices ranging from 1 to $i^* \equiv i^*(n, k, \alpha)$, which is an integer close to $(n - k)/(1 - \alpha)$.

For the derivation of an appropriate iteration scheme, we rewrite the initial critical values in the form

$$\alpha_{i:n} = \frac{ic_i}{n - i(1 - c_i)} = f_{c_i}^{-1}(i/n),\ i \in I_n, \tag{3.38}$$

which formally equals (3.1) with a vector of "local FDR levels" $c = (c_1, \ldots, c_n)$. Moreover, we make use of the notation $\text{FDR}_{n,n_0}(c)$ for $\text{FDR}_{n,n_0}(\varphi^n)$, where φ^n is defined via the critical values given in (3.38). Now, let $n_0(i)$ be an integer closest to $n - i(1 - \alpha)$ and consider the mapping

$$c \mapsto u(c) = (u_1(c), \ldots, u_{i^*}(c), c_{i^*+1}, \ldots, c_n),$$

where

$$u_i(c) = \alpha \frac{c_i}{\text{FDR}_{n,n_0(i)}(c)},\ i = 1, \ldots, i^*.$$

We note that $\mathrm{FDR}_{n,n_0}(u(c)) = \mathrm{FDR}_{n,n_0}(c)$ for $n_0 = 1, \ldots, n - i^*$. Assume that for fixed, given constants c_i, $i = i^* + 1, \ldots, n$, there exist some c_i^*, $i = 1, \ldots, i^*$, such that the vector $c^* = (c_1^*, \ldots, c_{i^*}^*, c_{i^*+1}, \ldots, c_n)$ fulfils the fixed point property $c^* = u(c^*)$. Then we obtain $c_i^* = u_i(c^*) = \alpha c_i^* / \mathrm{FDR}_{n,n_0}(c^*)$, $i = 1, \ldots, i^*$, which is equivalent to $\mathrm{FDR}_{n,n_0}(c^*) = \alpha$, $n_0 = k, \ldots, n$. Therefore, an iteration scheme for the vector of local FDR levels c, i.e., setting $c^{(j)} = u(c^{(j-1)})$, seems to be a promising approach. Clearly, there is no fixed-point theorem at hand guaranteeing convergence. Moreover, as mentioned in Section 3.2, for a given FDR-bounding curve (or pre-specified FDR-values) there are not necessarily corresponding feasible critical values, that is, the formal solution of the target equation (3.22) is not necessarily feasible. Finally, resulting FDR-values can slightly exceed the given α-level, because the mapping $u(c)$ does not guarantee that c^* is approached from below. However, the method seems to work well and the distances $|\alpha - \mathrm{FDR}_{n,n_0}(\varphi^n)|$, $n_0 = k, \ldots, n$, on average get reduced by the outlined iteration method for a suitable number of iterations $J \in \mathbb{N}$ (say).

In order to describe the method more formally, we set without restriction

$$i^* \equiv i^*(n, k, \alpha) = \lfloor (n - k)/(1 - \alpha) \rfloor,$$

and let $(\alpha_{1:n}^{(0)}, \ldots, \alpha_{n:n}^{(0)})$ be feasible start critical values and $\alpha_{i:n}^{(j)} \equiv \alpha_{i:n}^{(0)}$ for $i = i^* + 1, \ldots, n$, $j = 1, \ldots, J$, that is, only the smaller i^* critical values will be changed. For a modification of critical values with indices ranging from 1 to i^*, we proceed as given in the following algorithm.

For j from 1 to J do:

1. For i from i^* to 1 by -1 do:
   (a) Determine $c_i^{(j-1)}$ from $\alpha_{i:n}^{(j-1)} = i c_i^{(j-1)}/(n - i(1 - c_i^{(j-1)}))$.
   (b) Put $c_i^{(j)} = \alpha c_i^{(j-1)} / \mathrm{FDR}_{n,n_0(i)}(c^{(j-1)})$.
   (c) Calculate $\alpha_{i:n}^{(j)} = i c_i^{(j)}/(n - i(1 - c_i^{(j)}))$.
   (d) If (A1) is not fulfilled, then put $\alpha_{i:n}^{(j)} = i \alpha_{i+1:n}^{(j)}/(i + 1)$.
2. Calculate $\mathrm{FDR}_{n,n_0}(c^{(j)})$, $n_0 = n - i^* + 1, \ldots, n$.

Notice that in the latter algorithm the number $n_0(i)$ in the expression $\mathrm{FDR}_{n,n_0(i)}(c^{(j-1)})$ is only loosely defined by setting $n_0(i)$ as the integer "closest to $n - i(1 - \alpha)$". To be more precise, one can replace $\mathrm{FDR}_{n,n_0(i)}(c^{(j-1)})$ by a linear interpolation of the two adjacent values $\mathrm{FDR}_{n,\lfloor n-i(1-\alpha)\rfloor}(c^{(j-1)})$ and $\mathrm{FDR}_{n,\lceil n-i(1-\alpha)\rceil}(c^{(j-1)})$.

As a demonstrating example, we choose $n = 100$, $\alpha = 0.05$, and the critical values resulting from the simultaneous AORC-adjustment with $\beta_{100} = 1.76$ as starting values. The FDR of the SU test with these starting values takes its maximum in the point $n_0 = 15$, so we choose $k = 15$ and consequently $i^* = \lfloor 85/0.95 \rfloor = 89$. Moreover, we perform $J = 50$ iterations. Figure 3.12 shows the resulting FDR values of the SU test with critical values $\alpha_{1:100}^{(50)}, \ldots, \alpha_{89:100}^{(50)}$, $\alpha_{90:100}^{(0)}, \ldots, \alpha_{100:100}^{(0)}$ under DU configurations. Here the improvement obtained by the iterative method becomes obvious.

We tested the iterative method for a series of values of n and α. As initial critical values we took simultaneous β_n-adjusted as well as β_n^*-adjusted critical values (cf. Section 3.4). For exam-

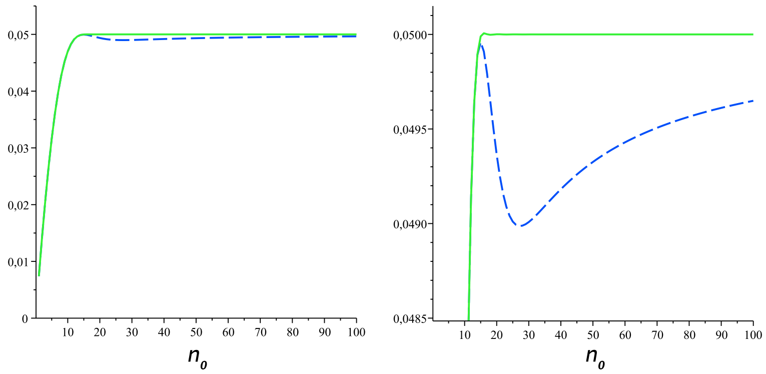

Figure 3.12: FDR-values for SU tests with $n = 100$ under DU configurations, which are based on iteratively modified critical values with 50 iterations (solid line) and simultaneously β-adjusted ones with $\beta_n = 1.76$ (dashed line), the right graph is zoomed.

ple, for $n = 100, 300, 1000$, $J = 20, 10, 10$ iterations based on initial simultaneous β_n-adjustment and $J = 10, 2, 1$ iterations based on initial β_n^*-adjustment gave satisfying results. Typically, β_n^*-adjusted critical values need fewer number of iterations than β_n-adjusted critical values. It seems that the closer starting FDR-values are to α the better the iterative method woks. Unfortunately, the resulting realised FDR-values under Dirac-uniform configurations typically exceed the given α-level for some $n_0 \geq k$. But the actual differences $|\alpha - \mathrm{FDR}_{n,n_0}(\varphi^n)|$, $n_0 \geq k$, seem of negligible magnitude, i.e., for a suitable number of iterations the observed differences were never greater than 5×10^{-5}. Clearly, in a final step we can decrease the resulting critical values in a suitable way by a suitable small amount such that all FDR-values are smaller than α.

3.6 Concluding remarks

We have implemented various approaches to construct critical values, which exhaust the given α-level. Thereby, different methods lead to different sets of critical values and no set uniformly dominates the others such that no method can be definitively preferred. The choice of the method may depend on previous knowledge and computational resources.

The FDR bounding curve method described in Section 2.4 seems to be the most attractive. Because it is a method for which the FDR (or the upper bound for SUD procedure) is explicitly given, so that, we only have to calculate critical values with the recursive formula (3.19). But the question, whether these critical values are feasible for a given n, is still open. Nevertheless, this approach seems to approximate $g^*(\zeta) = \min(\alpha, \zeta)$ very well and computations (especially for a fixed $n \leq 2000$) can be made in reasonable time for SU tests, the critical values of which are also valid for all corresponding SUD test procedures.

For the other methods we do not have any theoretical proof that the resulting FDR's are close to α, but we observed it in all simulations and the asymptotic FDR is equal to α, cf. Section 3.3.

Moreover, for the adjustment methods in Section 3.4 and an iterative method in Section 3.5 we can always construct feasible critical values, i.e. we can always find adjusting parameters such that the FDR is controlled. Note that all theses methods can be combined with exact solving (cf. Subsection 3.4.5) in order to improve smallest critical values.

Although the FDR of an individual $\beta_{i,n}$-adjustment is typically closer to α than the FDR of a β_n- or β_n^*-adjustment, computations in this case are distinctively slow, so that we do not recommend this method for $n > 200$. Since simultaneous β_n- or β_n^*-adjustment procedures are very easy to implement if a suitable β_n (or β_n^*) is computed, these approaches can be a good alternative to the FDR bounding curve method. Our investigations show that the FDR of simultaneous β_n- or β_n^*-adjustment is close to α if n is large (for example $n \geq 1000$ for a β_n-adjustment method and $n \geq 300$ for a β_n^*-adjustment method) and the computation time thereby is acceptable. Moreover, for the β_n-adjustment, the larger critical values there seem to be larger than the ones for the other procedures, that is, if it is known that the proportion of true null hypotheses is small, then a β_n-adjustment method can be the best.

For the application of an iterative method in Section 3.5 we first have to calculate a β_n (or β_n^*) with a simultaneous β_n- (or β_n^*-) adjustment procedure, which may result in an extended computation time. But for $n \in \mathbb{N}$ not too large (for example, $n \leq 1000$) the iterative method is reasonable (especially if the proportion of true null hypotheses is known to be small) and calculation time is reasonable, too. We recommend this method with simultaneous β_n-adjustment critical values as starting values for smaller values of n (for example $n \leq 300$). For $300 \leq n \leq 2000$, we recommend to use the iterative method in connection with β_n^*-adjusted initial critical values, because it seems that only a few iterations are needed in this case.

If we compare the critical values generated with the methods described before, we observe that the differences are negligible for most of the critical values except for a small proportion of the larger ones. Typically, large critical values come into play only if a large proportion of hypotheses is extremely false with p-values close to zero which is not often the case in practice. Therefore, we expect that the choice of the method for the determination of critical values should have nearly no influence on the final results of the test procedure.

For the computation of critical values according to the given procedures, we provide Maple worksheets under the URL `http://www.helmut-finner.de`, which can be executed in reasonable time on a standard desktop computer for $n \leq 2000$. Moreover, for SU and SUD(λ) tests with critical values (3.28) and SU tests based on (3.29), we tabulated the constants β_n (β_n^* respectively) for $n \leq 2000$, $\alpha = 0.01, 0.05, 0.1$ and $\lambda_n \leq 0.9n, 0.7n, 0.4n$.

Finally, we would like to give a recommendation for practical application if the number of hypotheses n is large ($n > 2000$). Computing time in this case can be enormous, so that we recommend the β_n- or β_n^*-adjustments with some fixed parameter β (or β^*). For example, for $\alpha = 0.05$ one may choose $\beta_n \in [\beta_{2000}, 2] = [1.58, 2]$ and $\lambda_n \approx 0.7n$ for an SUD(λ_n) test and $\beta_n^* \in [\beta_{2000}, 2] = [1.45, 2]$ for an SU test (for $k \approx n(1 - 2\alpha)$) with critical values (3.29) for $k \approx n(1 - 2\alpha)$. Although the upper FDR bound can exceed the α-level for these tests for some DU configurations, the possible exceedance should be negligible. As mentioned before, the FDR is asymptotically controlled such that the possible exceedance of the α-level converges to 0 as n increases.

Chapter 4

Dependent p-values and multiple test procedures

Up to now we have considered p-values that fulfil (I1) and optionally (D1) and/or (I2). In statistical applications these assumptions often do not apply. Especially if independence requirements are not satisfied, the pre-specified FWER- and/or FDR-level are possibly exceeded.

In this chapter we investigate various types of dependence of test statistics, for which the FWER and/or the FDR can be be controlled at least asymptotically. In Section 4.1 we review different types of dependence between test statistics that are commonly used in the literature on multiple tests controlling the FDR. Then we investigate a somewhat relaxed version of "weak dependence". In Section 4.2 we consider a BPI procedure with the threshold (2.4) and an SDPI test with the thresholds (2.33) based on a plug-in estimator $\hat{n}_0$ (cf. Chapter 2) and give a condition on $\hat{n}_0$, for which asymptotic FWER control is ensured. We introduce assumptions concerning the ecdf of p-values corresponding to true null hypotheses and show that under these assumptions BPI tests control the FWER at least asymptotically. In Section 4.3 we show that "weak dependence" guarantees that a broad class of SUD test procedures (cf. Chapter 3) control the FDR asymptotically under specific conditions. We discuss various power requirements ensuring asymptotic FDR control. Section 4.4 deals with the question how weak dependence conditions and/or convergence of the ecdf of p-values can be proved. We give a condition on covariances of p-values corresponding to true/false null hypotheses which is equivalent to the convergence of the ecdf of these p-values in the sense of the Glivenko-Cantelli Theorem, which yields in the case of p-values under nulls some especial type of weak dependence. Moreover, we discuss different types of dependence fulfilling this condition. In Section 4.5 we consider an important example of "weak dependence", that is, block-dependent p-values. In Section 4.6 we are concerned with so-called pairwise comparisons, one of the most famous multiple hypotheses testing problems. We show that the concept of weak dependence applies to this problem yielding asymptotic FWER/FDR control. We conclude this chapter with some simulations for dependent p-values, cf. Section 4.7.

In the case of FWER control under dependence, it is sometimes necessary to restrict attention to situations, where the number of true hypotheses $n_0 = n_0(n)$ tends to infinity with n tending to infinity. In other words, asymptotic FWER control can only be guaranteed on the restricted parameter space

$$\Theta^* = \{\vartheta \in \Theta : \lim_{n\to\infty} |I_{n,0}(\vartheta)| = \infty\}.$$

4.1 Weak dependence

In recent time, some results have been obtained for different types of dependence. For example, Benjamini and Yekutieli [2001] introduced the concept of so-called positive regression dependence on subsets (PRDS) as follows.

Definition 4.1
Let $X = (X_1, \dots, X_n)$ be a vector of random variables with $n \geq 2$. The joint distribution of $X_1, \dots, X_n$ is called ***positive regression dependent on each one from a subset*** *$I' \subseteq I_n$, or* ***PRDS on*** *I', if $\mathbb{P}(X \in D|X_i = x)$ is non-decreasing in x for any increasing set $D \in Im(X)$ (i.e. $x \in D$ and $y \geq x$ imply $y \in D$) and for each $i \in I'$.*

Multivariate normal distributions with positive correlations belong to the set of distributions satisfying this property. Benjamini and Yekutieli [2001] proved that an LSU test procedure (cf. Section 1.3) controls the FDR when test statistics are PRDS on each of the test statistics corresponding to true null hypotheses.

A weaker condition than PRDS was given in Finner *et al.* [2009], that is,

(D2) $\forall\, \vartheta \in \Theta : \forall\, j \in I_n : \forall\, i \in I_{n,0}(\vartheta) : \mathbb{P}_\vartheta(R_n \geq j|p_i \leq t)$ is non-increasing in $t \in (0, \alpha_{j:n}]$.

Among others things, the authors showed that an LSU test controls the FDR if (D2) applies, cf. Theorem 4.1 in that paper.

Another interesting result concerning FDR control for dependent test statistics can be found in Storey *et al.* [2004]. The authors defined **weak dependence** for p-values in the following way.

(WD1) $\forall\, t \in (0,1) : \lim_{n\to\infty} \hat{F}_{n,0}(t) = F_0(t)$ and $\lim_{n\to\infty} \hat{F}_{n,1}(t) = F_1(t)$ almost surely

and $0 < F_0(t) \leq t$,

where $\hat{F}_{n,0}$ denotes the ecdf of p-values corresponding to true null hypotheses and $\hat{F}_{n,1}$ is the ecdf of p-values corresponding to alternatives. Storey *et al.* [2004] also introduced a modified LSU test based on a plug-in estimator for the proportion of true null hypotheses, which works as follows. In the first step estimate the proportion of true null hypotheses $\zeta_n = n_0/n$ by e.g.

$$\hat{\zeta}_n = \frac{1 - \hat{F}_n(\lambda)}{1 - \lambda},$$

where $\lambda \in (0,1)$ is arbitrary but fixed. Then apply an LSU test with α replaced by $\alpha/\hat{\zeta}_n$, that is, an SU test with critical values $\hat{\alpha}_{i:n} = i\alpha/(\hat{\zeta}_n n)$, $i \in I_n$. It was proved that the described **LSU plug-in tests** control the FDR asymptotically under certain additional assumptions if (WD1) is fulfilled.

In general, weak dependence in multiple testing problems can be often characterised by the requirement

$$\forall\, t \in (0,1):\ \hat{F}_{n,0}(t) \xrightarrow{\mathcal{C}} F_0(t) \leq t \ \text{ and } \ \hat{F}_{n,1}(t) \xrightarrow{\mathcal{C}} F_1(t) \tag{4.1}$$

for some cdfs $F_i : [0,1] \to [0,1]$, $i = 0,1$. Thereby, $\xrightarrow{\mathcal{C}}$ denotes some type of convergence for $n \to \infty$ like almost surely ($\mathcal{C} = a.s.$), complete convergence ($\mathcal{C} = c.c.$), in probability ($\mathcal{C} = P$), in the $\mathbb{L}_p$ norm ($\mathcal{C} = \mathbb{L}_p$) or in the sense of the Glivenko-Cantelli theorem ($\mathcal{C} = GC$). Given a fixed value $\vartheta \in \Theta$, the proportion of true null hypotheses will be denoted by $\zeta_n = n_0/n$. Thereby it is assumed that $\lim_{n\to\infty} \zeta_n = \zeta \in [0,1]$. A further simplification appears by assuming $F_0(t) = t$ for all $t \in [0,1]$, which is appropriate if p_i, $i \in I_{n,0}$, are independently and uniformly distributed on $[0,1]$, that is, condition (D1) applies. The nice point by assuming (4.1) is that we are asymptotically in a mixture model case

$$F = \zeta F_0 + (1-\zeta)F_1, \tag{4.2}$$

which is also referred to as a random effects model. As a consequence, asymptotically the p-values may be reinterpreted as iid variables with marginal cdf F. This argumentation may be considered as the main reason why many authors restrict attention to a mixture model for p-values defined via (4.2). Moreover, assuming iid p-values with marginal cdf given by (4.2) and ignoring any kind of weak dependence makes life much easier with respect to any error rate control criterion.

In order to get some asymptotic error control it will be shown in this chapter that it often suffices to relax the weak dependence condition (4.1) to

(WD2) $\quad \forall\, t \in [0,1]:\ \forall\, \epsilon > 0:\ \lim\limits_{n\to\infty} \mathbb{P}_\vartheta(\hat{F}_{n,0}(t) > t + \epsilon) = 0.$

This condition allows that p-values corresponding to true null hypotheses may be dependent but the limiting ecdf of these p-values is bounded by the cdf of the uniform distribution $F = Id$. A random variable Y such that $\lim_{n\to\infty} \mathbb{P}_\vartheta(\hat{F}_{n,0}(t) > Y) = 0$ is called **asymptotically larger than $\hat{F}_{n,0}(t)$ in probability**, cf. Edgar and Sucheston [1992], p. 117. Then $F = Id$ is the **stochastic upper limit of $\{\hat{F}_{n,0}(t)\}_{n\in\mathbb{N}}$**, written $s \limsup_{n\to\infty} \hat{F}_{n,0}(t)$, that is, $F = Id$ is the essential infimum of the set of all random variables which are asymptotically greater than $\hat{F}_{n,0}(t)$ in probability.

Similarly as in Lemma A.7 it can be proved that (WD2) is equivalent to

$$\forall\, \epsilon > 0:\ \lim_{n\to\infty} \mathbb{P}_\vartheta\left(\sup_{t\in[0,1]} (\hat{F}_{n,0}(t) - t) > \epsilon\right) = 0. \tag{4.3}$$

Condition (4.3) says that $\hat{F}_{n,0}$ is asymptotically stochastically uniformly bounded by $F = Id$. An important special case of (WD2) and/or (4.3) given by

$$\sup_{t\in[0,1]} |\hat{F}_{n,0}(t) - t| \to 0,\ n \to \infty, \ \text{ in probability} \tag{4.4}$$

is often least favourable for the FDR and/or FWER. An extended version of (4.4) is given as follows. Suppose $p_i \sim G_{i,0}$ for $i \in I_{n,0}$ with $G_{i,0}(t) \leq t$ for all $t \in [0,1]$. Then the condition

(WD3) $\quad \forall\, t \in [0,1]:\ \hat{F}_{n,0}(t) - \dfrac{1}{n_0} \sum\limits_{i\in I_{n,0}} G_{i,0}(t) \to 0,\ n \to \infty,\ \text{ in probability},$

also implies (WD2). Obviously, if $G_{i,0}(t) = t$ for $i \in I_{n,0}$, i.e. $p_i \sim U([0,1])$ for $i \in I_{n,0}$, then (WD3) is equivalent to (4.4). In view of (4.4) and (WD3) weak dependence typically means that the asymptotic ecdf of a set of dependent p-values $p_i \sim G_{i,0},\ i \in I_{n,0}$, coincides with the asymptotic ecdf of the corresponding set of independent p-values $p_i \sim G_{i,0},\ i \in I_{n,0}$.

Sometimes it even suffices to find a unique point $t_0 \in (0,1)$ such that

$$\forall\, \epsilon > 0 :\ \lim_{n\to\infty} \mathbb{P}_\vartheta(\hat{F}_{n,0}(t_0) > t_0 + \epsilon) = 0. \tag{4.5}$$

4.2 Plug-in tests and asymptotic control of the FWER under weak dependence

This section deals with asymptotic FWER control of a plug-in test procedure based on an estimator for the number n_0 of true null hypotheses (cf. Chapter 2) for dependent and not necessarily uniformly distributed p-values. Violation of conditions (D1) and/or (I1) may result in a poor estimation of n_0, exceedance of the FWER-level and/or low power. In the case of independent p-values being stochastically larger than a uniform variate, estimators for the number n_0 of true null hypotheses tend to be too large such that FWER of a BPI test or an SDPI procedure is controlled, while the power may be rather small. For example, interval hypotheses or discrete test statistics yield such kind of p-values. The problem of dependence between null p-values is generally more serious in terms of FWER control.

Remember that a multiple test procedure φ controls the FWER at level α with respect to Θ if $\sup_{\vartheta\in\Theta} \mathbb{P}_\vartheta(V_n > 0) \leq \alpha$. We say that **FWER is asymptotically controlled** at level α with respect to Θ^* if

$$\forall\, \vartheta \in \Theta^* :\ \limsup_{n\to\infty} \mathbb{P}_\vartheta(V_n > 0) \leq \alpha.$$

For iid uniformly distributed p-values corresponding to true null hypotheses the SLLN implies that $\hat{F}_{n,0}(z) \to z$ for $n \to \infty$ almost surely, uniformly in $z \in [0,1]$. This yields that estimators $\hat{n}_0$ given in (2.6) or (2.9) are asymptotically not smaller than n_0 and consequently the FWER is asymptotically controlled. If p-values are dependent, then $\hat{F}_{n,0}$ does not necessarily converge and estimates for n_0 may behave rather irregularly. It will be shown that the condition

$$\forall\, \vartheta \in \Theta^* :\ \forall\, \epsilon > 0 :\ \lim_{n\to\infty} \mathbb{P}_\vartheta\left(\frac{\hat{n}_0}{n_0} < 1 - \epsilon\right) = 0 \tag{4.6}$$

is sufficient for asymptotic FWER control with respect to Θ^* for some plug-in tests under weak dependence. The main result is given in the next theorem.

Theorem 4.2
Let $\hat{n}_0$ be an estimator of n_0 satisfying condition (4.6). Then a BPI procedure with threshold (2.4) *and/or an SDPI procedure with critical values* (2.33) *asymptotically control the FWER on Θ^* at the prespecified level α.*

Proof: Let $\vartheta \in \Theta^*$. For an SDPI test procedure the critical value that has to be compared with the smallest p-value corresponding to true null hypotheses is not greater than $\hat{\alpha}^{(1)}_{n_1+1:n} = \max(\alpha/\hat{n}_0, \alpha/n_0)$, where $n_1 = n - n_0$. Moreover, the threshold $\alpha/\hat{n}_0$ of a BPI test with the same estimator $\hat{n}_0$ as in an SDPI test is clearly not greater than $\hat{\alpha}^{(1)}_{n_1+1:n}$ either. Hence, $\mathbb{P}_\vartheta(V_n = 0) \geq \mathbb{P}_\vartheta(\min_{i \in I_{n,0}} p_i > \hat{\alpha}^{(1)}_{n_1+1:n})$ for both procedures. Since assumption (4.6) yields

$$\forall\, \epsilon > 0 :\ \forall\, \delta > 0 :\ \exists\, N_{\epsilon,\delta} \in \mathbb{N} :\ \forall\, n \geq N_{\epsilon,\delta} :\ \mathbb{P}_\vartheta\left(\frac{\hat{n}_0}{n_0} \geq 1 - \epsilon\right) \geq 1 - \delta,$$

we obtain for a BPI test and an SDPI procedure with the same $\hat{n}_0$ as in a BPI test

$$\begin{aligned}
\mathbb{P}_\vartheta(V_n = 0) &\geq \mathbb{P}_\vartheta\left(\left\{\min_{i \in I_{n,0}} p_i \geq \hat{\alpha}^{(1)}_{n_1+1:n}\right\} \cap \left\{\frac{\hat{n}_0}{n_0} \geq 1 - \epsilon\right\}\right) \\
&= \mathbb{P}_\vartheta\left(\left\{\min_{i \in I_{n,0}} p_i \geq \hat{\alpha}^{(1)}_{n_1+1:n}\right\} \cap \{\hat{n}_0 \in [(1-\epsilon)n_0, n_0)\}\right) \\
&\quad + \mathbb{P}_\vartheta\left(\left\{\min_{i \in I_{n,0}} p_i \geq \hat{\alpha}^{(1)}_{n_1+1:n}\right\} \cap \{\hat{n}_0 \geq n_0\}\right) \\
&= A \text{ (say)}.
\end{aligned}$$

If $\hat{n}_0 \in [(1-\epsilon)n_0, n_0)$, then $\hat{\alpha}^{(1)}_{n_1+1:n} = \alpha/\hat{n}_0 \leq \alpha/((1-\epsilon)n_0)$. For $\hat{n}_0 \geq n_0$ we obtain that $\hat{\alpha}^{(1)}_{n_1+1:n} = \alpha/n_0$. Hence, we get for a BPI test and/or an SDPI procedure

$$\begin{aligned}
A &\geq \mathbb{P}_\vartheta\left(\left\{\min_{i \in I_{n,0}} p_i \geq \frac{\alpha}{(1-\epsilon)n_0}\right\} \cap \{\hat{n}_0 \in [(1-\epsilon)n_0, n_0)\}\right) \\
&\quad + \mathbb{P}_\vartheta\left(\left\{\min_{i \in I_{n,0}} p_i \geq \frac{\alpha}{n_0}\right\} \cap \{\hat{n}_0 \geq n_0\}\right) \\
&\geq \mathbb{P}_\vartheta\left(\left\{\min_{i \in I_{n,0}} p_i \geq \frac{\alpha}{(1-\epsilon)n_0}\right\} \cap \left\{\frac{\hat{n}_0}{n_0} \geq 1 - \epsilon\right\}\right) \\
&\geq \mathbb{P}_\vartheta\left(\min_{i \in I_{n,0}} p_i \geq \frac{\alpha}{(1-\epsilon)n_0}\right) - \delta \\
&= 1 - \mathbb{P}_\vartheta\left(\exists\, i \in I_{n,0} :\ p_i \leq \frac{\alpha}{(1-\epsilon)n_0}\right) - \delta \\
&\geq 1 - \sum_{i \in I_{n,0}} \mathbb{P}_\vartheta\left(p_i \leq \frac{\alpha}{(1-\epsilon)n_0}\right) - \delta \\
&\geq 1 - \frac{\alpha}{1-\epsilon} + \delta.
\end{aligned}$$

Letting $\epsilon \to 0$ and $\delta \to 0$ yields $\mathbb{P}_\vartheta(V_n = 0) = 1 - \text{FWER}_\vartheta \geq 1 - \alpha$ and consequently the assertion follows. ∎

The next lemma shows that estimators defined in (2.6) or (2.9) fulfil condition (4.6).

Lemma 4.3

(a) If the ecdf $\hat{F}_{n,0}$ of all null p-values fulfils (4.5) *for $t_0 = \lambda \in (0,1)$, then condition (4.6) holds for the estimators defined in (2.6) for any fixed $\kappa \in \mathbb{R}$.*

(b) Let $k = k(n) \in I_n$ be such that

$$\limsup_{n\to\infty} \frac{k}{n} < 1. \tag{4.7}$$

If the ecdf $\hat{F}_{n,0}$ of all null p-values fulfils (WD2), then condition (4.6) holds for the estimators defined in (2.9) *for any fixed $\kappa \in \mathbb{R}$.*

Proof: Let $\vartheta \in \Theta^*$.

(a) Condition (4.5) can be rewritten as

$$\forall\, \epsilon > 0 :\ \forall\, \delta > 0 :\ \exists\, N_{\epsilon,\delta} \in \mathbb{N} :\ \forall\, n \geq N_{\epsilon,\delta} :\ \mathbb{P}_\vartheta(\hat{F}_{n,0}(\lambda) < \lambda + \epsilon) > 1 - \delta.$$

Estimators $\hat{n}_0$ as defined in equation (2.6) satisfy

$$\frac{\hat{n}_0}{n_0} \geq \frac{n_0 - V_n(\lambda) + \kappa}{n_0(1-\lambda)} = \frac{1 - \hat{F}_{n,0}(\lambda) + \kappa/n_0}{1-\lambda},$$

hence for fixed $\epsilon > 0$ and $\delta > 0$ we obtain for $n \geq N_{\epsilon,\delta}$ that

$$\begin{aligned} 1-\delta &< \mathbb{P}_\vartheta(\hat{F}_{n,0}(\lambda) < \lambda + \epsilon) \\ &= \mathbb{P}_\vartheta\left(\frac{1 - \hat{F}_{n,0}(\lambda) + \kappa/n_0}{1-\lambda} > 1 - \frac{\epsilon - \kappa/n_0}{1-\lambda}\right) \\ &\leq \mathbb{P}_\vartheta\left(\frac{\hat{n}_0}{n_0} > 1 - \frac{\epsilon - \kappa/n_0}{1-\lambda}\right). \end{aligned}$$

Obviously, for each $\epsilon_1 > 0$ there exist $\epsilon > 0$ and $N_\epsilon \in \mathbb{N}$ such that $\epsilon_1 > (\epsilon - \kappa/n_0)/(1-\lambda)$ for all $n \geq N_\epsilon$, since $\vartheta \in \Theta^*$, i.e. $n_0 = n_0(n) \to \infty$ for $n \to \infty$. Then for $n \geq \max(N_{\epsilon,\delta}, N_\epsilon)$ we get

$$1-\delta < \mathbb{P}_\vartheta\left(\frac{\hat{n}_0}{n_0} > 1 - \frac{\epsilon - \kappa/n_0}{1-\lambda}\right) \leq \mathbb{P}_\vartheta\left(\frac{\hat{n}_0}{n_0} > 1 - \epsilon_1\right),$$

which implies (4.6).

(b) We divide the proof in two steps, i.e. (i) $\liminf_{n\to\infty}(n-k)/n_0 \geq 1$; and (ii) $\liminf_{n\to\infty}(n-k)/n_0 < 1$.

(i) For $\hat{n}_0$ defined in (2.9) we immediately get $\hat{n}_0/n_0 \geq (n - k + \kappa)/n_0$. Hence, (4.6) is fulfilled.

(ii) W.l.o.g. let $k = k(n) = n - n_0 + s$ for an $s = s(n) \in I_{n_0}$ and let

$$\lim_{n\to\infty} \frac{s}{n_0} = \eta. \tag{4.8}$$

Thereby, the limiting value η is always greater than 0 and smaller than 1, because

$$1 > \liminf_{n\to\infty} \frac{n-k}{n_0} = \lim_{n\to\infty} \frac{n_0 - s}{n_0} = 1 - \lim_{n\to\infty} \frac{s}{n_0} = 1 - \eta$$

implies $\eta > 0$ and

$$\limsup_{n\to\infty} \frac{k}{n} = \limsup_{n\to\infty} \frac{n - n_0 + s}{n} = \limsup_{n\to\infty} \frac{n - (1-\eta)n_0}{n} = 1 - (1-\eta) \liminf_{n\to\infty} \frac{n_0}{n},$$

which together with (4.7) leads to $(1-\eta)\liminf_{n\to\infty} n_0/n > 0$ and hence $\eta < 1$. Moreover, $p_{k:n} \geq p^0_{s:n_0}$, where $p^0_{s:n_0}$ is the sth smallest p-value corresponding to true null hypotheses. Hence,

$$\frac{\hat{n}_0}{n_0} \geq \frac{1 - s/n_0 + \kappa/n_0}{1 - p^0_{s:n_0}}.$$

Note that $p^0_{s:n_0} = \hat{F}^{-1}_{n,0}(s/n_0)$, where $\hat{F}^{-1}_{n,0}(u) = \inf\{t \in [0,1] : \hat{F}_{n,0}(t) \geq u\}$. Condition (WD2) is equivalent to (4.3), which can be rewritten as

$$\forall\, \epsilon > 0 :\ \forall\, \delta > 0 :\ \exists\, N_{\epsilon,\delta} \in \mathbb{N} :\ \forall\, n \geq N_{\epsilon,\delta} :\ \mathbb{P}_\vartheta\left(\hat{F}_{n,0}(t) \leq t + \epsilon,\ \forall\, t \in (0,1)\right) > 1 - \delta.$$

Since $\hat{F}_{n,0}(t) \leq t + \epsilon$, for all $t \in (0,1)$ implies $\hat{F}^{-1}_{n,0}(u) \geq \inf\{t \in [0,1] : t + \epsilon \geq u\} = \max(0, u - \epsilon) \geq u - \epsilon$ for all $u \in (0,1)$, we get for each $u \in (0,1)$ that

$$\forall\, \epsilon > 0 :\ \forall\, \delta > 0 :\ \exists\, N_{\epsilon,\delta} \in \mathbb{N} :\ \forall\, n \geq N_{\epsilon,\delta} :\ \mathbb{P}_\vartheta\left(\hat{F}^{-1}_{n,0}(u) \geq u - \epsilon\right) > 1 - \delta.$$

Then

$$\begin{aligned} 1 - \delta &< \mathbb{P}_\vartheta\left(\hat{F}^{-1}_{n,0}(\eta) \geq \eta - \epsilon\right) \\ &= \mathbb{P}_\vartheta\left(\frac{1-\eta}{1 - \hat{F}^{-1}_{n,0}(\eta)} \geq 1 - \frac{\epsilon}{1 - \eta + \epsilon}\right). \end{aligned}$$

Since $\hat{F}^{-1}_{n,0}$ is non-decreasing and left-side continuous, $\lim_{n\to\infty} p_{s:n_0} = \lim_{n\to\infty} \hat{F}^{-1}_{n,0}(s/n_0) \geq F^{-1}_{n,0}(\eta)$. Then for a fixed $\kappa \in \mathbb{R}$ we obtain

$$\forall\, \epsilon_1 > 0 :\ \exists\, N_{\epsilon_1} \in \mathbb{N} :\ \forall\, n \geq N_{\epsilon_1} :\ \frac{1 - s/n_0 + \kappa/n_0}{1 - p^0_{s:n_0}} + \epsilon_1 \geq \frac{1-\eta}{1 - F^{-1}_{n,0}(\eta)}.$$

Hence, for all $n \geq \max(N_{\epsilon,\delta}, N_{\epsilon_1})$ we get

$$\begin{aligned} 1 - \delta &< \mathbb{P}_\vartheta\left(\frac{1-\eta}{1 - \hat{F}^{-1}_{n,0}(\eta)} \geq 1 - \frac{\epsilon}{1 - \eta + \epsilon}\right) \\ &\leq \mathbb{P}_\vartheta\left(\frac{1 - s/n_0 + \kappa/n_0}{1 - p^0_{s:n_0}} \geq 1 - \frac{\epsilon}{1 - \eta + \epsilon} - \epsilon_1\right) \\ &\leq \mathbb{P}_\vartheta\left(\frac{\hat{n}_0}{n_0} \geq 1 - \frac{\epsilon}{1 - \eta + \epsilon} - \epsilon_1\right). \end{aligned}$$

For each $\epsilon_2 > 0$ there exist $\epsilon > 0$ and $\epsilon_1 > 0$ such that $\epsilon_2 > \epsilon/(1 - \eta + \epsilon) - \epsilon_1$. Setting $N_{\epsilon_2,\delta} = \max(N_{\epsilon,\delta}, N_{\epsilon_1})$, we get

$$\forall\, \epsilon_2 > 0 :\ \forall\, \delta > 0 :\ \exists\, N_{\epsilon_2,\delta} \in \mathbb{N} :\ \forall\, n \geq N_{\epsilon_2,\delta} :\ \mathbb{P}_\vartheta\left(\frac{\hat{n}_0}{n_0} > 1 - \epsilon_2\right) > 1 - \delta.$$

This immediately yields (4.6). ■

Now we give a specific and maybe somewhat surprising example where condition (WD2) holds for $\lambda = 0.5$. Some simulations for this example can be found in Example 4.32 in Section 4.7.

Example 4.4
Let $X_i \sim N(\vartheta_i, \sigma^2)$, $i \in I_n$, be independent normal random variables and let $\nu S^2/\sigma^2 \sim \chi^2_\nu$ be independent of the X_i's. Consider the multiple-testing problem $H_i : \vartheta_i = 0$ versus $K_i : \vartheta_i > 0$, $i \in I_n$, with test statistics $T_i = X_i/S$, $i \in I_n$. Then $T = (T_1, \ldots, T_n)$ has a multivariate equicorrelated t-distribution. Denote the cdf of a univariate (central) t-distribution with ν degrees of freedom by F_{t_ν} and define p-values corresponding to T_i by $p_i = 1 - F_{t_\nu}(x_i/s)$. This model was studied extensively in Finner *et al.* [2007]; see Example 2.2 and Section 4 in Finner *et al.* [2007]. Among others, it follows from the derivations in Finner *et al.* [2007] that the ecdf $\hat{F}_{n,0}$ of the p-values under null hypotheses satisfies $\lim_{n\to\infty} \hat{F}_{n,0}(0.5) = 0.5$ almost surely. Hence, Lemma 4.3 applies for $\lambda = 0.5$. We note that $\hat{F}_{n,0}(x)$ does not converge for any $x \in (0,1)$, $x \neq 0.5$.

4.3 SUD tests and asymptotic FDR control under weak dependence

In Chapter 3 we introduced several multiple test procedures controlling the FDR under independence assumptions (I1) and (I2). In this section we consider various SUD tests for "weak dependent" p-values which control the FDR at least asymptotically. Unfortunately, if the asymptotic crossing point determined by a multiple test φ_n tends to 0, there is neither a positive nor a negative result concerning asymptotic FDR control. Therefore, we formulate results with respect to further restrictions on the parameter space Θ guaranteeing an asymptotic threshold larger than 0. Depending on the applied multiple test procedure we discuss different restrictions on Θ.

Remember that a multiple test procedure φ controls the FDR at level α with respect to Θ if $\sup_{\vartheta\in\Theta} \mathrm{FDR}_\vartheta(\varphi) \leq \alpha$, where $\mathrm{FDR}_\vartheta(\varphi) = \mathbb{E}_\vartheta[V_n/R_n \vee 1]$ denotes the actual FDR given $\vartheta \in \Theta$. We say that the **FDR is asymptotically controlled** at level α if

$$\forall\, \vartheta \in \Theta : \limsup_{n\to\infty} \mathrm{FDR}_\vartheta(\varphi) \leq \alpha.$$

Note that for $\vartheta \in \Theta$ with $n_0(\vartheta) = n$ we have $\mathrm{FDR}_\vartheta(\varphi) = \mathrm{FWER}_\vartheta(\varphi)$.

It is tempting to suggest that procedures with asymptotic FDR control if the p-values p_i, $i \in I_{n,0}$, are independent, also control the FDR under weak dependence. We consider two possible classes of multiple test procedures, for which weak dependence may allow asymptotic FDR control.

(i) Let φ_n, $n \in \mathbb{N}$, be SUD(λ_n) tests with $\lambda_n \in I_n$ based on some rejection curve $r : [0,b] \to [0,1]$ with $b \in (0,1]$ such that $r(t)/t$ is non-increasing in $t \in (0,b]$. Moreover, (a) for $b < 1$ we assume the existence of a unique crossing point $t_\zeta \in (0,1]$ with $r(t_\zeta) = F_\infty(t_\zeta|\zeta)$ for each $\zeta \in [0,1)$, where $F_\infty(t|\zeta) = 1 - \zeta + \zeta t$ is the limiting ecdf of p-values in DU(n, n_0) models with $n_0(n)/n \to \zeta$; or (b) if $b = 1$ let $\limsup_{n\to\infty} \lambda_n/n < 1$ and suppose that there exists a $\zeta_0 \in [0,1)$

such that for each $\zeta \in (\zeta_0, 1]$ there exists a unique crossing point t_ζ between r and $F_\infty(\cdot|\zeta)$ on $[0, 1)$ while the unique crossing point t_ζ on $[0, 1]$ is 1 for $\zeta \in [0, \zeta_0]$.

(ii) Let φ_n, $n \in \mathbb{N}$, be plug-in LSU tests introduced in Storey *et al.* [2004], that is, LSU tests based on a random rejection curve $r(t) = r(t|\hat{\zeta}_n(\lambda)) = \hat{\zeta}_n(\lambda)t/\alpha$ and $\hat{\zeta}_n(\lambda) = (1 - \hat{F}_n(\lambda))/(1 - \lambda)$ for a fixed $\lambda \in (0, 1)$. Thereby, a plug-in LSU test rejects all hypotheses if $\hat{\zeta}_n(\lambda) \leq \alpha$.

Note that LSU tests (cf. Section 1.3) and all SUD procedures based on the AORC considered in Chapter 3 (cf. Section 3.4) belong to (i). For example, φ_n may be SUD(λ_n) tests with critical values given in (3.28) (or (3.29)) for a fixed $\beta_n = \beta > 0$ (or $\beta_n^* = \beta > 0$, resp.), or φ_n may be SUD(λ_n) tests with $\limsup_{n\to\infty} \lambda_n/n < 1$ based on the AORC (i.e.critical values are given in (3.1)) or based on the rejection curve r given in Example 3.13 in Section 3.3. Thereby, these tests control the FDR at least asymptotically if p-values corresponding to true null hypotheses are independent.

The next theorem shows that tests from both classes (i) and (ii) asymptotically control the FDR under a suitable condition on a subset of Θ^* for "weak dependent" test statistics.

Theorem 4.5
Suppose (WD2) is fulfilled and let $\vartheta \in \Theta^*$ such that $n_0/n \to \zeta \in [0, 1)$. Consider a sequence of multiple test procedures φ_n, where either all φ_n, $n \in \mathbb{N}$, correspond to (i) or to (ii). If at least one of the conditions

$$\mathbb{P}_\vartheta \left(\liminf_{n\to\infty} \frac{R_n}{n} > 0 \right) = 1, \tag{4.9}$$

$$\exists\, \gamma > 0 : \lim_{n\to\infty} \mathbb{P}_\vartheta \left(\frac{R_n}{n} > \gamma \right) = 1 \tag{4.10}$$

holds , then

$$\limsup_{n\to\infty} \mathit{FDR}_\vartheta(\varphi_n) \leq \lim_{n\to\infty} \mathit{FDR}_{n,n_0}(\varphi_n), \tag{4.11}$$

where $\mathit{FDR}_{n,n_0}(\varphi_n)$ is the FDR of φ_n in a DU(n, n_0) model. Hence, asymptotic FDR control in DU models implies asymptotic FDR control under ϑ.

Proof: First, we consider φ_n given in (ii) in case $\hat{\zeta}_n(\lambda) \geq \alpha$ and φ_n given in (i). Define $B_{\gamma,n} = \{R_n/n \geq \gamma\}$ and $C_{\delta,n} = \{\sup_{t\in[0,1]}(\hat{F}_{n,0}(t) - t) \leq \delta\}$ for $n \in \mathbb{N}$, $\gamma > 0$ and $\delta > 0$. Condition (4.9) and (4.10) imply

$$\forall\, \epsilon_1 > 0 \ : \ \exists\, \gamma > 0 \ : \ \exists\, N_1 = N_1(\epsilon_1) \in \mathbb{N} \ : \ \forall\, n \geq N_1 \ : \ \mathbb{P}_\vartheta(B_{\gamma,n}) > 1 - \epsilon_1, \tag{4.12}$$

where $\gamma = \gamma(\epsilon_1)$ if (4.9) is fulfilled, and γ is fixed and given in (4.10) if (4.10) applies. The assumption (WD2) together with Lemma A.7 imply that

$$\forall\, \epsilon_2 > 0 \ : \ \forall\, \delta > 0 \ : \ \exists\, N_2 = N_2(\epsilon_2, \delta) \in \mathbb{N} \ : \ \forall\, n \geq N_2 \ : \ \mathbb{P}_\vartheta(C_{\delta,n}) > 1 - \epsilon_2.$$

Hence, for arbitrary but fixed $\epsilon > 0$ and $\delta > 0$ there exists a $\gamma > 0$ and an $N = N(\epsilon, \delta) \in \mathbb{N}$ such that for all $n \geq N$ we obtain

$$\mathrm{FDR}_\vartheta(\varphi_n) \leq \int\limits_{B_{\gamma,n} \cap C_{\delta,n}} \frac{V_n}{R_n} \, d\mathbb{P}_\vartheta + \epsilon.$$

Let $t_n \in [0, 1]$ denote the random crossing point between r and the ecdf of p-values $\hat{F}_n$ determined by φ_n, i.e. $r(t_n) = \hat{F}_n(t_n) = R_n/n$ and φ_n rejects H_i, $i \in I_n$, if and only if $p_i \leq t_n$. Then $V_n = n_0 \hat{F}_{n,0}(t_n)$ and $R_n = nr(t_n)$. Note that the latter equality does not apply for φ_n defined in (ii) in case $\hat{\zeta}_n(\lambda) < \alpha$. Since $\hat{F}_{n,0}(t) \leq t + \delta$ in $C_{\delta,n}$, $r(t_n) \geq \gamma$ in B_γ and r is increasing, it follows for all $n \geq N$ that

$$\mathrm{FDR}_\vartheta(\varphi_n) \leq \frac{n_0}{n} \int\limits_{B_{\gamma,n} \cap C_{\delta,n}} \frac{t_n}{r(t_n)} \, d\mathbb{P}_\vartheta + \epsilon + \frac{\delta}{\gamma}. \tag{4.13}$$

Now we are looking for a non-random upper bound for the function $t_n/r(t_n)$. In case (i) we obtain $F_\infty(t|\zeta) < r(t)$ for all $t \in (t_\zeta, b] \cap (0, 1)$. On the other hand it holds $\hat{F}_n(t) \leq F_\infty(t|n_0/n) + n_0\delta/n$ in $C_{\delta,n}$ for all $t \in [0, 1]$. Altogether we get for a $\beta > 0$ and sufficiently small δ-values that $F_\infty(t|n_0/n) + n_0\delta/n < r(t)$ for $t \in (t_\zeta + O(\delta), b] \cap (0, 1 - \beta)$, hence $t_n \leq t_\zeta + O(\delta)$ in $C_{\delta,n}$. By noting that $t/r(t)$ is non-decreasing we obtain $t_n/r(t_n) \leq t_\zeta/r(t_\zeta) + O(\delta)$ in $C_{\delta,n}$. Remark 3.15 implies $\zeta t_\zeta/r(t_\zeta) = g(\zeta)$, where $g(\zeta)$ is the corresponding asymptotic FDR bounding curve. Moreover, Theorem 3.14 yields $g(\zeta) = \lim_{n\to\infty} \mathrm{FDR}_{n,n_0}(\varphi_n)$. Then for φ_n as described in (i) we obtain

$$\frac{t_n}{r(t_n)} \leq \frac{1}{\zeta} \lim_{n\to\infty} \mathrm{FDR}_{n,n_0}(\varphi_n) + O(\delta) \ \text{ in } \ C_{\delta,n}. \tag{4.14}$$

In case (ii) we get $t/r(t) = t/r(t|\hat{\zeta}_n(\lambda)) = \alpha/\hat{\zeta}_n(\lambda) = \alpha(1-\lambda)/(1-\hat{F}_n(\lambda))$ for all $t > 0$ and $t_n/r(t_n) \leq 1$ since $\hat{\zeta}_n(\lambda) \geq \alpha$. Obviously, $1 - \hat{F}_n(\lambda) \geq n_0/n(1 - \hat{F}_{n,0}(\lambda))$ and $\hat{F}_{n,0}(\lambda) \leq \lambda + O(\delta)$ in $C_{\delta,n}$. Then $t_n/r(t_n) \leq \min(\alpha n/n_0, 1) + O(\delta)$ in $C_{\delta,n}$. By noting that $n_0/n = \zeta + o(1)$ we obtain $t_n/r(t_n) \leq \min(\alpha/\zeta, 1) + O(\delta)$ in $C_{\delta,n}$. Since $\hat{\zeta}_n(\lambda) \to \zeta$ for $n \to \infty$ in $\mathrm{DU}(n, n_0)$ models, we get $\lim_{n\to\infty} \mathrm{FDR}_{n,n_0}(\varphi_n) = \min(\alpha, \zeta)$, that is, also for φ_n described in (ii) the inequality (4.14) is fulfilled if $\hat{\zeta}_n(\lambda) \geq \alpha$.

Then (4.13) and (4.14) imply for (i) and for (ii) (with $\hat{\zeta}_n(\lambda) \geq \alpha$) that

$$\mathrm{FDR}_\vartheta(\varphi_n) \leq \frac{n_0}{n} \mathbb{P}_\vartheta(B_{\gamma,n} \cap C_{\delta,n}) \frac{1}{\zeta} \lim_{n\to\infty} \mathrm{FDR}_{n,n_0}(\varphi_n) + \epsilon + \frac{\delta}{\gamma} + O(\delta).$$

Hence,

$$\limsup_{n\to\infty} \mathrm{FDR}_\vartheta(\varphi_n) \leq \lim_{n\to\infty} \mathrm{FDR}_{n,n_0}(\varphi_n) + \epsilon + \frac{\delta}{\gamma} + O(\delta).$$

Now (4.11) follows by letting $\epsilon, \delta \to 0$ and choosing $\delta = o(\gamma)$.

Finally we consider φ_n given in (ii) in case $\hat{\zeta}_n(\lambda) \leq \alpha$. This immediately implies $t_n = 1$, i.e. all hypotheses are rejected and $\mathrm{FDR}_\vartheta = \zeta$. Similar as before we get $\hat{\zeta}_n(\lambda) \geq \zeta + O(\delta)$ in $C_{\delta,n}$ for all $n \geq N$, which implies $\zeta \leq \alpha$. Hence, $\lim_{n\to\infty} \mathrm{FDR}_{n,n_0}(\varphi_n) = \zeta$ and consequently (4.11) is fulfilled. ∎

Remark 4.6

Let φ_n be a sequence of multiple tests as described in Theorem 4.5. A parameter $\vartheta \in \Theta^*$ such that (4.4) is fulfilled and p-values corresponding to alternatives follow a Dirac distribution with point mass 1 in 0 is an asymptotic LFC for the FDR if all assumptions of Theorem 4.5 apply. Moreover, for this ϑ equality holds in condition (4.11), that is, $\lim_{n\to\infty} \mathrm{FDR}_\vartheta(\varphi_n) = \lim_{n\to\infty} \mathrm{FDR}_{n,n_0}(\varphi_n)$.

Remark 4.7

Conditions (4.9) and (4.10) in Theorem 4.5 do not imply each other. Moreover, both can be replaced by the weaker condition (4.12). Another possibility to define a sufficient condition for asymptotic FDR control is to rewrite (4.9) and/or (4.10) in terms of the crossing point t_n determined by φ_n, that is,

$$\mathbb{P}_\vartheta\left(\liminf_{n\to\infty} t_n > 0\right) = 1 \text{ and/or } \exists\, t^* > 0 : \lim_{n\to\infty} \mathbb{P}_\vartheta\left(t_n > t^*\right) = 1,$$

respectively. Note that results in Storey *et al.* [2004] based on a condition ensuring that the asymptotic threshold is larger than 0, cf. Theorem 4 in that paper. In case of SUD(λ_n) procedures with $\liminf_{n\to\infty} \lambda_n/n = \kappa > 0$ condition (4.9) and/or (4.10) also can be replaced by power requirements like

$$\mathbb{P}_\vartheta\left(\exists\, t \in (0,\kappa) : \liminf_{n\to\infty} \hat{F}_n(t) > r(t)\right) = 1 \tag{4.15}$$

and/or

$$\exists\, t \in (0,\kappa) : \lim_{n\to\infty} \mathbb{P}_\vartheta\left(\hat{F}_n(t) > r(t)\right) = 1. \tag{4.16}$$

For SUD(λ_n) procedures with $\liminf_{n\to\infty} \lambda_n/n = 0$ (SD tests belong to this class of tests) conditions given in (4.15) and (4.16) seem to be insufficient, because the asymptotic behaviour of the ecdf $\hat{F}_n$ does not characterise the behaviour of smallest p-values. Even if $\hat{F}_n(t)$ converges to some cdf $F(t)$ almost surely for all $t \in (0,1)$, the smallest ordered p-values may be stochastically considerably larger/smaller than the corresponding ordered statistics coming from n iid random variables with the marginal cdf F. Smallest p-values are typically irrelevant for SUD tests with $\liminf_{n\to\infty} \lambda_n/n > 0$ but they may be crucial for SUD tests with $\liminf_{n\to\infty} \lambda_n/n = 0$. Unfortunately, even in the case of SU test procedures it is not clear how (4.9), (4.10) or (4.12) can be proved for SUD(λ_n) procedures with $\liminf_{n\to\infty} \lambda_n/n = 0$.

Remark 4.8

If condition (WD2) is fulfilled and smallest ordered p-values with respect to true nulls are stochastically larger than the corresponding ordered p-values coming from iid uniform p-values, then the FDR may be controlled even if $\lim_{n\to\infty} t_n = 0$. However, it seems there are no arguments for asymptotic FDR control if the asymptotic crossing point is not bounded away from 0. This is still an open problem. In any case, if the asymptotic threshold of multiple tests φ_n is 0, φ_n may have only a negligible gain in power compared to a procedure that controls the FWER.

For SUD(λ_n) procedures with $\lim_{n\to\infty} \lambda_n/n = \kappa > 0$, sufficient conditions for (4.9) and/or (4.10) can be given as follows. Let $\hat{F}_{n,1}$ denote the ecdf of p-values corresponding to false hypotheses and let $\lim_{n\to\infty} n_0/n = \zeta < 1$. Then the conditions

$$\mathbb{P}_\vartheta \left(\exists\, t \in (0,\kappa) : \liminf_{n\to\infty} \hat{F}_{n,1}(t) > \frac{r(t)}{1-\zeta} \right) = 1 \tag{4.17}$$

and

$$\exists\, t \in (0,\kappa) : \lim_{n\to\infty} \mathbb{P}_\vartheta \left(\hat{F}_{n,1}(t) > \frac{r(t)}{1-\zeta} \right) = 1 \tag{4.18}$$

imply (4.15) and (4.16), respectively.

If p-values corresponding to alternatives, i.e. p_i, $i \in I_{n,1}$, are independent, then the extended Glivenko-Cantelli Theorem (cf. Shorack and Wellner [1986], p. 105) holds, that is,

$$\sup_{t\in[0,1]} |\hat{F}_{n,1}(t) - \frac{1}{n-n_0} \sum_{i\in I_{n,1}} G_{i,1}(t)| \to 0 \text{ almost surely,} \tag{4.19}$$

where $G_{i,1}$ denotes the cdf of a p_i, $i \in I_{n,1}$. Therefore, for independent p-values under alternatives, (4.17) and (4.18) are fulfilled if and only if

$$\mathbb{P}_\vartheta \left(\exists\, t \in (0,\kappa) : \liminf_{n\to\infty} \frac{1}{n-n_0} \sum_{i\in I_{n,1}} G_{i,1}(t) > \frac{r(t)}{1-\zeta} \right) = 1 \tag{4.20}$$

and

$$\exists\, t \in (0,\kappa) : \lim_{n\to\infty} \mathbb{P}_\vartheta \left(\frac{1}{n-n_0} \sum_{i\in I_{n,1}} G_{i,1}(t) > \frac{r(t)}{1-\zeta} \right) = 1, \tag{4.21}$$

respectively. Since the cdfs $G_{i,1}$, $i \in I_{n,1}$, are typically unknown, it is hard to envisage how to verify these conditions in practice. Nevertheless, we give an example where (4.20) and (4.21) hold, which implies that the lower bound of the asymptotic threshold t_n determined by φ_n in Theorem 4.5 is larger than 0. Moreover, in Section 4.4 we provide a condition under which the ecdf $\hat{F}_{n,1}$ convergences in probability in the sense of the Glivenko-Cantelli Theorem. In Sections 4.5 and 4.6 we discuss convergence of the ecdf $\hat{F}_{n,1}$ for block-dependent p-values and p-values corresponding to pairwise comparisons, respectively.

Example 4.9

Let $X_{ij} \sim N(\vartheta_i, \sigma^2)$, $i \in I_n$, $j \in I_m$, be normally distributed random variables with unknown mean ϑ_i and known variance $\sigma^2 > 0$. Suppose

$$\exists\, \vartheta > 0 : \exists\, \eta > 0 : \liminf_{n\to\infty} \frac{\#\{i \in I_{n,1} : \vartheta_i \geq \vartheta\}}{n-n_0} = \eta. \tag{4.22}$$

We consider the multiple-testing problem $H_i : \vartheta_i = 0$ versus $K_i : \vartheta_i > 0$, $i \in I_n$, with test statistics $T_i = \sum_{j=1}^m X_{ij}/(\sigma\sqrt{m})$, $i \in I_n$, and p-values $p_i = p_i(t_i) = 1 - \Phi(t_i)$, $i \in I_n$, where t_i is a realisation of T_i and Φ is the standard Gaussian cdf with density ϕ. For given ϑ_i, σ and m, the distribution function of p_i is $G_i(t) = 1 - \Phi(U_{1-t} - \sqrt{m}\vartheta_i/\sigma)$ and the density of p_i is $g_i(t) = \phi(U_{1-t} - \sqrt{m}\vartheta_i/\sigma)/\phi(U_{1-t})$, where U_t is the tth percentile of the standard Gaussian distribution and $0 < t < 1$ (similar investigations can be found in Hung *et al.* [1997]). Note that G_i has an infinite right-hand derivative in 0 if the corresponding $\vartheta_i > 0$. By setting $G_\vartheta(t) = 1 - \Phi(Z_t - \sqrt{m}\vartheta/\sigma)$ we obtain from (4.22)

$$\liminf_{n\to\infty} \frac{1}{n-n_0} \sum_{i\in I_{n,1}} G_i(t) \geq \eta G_\vartheta(t) \text{ for all } t \in [0,1]. \tag{4.23}$$

Obviously, for any $\gamma \in (0,1]$ there exists some $t' \in (0,1]$ such that $G_\vartheta(t) > t/\gamma$ for all $t \in (0,t')$. Therefore, if (4.19) holds for $\hat{F}_{n,1}$ we obtain (4.20) and (4.21) for each rejection curve r of the type (i) if $r'(0) < \infty$. Hence, conditions (4.9) and (4.10) follow for SUD(λ_n) procedures of the type (i) with $\liminf_{n\to\infty} \lambda_n/n > 0$, and Theorem 4.5 applies if (4.22) is fulfilled.

4.4 Sufficient conditions for convergence of ecdfs

An interesting question is how weak dependence conditions and/or convergence of an ecdf of p-values can be proved. As seen before, the FWER and/or FDR of multiple tests typically become larger if p-values corresponding to true null hypotheses become stochastically smaller. Therefore, the case when $\hat{F}_{n,0}$ converges to an identity function $F = Id$ in some sense is most interesting with respect to FWER and/or FDR control. Below, we sometimes restrict attention to the case that p_i, $i \in I_{n,0}$, are uniformly distributed in $[0,1]$, i.e. condition (D1) is fulfilled.

The following quadratic mean law of large numbers for dependent random variables given in Parzen [1960], p. 419, is a useful tool for proving convergence of an ecdf $\hat{F}_{n,i}$, $i = 1, 2$, and may lead to (WD3), (4.9) and/or (4.10), which imply asymptotic FWER and/or FDR control.

Theorem 4.10 (Parzen [1960])
Let $\{X_n\}_{n\in\mathbb{N}}$ be a sequence of real valued random variables with mean 0 and uniformly bounded variances and let $Z_n = \sum_{i=1}^n X_i/n$. Then $Z_n \to 0$ for $n \to \infty$ in the $\mathbb{L}_2$ norm if and only if $\mathbb{E}[X_n Z_n] \to 0$ for $n \to \infty$.

Note that for bounded random variables, convergence in the $\mathbb{L}_2$ norm is equivalent to convergence in the $\mathbb{L}_1$ norm, which is equivalent to convergence in probability. Moreover, if it is known that $\mathbb{E}[X_n Z_n] = O(n^{-q})$ for some $q > 0$, then we can conclude that $Z_n \to 0$, $n \to \infty$, with probability 1, cf. Parzen [1960], p. 420. This result is given in the next theorem.

Theorem 4.11 (Parzen [1960])
Under the assumptions of Theorem 4.10 the sequence $\{X_n\}_{n\in\mathbb{N}}$ obeys the SLLN (in the sense that $\mathbb{P}(\lim_{n\to\infty} Z_n = 0) = 1$) if there exist $M > 0$ and $q > 0$ such that $|\mathbb{E}[X_n Z_n]| \leq M/n^q$ for all $n \in \mathbb{N}$.

Now we apply these two theorems in order to obtain sufficient conditions for the convergence of the ecdf of p-values corresponding to true null hypotheses. Let $t \in (0,1)$ and $G_{i,0}$ be the cdf of p_i, $i \in I_{n,0}$ such that $0 \leq G_{i,0}(t) \leq t$ for $t \in (0,1)$. Setting $X_i = I(p_i \leq t) - G_{i,0}(t)$,

$i \in I_{n,0}$, for $t \in (0,1)$, we get $\mathbb{E}X_i = 0$, $\text{Var}(X_i) = G_{i,0}(t)(1 - G_{i,0}(t)) \leq 1$ and $Z_n = \hat{F}_{n,0}(t) - (1/n_0)\sum_{i \in I_{n,0}} G_{i,0}(t)$. Then Theorem 4.10 implies that

$$|\hat{F}_{n,0}(t) - \frac{1}{n_0} \sum_{i \in I_{n,0}} G_{i,0}(t)| \to 0,\ n \to \infty, \tag{4.24}$$

in probability (and hence (WD3) is fulfilled) if and only if

$$\frac{1}{n_0} \sum_{i=1}^{n_0} \text{Cov}\left(I(p_i^0 \leq t), I(p_{n_0}^0 \leq t)\right) \to 0 \ \text{ for } \ n \to \infty, \tag{4.25}$$

where $p_1^0, \ldots, p_{n_0}^0$ are p-values corresponding to true null hypotheses and $n_0 = n_0(n) \to \infty$ for $n \to \infty$. Note that condition (4.25) is equivalent to $\text{Var}(\hat{F}_{n,0}(t)) \to 0$, $n_0 \to \infty$. Theorem 4.11 implies almost sure convergence in (4.24) if

$$\frac{1}{n_0} | \sum_{i=1}^{n_0} \text{Cov}\left(I(p_i^0 \leq t), I(p_{n_0}^0 \leq t)\right) | \leq O\left(\frac{1}{n_0^q}\right) \ \text{ for some } \ q > 0. \tag{4.26}$$

If for each $n_0 \in \mathbb{N}$ the proportion of p_i^0, $i \in I_{n_0}$, for which $\text{Cov}(I(p_i^0 \leq t), I(p_{n_0}^0 \leq t)) \neq 0$, converges to 0 for $n \to \infty$, then we get convergence in probability in (4.24) and consequently (WD3) follows. This may happen in the case of block-dependent p-values, which will be investigated in Section 4.5 or in case of p-values corresponding to pairwise comparisons problems, see Section 4.6. Moreover, condition (4.25) allows that all p-values corresponding to true null hypotheses may be mutually dependent, but the corresponding covariances have to be small enough. Autocorrelated test statistics yield such kind of p-values.

Example 4.12

Let $X = (X_1, \ldots, X_n)$ be a multivariate normally distributed random variable with $\mathbb{E}_\vartheta X_i = \vartheta_i = 0$ and a covariance matrix $\Sigma = (\sigma_{ij})_{i,j \in I_n}$ such that $\sigma_{ij} = \rho^{|i-j|}$ for some $\rho \in (0,1)$. Then

$$\frac{1}{n} \sum_{i=1}^{n} \text{Cov}\left(X_i, X_n\right) = \frac{1}{n} \sum_{i=1}^{n} \rho^{n-i} = \frac{1 - \rho^{n-1}}{n(1-\rho)} \to 0,\ n \to \infty.$$

For testing $H_i : \vartheta_i = 0$ versus $K_i : \vartheta_i \neq 0$, $i \in I_n$, with test statistics given by $X_1, \ldots, X_n$ the latter condition yields (4.25) (and under specific conditions on the marginal cdf of p-values is even equivalent to (4.25)) if each p-value $p_i = p(x_i)$, $i \in I_{n,0}$, depends only on the realisation of X_i, cf. Giraitis and Surgailis [1985], Lemma 5 in that paper.

For a set of test statistics (or p-values) with a sparse covariance matrix, i.e. a covariance matrix populated primarily with zeros, condition (4.25) is satisfied. A further class of p-values fulfilling (4.25) is given by sets of various mixing random variables like φ-, ψ- or α-mixing. Mixing random variables often ensures that the SLLN holds, which then leads to almost sure convergence in (4.24). For more information see, for example, Tuyen [1999] or Billingsley [1968], p. 166.

Typically, if we have a set of p-values fulfilling a Central Limit Theorem (CLT) and/or the SLLN it seems easy to prove asymptotic FWER and/or FDR control. For example, Farcomeni

[2006] proved that an oracle LSU test, i.e. an SU test with critical values $\alpha_{i:n} = \alpha i/n_0$, $i \in I_n$, and a plug-in LSU test asymptotically control the FDR for stationary (and probably associated) p-values fulfilling different conditions on mixing coefficients of these p-values. The author investigated the distribution of FDP $= V_n/R_n$ for dependent and independent p-values and proved that the FDP fulfils a CLT. Although there are no comments about the asymptotic threshold of an oracle LSU procedure, it seems that Farcomeni [2006] proved asymptotic FDR control assuming that the asymptotic threshold is bounded away from 0, cf. proof of Theorem 1 in that paper, p.296.

Now we consider the ecdf $\hat{F}_{n,1}$ of p-values under alternatives.

Remark 4.13
Let $n_1 = n_1(\vartheta) = n - n_0(\vartheta)$ and p_i^1, $i \in I_{n_1}$, be p-values corresponding to alternatives, i.e. p_i, $i \in I_{n,1}$. Let $G_{i,1}$ denote the cdf of p_i^1, $i \in I_{n,1}$,. If conditions (4.25) and/or (4.26) apply for p-values corresponding to false hypotheses (i.e. replace n_0 by n_1 and p_i^0 by p_i^1 in (4.25) and/or (4.26)), then the ecdf $\hat{F}_{n,1}$ of p-values under alternatives converges in the sense of the extended Glivenko-Cantelli Theorem (cf. Shorack and Wellner [1986], p. 105), that is,

$$\sup_{t\in[0,1]} |\hat{F}_{n,1}(t) - \frac{1}{n_1}\sum_{i\in I_{n,1}} G_{i,1}(t)| \to 0,\ n_1 \to \infty, \tag{4.27}$$

in probability and/or almost surely, respectively.

Convergence given in (4.24) can be sometimes proved by means of U-statistics. U-statistics were introduced in Hoeffding [1948] as follows.

Definition 4.14
Let $X_1, \dots, X_n$ be n independent random variables with values in a measurable space $(\mathcal{X}, \mathcal{A})$ and let $\psi : \mathcal{X}^m \to \mathbb{R}$ be a symmetric function of $m(\le n)$ arguments, i.e. $\psi(x_1, \dots, x_m) = \psi(x_{i_1}, \dots, x_{i_m})$ for all $(i_1, \dots, i_m) \in \Im_m$, where $\Im_m$ denotes the group of $m!$ permutations of $\{1, \dots, m\}$. A statistic of the form

$$U_n = \binom{n}{m}^{-1} \sum_{1\le i_1\le \dots \le i_m \le n} \psi(X_{i_1}, \dots, X_{i_m})$$

is called an U-statistic with kernel ψ of order m.

If $X_1, \dots, X_n$ have the same (cumulative) distribution function F, the corresponding U-statistic U_n is an unbiased estimate of $\vartheta(F)$ defined by

$$\vartheta(F) = \mathbb{E}_F(\psi(X_1, \dots, X_m)) = \int \dots \int \psi(x_1, \dots, x_m) dF(x_1) \dots dF(x_m).$$

An important result for U-statistics is the following SLLN introduced in Berk [1966].

Theorem 4.15 (SLLN of U-statistics, Berk [1966])
Let $X_1, \dots, X_n$ be independent random variables on a measurable space $(\mathcal{X}, \mathcal{A})$ with the same distribution function F. Let $\psi : \mathcal{X}^m \to \mathbb{R}$ be a symmetric function of $m(\le n)$ arguments and

$\psi \in \mathbb{L}_1(F^m)$, i.e. $\vartheta(F) = \mathbb{E}_F\psi(X_1,\dots,X_m) < \infty$. Then $U_n \to \vartheta(F)$ almost surely for $n \to \infty$.

For special pairwise comparisons problems we obtain under suitable assumptions on the distribution of test statistics that the ecdf $\hat{F}_{n,0}(t)$ of p-values corresponding to true null hypotheses can be represented as a sum of U-statistics with kernels of the form $\psi(X_i, X_j) = I(p_{ij} \le t)$, where $p_{ij} = p_{ij}(X_i, X_j)$ denotes a p-value for testing H_{ij}, $i \neq j$. Moreover, if the global null hypothesis $H_0 = \cap_{i=1}^n H_i$ is true the ecdf $\hat{F}_{n,0}$ may be a U-statistic. The SLLN of U-statistics implies almost sure convergence in (4.24) if the ecdf $\hat{F}_{n,0}(t)$ is a sum of a finite number of U-statistics and the number of independent random variables corresponding to at least one U-statistic increases. An application of U-statistics will be considered in Section 4.6.

4.5 Block-dependent p-values

In this section we present an important type of dependence of p-values such that the weak dependence condition (WD2) is fulfilled. This allows asymptotic FWER and/or FDR control by a multiple test, cf. Theorem 4.2 and Theorem 4.5. First, we consider a motivating example. Duncan [2004] wrote

> *" It is possible to identify the approximate chromosomal location of major gene as a result of the phenomenon of recombination. The first principle is that genes on different chromosomes segregate independently, so there can be no linkage between them. The second principle is that the probability of recombination between two loci on the same chromosome increases with the physical distance between them, eventually reaching the limiting value of $1/2$, the same probability as for two separate chromosomes. Thus, if a genetic marker is found to have a low recombination rate with a disease gene, one can infer that the disease gene must be close to that markers. The basic idea is then to determine the genotypes of various marker (whose location are known) for various members multiple case families. "*

Genome-wide association studies sometimes involve testing hundreds of thousands of single-nucleotide polymorphisms (SNPs), cf. Finner *et al.* [2010]. One may divide the whole genome into small blocks; for instance, each block may include some hundreds of SNPs. It will be assumed that test statistics within each block may be dependent, and that test statistics from different blocks are independent, cf. Kang et al. [2009].

A formal description of block-dependence is given as follows. Let $k \in \mathbb{N}$ and let q_{ki}, $k \in \mathbb{N}$, $i \in I_k$, with $q_{ki} \in \mathbb{N}$ be a double array of indexes such that $q_{ki} \le q_{k+1i}$ for all $i \in I_k$ and $k \in \mathbb{N}$. Let

$$\underbrace{p_{11},\dots,p_{1q_{k1}}}_{1}, \underbrace{p_{21},\dots,p_{2q_{k2}}}_{2}, \dots, \underbrace{p_{k1},\dots,p_{kq_{kk}}}_{k}, \tag{4.28}$$

be a set of p-values such that for $i, j \in I_k,\ \ s \in I_{q_{ki}}$ and $t \in I_{q_{kj}}$ we get that p_{is} and p_{jt} are independent if $i \neq j$ and they may be dependent for $i = j$.

First, we will investigate whether condition (WD2) is fulfilled for block-dependent p-values described in (4.28). W.l.o.g. we suppose that all p-values correspond to true null hypotheses, that is, $n = n_0$. Then the ecdf of p-values corresponding to true null hypotheses is given by

$$\hat{F}_{n,0}(z) = \frac{1}{\sum_{i=1}^{k} q_{ki}} \sum_{i=1}^{k} \sum_{s=1}^{q_{ki}} I(p_{is} \leq z),$$

where $n_0 = n_0(k) = \sum_{i=1}^{k} q_{ki}$ is the number of all null p-values, i.e. the number of true null hypotheses. If the number k of blocks is fixed, this may lead to a violation of (WD2). For example, if $p_{is} = p_s$, $i \in I_{q_{ks}}$, $s \in I_k$, then $\hat{F}_{n,0}(z) = (1/\sum_{i=1}^{k} q_{ki}) \sum_{s=1}^{k} I(p_s \leq z)$ does not converge to z for a fixed k. Hence, we restrict our attention to the case $k \to \infty$.

The next theorem provides sufficient conditions on block sizes q_{ki}, $i \in I_k$, such that the set of p-values fulfils the weak dependence condition given in (WD3).

Theorem 4.16
If block sizes q_{ki}, $i \in I_k$, $k \in \mathbb{N}$, fulfil the condition

$$\frac{\max_{i \in I_k} q_{ki}}{\sum_{i=1}^{k} q_{ki}} \longrightarrow 0 \ \textit{for} \ n_0 \to \infty, \tag{4.29}$$

then we get convergence in probability in (4.24) *and consequently the ecdf $\hat{F}_{n,0}$ of p-values given in* (4.28) *fulfils (WD3). Moreover, if there exists some $q \in [0, 1)$ such that*

$$\max_{i \in I_k} q_{ki} \leq O\left(n_0^q\right), \tag{4.30}$$

then we obtain almost sure convergence in (4.24) *and hence in (WD3).*

Proof: Since condition (4.29) and/or (4.30) immediately imply conditions (4.25) and/or (4.26), Theorem 4.10 and/or Theorem 4.11 yield the assertions. ■

Remark 4.17
If p-values given in (4.28) correspond to false hypotheses (i.e. p_i, $i \in I_{n,1}$, are block-dependent), then conditions (4.29) and/or (4.30) yield convergence of the ecdf $\hat{F}_{n,1}$ of p-values under alternatives in the sense of the extended Glivenko-Cantelli Theorem, that is, (4.27) applies in probability and/or almost surely, respectively.

Condition (4.30) in Theorem 4.16 is equivalent to

$$\frac{\max_{i \in I_k} q_{ki}}{\sum_{i=1}^{k} q_{ki}} \leq O\left(\frac{1}{n_0^{\gamma}}\right) \quad \text{for some} \ \ \gamma > 0. \tag{4.31}$$

The next theorem gives a weaker condition on block sizes than condition (4.31), which also ensures the almost sure and even complete convergence in (4.24).

Theorem 4.18
Suppose there exists some $\gamma > 0$ such that

$$\frac{\max_{i \in I_k} q_{ki}}{\sum_{i=1}^{k} q_{ki}} \leq O\left(\frac{1}{k^{\gamma}}\right). \tag{4.32}$$

Then we obtain complete convergence in (4.24) *and hence (WD3) for* $k \to \infty$.

Proof: We will prove the assertion by applying Theorem A.8, that is, we have to check conditions (A.2)-(A.5). W.l.o.g. let p_i, $i \in I_{n,0}$, be uniformly distributed in $[0,1]$, that is, (D1) is fulfilled. Define new random variables by

$$X_{ki} = \sum_{s=1}^{q_{ki}} I(p_{is} \leq z) - zq_{ki},\ z \in [0,1],\ k \in \mathbb{N},\ i \in I_k.$$

It follows that X_{ki}, $i \in I_k$, are independent and $|X_{ki}| \leq q_{ki}$ for all $k \in \mathbb{N}$. Setting $a_k = \sum_{s=1}^{k} q_{ki}$, $k \in \mathbb{N}$, we get

$$\hat{F}_{n,0}(z) = \frac{1}{a_k}\sum_{i=1}^{k} X_{ki} + z.$$

Moreover, we define $\psi(t) = |t|^{p+1/s}$ for some $p \in \mathbb{N}$, $p \geq 2$ and $s \geq 1$. Then the condition (A.2) in Theorem A.8 is fulfilled. Since $\mathbb{E}X_{ki} = 0$, we get (A.3) in Theorem A.8. Now we prove (A.4). It holds

$$\begin{aligned}
\sum_{k=1}^{\infty}\sum_{i=1}^{k}\mathbb{E}\left(\frac{|X_{ki}|^{p+1/s}}{a_k^{p+1/s}}\right) &\leq \sum_{k=1}^{\infty}\sum_{i=1}^{k}\left(\frac{m_{ki}}{\sum_{s=1}^{k} q_{ki}}\right)^{p+1/s} \\
&\leq \sum_{k=1}^{\infty}\frac{(\max_{i\in I_k} q_{ki})^{p-1+1/s}\sum_{i=1}^{k} q_{ki}}{\left(\sum_{s=1}^{k} q_{ki}\right)^{p+1/s}} \\
&= \sum_{k=1}^{\infty}\left(\frac{\max_{i\in I_k} q_{ki}}{\sum_{s=1}^{k} q_{ki}}\right)^{p-1+1/s}.
\end{aligned}$$

Assumption (4.32) implies that the latter expression is finite if

$$\sum_{k=1}^{\infty}\frac{1}{k^{\gamma(p-1+1/s)}} < \infty. \tag{4.33}$$

Obviously, for each $\gamma > 0$ there exists a $p \in \mathbb{N}$, $p \geq 2$, such that $\gamma(p-1+1/s) > 1$. Hence, condition (A.4) is fulfilled. It remains to check (A.5). For $r > 0$ we get

$$\begin{aligned}
\sum_{k=1}^{\infty}\left(\sum_{i=1}^{k}\frac{\mathbb{E}(X_{ki}^2)}{a_k^2}\right)^{2r} &\leq \sum_{k=1}^{\infty}\left(\frac{\sum_{i=1}^{k} q_{ki}^2}{\left(\sum_{s=1}^{k} q_{ki}\right)^2}\right)^{2r} \\
&\leq \sum_{k=1}^{\infty}\left(\frac{\max_{i\in I_k} q_{ki}\sum_{i=1}^{k} q_{ki}}{\left(\sum_{s=1}^{k} q_{ki}\right)^2}\right)^{2r} \\
&\leq \sum_{k=1}^{\infty}\left(\frac{\max_{i\in I_k} q_{ki}}{\sum_{s=1}^{k} q_{ki}}\right)^{2r}.
\end{aligned}$$

Inequality (4.33) implies that there exists a $r > 0$ such that the latter expression is finite. Since all assumptions of Theorem A.8 are fulfilled, it follows that $\hat{F}_{n,0}(z)$ converges in the sense of the Glivenko-Cantelli Theorem almost surely, while the remaining assertions are an immediate consequence of Theorem A.9. ■

Theorem 4.18 implies the convergence of the ecdf $\hat{F}_{n,0}$ in the sense of the Glivenko-Cantelli Theorem if the number of independent blocks k increases. The next remark shows the convergence of $\hat{F}_{n,0}$ if n_0 increases (note that $n_0 = n_0(n) \to \infty$, $n \to \infty$, for all $\vartheta \in \Theta^*$).

Remark 4.19
Suppose the block lengths q_i, $i \in I_k$, are fixed in advance and consider the sequence of p-values given in (4.28). Consider the ecdf of the first $n = \sum_{i=1}^{k-1} q_i + j$ p-values with $j \in I_{q_k}$ and $k = k(n)$, that is,

$$\hat{F}_{n,0}(z) = \frac{\sum_{i=1}^{k-1} q_i}{\sum_{i=1}^{k-1} q_i + j} \hat{F}_{n',0}(z) + \frac{1}{\sum_{i=1}^{k-1} q_i + j} \sum_{t=1}^{j} I(p_{kt} \le z),$$

where $n' = \sum_{i=1}^{k-1} q_i$. By noting that the second summand is bounded by $q_k / \sum_{i=1}^{k} q_i$, which converges to 0 if (4.32) is fulfilled, we get almost sure convergence in (4.24) for $n \to \infty$.

In the next remark we give some examples for sets of block sizes fulfilling (4.29) and (4.32).

Remark 4.20
If

$$\max_{i \in I_k} q_{ki} = O(\min_{i \in I_k} q_{ki}) \text{ or } \max_{i \in I_k} q_{ki} = o(k \min_{i \in I_k} q_{ki}),$$

then condition (4.29) follows. Obviously, if the block sizes are bounded, i.e. there exists some $q \in \mathbb{N}$ such that $q_{ki} \le q$ for $i \in I_k$ and $k \in \mathbb{N}$, then we get condition (WD3).

Finally, we consider various sets of block-dependent p-values and perform some multiple test procedures.

Example 4.21
Let

$$\vartheta = 1_k \otimes \begin{pmatrix} 1 \\ 0 \\ 1 \\ 0 \end{pmatrix} \text{ and } \Sigma = J_k \otimes \begin{pmatrix} 1 & \rho & \rho^2 & \rho^3 \\ \rho & 1 & \rho & \rho^2 \\ \rho^2 & \rho & 1 & \rho \\ \rho^3 & \rho^2 & \rho & 1 \end{pmatrix}, \quad \rho \in (0,1),$$

where 1_k denotes a column vector of length k with entries 1 and J_k is the $k \times k$ identity matrix. Let $X_j \sim N_n(\vartheta, \Sigma)$, $j \in I_m$, be independent and identically distributed with $n = n(k) = 4k$ and block sizes $q_{ik} = 4$ (if we consider only null p-values, then $n_0 = 2k$ and $q_{ki} = 2$). Consider the multiple testing problem

$$H_i : \vartheta_i = 0 \text{ versus } K_i : \vartheta_i \neq 0, \quad i = 1, \dots, n.$$

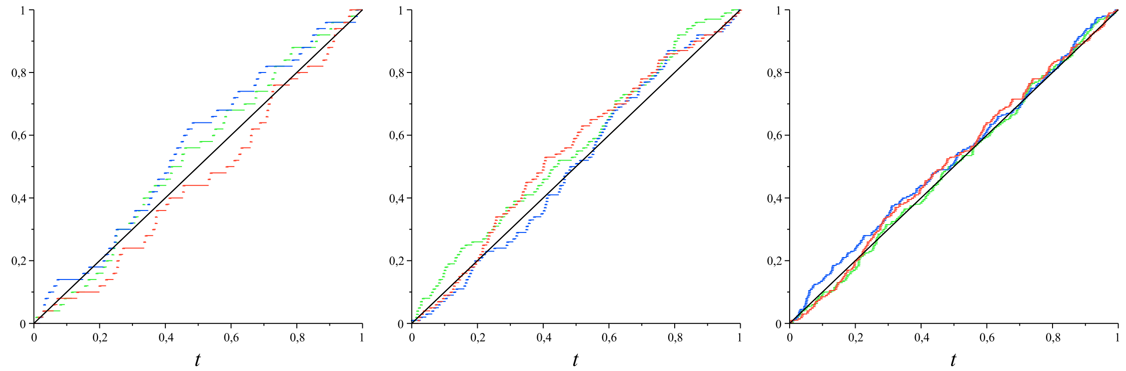

Figure 4.1: Simulated ecdfs $\hat{F}_{n,0}$s of p-values based on block-dependent autocorrelated normal random variables corresponding to true null hypotheses with $m = 10$, $n_0 = 50$ (left graph), $n_0 = 100$ (middle graph) and $n_0 = 200$ (right graph). Green curves correspond to $\rho = 0.1$, blue curves correspond to $\rho = 0.5$ and red curves correspond to $\rho = 0.9$ in each graph.

		Tests					
		β_n-adjustment	LSU	plug-in LSU	BPI	oracle Bonferroni	Bonferroni
$\rho = 0.1$	R_n	47	38	48	9	9	6
	V_n	3	1	3	0	0	0
$\rho = 0.5$	R_n	36	34	36	11	11	8
	V_n	2	2	2	0	0	0
$\rho = 0.9$	R_n	36	28	38	5	5	3
	V_n	1	0	2	0	0	0

Table 4.1: Simulation study for block-dependent test statistics in Example 4.21.

In this example, we do not assume that all variances are equal (although we choose all variances equal to 1 in the simulations) and choose the test statistics $T_i = \sqrt{m}\bar{X}_i/s_i$ where $\bar{X}_i = (1/m)\sum_{j=1}^m X_{ij}$ and $s_i^2 = 1/(m-1)\sum_{j=1}^m (X_{ij} - \bar{X}_i)^2$. We define p-values corresponding to T_i by $P_i = 2F_{t_{m-1}}(-|T_i|)$, where F_{t_ν} denotes the cdf of a univariate (central) t-distribution with ν degrees of freedom.

Figure 4.1 illustrates different realisations of the ecdf $\hat{F}_{n,0}$ of p-values corresponding to true null hypotheses for $n_0 = 50, 100, 200$ (left, middle and right pictures). We simulate this model for $m = 10$ and $\rho = 0.1$ (almost independence, green curves), $\rho = 0.5$ (moderate dependence, blue curves) and $\rho = 0.9$ (strong dependence, red curves).

Figure 4.2 displays simulated ecdfs of all p-values with $n = 100$, $n_0 = 50$, $m = 10$, $m_i = 4$ and $\rho = 0.1$ (green curve), $\rho = 0.5$ (blue curve) and $\rho = 0.9$ (red curve).

Table 4.1 shows the number of all rejected hypotheses R_n and the number of rejected true null hypotheses V_n for the following tests at the pre-specified level $\alpha = 0.05$: the β_n-adjustment SU procedure based on (3.28) with $\beta_{100} = 1.76$ (cf. Subsection 3.4.1), the LSU test (cf. Section

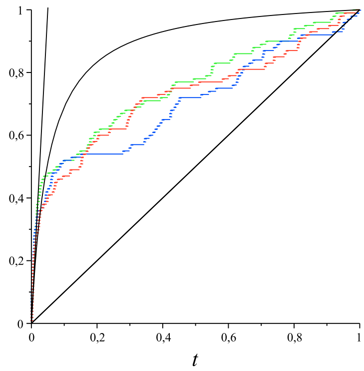

Figure 4.2: Simulated ecdfs $\hat{F}_n$s of all p-values with $m = 10$, $n = 100$ and $n_0 = 50$. The green curve corresponds to $\rho = 0.1$, the blue curve corresponds to $\rho = 0.5$ and the red curve corresponds to $\rho = 0.9$. The black curve is the AORC with $\alpha = 0.05$ and the black line is a rejection curve corresponding to the LSU test with $\alpha = 0.05$.

1.3), the plug-in LSU test with $\lambda = 0.5$, the BPI test with the threshold (2.4) based on (2.6) with $\lambda = 0.5$, the oracle Bonferroni and Bonferroni tests. For example, the β_n-adjustment test (LSU test or plug-in LSU test, resp.) rejects 47 (38 or 48, resp.) hypotheses for $\rho = 0.1$, 36 (34 or 36, resp.) hypotheses for $\rho = 0.5$ and 36 (28 or 38, resp.) for $\rho = 0.9$.

4.6 Pairwise comparisons

Pairwise comparisons provide further sets of p-values for which the weak dependence condition (WD2) is fulfilled. An example for a pairwise comparisons problem can be found in Keuls [1952]. He wrote

"In breeding agricultural and horticultural crops it is, in many cases, of much importance to compare the different selections obtained, e.g. in regard to their productive capacity. This is usually done in field trials involving these selections. The different plot yields will give us an impression of the productivity of the selections grown. In order to find out how far such impressions are reliable, the yield figures are mathematically worked out."

Keuls [1952] considered a trial on white cabbage carried out in 1950 and described the trial as follows:

"A trial field had been divided into 39 plots, grouped into 3 blocks of 13 plots each. In each block the 13 varieties to be investigated were planted out (randomized blocks design). During this trial all plots were treated in exactly the same way. The purpose was to learn which variety would give the highest gross yield per head of cabbage and which the lowest, in other words to find approximately the order of the varieties according to gross yield per cabbage."

Some investigations concerning FDR control for pairwise comparisons can be found in Yekutieli [2008]. Now we give a formal definition of the pairwise comparisons problem.

Let $X_i : \Omega \to \mathcal{X}_i$, $i \in I_k$, denote a sequence of independent random variables and let $\vartheta_i \in \tilde{\Theta}$ be a suitable parameter corresponding to X_i, $i \in I_k$, as for example $\vartheta_i = \mathbb{E}(X_i)$ or $\vartheta_i = \mathrm{Var}(X_i)$. Thereby, $\tilde{\Theta}$ may be multidimensional, e.g., $\tilde{\Theta} \subseteq \mathbb{R}^p$ or $\tilde{\Theta}$ may be the set of all positive definite $p \times p$ matrices, or $\tilde{\Theta}$ may be non-parametric, e.g., $\tilde{\Theta}$ may denote all continuous distribution functions on $\mathbb{R}$. The entire parameter space is $\Theta^* = \tilde{\Theta}^k$. A classical model is the k-sample model with $X_i = (X_{ij} : 1 \leq j \leq n_i)$, $i \in I_k$, and n_i denoting the sample size in group i.

Formally, a pairwise comparisons problem can be written as

$$H_{ij} : \vartheta_i = \vartheta_j \;\text{ versus }\; K_{ij} : \vartheta_i \neq \vartheta_j,\; 1 \leq i < j \leq k. \tag{4.34}$$

We restrict attention to tests based on p-values $p_{ij} = p_{ij}(x_i, x_j)$ depending only on the realisations of X_i and X_j, $1 \leq i < j \leq k$.

To investigate conditions under which null p-values fulfil the weak dependence condition (WD3), we consider (4.34) more precisely. Let $\vartheta \in \Theta^*$ be fixed for the moment such that for an arbitrary but fixed $r \in \mathbb{N}$ there are exactly r different parameters in the multiple-testing problem (4.34), i.e. there exist $\eta^1, \ldots, \eta^r$ such that $\vartheta_i \in \{\eta^1, \ldots, \eta^r\}$ for all $i \in I_k$. For a fixed $k \in \mathbb{N}$ with $r \leq k$, let $Q_{k1}, \ldots, Q_{kr}$ be a partition of the index set I_k such that $\vartheta_i = \eta^s$ if and only if $i \in Q_{ks}$. Hence, H_{ij} is true if and only if $i, j \in Q_{ks}$ for some $s \in I_r$. Let $q_{ks} = |Q_{ks}|$ and $q_{ks} \leq q_{k+1,s}$ for all $s \in I_r$, $r \in \mathbb{N}$ and $k \in \mathbb{N}$. Furthermore, $k = \sum_{s=1}^r q_{ks}$. Note that p_{ij} with $i, j \in Q_{ks}$ and $s \in I_r$ are p-values corresponding to true null hypotheses. The ecdf of all p-values is given by

$$\hat{F}_n(z) = \zeta_n \hat{F}_{n,0}(z) + (1 - \zeta_n)\hat{F}_{n,1}(z),$$

where

$$n = \frac{k(k-1)}{2}$$

is the number of all p-values,

$$\zeta_n = \frac{\sum_{s=1}^r q_{ks}(q_{ks} - 1)}{k(k-1)}$$

is the proportion of true null hypotheses,

$$\hat{F}_{n,0}(z) = \frac{2}{\sum_{s=1}^r q_{ks}(q_{ks} - 1)} \sum_{s=1}^r \sum_{i,j \in Q_{ks}} I(p_{ij} \leq z)$$

is the ecdf of p-values corresponding to true null hypotheses and

$$\hat{F}_{n,1}(z) = \frac{1}{\sum_{1 \leq s < t \leq r} q_{ks} q_{kt}} \sum_{1 \leq s < t \leq r} \sum_{i \in Q_{ks}} \sum_{j \in Q_{kt}} I(p_{ij} \leq z)$$

is the ecdf of p-values corresponding to alternatives.

In the next remark we study the asymptotic behaviour of the proportion ζ_n of true null hypotheses under suitable assumptions.

Remark 4.22
The proportion ζ_n of true null hypotheses can be rewritten as

$$\begin{aligned}\zeta_n &= \frac{\sum_{s=1}^r q_{ks}^2 - k}{\sum_{s=1}^r q_{ks}^2 - k + 2\sum_{1\le s<t\le r} q_{ks}q_{kt}} \\ &= \left(1 + 2\frac{\sum_{1\le s<t\le r} q_{ks}q_{kt}}{\sum_{s=1}^r q_{ks}^2 - k}\right)^{-1}.\end{aligned}$$

By noting that $k/(\sum_{s=1}^r q_{ks}^2) \to 0$ for $k \to \infty$ if the number r of blocks is fixed, we obtain that the proportion ζ_n of true null hypotheses converges to $1/r$ if $\max_{s\in I_r} q_{ks} = (1+o(1))\min_{s\in I_r} q_{ks}$, $k\to\infty$. Moreover, if there exists a $\gamma > 0$ such that

$$\lim_{k\to\infty} \frac{\sum_{1\le s<t\le r} q_{ks}q_{kt}}{\sum_{s=1}^r q_{ks}^2} = \gamma,$$

then $\zeta_n \to (1+2\gamma)^{-1}$ for $k \to \infty$. For example, if $r \in \mathbb{N}$ is fixed, $\max_{s\in I_r} q_{ks} = (1 + o(1))\min_{s\in I_r} q_{ks}$, $k \to \infty$, then $\gamma = (r-1)/2$. Another example with $r \to \infty$ is given as follows. Let $q_{k1} = q_{k2} = q_k$, $\lim_{k\to\infty} r(k)/q_k = 0$ and let $q_{ks} = q \in \mathbb{N}$ be fixed for all $s \ge 3$ and $k \in \mathbb{N}$. Then

$$\frac{\sum_{1\le s<t\le r} q_{ks}q_{kt}}{\sum_{s=1}^r q_{ks}^2} = \frac{q_k^2 + 2(r-2)q_k q + \binom{r-2}{2}q^2}{2q_k^2 + (r-2)q^2} = \frac{1 + O(r/q_k) + O(r^2/q_k^2)}{2 + O(r/q_k^2)}.$$

The latter converges to $\gamma = 1/2$ for $k \to \infty$ and consequently $\zeta_n \to 1/2$, $k \to \infty$.

The main result of this section shows that the ecdf of p-values corresponding to true null hypotheses of a pairwise comparisons problem fulfils the weak dependence condition (WD3), which allows asymptotic FWER and/or FDR control, cf. Theorem 4.2 and Theorem 4.5.

Theorem 4.23
Let q_{ks}, $k \in \mathbb{N}$, $s \in I_r$, $r \in \mathbb{N}$, be a double array of natural numbers with $2 \le q_{ks} \le q_{k+1s}$ for all $s \in I_r$, $r \in \mathbb{N}$ and $k \in \mathbb{N}$. Then we obtain convergence in probability in (4.24) *and hence (WD3) applies for the pairwise comparisons problem given in* (4.34) . *Moreover, if there exists a $q \in \mathbb{N}$ such that $\max_{s\in I_r} q_{ks} \le q$ for all $k \in \mathbb{N}$ and $r \in \mathbb{N}$, then we even get almost sure convergence in* (4.24).

Proof: Convergence in probability, i.e. the first assertion in Theorem 4.23, can be proved (a) by means of (4.25) or alternatively (b) by proving Var$(\hat{F}_{n,0}(z)) \to 0$ for $n_0 \to \infty$. As mentioned in Section 4.4, both conditions are equivalent.
(a) W.l.o.g. let the block size $q_{k1} = q_{k-1,1} + 1$ for a fixed $k \in \mathbb{N}$, i.e. $\theta_k = \theta_i = \eta^1$ for $i \in Q_{k1}$ and let $Q_{k1} = Q_{k-1,1} \cup \{k\} = \{1, \dots, q_{k1} - 1, k\}$. Then

$$n_0(i) = \sum_{s=2}^r \binom{q_{ks}}{2} + \binom{q_{k1}-1}{2} + i$$

denotes the number of p-values corresponding to true null hypotheses related to all comparisons between $X_1, \ldots, X_{k-1}$ and comparisons of X_k with $X_1, \ldots, X_i$ for $i \in Q_{k1} \setminus \{k\} = I_{q_{k1}-1}$, that is, p-values corresponding to true nulls are given by

$$p_{uv} : \; u, v \in Q_{ks}, \; s \in \{2, \ldots, r\} \;\text{ or }\; u, v \in Q_{k1} \setminus \{k\} \;\text{ and }\; p_{jk} : \; j \in \{1, \ldots, i\}. \tag{4.35}$$

For p_{ik}, $i \in Q_{k1} \setminus \{k\}$, there are exactly $q_{k1} - 2 + i - 1$ p-values p_l (say) in (4.35), for which $\mathrm{Cov}(p_{ik}, p_l) \neq 0$ is possible, that is, p_{ij}, $j \in Q_{k1} \setminus \{i, k\}$ and p_{jk}, $j \in \{1, \ldots, i-1\}$. Hence, setting $n_0 = n_0(i)$ for a fixed $i \in Q_{k1} \setminus \{k\}$ we get

$$\frac{1}{n_0} \sum_{j=1}^{n_0} \mathrm{Cov}\left(I(p_j^0 \leq t), I(p_{n_0}^0 \leq t)\right) \leq \frac{q_{k1} + i - 3}{\sum_{s=2}^{r} \binom{q_{ks}}{2} + \binom{q_{k1}-1}{2} + i},$$

where p_j^0, $j \in I_{n_0}$, denote p-values corresponding to true null hypotheses and $p_{n_0}^0 = p_{ik}$. Noting that the right-hand side of this expression is maximum for $i = q_{k1} - 1$ and $n_0 = n_0(q_{k1} - 1) = \sum_{s=1}^{r} q_{ks}(q_{ks} - 1)/2$, we obtain that the condition

$$\frac{4(q_{k1} - 2)}{\sum_{s=1}^{r} q_{ks}(q_{ks} - 1)} \to 0 \;\text{ for }\; k \to \infty \tag{4.36}$$

implies (4.25). Condition (4.36) can be proved by making use of the following consideration. If $\max_{s \in I_r}(q_{ks}) \to \infty$ for $k \to \infty$, then

$$\frac{4(q_{k1} - 2)}{\sum_{s=1}^{r} q_{ks}(q_{ks} - 1)} \leq \frac{4 \max_{s \in I_r} q_{ks}}{\sum_{s=1}^{r} q_{ks}(q_{ks} - 1)} \leq \frac{4}{\max_{s \in I_r} q_{ks} - 1} = \mathrm{O}\left(\frac{1}{\max_{s \in I_r} q_{ks}}\right) \to 0$$

for $k \to \infty$. If there exists some $q \in \mathbb{N}$ such that $\max_{s \in I_r} q_{ks} \leq q$ for all $k \in \mathbb{N}$ and $r \in \mathbb{N}$, i.e. $r \to \infty$ for $k \to \infty$, then

$$\frac{4(q_{k1} - 2)}{\sum_{s=1}^{r} q_{ks}(q_{ks} - 1)} \leq \frac{4 \max_{s \in I_r} q_{ks}}{\sum_{s=1}^{r} q_{ks}(q_{ks} - 1)} \leq \frac{2q}{n_0} \leq \frac{4q}{r} = \mathrm{O}\left(\frac{1}{r}\right) \to 0$$

for $k \to \infty$, which yields conditions (4.25) and/or (4.26) and hence completes the proof.
(b) Convergence $\mathrm{Var}(\hat{F}_{n,0}(z)) \to 0$ for $n \to \infty$ yields that (WD3) is satisfied. Since p_{ij} and p_{uv} are independent if $i, j \in Q_{ks_1}$, $u, v \in Q_{ks_2}$ and $s_1 \neq s_2$, we obtain

$$\begin{aligned} \mathrm{Var}(\hat{F}_{n,0}(z)) &= \frac{1}{n_0} \sum_{s=1}^{r} \mathrm{Var}\left(\sum_{i,j \in Q_{ks}} I(p_{ij} \leq z)\right) \\ &= \frac{1}{\left(\sum_{s=1}^{r} \binom{q_{ks}}{2}\right)^2} \sum_{s=1}^{r} \sum_{i,j \in Q_{ks}} \sum_{u,v \in Q_{ks}} \mathrm{Cov}\left(I(p_{ij} \leq z), I(p_{uv} \leq z)\right). \end{aligned}$$

For a fixed $s \in I_r$ exactly $6\binom{q_{ks}}{4}$ covariances in the expression above are equal to zero. Therefore,

$$\begin{aligned}\text{Var}(\hat{F}_{n,0}(z)) &\leq \frac{\sum_{s=1}^{r}\binom{q_{ks}}{2}^2 - 6\sum_{s=1}^{r}\binom{q_{ks}}{4}}{\left(\sum_{s=1}^{r}\binom{q_{ks}}{2}\right)^2} \\ &= \frac{\sum_{s=1}^{r} q_{ks}(q_{ks}-1)(4q_{ks}-6)}{\left(\sum_{s=1}^{r} q_{ks}(q_{ks}-1)\right)^2} \\ &\leq \frac{4\max_{s\in I_r} q_{ks}\sum_{s=1}^{r} q_{ks}(q_{ks}-1)}{\left(\sum_{s=1}^{r} q_{ks}(q_{ks}-1)\right)^2} \\ &= \frac{4\max_{s\in I_r} q_{ks}}{\sum_{s=1}^{r} q_{ks}(q_{ks}-1)}.\end{aligned}$$

Obviously, the latter converges to 0, since (4.36) is fulfilled. This implies the desired converges in probability for $n \to \infty$.

■

The next result corresponds to convergence of the ecdf $\hat{F}_{n,1}$ of p-values under alternatives.

Theorem 4.24
Let q_{ks}, $k \in \mathbb{N}$, $s \in I_r$, $r \in \mathbb{N}$, be a double array of natural numbers with $1 \leq q_{ks} \leq q_{k+1,s}$ for all $s \in I_r$, $r \in \mathbb{N}$ and $k \in \mathbb{N}$. Let $n_1 = n_1(\vartheta) = n - n_0(\vartheta) \to \infty$ if $n \to \infty$. Then condition

$$\frac{k}{\sum_{1\leq s<t\leq r} q_{ks}q_{kt}} \to 0 \ \textit{for} \ k \to \infty, \tag{4.37}$$

implies (4.27) *with convergence in probability.*

Proof: We prove the statement in Theorem 4.24 (a) by means of (4.25); and (b) by proving $\text{Var}(\hat{F}_{n,1}(z)) \to 0$ for $n_1 \to \infty$.
(a) It suffices to prove condition (4.25) applying to p-values under alternatives, i.e.

$$\frac{1}{n_1}\sum_{i=1}^{n_1} \text{Cov}\left(I(p_i^1 \leq t), I(p_{n_1}^1 \leq t)\right) \to 0 \ \text{ for } \ n_1 \to \infty, \tag{4.38}$$

where p_i^1, $i \in I_{n_1}$, are p-values under alternatives, i.e. p_i, $i \in I_{n,1}$. W.l.o.g. let for a fixed $k \in \mathbb{N}$ the block size q_{k1} be equal to $q_{k-1,1}+1$, i.e. $\theta_k = \theta_i = \eta^1$ for $i \in Q_{k1}$ and let $Q_{k1} = Q_{k-1,1} \cup \{k\} = \{1, \ldots, q_{k1}-1, k\}$. Moreover, let $Q_{ks} = \{\sum_{v=1}^{s-1} q_{kv}, \ldots, \sum_{v=1}^{s-1} q_{kv} - 1 + q_{ks}\}$ for $s \in \{2, \ldots, r\}$. Then for $i \in \{1, \ldots, \sum_{s=2}^{r} q_{ks}\}$

$$n_1(i) = \sum_{2\leq s<t\leq r} q_{ks}q_{kt} + \sum_{s=2}^{r}(q_{k1}-1)q_{ks} + i$$

denotes the number of p-values corresponding to false hypotheses related to all comparisons between $X_1, \ldots, X_{k-1}$ and comparisons of X_k with X_j, $j \in \{q_{k1}, \ldots, q_{k1}-1+i\}$, that is, p-values corresponding to false hypotheses are given by

$$p_{uv}: \ u \in Q_{ks}, \ v \in Q_{kt}, \ 2 \leq s < t \leq r, \ \text{ or } \ u \in Q_{k1} \setminus \{k\}, \ v \in Q_{ks}, \ 2 \leq s \leq r \tag{4.39}$$

$$\text{and } p_{kj} : j \in \{q_{k1}, \dots, q_{k1} - 1 + i\}.$$

Let $h \in \{2, \dots, r-1\}$ and $b \in \{1, \dots, q_{k,h+1}\}$ be such that $i = \sum_{s=2}^{h} q_{ks} + b$, that is, $h = h(i)$ and $b = b(i)$. For $p^1_{n_1(i)} = p_{k, q_{k1}-1+i}$, $i \in \{1, \dots, \sum_{s=2}^{r} q_{ks}\}$, there exist exactly

$$q_{k1} + 2\sum_{s=2}^{h} q_{ks} + \sum_{s=h+2}^{r} q_{ks} + b - 2$$

p-values p_l (say) in (4.39), for which $\mathrm{Cov}(p_{ik}, p_l) \neq 0$ is possible, that is, $p_{j\ q_{k1}-1+i}$ with $j \in Q_{ks}$ for $s \in \{2, \dots, r\} \setminus \{h+1\}$ or $j \in Q_{k1} \setminus \{k\}$ and p_{kj}, $j \in Q_{ks}$ with $s \in \{2, \dots, h\}$ or $j \in \{\sum_{s=1}^{h} q_{ks}, \dots, \sum_{s=1}^{h} q_{ks} - 1 + b\} \subseteq Q_{k,h+1}$. Hence, setting $n_1 = n_1(i)$ for a fixed $i \in \{1, \dots, \sum_{s=2}^{r} q_{ks}\}$ we get

$$\begin{aligned}
\frac{1}{n_1}\sum_{j=1}^{n_1} \mathrm{Cov}\left(I(p^1_j \le t), I(p^1_{n_1} \le t)\right) &\le \frac{q_{k1} + 2\sum_{s=2}^{h} q_{ks} + \sum_{s=h+2}^{r} q_{ks} + b - 2}{\sum_{2\le s<t\le r}^{r} q_{ks}q_{kt} + \sum_{s=2}^{r}(q_{k1}-1)q_{ks} + i} \\
&= \frac{q_{k1} + \sum_{s=2}^{h} q_{ks} + \sum_{s=h+2}^{r} q_{ks} - 2 + i}{\sum_{2\le s<t\le r}^{r} q_{ks}q_{kt} + \sum_{s=2}^{r}(q_{k1}-1)q_{ks} + i}.
\end{aligned}$$

Noting that the right-hand side of the expression before is maximum for $i = \sum_{s=2}^{r} q_{ks}$ (i.e. $h = r-1$ and $b = q_{kr}$) and $n_1 = n_1(\sum_{s=2}^{r} q_{ks}) = \sum_{1\le s<t\le r}^{r} q_{ks}q_{kt}$, we obtain that the condition

$$\frac{q_{k1} + 2\sum_{s=2}^{r-1} q_{ks} + q_{kr} - 2}{\sum_{1\le s<t\le r}^{r} q_{ks}q_{kt}} \to 0 \text{ for } k \to \infty$$

implies (4.38). Noting that $k = \sum_{s=1}^{r} q_{ks}$ we get (4.37), which completes the proof.
(b) Now we prove that $\mathrm{Var}(\hat{F}_{n,1}(z)) \to 0$ for $n \to \infty$, which implies the assertion in Theorem 4.24. For $1 \le s < t \le r$, $i \in Q_{ks}$ and $j \in Q_{kt}$ there are

$$q_{ks} + q_{kt} - 2 + 2\sum_{v \in I_r \setminus \{s,t\}} q_{kv}$$

p-values p_i, $i \in I_{n,1}$, for which $\mathrm{Cov}(p_{ik}, p_i) \neq 0$ is possible. Then

$$\mathrm{Var}(\hat{F}_{n,1}(z)) \le \frac{2\sum_{1\le s<t\le r} q_{ks}q_{kt}(\sum_{s=1}^{r} q_{ks})}{\left(\sum_{1\le s<t\le r} q_{ks}q_{kt}\right)^2} = \frac{2\sum_{s=1}^{r} q_{ks}}{\sum_{1\le s<t\le r} q_{ks}q_{kt}} = \frac{2k}{\sum_{1\le s<t\le r} q_{ks}q_{kt}}.$$

Condition (4.37) implies that $\mathrm{Var}(\hat{F}_{n,1}(z)) \to 0$ for $n_1 \to \infty$ and hence, we get the convergence in probability in (4.27). ■

Example 4.25
If $\max_{s\in I_r} q_{ks} = \min_{s\in I_r} q_{ks}(1+o(1))$ or $\max_{s\in I_r} q_{ks} = o(r(\min_{s\in I_r} q_{ks})^2)$, then condition (4.27) is always fulfilled. Note that for the case that $\max_{s\in I_r} q_{ks} = \min_{s\in I_r} q_{ks}(1+o(1))$ and r is fixed we get convergence of the proportion ζ_n of true null hypotheses to $1/r$ (cf. Remark 4.22) as well as convergence of $\hat{F}_{n,1}$ in the Glivenko-Cantelli sense.

Remark 4.26

Let the number r of different parameters in the pairwise comparisons problem (4.34) be fixed for all $k \in \mathbb{N}$. Let $q_{k1} \to \infty$ for $k \to \infty$ and q_{ks}=1 for all $s \in \{2, \ldots, r\}$ and $k \in \mathbb{N}$, i.e. we get r many-one comparisons. If X_i, $i \in Q_{k1}$, are iid for all $k \in \mathbb{N}$, then there exists $C \in (0,1)$ such that

$$\frac{1}{n_1}\sum_{i=1}^{n_1} \mathrm{Cov}\left(I(p_i^1 \leq t), I(p_{n_1}^1 \leq t)\right) \quad \geq \quad C\ \frac{q_{k1} + 2\sum_{s=2}^{r-1} q_{ks} + q_{kr} - 2}{\sum_{1\leq s<t\leq r}^{r} q_{ks}q_{kt}}.$$

The latter converges to $C/(r-1) > 0$ for $k \to \infty$, that is, condition (4.38) is not fulfilled and consequently the ecdf $\hat{F}_{n,1}$ does not converge in the sense of the Glivenko-Cantelli Theorem.

Although the convergence in probability of the ecdf $\hat{F}_{n,0}$ is sufficient for weak dependence, sometimes it is interesting to know that $\hat{F}_{n,0}$ converges not only in probability but also almost surely. The next theorem gives conditions which allow the almost sure convergence by means of the U-statistics theory.

Theorem 4.27

Let $r \in \mathbb{N}$ be fixed and q_{ks}, $k \in \mathbb{N}$, $s \in I_r$, be a double array of natural numbers with $2 \leq q_{ks} \leq q_{k+1s}$ for all $s \in I_r$ and $k \in \mathbb{N}$. Let X_i, $i \in Q_{ks}$, be iid for all $s \in I_r$ and let $p_{ij} = h(X_i, X_j)$ be the corresponding p-values. Then for each $z \in [0,1]$ we obtain almost sure convergence in (4.24).

Proof: W.l.o.g. let p_i, $i \in I_{n,0}$, be uniformly distributed in $[0,1]$, i.e. (D1) is fulfilled. Let $a(q) = q(q-1)$, $q \geq 2$. The almost sure convergence can be proved by means of U-statistics. By setting

$$U_{ks}(z) = \frac{2}{a(q_{ks})}\sum_{i,j \in Q_{ks}} I(p_{ij} \leq z),$$

we obtain

$$\hat{F}_{n,0}(z) = \frac{1}{\sum_{s=1}^{r} a(q_{ks})}\sum_{s=1}^{r} a(q_{ks})U_{ks}(z).$$

Note that $U_{ks}(z)$, $s \in I_r$, are independent U-statistics. Let $I_r' \subset I_r$ be such that q_{ks}, $k \in \mathbb{N}$, are bounded for all $s \in I_r'$, that is, there exists $q \in \mathbb{N}$ with $q_{ks} \leq q$ for all $k \in \mathbb{N}$ and $s \in I_r'$. Then $a(q_{ks})/\sum_{s=1}^{r} a(q_{ks}) \to 0$, $s \in I_r'$. Obviously, it holds

$$\hat{F}_{n,0}(z) \geq \frac{\sum_{s\in I_r\setminus I_r'} a(q_{ks})}{\sum_{s=1}^{r} a(q_{ks})}\min_{s\in I_r\setminus I_r'} U_{ks}(z) = A(z)\ \ \text{(say)}$$

and

$$\hat{F}_{n,0}(z) \leq \frac{ra(q)}{\sum_{s=1}^{r} a(q_{ks})} + \frac{\sum_{s\in I_r\setminus I_r'} a(q_{ks})}{\sum_{s=1}^{r} a(q_{ks})}\max_{s\in I_r\setminus I_r'} U_{ks}(z) = B(z)\ \ \text{(say)}.$$

Note that

$$\frac{\sum_{s\in I_r\setminus I_r'} a(q_{ks})}{\sum_{s=1}^{r} a(q_{ks})} \to 1 \ \text{ and } \ \frac{ra(q)}{\sum_{s=1}^{r} a(q_{ks})} \to 0 \ \text{ for } \ k \to \infty.$$

Moreover, the SLLN of U-statistics (cf. Theorem 4.15) yields for $s \in I_r \setminus I_r'$ (i.e. $\lim_{k\to\infty} q_{ks} = \infty$) that $U_{ks}(z) \to z$, $k \to \infty$, almost surely. Since maximum and minimum of a finite number of variables are continuous functions, the random variables $A(z)$ and $B(z)$ converge to z almost surely. Hence, $A(z) \leq \hat{F}_{n,0}(z) \leq B(z)$ implies $\hat{F}_{n,0}(z) \to z$ for $k \to \infty$ almost surely. ■

Finally, we consider a simulation study of p-values corresponding to pairwise comparisons problems.

Example 4.28

Let X_{ij}, $i \in I_k$, $j \in I_m$, be independent normally distributed random variables with unknown mean ϑ_i and unknown variance $\sigma_i^2 > 0$. We choose $\sigma_i = 1$, $i \in I_k$, in the simulation. We consider the pairwise comparisons problem given in (4.34) for various scenarios of means. We utilise (a) t-tests with a pooled variance, (b) Welch approximate t-tests and (c) Wilcoxon-Mann-Whitney tests to perform individual tests.

(a) The test statistics of the t-tests are given by $T_{uv}^{(1)} = \sqrt{m/2}(\bar{X}_u - \bar{X}_v)/s$, where $\bar{X}_i = \frac{1}{m}\sum_{j=1}^m X_{ij}$ and $s^2 = \frac{1}{k(m-1)}\sum_{i=1}^k \sum_{j=1}^m (X_{ij} - \bar{X}_i)^2$. Hence, the test statistics have a $t_{k(m-1)}$-distribution given that $\sigma_1^2 = \ldots = \sigma_k^2$. Denote the cdf of a univariate (central) t-distribution with ν degrees of freedom by F_{t_ν} and define p-values corresponding to the test statistic $T_{ij}^{(1)}$ by $P_{ij}^{(1)} = 2F_{t_{k(m-1)}}(-|T_{ij}^{(1)}|)$.

(b) The test statistics of the Welch approximate t-test are given by $T_{ij}^{(2)} = \sqrt{m}(\bar{X}_i - \bar{X}_j)/\sqrt{s_i^2 + s_j^2}$ with $s_i^2 = \frac{1}{m-1}\sum_{j=1}^m (X_{ij} - \bar{X}_i)^2$. Under null hypotheses of equal expectations the distribution of the Behrens Fisher statistics $T_{ij}^{(2)}$, $1 \leq i < j \leq k$, could be approximated by Student's t-distribution with

$$\nu = \frac{(\gamma_i + \gamma_j)^2}{\gamma_i^2/(m-1) + \gamma_j^2/(m-1)}$$

degrees of freedom, where $\gamma_i = \sigma_i^2/m$. Since σ_i^2, $i \in I_k$, are typically unknown, ν will be replaced by the following estimate

$$\hat{\nu} = \frac{(g_i + g_j)^2}{g_i^2/(m-1) + g_j^2/(m-1)},\ g_i = s_i^2/m,$$

cf. Welch [1947]. Then p-values corresponding to $T_{ij}^{(2)}$ are defined by $P_{ij}^{(2)} = 2F_{t_{\hat{\nu}}}(-|T_{ij}^{(2)}|)$.

(c) The test statistics of the Wilcoxon-Mann-Whitney test (also called Wilcoxon rank-sum test) are given by $T_{ij}^{(3)} = \min\left(\sum_{r=1}^m \sum_{f=1}^m I(X_{ir} < X_{jf}), \sum_{r=1}^m \sum_{f=1}^m I(X_{ir} > X_{jf})\right)$. The exact distribution of $U_{m,m} = \sum_{r=1}^m \sum_{f=1}^m I(X_{ir} < X_{jf})$ can be calculated with the following formula

$$P(U_{m,r} = u) = P(U_{m-1,r} = u - r)\frac{m}{m+r} + P(U_{m,r-1} = u)\frac{r}{m+r},$$

$$P(U_{m,r} < 0) = P(U_{m,r} > mr) = 0 \text{ for } r, m \geq 1,$$

$$P(U_{m,0} = 0) = P(U_{0,r} = 0) = 1 \text{ and } P(U_{m,0} > 0) = P(U_{0,r} > 0) = 0 \text{ for } r, m \geq 1,$$

cf. Mann and Whitney [1947]. Denote the cdf of $\min(U_{m,m}, m^2 - U_{m,m})$ by F^U. Thereby, m^2 is the maximal value of $U_{m,m}$. The p-values corresponding to $T_{ij}^{(3)}$ are given by $P_{ij}^{(3)} = F^W(T_{ij}^{(3)})$. We also consider randomised p-values which are given by $P_{ij}^{(4)} = F^W(T_{ij}^{(3)} - 1) + Y_{ij}[F^W(T_{ij}^{(3)}) -$

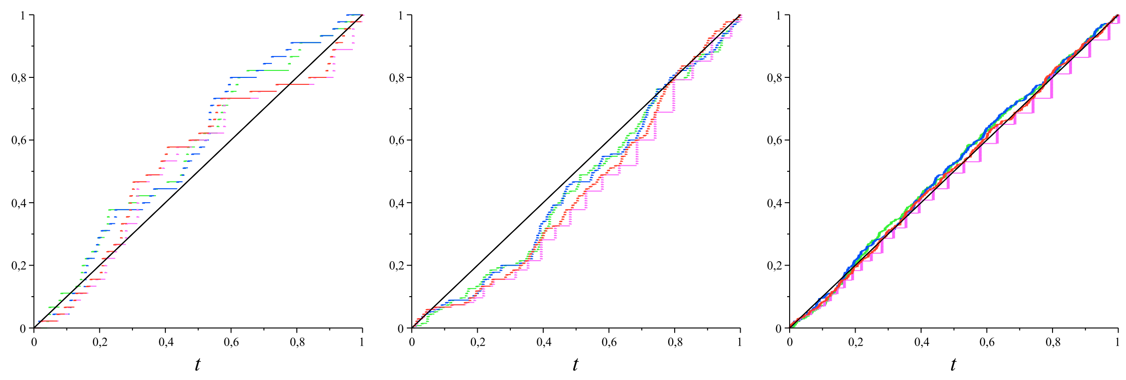

Figure 4.3: Simulated ecdfs $\hat{F}_{n,0}$s of p-values corresponding to true null hypotheses with $m = 10$, scenario $\{0_6, 1_6, 2_6\}$ and $n_0 = 45$ (left picture), $\{0_{10}, 1_{10}, 2_{10}\}$ and $n_0 = 135$ (picture in the middle), $\{0_{16}, 1_{16}, 2_{16}\}$ and $n_0 = 360$ (right picture). The ecdf of p-values corresponding to the t-test is green, to the Welch t-test is blue, to the Wilcoxon-Mann-Whitney test is magenta and to the Wilcoxon-Mann-Whitney test with randomised p-values is red in each graph.

$F^W(T_{ij}^{(3)} - 1)]$, where Y_{ij} are iid uniformly distributed random variables independent of X_{ij}, $i \in I_n$, $j \in I_m$. More information about randomised p-values can be found in Finner *et al.* [2010].

Setting $t_0 = 0$, a scenario $\{\eta_{q_1}^1, \ldots, \eta_{q_r}^r\}$ means $\vartheta_{q_{i-1}+1} = \ldots = \vartheta_{q_{i-1}+q_i} = \eta^i$ for $i = 1, \ldots, r$. Hence, the case $\vartheta_1 = \vartheta_2 = \vartheta_3 = \vartheta_4 = \vartheta_5 = \vartheta_6 = 0$, $\vartheta_7 = \vartheta_8 = \vartheta_9 = \vartheta_{10} = \vartheta_{11} = \vartheta_{12} = 1$ and $\vartheta_{13} = \vartheta_{14} = \vartheta_{15} = \vartheta_{16} = \vartheta_{17} = \vartheta_{18} = 2$ corresponds to $\{0_6, 1_6, 2_6\}$.

Figure 4.3 shows simulated ecdfs of p-values corresponding to true null hypotheses for different tests and scenarios. Although, in the case of the t-test, all p-values are dependent, because of the pooled variance estimate, the ecdf of these p-values (green curves) seems to converge to the identity function $F(t) = t$, $t \in [0, 1]$. Figure 4.4 displays simulated ecdfs of all p-values corresponding to the t-test (green curve), to the Welch t-test (blue curve), to the Wilcoxon-Mann-Whitney test (magenta curve) and to the Wilcoxon-Mann-Whitney test based on randomised p-values (red curve). The considered scenario is given by $\{0_6, 1_6, 2_6\}$ with $n = 153$, $n_0 = 45$ and $m = 10$. Table 4.2 shows the number of all rejected hypotheses R_n and the number of rejected true null hypotheses V_n for the following tests at the pre-specified level $\alpha = 0.05$: the β_n-adjustment SU procedure based on (3.28) with $\beta_{153} = 1.93$ (cf. Section 3.4.1), the LSU test (cf. Section 1.3), the plug-in LSU test with $\lambda = 0.5$, the BPI test with the threshold (2.4) based on (2.6) with $\lambda = 0.5$, the oracle Bonferroni and Bonferroni tests. Thereby, BPI, oracle Bonferroni and Bonferroni tests reject considerable less null hypotheses than the considered SU procedures. For example, the LSU test (β_n-adjustment test resp.) rejects 70 (78 resp.) hypotheses if p-values correspond to the t-tests, 70 (80 resp.) hypotheses if p-values correspond to the Welch t-tests, 66 (74 resp.) if p-values correspond to the Wilcoxon-Mann-Whitney tests and 71 (74 resp.) in the case of randomised p-values based on the Wilcoxon-Mann-Whitney tests.

Typically, parametric tests (t-tests, for example) have larger power than non-parametric tests

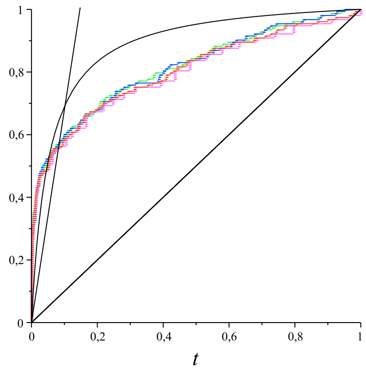

Figure 4.4: Simulated ecdfs $\hat{F}_n$ of all p-values with $m = 10$, scenario $\{0_6, 1_6, 2_6\}$ and $n = 153$ hypotheses. The ecdf of p-values corresponding to the t-test is green, to the Welch t-test is blue, to the Wilcoxon-Mann-Whitney test is magenta and to the Wilcoxon-Mann-Whitney test with randomised p-values is red. The Simes line is given by the black line and the black curve shows the AORC.

Test	p-values based on t-tests		Welch t-tests		Wilcoxon tests		Wilcoxon tests (random. p-val.)	
	R_n	V_n	R_n	V_n	R_n	V_n	R_n	V_n
β_n-adjustment	78	2	80	3	74	2	74	2
LSU	70	1	70	2	66	2	71	2
plug-in LSU	87	3	86	5	85	5	85	5
BPI	36	0	29	0	27	0	29	0
oracle Bonferroni	36	0	30	0	30	0	32	0
Bonferroni	27	0	22	0	18	0	18	0

Table 4.2: Simulation study for the pairwise comparisons problem in Example 4.28.

(Wilcoxon-Mann-Whitney tests, for example). On the other hand, a parametric test may lead to a large number of false rejections if test statistics are not normally distributed. Figure 4.4 and Table 4.2 show that randomised p-values based on Wilcoxon-Mann-Whitney tests (red curve in Figure 4.4) seem to lead to a power that is almost as large as the power of the corresponding p-values based on the parametric t-tests.

4.7 Simulations of FWER and power for BPI tests

In this section we conduct a simulation study to investigate numerically the FWER control level and the power of the BPI test in the case of dependent test statistics, cf. Sections 4.5 and 4.6. We restrict our attention to the BPI test with $\kappa = 1$ and critical value $\alpha/\hat{n}_0$ based on the estimator (2.6). Thereby the BPI test will be compared with the classical Bonferroni test, the corresponding SD Bonferroni-Holm test and the OB test.

To demonstrate the behaviour of the BPI procedure for dependent p-values we simulate four different models. In the first two models (block-dependence and pairwise mean comparisons) we simulate the FWER and the power β as defined in (2.39). In the third example we simulate an equi-correlated normal model and show that FWER is typically not controlled by the BPI procedure. The fourth example picks up the situation that is described in Example 4.4. In all cases the simulations are based on 100000 repetitions for $\alpha = 0.05$, $\lambda = 0.5$, $\kappa = 1$, and the Bonferroni-type critical values that are defined in expression (2.5). Note that the variance of $\hat{n}_0$ typically tends to be larger (possibly much larger) under dependence than under independence. Moreover, the chance for a type I error heavily depends on the estimate $\hat{n}_0$.

Example 4.29 (Block-dependence, cf. Section 4.5)
Let

$$\vartheta = 1_{25} \otimes \begin{pmatrix} 1 \\ 1 \\ 0 \\ 0 \end{pmatrix} \text{ and } \Sigma = \sigma^2 J_{25} \otimes \left[(1-\rho)J_4 + \rho 1_{4\times 4}\right], \quad \rho \in (0,1),$$

where 1_k denotes a column vector of length k with entries 1, $1_{q\times q}$ denotes a $q \times q$-matrix with entries 1 and J_k is the identity matrix. We choose $\sigma = 1$ in the simulations. Let $X_j \sim N_{100}(\vartheta, \Sigma)$, $j \in I_m$, be independent and identically distributed. Consider the multiple-testing problem

$$H_i : \vartheta_i = 0 \text{ versus } K_i : \vartheta_i \neq 0, \quad i = 1, \ldots, 100.$$

In this example, we do not assume that all variances are equal (although we choose all variances equal to 1 in the simulations) and choose the test statistics $T_i = \sqrt{m}\bar{X}_i/s_i$ with $\bar{X}_i = \frac{1}{m}\sum_{j=1}^m X_{ij}$ and $s_i^2 = \frac{1}{m-1}\sum_{j=1}^m (X_{ij} - \bar{X}_i)^2$. We define p-values corresponding to T_i by $P_i = 2F_{t_{m-1}}(-|T_i|)$, where F_{t_ν} denotes the cdf of a univariate (central) t-distribution with ν degrees of freedom.

For illustration, we simulate this model for $m = 10, 15, 20$ and only three values of ρ, i.e. 0.1 (almost independence), 0.5 (moderate dependence) and 0.9 (strong dependence). Table 4.3 indicates that the BPI procedure controls the FWER at the pre-specified level α for each ρ. Moreover,

ρ	Test	Results for $m = 10$		Results for $m = 15$		Results for $m = 20$	
		FWER	β	FWER	β	FWER	β
0.1	Bonferroni	0.025	0.103	0.025	0.340	0.025	0.613
	OB	0.049	0.160	0.049	0.442	0.049	0.707
	BPI	0.049	0.156	0.049	0.437	0.049	0.703
	SD	0.026	0.107	0.031	0.368	0.037	0.667
0.5	Bonferroni	0.024	0.103	0.024	0.341	0.025	0.613
	OB	0.048	0.160	0.048	0.442	0.049	0.707
	BPI	0.048	0.157	0.049	0.437	0.049	0.703
	SD	0.026	0.107	0.030	0.369	0.037	0.668
0.9	Bonferroni	0.022	0.103	0.022	0.341	0.021	0.613
	OB	0.043	0.160	0.043	0.442	0.042	0.707
	BPI	0.043	0.158	0.044	0.437	0.043	0.702
	SD	0.023	0.107	0.027	0.369	0.032	0.668

Table 4.3: Simulation study for the block-dependence model in Example 4.29.

FWER and power seem to be nearly independent of ρ. The results for the OB and BPI tests nearly coincide and there is some gain in power of the BPI compared with the Bonferroni and SD tests.

Example 4.30 (Pairwise mean comparisons, cf. Section 4.6)
Let X_{ij}, $i \in I_k$, $j \in I_m$, be independent normally distributed random variables with unknown mean ϑ_i and unknown variance $\sigma^2 > 0$. We choose $\sigma = 1$ in the simulations. We consider the pairwise comparisons problem

$$H_{ij} : \vartheta_i = \vartheta_j \text{ versus } K_{ij} : \vartheta_i \neq \vartheta_j, \;\; 1 \leq i < j \leq k,$$

for various scenarios of means. The test statistics are given by $T_{ij} = \sqrt{m/2}(\bar{X}_i - \bar{X}_j)/s$ where $\bar{X}_i$ is defined as in Example 4.29 and $s^2 = \frac{1}{k(m-1)} \sum_{i=1}^k \sum_{j=1}^m (X_{ij} - \bar{X}_i)^2$. Hence, the test statistics have a $t_{k(m-1)}$-distribution. Denote the cdf of a univariate (central) t-distribution with ν degrees of freedom by F_{t_ν} and define p-values corresponding to T_{ij} by $P_{ij} = 2F_{t_{k(m-1)}}(-|T_{ij}|)$.

Setting $t_0 = 0$, a scenario $\{\eta^1_{q_1}, \ldots, \eta^k_{q_k}\}$ means $\vartheta_{q_{i-1}+1} = \ldots = \vartheta_{q_{i-1}+q_i} = \eta^i$ for $i = 1, \ldots, k$. Hence, the case $\vartheta_1 = \vartheta_2 = \vartheta_3 = 0$, $\vartheta_4 = \vartheta_5 = \vartheta_6 = \vartheta_7 = 2$ and $\vartheta_8 = \vartheta_9 = \vartheta_{10} = 4$ corresponds to $\{0_3, 2_4, 4_3\}$. The simulation results (Table 4.4) provide that the BPI procedure apparently controls FWER. Note that because of the pooled variance estimate all p-values are dependent. Although the power of the OB procedure is always larger than the power of the BPI procedure, the power of the BPI is considerably lager than the power of the Bonferroni test and SD procedure. Clearly, the gain in power is due to small proportions of null hypotheses.

Example 4.31 (Equi-correlation)
Let $X_{ij} = \vartheta_i + \sqrt{1-\rho}Y_{ij} + \sqrt{\rho}Y_{0j}$, $i \in I_n$, $j \in I_m$, $\rho \in (0, 1)$, where $Y_{ij} \sim N(0, \sigma^2)$,

ϑ-scenario	Test	Results for $m=3$		Results for $m=5$		Results for $m=7$	
		FWER	β	FWER	β	FWER	β
$\{0_3, 2_4, 4_3\}$,	Bonferroni	0.011	0.323	0.012	0.548	0.012	0.726
$k = 10, n = 45, n_0 = 12$	OB	0.040	0.440	0.042	0.675	0.043	0.835
	BPI	0.040	0.417	0.047	0.657	0.050	0.822
	SD	0.019	0.348	0.024	0.601	0.029	0.795
$\{0_4, 1_1, 2_4, 3_1\}$,	Bonferroni	0.011	0.130	0.012	0.316	0.012	0.466
$k = 10, n = 45, n_0 = 12$	OB	0.038	0.226	0.040	0.432	0.043	0.571
	BPI	0.029	0.185	0.035	0.394	0.041	0.546
	SD	0.014	0.138	0.017	0.339	0.019	0.500
$\{0_4, 1_4, 2_4, 3_4, 4_4\}$,	Bonferroni	0.007	0.185	0.008	0.349	0.007	0.455
$k = 20, n = 190, n_0 = 30$	OB	0.042	0.285	0.044	0.447	0.043	0.546
	BPI	0.022	0.237	0.028	0.414	0.030	0.523
	SD	0.009	0.193	0.012	0.367	0.013	0.478

Table 4.4: Simulation study for the pairwise mean comparisons problem in Example 4.30.

$i = 0, \ldots, n$, $j \in I_m$, are independent standard normal random variables. We test again

$$H_i : \vartheta_i = 0 \text{ versus } K_i : \vartheta_i \neq 0, \quad i \in I_n.$$

The test statistics and corresponding p-values are defined as in Example 4.29. The simulation is performed for $n = 100$, $m = 10$, $\sigma^2 = 1$ and $\vartheta_i = 0$ for $1 \leq i \leq 50$ and $\vartheta_i = 1$ otherwise.

Figure 4.5 illustrates the dependence of FWER on ρ for the BPI procedure. The BPI test controls FWER at most for very small or large values of ρ. For most of the ρ values the FWER exceeds the pre-specified α-level. The reason is that the variance of the estimator $\hat{n}_0$ seems to be increasing in ρ. For $\rho = 1$, it can be easily checked that the FWER is controlled.

Example 4.32 (Multivariate equi-correlated t-distribution)
Finally, we consider the situation described in Example 4.4 with equi-correlated t-distributed test statistics. As pointed out before, in this special case the empirical distribution function of all p-values corresponding to true null hypotheses converges in 0.5, i.e. $\lim_{n\to\infty} \hat{F}_{n,0}(0.5) = 0.5$ almost surely. Therefore, $\lambda = 0.5$ is the best choice in order to estimate $\hat{\pi}_0$. Here we illustrate the behaviour of the BPI test especially for the situation where nearly all hypotheses are true. Let $n = 50$, $\nu = 15$, $\sigma^2 = 1$, $\vartheta_i = 0$, $i = 1, \ldots, n_0$ and $\vartheta_i = 3$, $i = n_0{+}1, \ldots, n$, $n_0 = 20, 48, 49, 50$. Table 4.5 demonstrates that FWER is obviously controlled for all values of n_0 that are considered here. Moreover, the differences between the four tests in FWER and power are more or less negligible for large values of n_0. For $n_0 = 20$ the power of the BPI procedure is close to that of the OB test and considerably larger than for the Bonferroni and SD tests. Especially for $n_0 = 49$ the power of the BPI test seems a little smaller than for the Bonferroni test. This seems due to the bias of $\hat{n}_0$, i.e. the expectation of $\hat{n}_0$ is somewhat larger than n_0.

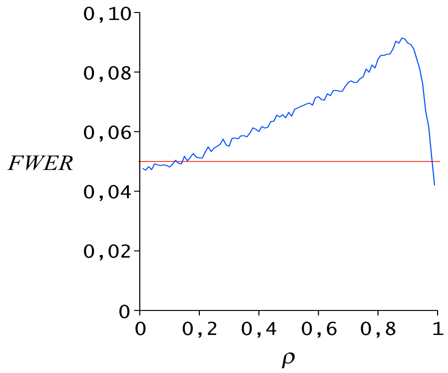

Figure 4.5: Simulated FWER of the BPI test in terms of the correlation coefficient ρ for equi-correlated normal random variables defined in Example 4.31.

Test	Results for $n_0 = 20$		Results for $n_0 = 48$		Results for $n_0 = 49$		Results for $n_0 = 50$
	FWER	β	FWER	β	FWER	β	FWER
Bonferroni	0.018	0.290	0.039	0.290	0.041	0.291	0.041
OB	0.044	0.422	0.041	0.295	0.042	0.294	0.041
BPI	0.044	0.411	0.041	0.292	0.042	0.290	0.041
SD	0.029	0.321	0.040	0.291	0.042	0.292	0.041

Table 4.5: Simulation study for the multivariate equi-correlated t-distribution model in Example 4.32 (Example 4.4).

4.8 Summary

In this chapter we considered situations where some kind of weak dependence occurs. We investigated models for which the limiting ecdf of p-values corresponding to true null hypotheses is asymptotically bounded by the cdf of the uniform distribution $F = Id$ if the number of true null hypotheses tends to infinity.

In the case of BPI procedures (cf. Chapter 2) we gave a sufficient condition on an estimator of the number of true null hypotheses ensuring FWER control at least asymptotically. We showed that the estimators given in (2.6) and (2.9) fulfil the aforementioned condition so that BPI tests, which control the finite FWER for independent p-values, control the asymptotic FWER for weak dependent p-values. For estimators defined in (2.6), weak dependence condition (WD2) can be reduced to condition (4.5), that is, the limiting ecdf of p-values under nulls in the point λ is not larger than λ. We gave an example for a set of p-values that fulfil (4.5) only for the unique point $\lambda = 0.5$. Note that for FWER control we did not need any additional condition on p-values. Weak dependence is sufficient for asymptotic FWER control.

For asymptotic FDR control for SUD test procedures, we needed a so-called power assumption guaranteeing that the proportion of rejected hypotheses is asymptotically bounded away from 0. Conditions (4.9) and/or (4.10) yield that the asymptotic FDR of some SUD tests is asymptotically controlled under weak dependence. Unfortunately, it seems to be difficult to prove such conditions. If the asymptotic crossing point converges to 0, i.e. the proportion of rejected hypotheses converges to 0, we do not have any arguments for asymptotic FDR control in this case. It remains an open problem. However, we considered a specific set of p-values, for which SUD(λ_n) procedures with $\liminf_{n\to\infty} \lambda_n/n > 0$ (SU tests belong to this class) control the FDR asymptotically. On the other hand, SD tests and SUD(λ_n) tests with $\lim_{n\to\infty} \lambda_n/n = 0$ may violate the pre-specified FDR-level.

We investigated various methods how convergence of an ecdf of p-values and/or weak dependence can be proved. We gave simple conditions on correlations between p-values which are equivalent to weak dependence given in (WD3) and/or convergence of the ecdf of p-values under alternatives in the sense of the Glivenko-Cantelli Theorem. A slightly modified condition implies even almost sure convergence of the ecdf of p-values corresponding to true/false null hypotheses. Among others things, we considered different examples for weak dependent p-values like various mixing models or autocorrelations. Weak dependence can sometimes be proved by means of U-statistics. The SLLN for U-statistics may imply almost sure convergence of the ecdf of p-values under nulls, which immediately yields the weak dependence condition (WD3).

For block-dependent p-values, we introduced conditions on block sizes leading to weak dependence. The same conditions that apply to p-values under alternatives yield the convergence of the ecdf of p-values corresponding to false hypotheses. We considered numerical examples for block-dependent p-values and performed different multiple test procedures for a fixed set of p-values. Thereby, tests controlling the FDR are more powerful than procedures controlling the FWER. Moreover, the β_n-adjustment method and the plug-in LSU test rejected the most hypotheses.

In the case of a pairwise comparison we investigated the behaviour of the proportion of true null hypotheses. We gave some examples for scenarios of parameters leading to a fixed asymptotic proportion of true nulls. We proved that the ecdf of true null hypotheses corresponding to a comparison problem always fulfils the weak dependence condition (WD3). If the number of different parameters is bounded and random variables with the same parameter are iid, then we get the almost sure convergence of the ecdf of true nulls. We presented conditions under which the ecdf of p-values under alternatives converges in the sense of the Glivenko-Cantelli Theorem. We considered a numerical example for p-values of a pairwise comparison problem based on different singular tests. We also performed various multiple tests for a fixed set of p-values. As anticipated, test procedure controlling the FDR rejected considerably more hypotheses than procedures controlling the FWER.

For BPI procedures with the threshold (2.4) based on the estimator given in (2.6) we implemented four numerical examples for dependent p-values, which do not necessarily fulfil the weak dependence condition. In the case of equi-correlated p-values FWER control failed if the correlation coefficient is not too large and not too small. However, for other dependence structures the FWER is always controlled.

Appendix A

Types of convergence

In probability theory, there exist several different notions of convergence of random variables. The convergence (in one of the senses presented below) of sequences of random variables to some limit random variable is an important concept in probability theory and its applications to statistics and stochastic processes.

Throughout the following, we assume that $\{X_n\}_{n\in\mathbb{N}}$ is a sequence of random variables, and X is a random variable, and all of them are defined on the same probability space $(\Omega, \mathcal{A}, \mathbb{P})$.

Definition A.1
*A sequence $\{X_n\}_{n\in\mathbb{N}}$ of random variables converges **in probability** to X if*

$$\forall\, \epsilon > 0: \quad \lim_{n\to\infty} \mathbb{P}\big(|X_n - X| > \varepsilon\big) = 0.$$

Definition A.2
*A sequence $\{X_n\}_{n\in\mathbb{N}}$ converges **almost surely** to X if*

$$\mathbb{P}\left(\lim_{n\to\infty} X_n = X\right) = 1.$$

Definition A.3
*A sequence $\{X_n\}_{n\in\mathbb{N}}$ converges **in the r-th mean** or **in the $\mathbb{L}_r$ norm** to X if $\mathbb{E}|X_n|^r < \infty$, $r \geq 1$, for all $n \in \mathbb{N}$ and*

$$\lim_{n\to\infty} \mathbb{E}\left(|X_n - X|^r\right) = 0.$$

Definition A.4 (Hsu and Robbins [1947])
*We say that a sequence $\{X_n\}_{n\in\mathbb{N}}$ converges **completely** to a random variable X if*

$$\forall\, \epsilon > 0: \quad \sum_{n=1}^{\infty} \mathbb{P}(|X_n - X| > \epsilon) < \infty.$$

Corollary A.5
The following implications between the various notions of convergence apply:

- *convergence almost surely implies convergence in probability,*

- *convergence in the $\mathbb{L}_r$ norm implies convergence in probability,*
- *convergence in the $\mathbb{L}_r$ norm implies convergence in the $\mathbb{L}_s$ norm, provided that $r \geq s \geq 1$,*
- *convergence completely implies convergence almost surely.*

Consider a sequence of iid random variables $\{X_n\}_{n\in\mathbb{N}}$ defined on a probability space (Ω,$\mathcal{A}$,$\mathbb{P}$) such that $X_n : \Omega \to \mathbb{R}$ with cumulative distribution function (cdf) F. The **empirical cumulative distribution function (ecdf)** for $X_1, \ldots, X_n$ is defined by

$$\hat{F}_n(x) = \hat{F}_n(x, \omega) = \frac{1}{n} \sum_{i=1}^{n} I_{(-\infty,x]}(X_i).$$

Note that for an arbitrary but fixed $x \in \mathbb{R}$, $\hat{F}_n(x, \cdot)$ is a sequence of random variables which converges to $F(x)$ almost surely by the strong law of large numbers (SLLN), i.e. $\hat{F}_n(x, \cdot) \to F(x)$ pointwise. The Glivenko–Cantelli theorem strengthens this result by proving uniform convergence of $\hat{F}_n$ to F.

Theorem A.6 (Glivenko-Cantelli)
Let $\{X_n\}_{n\in\mathbb{N}}$ be an iid sequence of random variables with distribution function F on $\mathbb{R}$. Then,

$$\|\hat{F}_n - F\|_\infty = \sup_{x\in\mathbb{R}} |\hat{F}_n(x, \cdot) - F(x)| \to 0 \quad \textit{almost surely}.$$

Now we consider a sequence of random variables $X_1, X_2, \ldots$ defined on $[0, 1]$, which are not necessarily iid, but for which we get $\hat{F}_n(x, \cdot) \to F(x)$ in probability for a fixed $x \in [0, 1]$. The classical Glivenko-Cantelli theorem can be generalised to the case of convergence in probability in the following way.

Lemma A.7 (Glivenko-Cantelli: convergence in probability)
Let $F_n : ([0, 1], \Omega) \to [0, 1]$, $n \in \mathbb{N}$, be such that $F_n(z, \cdot)$ is non-decreasing in z for each $n \in \mathbb{N}$ and $F_n(z, \cdot) \to z$, $n \to \infty$, in probability for all $z \in [0, 1]$. Then,

$$\sup_{z\in[0,1]} |F_n(z, \cdot) - z| \to 0 \;\; \textit{for} \;\; n \to \infty \quad \textit{in probability}. \tag{A.1}$$

Proof: From $F_n(z, \cdot) \to z$ in probability for all $z \in [0, 1]$, we get

$$\forall\, z \in [0, 1] : \;\; \forall\, \epsilon > 0 : \;\; \forall\, \epsilon_1 > 0 : \;\; \exists\, n_{z,\epsilon,\epsilon_1} \in \mathbb{N} : \;\; \forall\, n \geq n_{z,\epsilon,\epsilon_1} :$$

$$\mathbb{P}\left(|F_n(z, \cdot) - z| < \epsilon\right) \geq 1 - \epsilon_1.$$

For an arbitrary but fixed $k \in \mathbb{N}$ and $i \in \{0, \ldots, k\}$ we obtain

$$\forall\, \epsilon > 0 : \;\; \forall\, \epsilon_1 > 0 : \;\; \exists\, n_{i/k,\epsilon,\epsilon_1} \in \mathbb{N} : \;\; \forall\, n \geq n_{i/k,\epsilon,\epsilon_1} :$$

$$\mathbb{P}\left(\left|F_n\left(\frac{i}{k}, \cdot\right) - \frac{i}{k}\right| < \epsilon\right) \geq 1 - \epsilon_1.$$

Setting $n^k_{\epsilon,\epsilon_1} = \max_{i\in\{0,\dots,k\}} n_{i/k,\epsilon,\epsilon_1}$, it follows by applying the Bonferroni inequality that

$$\forall\, n \geq n^k_{\epsilon,\epsilon_1} :\ \mathbb{P}\left(\max_{i\in\{0,\dots,k\}} \left| F_n\left(\frac{i}{k},\cdot\right) - \frac{i}{k}\right| < \epsilon\right) \geq 1-(k+1)\epsilon_1.$$

Furthermore, for $i \in \{0,\dots,k-1\}$ we obtain

$$\begin{aligned}
\sup_{z\in\left[\frac{i}{k},\frac{i+1}{k}\right]} (F_n(z,\cdot)-z) &\leq F_n\left(\frac{i+1}{k},\cdot\right) - \frac{i}{k} \\
&= \left(F_n\left(\frac{i+1}{k},\cdot\right) - \frac{i+1}{k}\right) + \frac{1}{k}, \\
\sup_{z\in\left[\frac{i}{k},\frac{i+1}{k}\right]} (z-F_n(z,\cdot)) &\leq \frac{i+1}{k} - F_n\left(\frac{i}{k},\cdot\right) \\
&= \left(\frac{i}{k} - F_n\left(\frac{i}{k},\cdot\right)\right) + \frac{1}{k}
\end{aligned}$$

and

$$\sup_{z\in\left[\frac{i}{k},\frac{i+1}{k}\right]} |F_n(z,\cdot)-z| \leq \max\left(\left|F_n\left(\frac{i+1}{k},\cdot\right) - \frac{i+1}{k}\right|, \left|\frac{i}{k} - F_n\left(\frac{i}{k},\cdot\right)\right|\right) + \frac{1}{k},$$

$$\sup_{z\in[0,1]} |F_n(z,\cdot)-z| \leq \max_{i\in\{0,\dots,k\}} \left|F_n\left(\frac{i}{k},\cdot\right) - \frac{i}{k}\right| + \frac{1}{k}.$$

Hence,

$$\mathbb{P}\left(\sup_{z\in[0,1]} |F_n(z,\cdot)-z| < \epsilon\right) \geq \mathbb{P}\left(\max_{i\in\{0,\dots,k\}} \left|F_n\left(\frac{i}{k},\cdot\right) - \frac{i}{k}\right| < \epsilon - \frac{1}{k}\right).$$

Then

$$\forall\, \epsilon > 0 :\ \ \forall\, k > \frac{1}{\epsilon} :\ \ \forall\, \epsilon_1 > 0 :\ \ \exists\, n^k_{\epsilon,\epsilon_1} \in \mathbb{N} :\ \ \forall\, n \geq n^k_{\epsilon,\epsilon_1} :$$
$$\mathbb{P}\left(\sup_{z\in[0,1]} |F_n(z,\cdot)-z| < \epsilon\right) \geq 1-(k+1)\epsilon_1.$$

Finally, choosing $\epsilon_1 = \delta/(k+1)$ for a $\delta > 0$ we get

$$\forall\, \epsilon > 0 :\ \ \forall\, \delta > 0 :\ \ \exists\, n_{\epsilon,\delta} \in \mathbb{N} :\ \ \forall\, n \geq n_{\epsilon,\delta} :\ \ \mathbb{P}\left(\sup_{z\in[0,1]} |F_n(z,\cdot)-z| < \epsilon\right) \geq 1-\delta,$$

hence the assertion follows. ■

The next theorem taken from Hu and Taylor [1997] gives conditions on random variables, which imply almost sure convergence of weighted sums of these variables.

Theorem A.8 (Hu and Taylor [1997])
Let $\{X_{ni}\}_{i\in I_n, n\in\mathbb{N}}$ be real-valued random variables such that $X_{n1},\dots,X_{nn}$ are independent for all $n \in \mathbb{N}$. Let $\{a_n\}_{n\in\mathbb{N}}$ be a sequence of positive real numbers such that $a_{n+1} > a_n$ and

$\lim_{n\to\infty} a_n = \infty$. *Let* $\psi(t)$ *be a positive, even and continuous function such that* $\frac{\psi(|t|)}{|t|^p}$ *is an increasing function of* $|t|$, *and* $\frac{\psi(|t|)}{|t|^{p+1}}$ *is a decreasing function of* $|t|$, *i.e.,*

$$\frac{\psi(|t|)}{|t|^p} \uparrow \quad and \quad \frac{\psi(|t|)}{|t|^{p+1}} \downarrow \quad as \quad |t| \uparrow \tag{A.2}$$

for some integer $p \geq 2$. *Moreover, suppose that*

$$\mathbb{E}(X_{ni}) = 0, \tag{A.3}$$

$$\sum_{n=1}^{\infty} \sum_{i=1}^{n} \frac{\mathbb{E}\psi(|X_{ni}|)}{\psi(a_n)} < \infty, \tag{A.4}$$

$$\sum_{n=1}^{\infty} \left(\sum_{i=1}^{n} \frac{\mathbb{E}(X_{ni}^2)}{a_n^2} \right)^{2k} < \infty, \tag{A.5}$$

where k *is a positive integer. Then*

$$\frac{1}{a_n} \sum_{i=1}^{n} X_{ni} \to 0 \ \text{almost surely}. \tag{A.6}$$

The next theorem given in Sung [2000] shows that different types of convergence are equivalent under certain conditions.

Theorem A.9 (Sung [2000])
Let $\{X_{ni}\}_{i\in I_n, n\in\mathbb{N}}$ *and* $\{a_n\}_{n\in\mathbb{N}}$ *be defined as in Theorem A.8. Assume that (A.4) and (A.5) are fulfilled. Then the following statements are equivalent:*

1) $\frac{1}{a_n} \sum_{i=1}^{n} X_{ni} \to 0$ *in the* $\mathbb{L}_1$ *norm*,

2) $\frac{1}{a_n} \sum_{i=1}^{n} X_{ni} \to 0$ *completely*,

3) $\frac{1}{a_n} \sum_{i=1}^{n} X_{ni} \to 0$ *almost surely*,

4) $\frac{1}{a_n} \sum_{i=1}^{n} X_{ni} \to 0$ *in probability*.

The next theorem gives the Dvoretzky-Kiefer-Wolfowitz (DKW) inequality, cf. Massart [1990] and Dvoretzky *et al.* [1956].

Theorem A.10 (Dvoretzky *et al.* [1956])
Let $X_1, \ldots, X_n$ *be iid real valued random variables with the distribution function* F *and the ecdf* $\hat{F}_n$. *Then*

$$\exists\, K > 0 : \ \forall\, n \in \mathbb{N} : \ \forall\, \epsilon \geq 0 : \ \mathbb{P}\left(\sup_{z\in\mathbb{R}} |\hat{F}_n(z) - F(z)| \geq \epsilon \right) \leq K \exp(-2n\epsilon^2).$$

Moreover, Massart [1990] proved that K can be chosen to be equal to 2 and K cannot be further improved.

The next result concerning expected values of non-decreasing functions is taken from Tong [1980], p. 121.

Lemma A.11 (Tong [1980])
Let X_i, $i \in I_n$, be independent real valued random variables. Let F_i, $i \in I_n$, and G_i, $i \in I_n$, be cdfs such that

$$\forall\, i \in I_n :\ \forall\, x \in \mathbb{R} :\ \mathbb{P}^{F_i}(X_i \geq x) \geq \mathbb{P}^{G_i}(X_i \geq x),$$

where X_i, $i \in I_n$, has the cdf F_i (or G_i) under the measure $\mathbb{P}^{F_i}$ (or $\mathbb{P}^{G_i}$, resp.). Let $\phi : \mathbb{R}^n \to \mathbb{R}$ be such that $\phi(x_1, \ldots, x_n)$ is non-decreasing in each x_i. Then

$$\mathbb{E}^F[\phi(X_1, \ldots, X_n)] \geq \mathbb{E}^G[\phi(X_1, \ldots, X_n)],$$

where $\mathbb{E}^F$ (or $\mathbb{E}^G$) denotes the expected value of $\phi(X_1, \ldots, X_n)$ under the product measure $\otimes_{i=1}^n \mathbb{P}^{F_i}$ (or $\otimes_{i=1}^n \mathbb{P}^{G_i}$, resp.).

List of Tables

List of Figures

Bibliography

Abramovich, F., Benjamini, Y., Donoho, D. and Johnstone, I. M. (2006) Adapting to unknown sparsity by controlling the false discovery rate. *Ann. Statist.*, **34**, 584-653.

Anderson, T. W. (1958) An Introduction to multivariate Statistical analysis. *John Wiley and Sons. Inc., New York* .

Benjamini, Y. and Hochberg, Y. (1995) Controlling the false discovery rate: a practical and powerful approach to multiple testing. *J. R. Statist. Soc.* **B**, **57**, 289-300.

Benjamini, Y. and Hochberg, Y. (2000) On the adaptive control of the false discovery rate in multiple testing with independent statistics. *J. Educ. Behav. Statist.*, **25**, 60-83.

Benjamini, Y., Krieger, A. and Yekutieli, D. (2005) Adaptive linear step-up procedures that control the false discovery rate. *Biometrika*, **93**, 491-507.

Benjamini, Y. and Yekutieli, D. (2001) The control of the false discovery rate in multiple testing under dependency. *Ann. Statist.*, **29**, 1165-1188.

Berk, R. H. (1966) Limiting behavior of posterior distributions when the model is incorrect. *Ann. Math. Statist.*, **37**, 51-58.

Black, M. A. (2004) A note on the adaptive control of false discovery rates. *J. R. Statist. Soc.* **B**, **66**, 297-304.

Blanchard, G. and Roquain, E. (2008) Two simple sufficient conditions for FDR control. *Elect. J. Statis.*, **2**, 963-992.

Billingsley, P. (1968) Convergence of probability measures. *John Willey and Sons, Inc., New York.*

Duncan, C. T. (2004) Statistical methods in genetic epidemiology. *Oxford University Press, Inc.*

Dvoretzky, A., Kiefer, J. and Wolfowitz, J. (1956) Asymptotic minimax character of the sample distribution function and of the classical multinomial estimator. *Ann. Math. Statist.*, **27**, 642-669.

Edgar, G. A. and Sucheston, L. (1992) Stopping times and directed processes. *Cambridge University Press.*

Farcomeni, A. (2006) Some results on the control of the false discovery rate under dependence. *Scand. J. Statist.*, **32**, 275-297.

Finner, H., Dickhaus, T. and Roters, M. (2007) Dependency and false discovery rate: Asymptotics. *Ann. Statist.*, **35**, 1432-1455.

Finner, H., Dickhaus, T. and Roters, M. (2009) On the false discovery rate and an asymptotically optimal rejection curve. *Ann. Statist.*, **37**, 596-618.

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.

Finner, H., Hayter, A. J. and Roters, M. (1993) On the joint distribution function of order statistics with reference to step-up multiple test procedures. *Technical Report* 93-19, 1-11, *Universität Trier*.

Finner, H. and Roters, M. (2001) On the false discovery rate and expected type I errors. *Biom. J.*, **43**, 985-1005.

Finner, H. and Roters, M. (2002) Multiple hypotheses testing and expected number of type I errors. *Ann. Statist.*, **30**, 220-238.

Finner, H., Straßburger, K., Heid, I., Herde, C., Rathmann, W., Giani, G., Dickhaus, T., Lichtner, P., Meitinger, T., Wichmann, H-E., Illig, T. and Gieger, C. (2010) How to link call rate and p-values for Hardy–Weinberg equilibrium as measures of genome-wide SNP data quality. *Statistics in Medicine*, **29**, 2347–2358.

Frampton, P. (1986) Dual resonance models and superstrings. *World Scientific Publishing Co Pte Ltd.*

Gavrilov, Y., Benjamini, Y. and Sarkar, S. K. (2009) An adaptive step-down procedure with proven FDR control. *Ann. Statist.*, **37**, 619-629.

Genovese, C. and Wasserman, L. (2004) A stochastic process approach to false discovery control. *Ann. Statist.*, **32**, 1035-1061.

Girautis, L. and Surgailis, D. (1985) CLT and other limit theorems for functionals of Gaussian processes. *Z. Wahrscheinlichkeitstheorie verw. Gebiete*, **70**, 191-212.

Guo, W. (2009) A note on adaptive Bonferroni and Holm procedures under dependence. *Biometrika*, **94**, 1012-1018.

Hochberg, Y. (1988) A sharper Bonferroni procedure for multiple tests of significance. *Biometrika*, **75**, 800-802.

Hochberg, Y. and Benjamini, Y. (1990) More powerful procedures for multiple significance testing. *Statist. Med.*, **9**, 811-818.

Hoeffding, W. (1948) A class of statistics with asymptotically normal distribution. *Ann. of math. Statist.*, **19**, 293-325.

Hsu, P. L. and Robbins, H. (1947) Complete convergence and the law of large numbers. *Proc. Nat. Acad. Sci. USA*, **33**, 25-31.

Hsueh, H. M., Chen, J. J. and Kodell, R. L. (2003) Comparison of Methods for Estimating the Number of True Null Hypotheses in Multiplicity Testing. *J. Biopharm. Statist*, **13**, 675-689.

Hu, Tien-Chung and Taylor, R. L. (1997) On the strong law for arrays and for the bootstrap mean and variance. *Internat. J. Math. and Math. Sci.*, **20**, 375-382.

Hung, H. M. J., O'Neill, R. T., Bauer, P. and Köhne, K. (1997) The behavior of the p-value when the alternative hypothesis is true. *Biometrics*, **53**, 11-22.

Jin, J. (2008) Proportion of non-zero normal means: universal oracle equivalences and uniformly consistent estimators. *J. R. Statist. Soc.* **B**, **70**, 461-493.

Jin, J. and Cai, T. T. (2007) Estimating the null and the proportion of non-null effects in large-scale multiple comparisons. *J. Am. Statist. Ass.*, **102**, 495-506.

Kang, G., Childers D., Liu, N., Zhang, K. and Gao, G. (2009) Genome-wide association studies of rheumatoid arthritis data via multiple hypothesis testing methods for correlated tests. *BMC Proceedings*, **3** (Suppl. 7): S38.

Keuls, M. (1952) The use of the "studentized range" in connection with an analysis of variance. *Euphytica*, **1**, 112-122.

Kwong, K. -S. and Wong, E. -H. (2002) A more powerful step-up procedure for controlling the false discovery rate under independence. *Stat. Probab. Lett.*, **56**, 217-225.

Langaas, M., Lindqvist, B. H. and Ferkingstad E. (2005) Estimating the proportion of true null hypotheses, with application to DNA microarray data. *J. R. Statist. Soc.* **B**, **67**, 555-572.

Lehmann, E. L. and Romano, J. P. (2005) Generalizations of the familywise error rate. *Ann. Statist.*, **33**, 1138-1154.

Mann, H. and Whitney, D. (1947) On a test of whether one of two random variables is stochastically larger than the other. *Ann. of math. Statist.*, **18**, 50-60.

Massart, P. (1990). The tight constant in the Dvoretzky-Kiefer-Wolfowitz inequality. *Ann. Probab.*, **18**, 1269-1283.

Meinshausen, N. and Bühlmann, P. (2005) Lower bounds for the number of false null hypotheses for multiple testing of associations under general dependence structures. *Biometrika*, **92**, 893-907.

Meinshausen, N. and Rice, J. (2006) Estimating the proportion of false null hypotheses among a large number of independently tested hypotheses. *Ann. Statist.*, **34**, 373-393.

Parzen, E. (1960) Modern probability theory and its applications. *John Wiley and Sons, Inc., New York*.

Sarkar, S. K. (1998). Some probability inequalities for ordered MTP2 random variables: A proof of the Simes conjecture. *Ann. Statist.*, **26**, 494-504.

Sarkar, S. K. (2002) Some results on false discovery rate in stepwise multiple testing procedures. *Ann. Statist.*, **30**, 239-257.

Sarkar, S. K. and Guo, W. (2009) On a generalized false discovery rate. *Ann. Statist.*, **37**, 1545-1565.

Schweder, T. and Spjøtvoll, E. (1982) Plots of P-values to evaluate many tests simultaneously. *Biometrika*, **69**, 493-502.

Sen, P. K. (1999) Some remarks on Simes-type multiple test of significance. *J. Statist. Plann. Inference*, **82**, 139-145.

Shorack, G. P. and Wellner, J. A. (1986) Empirical Processes with Applications to Statistics. *John Wiley and Sons. Inc., New York* .

Simes, R. J. (1986) An improved Bonferroni procedure for multiple tests of significance. *Biometrika*, **73**, 751-754.

Storey, J. D. (2002) A direct approach to false discovery rates. *J. R. Statist. Soc.* **B**, **64**, 479-498.

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.

Sun, W. and Cai, T. T. (2007) Oracle and adaptive compound decision rules for false discovery rate control. *J. Am. Statist. Ass.*, **102**, 901-912.

Sung, Soo Hak (2000) Complete convergence for sums of arrays of random elements. *Internat. J. Math. and Math. Sci.*, **23**, 789-794.

Swanepoel, J. (1999) The limiting behavior of a modified maximal symmetric 2s-spacing with applications. *Ann. Statist.*, **27**, 24-35.

Tong, Y. L. (1980) Probability inequalities in multivariate distributions. *Academic Press, New York*.

Tuyen, D. Q. (1999) A strong law for mixing random variables. *Period. Math. Hungar.*, **38**, 131-136.

Welch, B. L. (1947) The generalization of "Student's" problem when several different population variances are involved. *Biometrika*, **34**, 28-35.

Witting, H. (1985) Mathematische Statistik I. *B. G. Teubner, Stuttgart*.

Witting, H. and Müller-Funk, U. (1995) Mathematische Statistik II. *B. G. Teubner, Stuttgart*.

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 in ähnlicher Form noch bei keiner anderen Institution eingereicht. Ich habe bisher keine erfolglosen Promotionsversuche unternommen.

Düsseldorf, den 27.10.2010

Veronika Gontscharuk