

Flexible Designs für klinische Studien  
in Verteilungssituationen mit Störparametern



Dissertation

zur Erlangung des Doktorgrades  
der Naturwissenschaften  
(Dr. rer. nat.)

dem Fachbereich Mathematik u. Informatik  
der Philipps-Universität Marburg  
vorgelegt von

Nina Timmesfeld  
aus Köln

Marburg, Januar 2008

Vom Fachbereich Mathematik und Informatik  
der Philipps-Universität Marburg als Dissertation angenommen.

Erstgutacher: Prof. Dr. Helmut Schäfer

Zweitgutachter: Prof. Dr. Volker Mammitzsch

Tag der mündlichen Prüfung: 11.02.2008

# Zusammenfassung

Kontrollierte klinische Studien sind eine experimentelle Methode der klinisch-epidemiologischen Forschung, mit deren Hilfe der kausale Nachweis therapeutischer Wirksamkeit erbracht und somit der Therapiefortschritt abgesichert wird. Im Laufe einer solchen Studie stellte sich nicht selten heraus, dass die Annahmen, die der Studienplanung zugrundeliegen, einer Korrektur bedürfen. Der zu erwartende Therapieunterschied oder die Varianz der Messgröße kann sich als kleiner oder größer herausstellen als ursprünglich angenommen. Mit konventionellen statistischen Verfahren, war es nur bedingt möglich eine einmal geplante Studie nachträglich zu ändern, da nach datenabhängige Änderungen die statistische Fehlerkontrolle nicht mehr gesichert ist. Dies führte dazu, dass viele Studien aufgrund falscher Planungsannahmen nicht zufriedenstellend beendet werden konnten oder die statistische Fehlerkontrolle nicht mehr in der gebotenen Strenge beachtet wurde. In den letzten Jahren wurden verschiedene Methoden entwickelt, die Änderungen in laufenden Studien ermöglichen, ohne die Integrität der Studien zu verletzen. Dies sind unter anderem die adaptiven Designs, welche erstmals von Bauer (1989) vorgeschlagen wurden. Später wurden diese Verfahren in flexible Designs umbenannt, um einen Namenskonflikt zu vermeiden. Flexible Designs ermöglichen es, an vorgegebenen Zeitpunkten im Verlaufe der Studie Designänderungen vorzunehmen und dabei das Fehlerniveau 1. Art einzuhalten. Ein weiterer Ansatz wurde von Müller und Schäfer (2004) vorgeschlagen,

das sogenannte CRP-Prinzip. Im Gegensatz zu den flexiblen Designs müssen hierbei die Zeitpunkte, an denen Designänderungen möglich sind, nicht vorher (zum Zeitpunkt der Studienplanung) festgelegt werden. Dennoch ist die Kontrolle des Fehlerniveaus 1. Art sichergestellt. Damit ist es möglich, noch zeitnaher Korrekturen des Studiendesigns vorzunehmen, wenn sich Planungsannahmen als falsch herausstellen.

Obwohl die Idee des CRP-Prinzip sehr allgemein formuliert ist, war eine exakte Anwendungen bisher nur in Verteilungssituationen mit nur einem unbekanntem Parameter möglich, wie im Falle der Normalverteilung mit unbekanntem Erwartungswert und bekannter Varianz. War die Varianz ebenfalls unbekannt ( $t$ -Test), so konnte dieses Verfahren nur noch approximativ angewendet werden. Insbesondere für kleine Fallzahlen ist daher die Kontrolle des Fehlerniveaus 1. Art nicht mehr sichergestellt.

Ziel dieser Arbeit ist eine Weiterentwicklung des von Müller und Schäfer vorgeschlagenen Verfahrens, so dass es auch in Verteilungssituationen mit mehreren unbekanntem Parametern, wie dem  $t$ -Test, exakt angewandt werden kann.

Im Kapitel 2 geben wir zunächst eine Einführung in die besondere Problematik klinischer Studien. Kapitel 3 enthält einen Literaturüberblick über die bisherigen Gruppensequentiellen und flexiblen Designs. Im Kapitel 4 wird ein allgemeines mathematisches Konzept zur formalen Beschreibung und Definition entwickelt, welches sowohl die flexiblen Designs als auch das CRP-Prinzip beinhaltet. Dieses Konzept stellt eine Erweiterung des bekannten Begriffs der statistischen Entscheidungsfunktion dar. Auf dieser Basis werden im Kapitel 5 die Hauptergebnisse dieser Arbeit entwickelt. Damit gelingt es, das CRP-Prinzip auf beliebige  $k$ -parametrische Exponentialfamilien zu verallgemeinern. Diese Verallgemeinerung wenden wir am Ende dieses Kapitels auf die Testung des Lageparameters einer Normalverteilung bei unbekannter

Varianz ( $t$ -Test) und auf den Vergleich zweier unbekannter Binomialparameter (Fisher's exakter Test) an.



# Inhaltsverzeichnis

<b>1</b>	<b>Einleitung</b>	<b>1</b>
<b>2</b>	<b>Besondere Aspekte von klinischen Studien</b>	<b>3</b>
2.1	Die Phasen einer Studie . . . . .	4
2.2	Zielsetzung einer Phase III Studie . . . . .	6
2.3	Die Bedeutung von Fehlentscheidungen . . . . .	7
2.4	Vermeidung von Fehlentscheidungen . . . . .	8
2.5	Einige ethische Aspekte . . . . .	9
2.6	Ziele der statistischen Studienplanung . . . . .	10
<b>3</b>	<b>Spezielle Studiendesigns</b>	<b>15</b>

3.1	Beschreibung eines statistischen Tests . . . . .	16
3.2	Designs mit adaptiver Zuteilungsregel . . . . .	19
3.3	Designs mit interner Pilotstudie . . . . .	20
3.3.1	Problem der Fallzahlbestimmung beim Vorliegen eines Störparameters . . . . .	20
3.3.2	Fallzahlneubestimmung im Verlauf einer Studie . . . . .	21
3.4	Gruppensequentielle Designs . . . . .	23
3.5	Flexible (adaptive) Designs . . . . .	24
3.5.1	Flexible Designs basierend auf einer Kombinationsregel . . . . .	26
3.5.2	Flexible Designs basierend auf einer bedingten Fehler- funktion . . . . .	27
3.5.3	Zusammenfassung . . . . .	29
3.6	Designänderungen nach dem CRP-Prinzip . . . . .	30
3.6.1	Eine vereinfachte Darstellung der Idee . . . . .	31

#### **4 Theorie der flexiblen Verfahren**



---

4.1	Entscheidungsprozeduren . . . . .	33
4.1.1	Klassische Entscheidungsprozeduren . . . . .	34
4.1.2	Sequentielle Entscheidungsprozeduren . . . . .	35
4.1.3	Entscheidungsprozeduren mit datenabhängiger Anpassung . . . . .	37
4.2	Theoretische Grundlagen der flexiblen Verfahren . . . . .	40
<b>5</b>	<b>Flexible Designs in Exponentialfamilien</b>	<b>49</b>
5.1	Modell und Eigenschaften . . . . .	49
5.1.1	Exponentialfamilien . . . . .	49
5.1.2	Voraussetzungen . . . . .	54
5.2	Flexible Designs in einparametrischen Exponentialfamilien . . .	56
5.2.1	Einseitige Testprobleme . . . . .	56
5.2.2	Zweiseitige Testprobleme . . . . .	61
5.3	Flexible Designs in mehrparametrischen Exponentialfamilien . .	66
5.3.1	Einseitige Testprobleme . . . . .	66

5.3.2	Zweiseitige Testprobleme . . . . .	72
5.4	Anwendung der Methode an Beispielen . . . . .	74
5.4.1	Fisher's exakter Test . . . . .	74
5.4.2	t-Test . . . . .	78
<b>6</b>	<b>Diskussion</b>	<b>83</b>
<b>A</b>	<b>Anhang</b>	<b>87</b>
A.1	Bedingte Unabhängigkeit . . . . .	87
	<b>Literaturverzeichnis</b>	<b>88</b>
	<b>Abbildungsverzeichnis</b>	<b>93</b>
	<b>Danksagung</b>	<b>95</b>
	<b>Lebenslauf</b>	<b>97</b>
	<b>Erklärung</b>	<b>99</b>

# 1

## Einleitung

Kontrollierte klinische Studien sind eine experimentelle Methode der klinisch-epidemiologischen Forschung, mit deren Hilfe der kausale Nachweis therapeutischer Wirksamkeit erbracht und somit der Therapiefortschritt abgesichert wird. Im Laufe einer solchen Studie stellte sich nicht selten heraus, dass die Annahmen, die der Studienplanung zugrundeliegen, einer Korrektur bedürfen. Der zu erwartende Therapieunterschied oder die Varianz der Messgröße kann sich als kleiner oder größer herausstellen als ursprünglich angenommen. Mit konventionellen statistischen Verfahren, war es nur bedingt möglich eine einmal geplante Studie nachträglich zu ändern, da nach datenabhängige Änderungen die statistische Fehlerkontrolle nicht mehr gesichert ist. Dies führte dazu, dass viele Studien aufgrund falscher Planungsannahmen nicht zufriedenstellend beendet werden konnten oder die statistische Fehlerkontrolle nicht mehr in der gebotenen Strenge beachtet wurde. In den letzten Jahren wurden verschiedene Methoden entwickelt, die Änderungen in laufenden Studien ermöglichen, ohne die Integrität der Studien zu verletzen.

Ziel dieser Arbeit ist eine Weiterentwicklung der von Müller und Schäfer

(2004) vorgeschlagenen Methode. Diese Methode (das sogenannte CRP-Prinzip) ermöglicht im Gegensatz zu anderen flexiblen Verfahren Änderungen an jedem beliebigen Zeitpunkt einer Studie.

Im Kapitel 2 geben wir zunächst eine Einführung in die besondere Problematik klinischer Studien. Kapitel 3 enthält einen Literaturüberblick über die bisherigen Gruppensequentiellen und flexiblen Designs. Im Kapitel 4 wird ein allgemeines mathematisches Konzept zur formalen Beschreibung und Definition solcher Verfahren entwickelt. Dieses Konzept stellt eine Erweiterung des bekannten Begriffs der statistischen Entscheidungsfunktion dar. Auf dieser Basis werden im Kapitel 5 die Hauptergebnisse dieser Arbeit entwickelt. Damit gelingt es, das CRP-Prinzip auf beliebige  $k$ -parametrische Exponentialfamilien zu verallgemeinern. Diese Verallgemeinerung wenden wir am Ende dieses Kapitels auf die Testung des Lageparameters einer Normalverteilung bei unbekannter Varianz (t-Test) und auf den Vergleich zweier unbekannter Binomialparameter (Fisher's exakter Test) an.

## 2

# Besondere Aspekte bei der Planung und Durchführung von klinischen Studien

Die in den nachfolgenden Kapiteln vorgestellten Verfahren wurden hauptsächlich für die Anwendung in klinischen Studien entwickelt. Für diese Verfahren existieren sicherlich noch weitere Anwendungsmöglichkeiten, aber der Haupteinsatzbereich liegt in einer flexibleren Gestaltung der Durchführung von klinischen Studien. Aus diesem Grund sind alle nachfolgenden Beispiele aus diesem Bereich gewählt.

Dieses Kapitel dient dazu, einen Überblick über die speziellen Anforderungen an klinische Studien im Vergleich zu sonstigen experimentellen Studien, wie sie in anderen Teilen der Wissenschaft, z.B. in den Agrarwissenschaften, der Soziologie und der Psychologie vorkommen, zu geben.

Dieses Kapitel beschränkt sich auf klinische Studien, deren Ziel die Prüfung neuer Therapien hinsichtlich der Wirksamkeit in der Behandlung von spezifischen Erkrankungen ist. Es handelt sich dabei um so genannte Therapiestudien. Weitere Studienarten in der klinischen Forschung dienen der Verbesserung der Diagnose und der Erstellung von Modellen zur Prognose und Risiken von Erkrankungen. Für diese Studienarten gelten viele der Aussagen in angepasster Form. Für eine ausführliche Einführung in klinische Studien sei z.B. auf Schumacher und Schulgen (2002) und Pocock (1983) verwiesen.

## 2.1 Die Phasen einer klinischen Prüfung

Ziel einer klinischen Prüfung ist die Untersuchung neu entwickelter Substanzen und Therapieformen hinsichtlich einer verbesserten Behandlung einer Erkrankung.

Nach einer vorklinischen Entwicklungsphase mit Laborexperimenten und Tierversuchen zur Untersuchung möglicher Wirkungsweisen einer neuen Substanz beginnt der Abschnitt der klinischen Prüfung. Sie wird unterteilt in vier Phasen (ICH E8, 1997):

**Phase I** An wenigen, meist gesunden Probanden wird hauptsächlich die Verträglichkeit und Pharmakokinetik (z.B. Verteilung im menschlichen Organismus) der Substanz untersucht.

**Phase II** An meist wenigen hundert Patienten wird versucht, eine wirksame Dosis zu finden und gleichzeitig die Wirksamkeit zu explorieren. Phase II Studien erstrecken sich in der Regel über einige Wochen bis hin zu einigen Monaten.

**Phase III** Die so genannte konfirmatorische Phase. In einer großen Studie (mehrere hundert bis tausend Patienten) soll der Nachweis der Wirksamkeit der neuen Substanz bzw. Therapieform erbracht werden. Die Wirksamkeit wird in diese Studie anhand einer für die Patienten relevanten Größe, wie Überlebenszeit, Anzahl der Todesfälle, etc. festgemacht. Gleichzeitig sollen in einer solchen Studie auch Daten über auftretende Nebenwirkungen gesammelt werden. Diese Studien sind Voraussetzung für die Zulassung eines neuen Arzneimittels und erstrecken sich mindestens über mehrere Monate, oftmals sogar über mehrere Jahre.

**Phase IV und Anwendungsbeobachtungen** Hierbei handelt es sich um die Studien nach der Einführung einer neuen Therapie. Sie dienen dazu, weitere Erkenntnisse über die zugelassene Therapie zu gewinnen. Dies betrifft insbesondere Informationen über seltene Nebenwirkungen und eventuelle Wechselwirkungen mit anderen Substanzen sowie zur Wirksamkeit in speziellen Subgruppen.

In allen Phasen ist die Teilnahme an den Studien freiwillig und setzt die Einwilligung des Patienten nach erfolgter Aufklärung voraus.

Die Entscheidung über eine Markteinführung einer Therapie am Ende der Phase III ist die schwerwiegendste im Verlaufe einer klinischen Prüfung. Zur Vermeidung von Fehlentscheidungen unterliegen die Phase III Studien auch den stärksten regulatorischen Bestimmungen, was sowohl ihre statistische Planung und Entscheidungsfindung als auch ihre allgemeine Durchführung betrifft. Die in den folgenden Kapiteln vorgestellten Verfahren sind primär für die Verwendung in Phase III Studien entwickelt worden. In neuerer Zeit gibt es allerdings auch immer mehr Vorschläge zur Anwendung dieser Verfahren in Phase II Studien zur Dosisfindung, z.B. Zhang et al. (2006) und König et al. (2006), und zur Kombination von Phase II und III Studien, z.B.

Todd und Stallard (2005) und Bretz et al. (2006). Im folgenden sollen diese Verfahren aber immer vor dem Anspruch einer Studie in der Phase III betrachtet werden.

## 2.2 Zielsetzung einer Phase III Studie

Wie bereits erwähnt, ist das Ziel einer Phase III Studie der Nachweis der Wirksamkeit einer neuen Therapie im Hinblick auf einen für die Patienten relevanten Endpunkt. Aufgrund des häufig zu beobachtenden Placeboeffektes, d.h. einer Verbesserung der Erkrankung im Rahmen einer Scheintherapie, bedingt z.B. durch das sich behandelte Fühlen des Patienten, und der natürlichen Schwankungen im Erkrankungsbild sollte der Nachweis der Wirksamkeit immer nur im Vergleich mit einer Placebothherapie oder einer bisherigen Standardtherapie erfolgen. Ein solcher Vergleich heißt kontrollierte Studie und die Vergleichstherapie wird häufig als Kontrolltherapie bezeichnet. In einer kontrollierten Studie gibt es verschiedene Ziele zum Nachweis eines Therapiefortschritts: Die häufigste Studienform ist die Überlegenheitsstudie. Sie hat zum Ziel, den Nachweis zu erbringen, dass die neue Therapie bezogen auf den Endpunkt der bisherigen überlegen ist, also z.B. eine Verlängerung der Überlebenszeit bewirkt. Eine weitere Form ist die Nichtunterlegenheitsstudie, deren Ziel es ist zu zeigen, dass die neue Therapie bezüglich des definierten Endpunktes der ursprünglichen höchstens in einem für den Patienten irrelevanten Maße unterlegen ist. Eine solche Studie wird häufig im Vergleich mit einer bisherigen Standardtherapie eingesetzt, wenn davon auszugehen ist, dass die neue Therapie z.B. weniger Nebenwirkungen hat oder einen anderen Vorteil besitzt. Im Rahmen einer so genannten Äquivalenzstudie soll der Nachweis der äquivalenten Wirksamkeit bezüglich des Endpunktes erbracht werden. Äquivalenzstudien unterscheiden sich von Nichtunterlegenheitsstudien insofern, dass bei letztgenannten auch der Nachweis der Überlegenheit



einer Therapie erbracht werden kann, während bei Ersteren das Studienziel in zwei gleichwertigen Therapien zu sehen ist.

## 2.3 Die Bedeutung von Fehlentscheidungen am Ende der Phase III

Fehlentscheidungen am Ende klinischer Therapiestudien haben oftmals gravierendere Folgen als in vielen andere Bereichen, in denen experimentelle Studien durchgeführt werden. Die möglichen Fehlentscheidungen und ihre Bedeutung für die Patienten sollen nun exemplarisch für Überlegenheitsstudien aufgezeigt werden.

Nach Beendigung einer Phase III Studie soll die Entscheidung über die Einführung einer neuen Therapie getroffen werden. Somit gibt es nur zwei mögliche Entscheidungen: Für oder gegen die Einführung der neuen Therapie.

Wird fälschlicherweise eine Entscheidung für die Therapie getroffen, obwohl diese nicht besser oder im ungünstigen Falle sogar schlechter ist als die Vergleichstherapie, so wird dies auch als Patientenrisiko bezeichnet. Im Falle eines Vergleiches mit einer Placebothherapie bedeutet dies, dass zukünftige Patienten mit einer Therapie behandelt werden, die nicht besser oder sogar schlechter als ein Placebo ist, zudem aber noch spezifische Risiken in Form von Nebenwirkungen besitzt. Somit wird deutlich, dass die Wahrscheinlichkeit für das Auftreten einer solchen Fehlentscheidung klein sein sollte.

Wird fälschlicherweise eine Entscheidung gegen die Therapie getroffen, obwohl diese der Vergleichstherapie überlegen ist, bedeutet dies, dass eine bessere Therapie den Patienten vorenthalten wird. Im Falle einer Therapie, deren

Ziel eine Verringerung der Todesfallrate einer Erkrankung ist, bedeutet z.B. eine Verringerung der Rate um 1% im Vergleich zur bisherigen Therapie, dass mit der neuen Therapie pro 100 Patienten ein Leben gerettet werden könnte. Damit wird offensichtlich, dass zumindest im Falle großer Therapieunterschiede die Wahrscheinlichkeit für eine solche Fehlentscheidung ebenfalls gering sein sollte. Hinzu kommt noch das Interesse des Produzenten, sein neues Produkt, sofern es denn überlegen ist, einzuführen, nachdem er eventuell etliche Millionen in die Forschung und Entwicklung investiert hat.

## 2.4 Vermeidung von Fehlentscheidungen durch geeignete Studiendurchführung

Es gibt zwei mögliche Ursachen, die zu einer Fehlentscheidung führen können: Zum einen, die zufällige Streuung der Daten und zum anderen, eine systematische Verzerrung der Daten.

Kommt als Ursache für mögliche Fehlentscheidungen nur die zufällige Streuung der Daten in Frage, können die Wahrscheinlichkeiten für ihr Eintreten durch statistische Methoden kontrolliert werden.

Liegt zusätzlich eine systematische Verzerrung der Daten vor, ist die Kontrolle der Wahrscheinlichkeiten nicht mehr gesichert. Deshalb wird versucht, durch geeignete Studiendurchführung mögliche Verzerrungen der Ergebnisse zu vermeiden. Dazu dienen:

**Randomisierung** Zur Vermeidung von Verzerrungen aufgrund bekannter und unbekannter prognostischer Faktoren sollten die Patienten nach dem Zufallsprinzip auf die Gruppen zugeteilt werden.

**Verblindung der Patienten und der Beurteiler** Zur Vermeidung von Unterschieden in der Erwartungshaltung der Patienten und der über den Therapieerfolg urteilenden Ärzte (Personen) sollten beide Personengruppen keine Kenntnis über die Gruppenzugehörigkeit besitzen.

**Verblindung der behandelnden Ärzte** Zur Vermeidung einer bevorzugten Behandlung von Patienten einer Therapiegruppe, insbesondere hinsichtlich Begleittherapien, sollten alle in die Behandlung eingebundenen Personen, besonders die behandelnden Ärzte und evtl. auch Schwestern, die Gruppenzugehörigkeit der Patienten nicht kennen.

## 2.5 Einige ethische Aspekte

Sofern es sich nicht vollständig vermeiden lässt, sollte es immer das Ziel sein, so wenig Patienten (Probanden) wie möglich einem erhöhten Risiko auszusetzen. Für klinische Studien bedeutet dies, eine Entscheidung zu dem frühestmöglichen Zeitpunkt zu treffen. In Phase II Studien zur Dosisfindung sollten mögliche Dosen mit schweren Nebenwirkungen früh erkannt werden und die entsprechenden Dosen dann nicht weiter verwendet werden.

In vergleichenden Studien sollte die Anzahl der Patienten, die mit einer unterlegenen Therapie behandelt werden, so klein wie möglich gehalten werden. Das Ziel einer geringen Anzahl an Patienten wird noch offensichtlicher, wenn berücksichtigt wird, dass viele Studien aufgrund der langsamen Patientenrekrutierung über viele Jahre laufen und dass an diesen Studien nur ein kleiner Teil der Patienten teilnimmt. Die Patienten außerhalb der laufenden Studie werden solange mit der bisherigen Therapie behandelt, bis eine Studie den eventuellen Nachweis der Überlegenheit der neuen Therapie erbracht hat. Im Falle der Überlegenheit einer neuen Therapie kann also allein die Verkürzung der Studiendauer einige Menschenleben retten.

Dem gegenüber steht die Notwendigkeit für eine vernünftige Absicherung der Entscheidung, so dass eine Fehlentscheidung nur in wenigen Fällen getroffen wird. Dies kann nur durch eine hinreichende Anzahl an Beobachtungen geschehen.

Zusammenfassend gilt damit, so viele Patienten wie für die statistische Absicherung der Entscheidung nötig, aber so wenige wie möglich, in eine Studie einzubringen.

## 2.6 Ziele der statistischen Studienplanung

Die Aufgaben der statistischen Planung von klinischen Studien werden in diesem Abschnitt zunächst für den Fall der Überlegenheitsstudien besprochen. Vieles lässt sich direkt auf die anderen Formen übertragen. Die wichtigsten Besonderheiten werden am Ende des Abschnittes erläutert.

Ziel einer Überlegenheitsstudie ist der Nachweis, dass die neue Therapie bezüglich einer definierten Zielgröße besser ist als die bisherige. Dies bedeutet, es soll zwischen der Nullhypothese der Unterlegenheit oder Gleichheit und der Alternative Überlegenheit der neuen Therapie entschieden werden. Aus ethischen Gründen sollte eine vorhandene Überlegenheit der bisherigen Therapie genauso entdeckt werden wie die Überlegenheit der neuen Therapie, so dass die Fragestellung zweiseitig zu untersuchen ist. Die zu testenden Hypothesen lauten dann: Nullhypothese, es gibt keinen Unterschied, versus der Alternativhypothese, es gibt einen Unterschied zwischen den Therapien. Da die beiden Alternativhypothesen, Überlegenheit der neuen bzw. der bisherigen Therapie, gleichberechtigt untersucht werden sollen, ist es eine Besonderheit, in klinischen Studien anstelle eines zweiseitigen Tests die Kombination von zwei einseitigen Tests zu verwenden. Im Falle von symmetrischen

Verteilungen, wie z.B. der Normalverteilung, ist die Kombination identisch mit einem zweiseitigem Test. Es gibt aber auch viele Beispiele, bei denen dies nicht der Fall ist, z.B. im Falle einer binären Zielgröße und der Verwendung von Fisher's exaktem Test.

Somit entspricht ein Fehler 1. Art, d.h. ein Verwerfen der wahren Nullhypothese, der fälschlichen Behauptung, dass eine Therapie der anderen überlegen ist. Ein Fehler 2. Art, Beibehalten der Nullhypothese, obwohl die Alternative wahr ist, entspricht dem nicht erkennen einer überlegenen Therapie.

Eine wichtige Aufgabe der statistischen Planung ist die Wahl eines geeigneten Testverfahrens, so dass die Einhaltung der vorgegebenen Wahrscheinlichkeit für den Fehler 1. Art sichergestellt ist (ICH E9, 1998). In klinischen Studien wird die Wahrscheinlichkeit für den Fehler 1. Art (Signifikanzniveau, Fehlerniveau 1. Art) häufig auf 5% festgesetzt, entsprechend für jeden der einseitigen Tests auf 2.5%. Im Gegensatz zu der Wahrscheinlichkeit für den Fehler 1. Art, die durch die Wahl eines geeigneten statistischen Testverfahrens kontrolliert werden kann, hängt die Wahrscheinlichkeit für einen Fehler 2. Art bekanntlich von der Größe des Therapieunterschiedes und von dem Stichprobenumfang (Fallzahl) ab. Bei der Planung einer Studie sollte die Fallzahl so festgelegt werden, dass ein vorhandener Therapieunterschied in festgelegter Größenordnung höchstens mit einer festgelegten Wahrscheinlichkeit nicht entdeckt wird, bzw. anders formuliert, mindestens mit einer festgelegten Wahrscheinlichkeit (Power) entdeckt wird. Typische Werte für die Power sind 80% oder 90%.

Eine Schwierigkeit bei der Wahl der geeigneten Fallzahl liegt in der Festlegung des Therapieunterschiedes, der mit der vorgegebenen Power entdeckt werden soll. Als zu entdeckender Unterschied sollte der minimale noch für den Patienten relevante Therapieunterschied gewählt werden. Dies bedeutet, der Unterschied, ab dem für den Patienten eine merkliche Verbesserung

seiner Situation eintritt. Allerdings ist es in der Realität häufig schwierig, diesen minimalen Unterschied zu bestimmen. So ist es offensichtlich, dass eine Verlängerung der Überlebenszeit um einen Tag nicht relevant ist, eine Verlängerung um ein Jahr hingegen schon. Die Frage ist, wo liegt die Grenze? Diese Frage kann nur in enger Zusammenarbeit mit den Medizinern beantwortet werden. Häufig werden verschiedene Gruppen zu unterschiedlichen Ergebnissen kommen, weshalb es durchaus vorkommen kann, dass verschiedene Studien zu derselben Erkrankung mit unterschiedlichen zu entdeckenden Therapieunterschieden geplant werden. Es kann sein, dass im Verlauf der Studie die ursprüngliche Einschätzung revidiert werden muss.

Zudem hängt die benötigte Fallzahl selbst in einfachen Situationen nicht nur von dem Therapieunterschied, sondern zusätzlich auch noch von der Streuung der Daten ab. Um die Streuung der Daten abzuschätzen, werden häufig die Informationen zur Streuung aus ähnlichen Studien oder aus so genannten Vorstudien, welches z.B. die Phase II Studien sein können, verwendet. Aufgrund eines unterschiedlichen Patientenkollektives, Verbesserung in der Begleittherapie, usw., weicht deren Streuung oftmals von der Streuung der zu planenden Studie ab. In Abbildung 2.1 ist für den Fall einer normalverteilten Teststatistik und Anwendung eines einseitigen Gausstests die Veränderung der Power für Abweichungen von den Planungsannahmen dargestellt.

Aufgrund der im vorherigen Abschnitt beschriebenen ethischen Gründe sollte die Fallzahl nicht zu groß gewählt werden. Deshalb erscheint es wünschenswert, dass es in klinischen Studien die Möglichkeit gibt, datenabhängig die Planung nochmals zu korrigieren oder zumindest zwischen mehreren, zur Planungsphase festgelegten Verläufen zu entscheiden.

Ähnliche Probleme bestehen auch bei der Planung von Nichtunterlegenheits- und Äquivalenzstudien. In diesen Studien liegt die Schwierigkeit nicht in der Festlegung des minimalen relevanten Therapieunterschiedes, sondern in

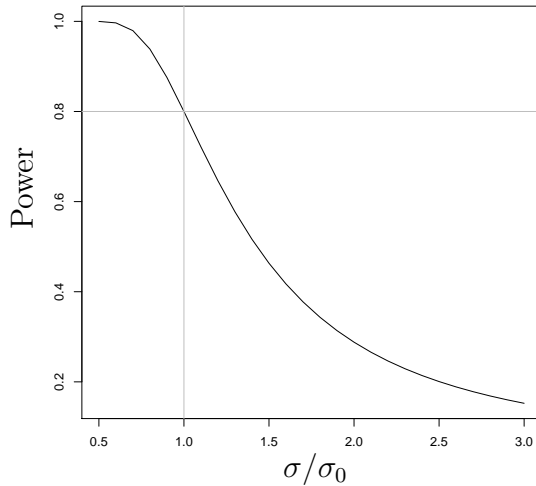


Abbildung 2.1: Power in Abhängigkeit vom Quotienten  $\sigma/\sigma_0$  eines einseitigen Gausstests zum Niveau 2.5%. Dieser ist geplant für ein festes  $\mu_0, \sigma_0$  mit einer Power von 80%.

der Festlegung der Grenzen für den noch tolerierbaren Unterschied, um von einer Gleichwertigkeit der Therapie zu sprechen. Dieses Problem ist eng verknüpft mit der Festlegung des relevanten Unterschiedes. Die Fallzahl einer Nichtunterlegenheits- oder Äquivalenzstudie hängt in demselben Maße von der Festlegung der Grenzen ab, wie in einer Überlegenheitsstudie von dem relevanten Unterschied. Allgemein kann gesagt werden, dass die Fallzahlen in den letzten beiden Studientypen im Durchschnitt größer sind als in einer Überlegenheitsstudie.





# 3

## Spezielle Studiendesigns

Im vorherigen Kapitel wurde dargestellt, warum es für klinische Studien mit langer Laufzeit oder großer benötigter Fallzahl sinnvoll sein kann, den Verlauf datenabhängig zu wählen. Allerdings ist es bei der Verwendung herkömmlicher statistischer Testverfahren, wie z.B. Gausstest oder  $t$ -Test, nicht möglich, datenabhängig Änderungen der Planungsannahmen vorzunehmen und dabei die Einhaltung des Fehlerniveaus 1. Art sicherzustellen. Dies wird an folgendem einfachem Beispiel offensichtlich: Angenommen, nach der Hälfte der geplanten Fallzahl wird auf die Daten ein Niveau- $\alpha$ -Test angewendet. Kann mit diesem die Nullhypothese verworfen werden, so wird die Studie gestoppt und die Nullhypothese verworfen. Anderenfalls wird die Studie bis zum geplanten Ende durchgeführt, mit den Gesamtdaten zum Niveau- $\alpha$  getestet und entsprechend entschieden. Bei einem derartigen Vorgehen ist die Gesamtwahrscheinlichkeit für einen Fehler 1. Art nahe an  $2\alpha$  und somit fast doppelt so hoch wie vorgesehen.

In diesem Kapitel werden nun einige Studiendesigns vorgestellt, bei denen die weitere Vorgehensweise in der Studie von den Zwischenergebnissen abhängt

und die dennoch die Einhaltung des Fehlerniveaus 1. Art garantieren. In diesen Designs kann z.B. die Fallzahl, die Randomisierungswahrscheinlichkeiten der Patienten zu den Therapiegruppen oder sogar das gesamte weitere Studiendesign von den Zwischenergebnissen abhängen.

Bevor mit der eigentlichen Vorstellung der Designs begonnen wird, sollen einige Begriffe aus der Statistik eingeführt werden. Der folgende Abschnitt versucht dabei, ohne exakte mathematische Definitionen auszukommen. Für diese sei auf die Lehrbücher der mathematischen Statistik, wie Witting (1985) oder Lehmann und Romano (2005), verwiesen. Ein Teil der Begriffe wird auch im Kapitel 4.1 definiert.

### 3.1 Beschreibung eines statistischen Tests

Ziel einer statistischen Entscheidungsprozedur ist es, Aussagen über die unbekannte Verteilung der Daten zu machen. Häufig kann die Menge der in Frage kommenden Verteilung durch Annahmen eingeschränkt werden, so dass nur noch eine Teilmenge  $\Theta$  untersucht werden muss. Diese Teilmenge lässt sich oftmals mit Hilfe von *Parametern* beschreiben. Ein klassisches Beispiel ist die Menge aller Normalverteilungen, die mit Hilfe des Erwartungswertes und der Varianz beschrieben werden kann.

Der nachfolgende Teil dieses Abschnittes wird vereinfachend nur für den Fall des unbekanntes Parameters  $\vartheta$  formuliert. Allerdings kann eine unbekannte Verteilung auch ein Parameter sein, so dass dies keine Einschränkung bedeutet.

Einem *statistischen Test* liegt eine spezielle Entscheidungsprozedur zugrunde. Die Menge der möglichen Parameter sei in zwei Teilmengen  $\Theta_0$  und  $\Theta_1$

aufgeteilt. Ziel eines statistischen Tests ist es zu entscheiden, in welcher der beiden Teilmengen der unbekannte Parameter liegt. Diese Teilmengen bilden die zu testende *Nullhypothese*  $H_0 : \vartheta \in \Theta_0$  und die *Alternativhypothese*  $H_1 : \vartheta \in \Theta_1$ . Ein Test trifft also eine Entscheidung zwischen der Nullhypothese und der Alternativhypothese. Hierbei sind, wie schon im Kapitel 2 erwähnt, zwei Fehlentscheidungen möglich:

**Fehler 1. Art** In Wahrheit gilt die Nullhypothese, d.h.  $\vartheta \in \Theta_0$ , aber der Test entscheidet für die Alternative

**Fehler 2. Art** In Wahrheit gilt die Alternativhypothese, d.h.  $\vartheta \in \Theta_1$ , aber der Test entscheidet für die Nullhypothese.

Aufgabe eines statistischen Tests ist es, die Wahrscheinlichkeit, einen Fehler 1. Art zu begehen, zu kontrollieren. Dies bedeutet, dass die Wahrscheinlichkeit mit diesem Test für einen Fehler 1. Art unter einer vorgegebenen, kleinen Schranke  $\alpha$  liegen soll. Diese Schranke wird auch als das  $\alpha$ -Niveau des Tests bezeichnet.

Die Wahrscheinlichkeit für einen Fehler 2. Art kann nicht gleichfalls kontrolliert werden. Sie hängt von der Größe des wahren Wertes des Parameters  $\vartheta \in \Theta_1$  ab. Häufig wird anstelle der Wahrscheinlichkeit für einen Fehler 2. Art ihre Gegenwahrscheinlichkeit betrachtet. Dies ist die Wahrscheinlichkeit unter der Alternativhypothese, dass sich der Test für diese entscheidet. Diese Wahrscheinlichkeit wird auch als *Power* eines Tests bezeichnet und da sie von dem Wert des Parameters abhängt, oft auch als *Powerfunktion*  $\beta(\vartheta)$ . Einen "guten" Test zeichnet aus, dass die Power umso größer wird, je weiter "entfernt" der Parameter von der Nullhypothese ist. Für einen festen Wert des Parameters sollte sie mit einer wachsenden Zahl an Beobachtungen, die in die Entscheidung einfließen, ebenfalls größer werden. Aus diesem Grund muss die Fallzahl geeignet gewählt werden, um die Power über bzw. die

Wahrscheinlichkeit für einen Fehler 2. Art unter eine vorgegebene Schranke zu bekommen.

Häufig wird nach Beobachtung der Daten die Entscheidung für das Verwerfen oder Beibehalten der Nullhypothese mit Hilfe einer (*Test-*)statistik  $T$  getroffen. Dies ist eine Größe, die die vielen Beobachtungen  $X_1, \dots, X_n$  zu einem Wert zusammenfasst. Ein Beispiel für eine Statistik ist der Mittelwert  $\bar{X}_n = \frac{1}{n} \sum_{i=1}^n X_i$ , der aber direkt in Tests selten Anwendung findet. Eine Variante der Formulierung der Entscheidungsregel eines Tests mit Hilfe der Teststatistik  $T$  ist

$$H_0 \text{ verwerfen} \Leftrightarrow T > c.$$

Hierbei ist  $c$  ein vorher festgelegter Wert, die so genannte *kritische Grenze*, der so gewählt wurde, dass der Test das Niveau  $\alpha$  einhält. Deshalb findet auch häufig die Schreibweise  $c_\alpha$  Verwendung.

Insbesondere in der Medizin wird die Entscheidungsregel häufig mit Hilfe des *p-Wertes* formuliert. Der *p-Wert* kann dabei als eine spezielle Teststatistik aufgefasst werden. Es gibt zwei Ansätze, ihn zu definieren:

- Als Wahrscheinlichkeit, unter der Nullhypothese noch "extremere" Beobachtungen zu erhalten als die aktuellen.
- Als das kleinstmögliche Niveau, für das die Nullhypothese mit den vorliegenden Daten noch abgelehnt werden kann.

Obwohl diese Definitionen sehr unterschiedlich klingen, stimmen sie in fast allen Fällen überein. Mit Hilfe des *p-Wertes* hat die Entscheidungsregel dann die folgende Form

$$H_0 \text{ verwerfen} \Leftrightarrow p \leq \alpha.$$

Bisher wurde davon ausgegangen, dass die Menge der Verteilungen nur von einem unbekanntem reellen Parameter abhängt. Dies ist in der Praxis häufig nicht der Fall, wie schon an dem oben erwähnten Beispiel der Menge aller Normalverteilungen deutlich wird. Hier ist es in der Regel so, dass nicht nur der Erwartungswert  $\delta$  unbekannt ist, sondern zusätzlich auch die Varianz  $\sigma^2$ . Trotzdem soll häufig nur über die Größe von  $\delta$  entschieden werden, z.B.  $\Theta_0 = \{\delta \leq 0\}$ . In einem solchen Fall wird  $\delta$  als der *interessierende Parameter* und  $\sigma$  als ein *Störparameter* bezeichnet. Der geeignete Test für die beschriebene Situation der Normalverteilung ist der *t*-Test.

## 3.2 Designs mit adaptiver Zuteilungsregel (Adaptive Allocation Designs)

Gewöhnlich werden die Randomisierungswahrscheinlichkeiten der Patienten zu den Gruppen so gewählt, dass diese möglichst gleich groß werden. Im Falle des Vergleiches von zwei Gruppen wird typischerweise ein Patient mit Wahrscheinlichkeit  $\frac{1}{2}$  der einen oder der anderen Gruppe zugeteilt.

Um in einer Studie möglichst wenig Patienten mit der unterlegenen Therapie zu behandeln, bestimmen die Designs mit adaptiver Zuteilung die Randomisierungswahrscheinlichkeiten zu den einzelnen Gruppen in Abhängigkeit der bisherigen Ergebnisse. Damit werden einer Therapiegruppe vermehrt Patienten zugeteilt, sobald sich die entsprechende Therapie als überlegen abzeichnet.

Unter den in diesem Kapitel vorgestellten, speziell auf klinische Studien ausgerichteten Designs gehören diese zu den ältesten. Für eine Einführung sei

z.B. auf den Übersichtsartikel von Rosenberger (1996) und die dortigen Referenzen verwiesen.

### 3.3 Designs mit interner Pilotstudie (Internal Pilot Study Designs)

#### 3.3.1 Problem der Fallzahlbestimmung beim Vorliegen eines Störparameters

Im Falle des Vorliegens von Störparametern ist die Power des Tests nicht nur eine Funktion des interessierenden Parameters, sondern hängt zusätzlich auch von den Störparametern ab. Aus diesem Grund ist es für die Bestimmung der notwendigen Fallzahl erforderlich, die Werte der Störparameter zu schätzen. Dies geschieht für klinische Studien häufig auf der Basis bisheriger Studien zu einer ähnlichen Fragestellung. Allerdings weichen die so ermittelten Werte oftmals stark von den wahren Werten ab, was zu Studien mit zu hoher Power (überpower) aufgrund einer unnötig großen Fallzahl oder zu Studien mit zu kleiner Power (unterpower) aufgrund einer zu geringen Fallzahl führt.

Ein typisches Beispiel für die Abhängigkeit von einem unbekanntem Störparameter ist der im Abschnitt 3.1 erwähnte, häufig verwendete  $t$ -Test. In dem Fall der einseitigen Nullhypothese  $H_0 : \delta \leq 0$  ist die Powerfunktion eines Niveau- $\alpha$ -Testes mit Fallzahl  $n$  gegeben durch

$$\beta(\delta, \sigma^2) = 1 - F_{t_{n-1; \sqrt{n} \frac{\delta}{\sigma}}}(t_{n-1; 1-\alpha}), \quad (3.1)$$

wobei  $F_{t_{df, \lambda}}$  die dezentrale  $t$ -Verteilung mit  $df$  Freiheitsgraden und Nichtzentralitätsparameter  $\lambda$  und  $t_{df; p}$  das  $p$ -Quantil einer zentralen  $t$ -Verteilung mit  $df$  Freiheitsgraden bezeichnet. Die Abhängigkeit der Power von der Schätzung

der Varianz verdeutlicht das folgende Beispiel eines einseitigen  $t$ -Tests zum Niveau 2.5% mit einer Fallzahl von  $n = 34$  Beobachtungen. Dieser Test besitzt für einen Effekt von  $\delta = 0.5$  und einer angenommenen Varianz von  $\sigma^2 = 1$  eine Power von ca. 81%. Ist die Varianz hingegen doppelt so groß ( $\sigma^2 = 2$ ) verringert sich die Power bei gleichem  $\delta$  auf nur noch 52%, ist die Varianz nur halb so groß ( $\sigma^2 = 0.5$ ) erreicht die Power fast 98%. Solche Bereiche für die Unsicherheit in der Schätzung der wahren Varianz sind in der Praxis durchaus üblich.

### 3.3.2 Fallzahlneubestimmung im Verlauf einer Studie

Um über- oder unterpowerete Studien aufgrund der beschriebenen Probleme bei der Fallzahlbestimmung zu Studienbeginn zu vermeiden, kam die Idee auf, den unbekanntem Störparameter im Verlauf der Studie aus den bisherigen Daten zu schätzen und mit diesem eine neue Gesamtfallzahl zu bestimmen. Erstmals wurde eine solche Vorgehensweise von Stein (1945) für die  $t$ -Test Situation beschrieben. Stein's Zwei-Stufen-Prozedur verwendet anstelle der  $t$ -Teststatistik  $T_n = \sqrt{n}\bar{X}_n/S_n$ , die  $t$ -verteilt mit  $n - 1$  Freiheitsgraden ist, die Statistik

$$T^* = \sqrt{n} \frac{\bar{X}_n}{S_0},$$

wobei  $S_0^2 = \frac{1}{n_0-1} \sum_{i=1}^{n_0} (X_i - \bar{X}_{n_0})^2$  der Varianzschätzer aus der ersten Stufe basierend auf  $n_0$ -Beobachtungen ist. Die Gesamtfallzahl  $n \geq n_0$  kann nach der ersten Stufe in Abhängigkeit vom beobachteten Wert von  $S_0$  gewählt werden. Allerdings darf sie nicht vom beobachteten Mittelwertschätzer der ersten Stufe abhängen. Unter diesen Voraussetzungen ist die Statistik unter der Nullhypothese  $t$  verteilt mit  $n_0 - 1$  Freiheitsgraden. Stein zeigt, dass für jedes  $\delta$  und jede Power  $1 - \beta$  eine Konstante  $c$  bestimmt werden kann, so dass bei Berechnung der Gesamtfallzahl  $n$  aus den Beobachtungen der 1. Stufe mit

der Regel

$$n = \max\{n_0, \lceil \frac{S_0^2}{c} \rceil\},$$

diese zweistufige Vorgehensweise einen Test liefert, dessen Powerfunktion  $\beta(\delta; \sigma)$  größer  $1 - \beta$  für alle  $\sigma$  ist.

Obwohl Stein's Zwei-Stufen Prozedur die Einhaltung einer vorgegebenen Power für alle Werte der unbekanntem Varianz garantiert, findet sie in klinischen Studien kaum Verwendung. Dies liegt vermutlich daran, dass die Varianz nur aus den Daten der 1. Stufe geschätzt wird und dieses zu einem vermeintlichen Powerverlust führt. Insbesondere in den 90er Jahren wurden weitere Prozeduren zur Neuberechnung der Fallzahl im Verlaufe klinischer Studien basierend auf einer Schätzung des Störparameters aus den bisherigen Daten vorgeschlagen. Diese betrachten hauptsächlich die  $t$ -Test Situation, z.B. Wittes et al. (1999); Zucker et al. (1999); Denne und Jennison (1999); Kieser und Friede (2000, 2003), und die Situation des Vergleichs zweier Gruppen mit binärem Endpunkt, z.B. Friede und Kieser (2004). In der  $t$ -Test Situation verwenden die meisten dieser Verfahren die herkömmliche  $t$ -Statistik. Sie unterscheiden sich in der Anpassung der kritischen Grenze zur Verringerung der durch die Neuberechnung entstandenen Inflation des  $\alpha$ -Niveaus, der Neuberechnungsregel sowie in der Verwendung der verblindeten oder unverblindeten Daten zur Neuberechnung. Eine Verwendung der verblindeten Daten beim Zwei-Gruppen-Vergleich bedeutet, dass z.B. die Varianz aus den Beobachtungen ohne Kenntnis der Gruppenzugehörigkeit geschätzt wird. Einzig das Verfahren von Kieser und Friede (2000) kontrolliert das Fehlerniveau 1. Art

Diese zweistufigen Verfahren werden häufig als Designs mit interner Pilot-Studie (internal pilot study designs) bezeichnet. Eine gute Übersicht über die unterschiedlichen Vorschläge gibt die Arbeit von Friede und Kieser (2006).



## 3.4 Gruppensequentielle Designs

Sequentielle Verfahren wären aufgrund ihrer Eigenschaften, Einhaltung der vorgegebenen Wahrscheinlichkeiten für die Fehler 1. und 2. Art und gleichzeitig Minimierung der erwarteten Fallzahl, gut geeignet für die Verwendung in klinischen Studien. Aus praktischen Gründen finden sie dort aber sehr selten Anwendung. Denn jede Auswertung einer Studie erfordert in der Regel die Zusammenkunft eines so genannten Data Monitoring Komitees und eine vorherige Überprüfung der Daten auf ihre Richtigkeit. Aus diesem Grunde wurde von Armitage et al. (1969) erstmals vorgeschlagen, nicht nach jeder neuen Beobachtung einen Signifikanztest auf den kumulierten Daten durchzuführen, sondern nach jeder neuen Gruppe von fester Größe, bis eine maximale Anzahl an Gruppen erreicht ist.

Generell lässt sich ein gruppensequentielles Design folgendermaßen beschreiben: Es sei  $K$  die Anzahl der betrachteten Gruppen (Auswertungen) und  $n_i, i = 1, \dots, K$  die Gesamtzahl an Beobachtungen bis zur  $i$ -ten Auswertung. Zum Testen der Signifikanz an der  $i$ -ten Auswertung dient die Teststatistik  $T_i$  basierend auf allen bisherigen Beobachtungen  $X_1, \dots, X_{n_i}$ . Angenommen, große Werte der Teststatistik  $T_i$  sprechen gegen ein Vorliegen der Nullhypothese, dann soll die Nullhypothese an der  $i$ -ten Auswertung verworfen werden, wenn gilt  $T_i > c_i$  für eine vorher festgelegte kritische Grenze  $c_i$ . Für  $T_i \leq c_i$  soll die Studie mit der nächsten Gruppe fortgesetzt werden ( $i = 1, \dots, K - 1$ ) bzw. die Nullhypothese angenommen werden ( $i = K$ ).

Die Wahrscheinlichkeit, die Nullhypothese bis zum Ende der Studie nicht zu verwerfen, ist gegeben durch  $P(T_i \leq c_i, i = 1, \dots, K)$ . Diese lässt sich bei Kenntnis der gemeinsamen Verteilung der  $T_i$  mit Hilfe numerischer Integration bestimmen. Insbesondere ist es dann möglich, die Wahrscheinlichkeit für einen Fehler 1. Art zu bestimmen und die kritischen Grenzen  $c_i$  so zu wählen,

dass der gruppensequentielle Test das Niveau  $\alpha$  einhält.

In der Praxis wird zur Vereinfachung der Berechnungen häufig angenommen, dass die Teststatistiken  $T_i$  zu den 'Informationszeiten'  $t_1 < \dots < t_i < \dots < t_K$  eine Brownsche Bewegung mit Drift  $\delta$  bilden, also

- $T_i \sim N(t_i\delta, t_i)$
- $T_i - T_{i-1}$  und  $T_{i-1}$  sind unabhängig

Die Informationszeiten  $t_i$  sind z.B. gegeben durch  $n_i/n_K$ .

Die Anwendung dieser Approximation ist aufgrund der häufig großen Fallzahl in Studien mit gruppensequentielle Design durchaus gerechtfertigt und von guter Güte. Für einige wenige Verteilungssituationen existieren aber auch exakte gruppensequentielle Verfahren, wie z.B. in der  $t$ -Test Situation (Jennison und Turnbull, 1991).

Für eine ausführliche Darstellung der gruppensequentiellen Verfahren sei auf das Buch von Jennison und Turnbull (2000) verwiesen.

### 3.5 Flexible (adaptive) Designs

Noch mehr Flexibilität in der Durchführung einer Studie als die bisherigen Designs erlauben die so genannten flexiblen, vormals auch adaptiven, Designs. Sie wurden erstmals von Bauer (1989) vorgestellt. Die Grundidee ist, die Studie in mehrere Abschnitte, so genannte Stufen, zu unterteilen, wobei jede neue Stufe mit Hilfe sämtlicher Informationen aus den vorhergehenden geplant werden kann. Diese Stufen werden dann ähnlich einer Meta-Analyse zu

einer Gesamtstudie zusammengebunden. Im folgenden sollen nur Designs mit zwei Stufen betrachtet werden. Alle mehrstufigen adaptiven Designs können auch als rekursive Anwendung zweistufiger Designs betrachtet werden, vgl. Brannath et al. (2002).

Allgemein kann die Vorgehensweise eines flexiblen Designs wie folgt dargestellt werden:

**Zu Studienbeginn** Planung der 1. Stufe, u.a.

- Festlegung der Fallzahl
- Festlegung des zu verwendenden Tests bzw. der Teststatistik,

sowie Festlegung, wie die Stufen kombiniert werden sollen

**Nach der 1. Stufe** Planung der 2. Stufe mit Hilfe der Informationen aus der ersten, insbesondere

- Bestimmung der Fallzahl mit Hilfe der Schätzung für die Störparameter

**Nach der 2. Stufe** Entscheidung für oder gegen die Nullhypothese aufgrund der kombinierten Informationen aus beiden Stufen

Für die flexiblen Designs gibt es zwei verschiedene Darstellungen: Zum einen die Kombination der beiden Stufen mit Hilfe einer Kombinationsregel, wie sie auch in der Meta-Analyse verwendet wird, und zum anderen die Verwendung von bedingten Fehlerfunktionen. In einer Arbeit von Vandemeulebroecke (2006) wurde gezeigt, dass diese beiden Darstellungen äquivalent sind. Trotzdem sollen hier beide Formen kurz vorgestellt werden.

### 3.5.1 Flexible Designs basierend auf einer Kombinationsregel

In Anlehnung an die in Meta-Analysen verwendete Fisher'sche Kombinationsregel schlug Bauer (1989) folgende Regel vor, um die Informationen der beiden Stufen zu einer gemeinsamen Entscheidung zu kombinieren:

$$H_0 \text{ verwerfen} \Leftrightarrow p_1 p_2 \leq c_\alpha, \quad (3.2)$$

dabei bezeichnen  $p_1$  und  $p_2$  die  $p$ -Werte aus der ersten bzw. zweiten Stufe. Die Grenze  $c_\alpha$  sollte so gewählt werden, dass dieser Gesamttest das Fehlerniveau  $\alpha$  einhält. Im obigen Fall sollte  $c_\alpha = \exp(-\frac{1}{2} \chi_{4,1-\alpha}^2)$  gewählt werden, wobei  $\chi_{df,p}^2$  das  $p$ -Quantil einer  $\chi^2$ -Verteilung mit  $df$  Freiheitsgraden bezeichne.

Üblicherweise werden die  $p$ -Werte der einzelnen Stufen dazu verwendet, um die gemeinsame Entscheidungsregel zu formulieren. Dies kann geschehen durch

$$H_0 \text{ verwerfen} \Leftrightarrow C(p_1, p_2) \leq c_\alpha.$$

Um "vernünftige" Entscheidungsregeln zu erhalten, sollte die Kombinationsregel  $C(p_1, p_2)$  monoton wachsend in beiden  $p$ -Werten sein.

Das Design von Bauer erlaubt es, bei kleinen  $p$ -Werten in der ersten Stufe die Studie an diesem Zeitpunkt mit Verwerfen der Nullhypothese zu beenden. Dies ist immer dann möglich, wenn gilt  $p_1 \leq c_\alpha$ . Nachfolgende flexible Designs erlauben oftmals ein Stoppen nach der ersten Stufe sowohl mit Verwerfen als auch mit Beibehalten der Nullhypothese in Analogie zu den gruppensequentiellen Designs. Diese Designs lassen sich mit folgender, allgemeinerer Entscheidungsregel darstellen:

$$H_0 \text{ verwerfen} \Leftrightarrow \begin{cases} p_1 \leq \alpha_1 & \text{nach der 1. Stufe,} \\ C(p_1, p_2) \leq c_\alpha & \text{nach der 2. Stufe} \end{cases}$$

und

$$H_0 \text{ beibehalten} \Leftrightarrow \begin{cases} p_1 \geq \alpha_0 & \text{nach der 1. Stufe,} \\ C(p_1, p_2) \geq c_\alpha & \text{nach der 2. Stufe.} \end{cases}$$

Hierbei sollten die Grenzen  $\alpha_0$ ,  $\alpha_1$  und  $c_\alpha$  so gewählt werden, dass die gesamte Entscheidungsregel das Niveau  $\alpha$  einhält. Es muss also gelten

$$\alpha_1 + P_{H_0}(C(p_1, p_2) \leq c_\alpha, \alpha_0 > p_1 > \alpha_1) \leq \alpha.$$

Eine weitere gebräuchliche Kombinationsregel basiert auf der Inverse-Normal-Funktion

$$C(p_1, p_2) = w_1 \Phi^{-1}(p_1) + w_2 \Phi^{-1}(p_2),$$

wobei  $\Phi^{-1}$  die Quantil-Funktion der Standardnormalverteilung bezeichne und die Gewichte  $w_1, w_2$  so gewählt werden sollten, dass  $w_1^2 + w_2^2 = 1$  gilt. Diese Kombinationsregel wurde für die flexiblen Designs erstmals von Lehman und Wassmer (1999) vorgeschlagen.

### 3.5.2 Flexible Designs basierend auf einer bedingten Fehlerfunktion

Die Idee zur Verwendung bedingter Fehlerfunktionen zur Darstellung der Entscheidungsregel in flexiblen Designs geht auf Proschan und Hunsberger (1995) zurück. In Übereinstimmung mit den Kombinationsregeln werden hier nur bedingte Fehlerfunktionen als Funktionen des  $p$ -Wertes  $p_1$  der ersten Stufe betrachtet. Dies ist auch die am häufigsten verwendete Form.

Die Idee zur Beschreibung eines zweistufigen Designs mit Hilfe einer bedingten Fehlerfunktion lässt sich folgendermaßen darstellen: Für jeden beobach-

teten  $p$ -Wert  $p_1$  aus der ersten Stufe kann mit Hilfe der bedingten Fehlerfunktion  $A$  das "bedingte" Niveau  $A(p_1)$  für den Test in der 2. Stufe bestimmt werden.

Um wieder "vernünftige" Entscheidungsregeln zu erhalten, sollte die bedingte Fehlerfunktion  $A : [0, 1] \rightarrow [0, 1]$  monoton fallend sein. Dies bedeutet, je kleiner der beobachtete  $p$ -Wert in der ersten Stufe, desto größer kann das Niveau für den Test in der zweiten Stufe gewählt werden. Um eine Einhaltung des Niveaus  $\alpha$  zu erreichen, muss gelten

$$\int_0^1 A(p_1) dp_1 = \alpha.$$

Insbesondere kann eine bedingte Fehlerfunktion  $A$  so gewählt werden, dass sie auch die Werte 1 und 0 annimmt. Im ersten Fall kann die Studie mit Verwerfen der Nullhypothese, im zweiten mit Beibehalten der Nullhypothese nach der ersten Stufe beendet werden.

Damit ergibt sich die Entscheidungsregel für ein zweistufiges flexibles Design

$$H_0 \text{ verwerfen} \Leftrightarrow \begin{cases} A(p_1) = 1 & \text{nach der 1. Stufe,} \\ p_2 \leq A(p_1) & \text{nach der 2. Stufe} \end{cases}$$

und

$$H_0 \text{ beibehalten} \Leftrightarrow \begin{cases} A(p_1) = 0 & \text{nach der 1. Stufe,} \\ p_2 \geq A(p_1) & \text{nach der 2. Stufe.} \end{cases}$$

### 3.5.3 Zusammenfassung

Es ist offensichtlich, dass sich jedes flexible Design, welches mit Hilfe einer Kombinationsregel beschrieben ist, leicht auch mit Hilfe einer bedingten Fehlerfunktion beschreiben lässt, sofern die Kombinationsregel gewisse Eigenschaften besitzt (siehe Vandemeulebroecke, 2006). Die Kombinationsregel von Bauer (3.2) führt zu der bedingten Fehlerfunktion

$$A(p_1) = \begin{cases} 1 & , \text{ falls } p_1 \leq c_\alpha, \\ \frac{c_\alpha}{p_1} & , \text{ falls } p_1 > c_\alpha. \end{cases}$$

Um das Fehlerniveau zu kontrollieren muss die Kombinationsregel bzw. die bedingte Fehlerfunktion vor Beginn der Studie festgelegt werden. Anderenfalls ist die Einhaltung des Fehlerniveaus nicht mehr gesichert. Dies gilt insbesondere dann, wenn deren Wahl datenabhängig geschieht. Somit muss die Entscheidung, ein solches flexibles Design zu wählen, immer am Studienbeginn getroffen werden. In einem solchen Design kann das Design für die zweite Stufe abhängig von den Daten bestimmt werden. Insbesondere kann die Fallzahl für die zweite Stufe von den Ergebnissen der ersten Stufe abhängig gewählt werden, aber auch viel weitreichendere Änderungen sind denkbar, wie z.B. die Wahl einer anderen zu testenden Hypothese in der zweiten Stufe, etc.. Allerdings sind solche Modifikationen mit Vorsicht zu betrachten, insbesondere was die Bedeutung der erhaltenen Aussage nach Ablehnung einer Nullhypothese angeht. Diese wurde auch schon in der Originalarbeit von Bauer (1989) ausführlich diskutiert.

Oftmals wird an den flexiblen Designs kritisiert, dass sie nicht die effizientesten Tests verwenden, sofern das initial angedachte Design nicht verändert wird. Das heißt, für die erhaltene Flexibilität, die durch die Wahl eines solchen Designs entsteht, muss ein Verlust an Power gegenüber dem optimalen Test mit fester Fallzahl hingenommen werden. Dieser Unterschied liegt in

der Regel in einem Bereich von wenigen Prozent. Demgegenüber steht der große Vorteil, auch noch eine hohe Power zu erreichen, wenn die initialen Planungsannahmen nicht zutreffen, und dies im Verlaufe der Studie offensichtlich wird.

Für eine weitere Einführung in die flexiblen Designs und eine breite Diskussion über die Vor- und Nachteile einer Anwendung dieser Designs in den verschiedenen Phasen einer klinischen Prüfung sei auf die zahlreich erschienenen Sonderbände zu diesem Thema in den verschiedenen Zeitschriften verwiesen, wie z.B. in *Pharmaceutical Statistics* (2006) oder im *Biometrical Journal* (2006).

### **3.6 Designänderungen nach dem CRP-Prinzip**

Die im vorherigen Abschnitt vorgestellten Verfahren erlauben datenabhängige Anpassungen des Studienverlaufs nur an vorgeplanten Zeitpunkten. Allerdings kann es in klinischen Studien trotz sorgfältigster Planung immer wieder vorkommen, dass es gute Gründe gibt, eine Studie zu ändern, obwohl dies initial zu diesem Zeitpunkt oder im gesamten Studiendesign nicht vorgesehen war. Mögliche Gründe können eine Abweichung der Größe des Störparameters von den Planungsannahmen oder eine geringere Rekrutierungsrate sein, verbunden mit einem höheren Effekt als geplant. Im ersten Fall sollte die Fallzahl der Studie erhöht werden, um die Power einzuhalten, im zweiten Fall kann die Fallzahl eventuell verringert werden, um so die Studie mit einer niedrigeren Fallzahl doch in vertretbarer Zeit zu Ende zu bringen. Letzteres Szenario tritt häufig bei Studien ein, die nicht von großen Pharmafirmen unterstützt werden, sondern deren Forschungsfragen aus der klinischen Praxis entstehen und von den entsprechenden Ärzten initiiert und getragen werden. In derartigen Studien werden meist seltene Erkrankungen untersucht, so dass



es deutlich schwieriger ist, viele Patienten in die Studie einzuschließen.

In früheren Zeiten geschahen solche Änderungen des Studiendesigns durch sogenannte Protokoll Amendments ohne Berücksichtigung, dass bei datenabhängiger Änderung eine zum Teil deutliche Überschreitung des  $\alpha$ -Niveaus möglich ist. Datenabhängig kann dabei bedeuten, dass die Entscheidung zur Änderung oder die Form der gewählten Modifikation von den bisherigen Daten abhängt. In diesem Abschnitt soll nun ein allgemeiner Ansatz vorgestellt werden, der es ermöglicht, eine laufende Studie jederzeit mit Kenntnis aller bisherigen Daten zu ändern und dabei trotzdem die Einhaltung des  $\alpha$ -Niveau zu sichern.

### 3.6.1 Eine vereinfachte Darstellung der Idee

Um auch ungeplante Designänderungen zu ermöglichen, verwendet das von Müller und Schäfer (2004) (in allgemeiner Form) und Müller und Schäfer (2001); Schäfer und Müller (2001) (für spezielle Verteilungen) vorgestellte Verfahren die natürliche, bedingte Fehlerfunktion, die implizit durch das initiale Design gegeben ist durch

$$P(H_0 \text{ im weiteren Verlaufe der Studie zu verwerfen} \mid \text{bisherige Daten}).$$

Im Englischen wird diese Wahrscheinlichkeit "Conditional Rejection Probability" genannt, weshalb dieses Verfahren häufig als CRP-Prinzip bezeichnet wird.

Die Einhaltung des  $\alpha$ -Niveaus für die modifizierte Studie soll sichergestellt werden durch die Einhaltung der folgenden Bedingung:

$$P_{H_0}(H_0 \text{ im Verlaufe der modifizierten Studie zu verwerfen} \mid \text{bisherige Daten}) \\ \leq P_{H_0}(H_0 \text{ im Verlaufe der initialen Studie zu verwerfen} \mid \text{bisherige Daten}).$$

Der Vorteil dieses Verfahrens besteht darin, dass Änderungen der Studie nicht nur an geplanten Zwischenauswertungen, sondern zu jedem Zeitpunkt erlaubt sind. Dies bedeutet, dass im Gegensatz zu den flexiblen Designs mögliche Designänderungen nicht zum Zeitpunkt der Studienplanung eingeplant werden müssen. Insbesondere kann als initiales Design ein herkömmliches Design, wie ein Design mit fester Stichprobengröße oder ein gruppensequentielles Design, gewählt werden. Sofern keine Änderungen an der Studie erfolgen, bleiben somit auch die Optimalitätseigenschaften eines entsprechend gewählten, initialen Designs erhalten. Flexible Designs besitzen häufig keine derartige Optimalitätseigenschaft, sondern ihre Power liegt in der Regel etwas unter der der entsprechenden optimalen Designs für die Anfangsplanung. Bei dem CRP-Prinzip kommt es erst im Falle einer Umplanung zu einem Powerverlust im Vergleich zu einem optimalen Design.

# 4

## Theorie der flexiblen Verfahren

In diesem Kapitel soll nun eine mathematische Theorie entwickelt werden, mit deren Hilfe die flexiblen Verfahren des vorherigen Kapitels, insbesondere aus Abschnitt 3.5 und 3.6, beschrieben werden können.

### 4.1 Entscheidungsprozeduren

Im ersten Teil dieses Abschnitt soll das Modell einer klassischen Entscheidungsprozedur, d.h. einer Entscheidungsprozedur mit fester Stichprobengröße, eingeführt werden. In den nachfolgenden Abschnitten soll dieses Modell erweitert werden zu einem Modell für die flexiblen Verfahren. Die nachfolgenden Definitionen für eine klassische Entscheidungsprozedur finden sich in ähnlicher Form in jedem Lehrbuch zur mathematischen Statistik, wie z.B. Lehmann und Romano (2005); Witting (1985).

### 4.1.1 Klassische Entscheidungsprozeduren

Die Beobachtungen werden beschrieben durch einen Zufallsvektor  $X$ . Somit ist  $X$  eine messbare Funktion von einem *Wahrscheinlichkeitsraum*  $(\Omega, \mathcal{A}, P)$  in einen messbaren Raum  $(\mathfrak{X}, \mathcal{F})$ , der sogenannte *Stichprobenraum*. Wir wollen im Folgenden annehmen, dass  $\mathfrak{X}$  ein euklidischer Raum ist. Der Stichprobenraum bildet zusammen mit dem durch  $X$  induzierten Wahrscheinlichkeitsmaß  $P$  ebenfalls einen Wahrscheinlichkeitsraum  $(\mathfrak{X}, \mathcal{F}, P)$ . Ziel einer statistischen Entscheidungsprozedur ist es, Aussagen über die unbekannte Verteilung  $P$  zu machen. In der Regel kann durch Vorinformationen die Menge der in Frage kommenden Verteilungen beschränkt werden, so dass nur eine Teilmenge  $\mathcal{P}$  der Menge aller Wahrscheinlichkeitsmaße auf  $(\mathfrak{X}, \mathcal{F})$  für eine spezielle Entscheidungsprozedur betrachtet wird. Die Menge  $\mathcal{P}$  lässt sich immer über einem *Parameterraum*  $\Theta$  parametrisieren, d.h. es existiert eine bijektive Abbildung von  $\Theta$  in  $\mathcal{P}$ , z.B.  $\vartheta = P$ . Damit erhalten wir die Darstellung  $\mathcal{P} = \{P_\vartheta | \vartheta \in \Theta\}$ .  $\mathcal{P}$  wird als *Verteilungsfamilie* und das Triple  $(\mathfrak{X}, \mathcal{F}, \mathcal{P})$  als *statistischer Raum* bezeichnet. Für eine über  $\Theta$  parametrisierte Verteilungsfamilie schreiben wir im Folgenden auch  $(P_\vartheta)_{\vartheta \in \Theta}$ .

**Definition 4.1** (klassische Entscheidungsregel). Es sei  $\mathcal{E}$  die Menge der möglichen Entscheidungen der Entscheidungsprozedur. Diese Menge  $\mathcal{E}$  bildet mit einer zugehörigen  $\sigma$ -Algebra  $\mathcal{E}$  den *Entscheidungsraum*  $(\mathcal{E}, \mathcal{E})$ . Eine messbare Abbildung  $\psi$  von dem Stichprobenraum  $(\mathfrak{X}, \mathcal{F})$  in den Entscheidungsraum  $(\mathcal{E}, \mathcal{E})$  wird als *Entscheidungsregel* bezeichnet.

**Bezeichnungen 4.2** (Testproblem). Im Falle eines Testproblems soll eine Entscheidung zwischen  $H_0$  und  $H_1$  getroffen werden. Deshalb kann die Menge der möglichen Entscheidungen gewählt werden als  $\mathcal{E} = \{0, 1\}$ , wobei 0 das Beibehalten und 1 das Verwerfen von  $H_0$  codiert. Die entsprechende Entscheidungsregel  $\psi$  wird als *Test* oder *Testfunktion* bezeichnet. Ein Test heißt *Niveau- $\alpha$ -Test* für ein Testproblem  $H_0 : \vartheta \in \Theta_0$  vs.  $H_1 : \vartheta \in \Theta_1$ , wenn gilt  $\mathbb{E}_\vartheta\{\psi\} \leq \alpha$  für alle  $\vartheta \in \Theta_0$ .

**Bezeichnungen 4.3** (Schätzproblem). Im Falle eines Schätzproblems soll die Größe eines unbekanntem Parameters ermittelt werden. In diesem Fall ist die Menge der Entscheidungen ein Teilraum des Parameterraumes, d.h.  $\mathcal{E} \subset \Theta$ . Eine entsprechende Entscheidungsregel wird als *Schätzer* oder *Schätzfunktion* bezeichnet. Ein Schätzer heißt *unverfälscht* oder *erwartungstreu*, wenn gilt  $\mathbb{E}_{\vartheta}\{\psi\} = \vartheta$ .

**Bezeichnungen 4.4.** Zur Zusammenfassung der Daten finden häufig messbare Abbildungen  $T$  von dem Stichprobenraum  $(\mathfrak{X}, \mathcal{F})$  in einen weiteren messbaren Raum, meist  $(\mathbb{R}^k, \mathcal{B}^k)$ , Anwendung. Diese werden als *Statistik* bzw. im Zusammenhang mit einem Testproblem auch als *Teststatistik* bezeichnet. Die Zufallsvariablen der Form  $T(X)$  bezeichnen wir häufig vereinfachend mit  $T$ .

Wir wollen im Folgenden annehmen, dass sowohl der Parameterraum  $\Theta$  als auch die Menge der Entscheidungen  $\mathcal{E}$  Teilmengen eines euklidischen Raumes sind.

### 4.1.2 Sequentielle Entscheidungsprozeduren

Im vorherigen Abschnitt wurde die Entscheidung nach Erhebung einer Stichprobe mit festem Umfang getroffen. In diesem Abschnitt sollen nun Entscheidungsprozeduren beschrieben werden, die nach Erhebung einer jeden Teilstichprobe (einzelnen Beobachtung) entweder eine endgültige Entscheidung treffen oder aber eine weitere, unabhängige Teilstichprobe nehmen. Die Größe der Gesamtstichprobe ist daher abhängig von den Daten. Ein Beispiel für eine sequentielle Entscheidungsprozedur sind die im Abschnitt 3.4 beschriebenen gruppensequentiellen Verfahren.

Sequentielle Verfahren zeichnen sich dadurch aus, dass die Daten nacheinander in Blöcken erhoben werden. Diese können durch Zufallsvariablen

$X_i : (\Omega, \mathcal{A}) \rightarrow (\mathfrak{X}_i, \mathcal{F}_i)$  beschrieben werden. Nach der Erhebung des Blockes  $i$  wird basierend auf allen bisherigen Daten  $X_1, \dots, X_i$  darüber entschieden, ob die Studie fortgesetzt oder gestoppt werden soll. Im Fall des Stoppens ist eine Entscheidung zu treffen, z.B. Beibehalten oder Verwerfen der Nullhypothese.

Eine Beschreibung liefert das folgende Model (Schmitz, 1991):

**Definition 4.5** (sequentielles Experiment). Sei  $(\mathfrak{X}, \mathcal{F})$  ein Stichprobenraum,  $\Theta$  ein Parameterraum und  $(P_\vartheta)_{\vartheta \in \Theta}$  eine Familie von Verteilungen auf  $(\mathfrak{X}, \mathcal{F})$ . Sei zusätzlich  $I \subset \mathbb{N}$  die Menge der möglichen Gesamtfallzahlen und  $(\mathcal{B}_i)_{i \in I}$  eine aufsteigende Familie von Sub- $\sigma$ -Algebren von  $\mathcal{F}$ , d.h. es gilt für alle  $i \in I$ ,  $\mathcal{B}_i \subset \mathcal{B}_{i+1} \subset \mathcal{F}$ . Dann heißt das Tupel  $((\mathfrak{X}, \mathcal{F}), \Theta, (P_\vartheta)_{\vartheta \in \Theta}, I, (\mathcal{B}_i)_{i \in I})$  ein *sequentielles Experiment*.

**Definition 4.6** (sequentielle Entscheidungsprozedur). Sei  $((\mathfrak{X}, \mathcal{F}), \Theta, (P_\vartheta)_{\vartheta \in \Theta}, I, (\mathcal{B}_i)_{i \in I})$  ein sequentielles Experiment und  $(\mathcal{E}, \mathcal{L})$  ein Entscheidungsraum. Eine messbare Abbildung  $\tau : \mathfrak{X} \rightarrow I$ , so dass gilt  $\{\tau = i\} \in \mathcal{B}_i$ , heißt *Stoppzeit*. Sei  $\psi = (\psi_i)_{i \in I}$  eine Familie von Entscheidungsregeln  $\psi_i : \mathfrak{X} \rightarrow \mathcal{E}$  so, dass  $\psi_i$  messbar bezüglich  $\mathcal{B}_i$  für alle  $i \in I$  ist. Das Tupel  $(\tau, \psi)$  wird als *sequentielle Entscheidungsprozedur* bezeichnet und  $\psi_\tau$  auch als *sequentielle Entscheidungsregel*.

*Bemerkung 4.1.* Aufgrund der Abzählbarkeit von  $I$  gilt:

$$\psi_\tau = \sum_{i \in I} 1_{\{\tau=i\}} \psi_i.$$

**Definition 4.7.** Eine sequentielle Entscheidungsprozedur  $(\tau, \psi)$  heißt *sequentieller Niveau- $\alpha$ -Test* für  $H_0 : \vartheta \in \Theta_0$  vs.  $H_1 : \vartheta \in \Theta_1$ , wenn gilt  $\mathbb{E}_\vartheta\{\psi_\tau(X)\} \leq \alpha$  für alle  $\vartheta \in \Theta_0$ .

**Beispiel 4.1** (einseitiger, gruppensequentieller Test ohne Futility Grenzen). Für einen einseitigen, gruppensequentiellen Test, wie im Abschnitt 3.4 beschrieben, ist die Stoppzeit gegeben durch  $\tau = \min\{i \in I | T_i > c_i\} \cup \{K\}$ .

Damit lässt sich dieser in Form einer Entscheidungsregel darstellen durch

$$\begin{aligned}\psi_\tau(X) &= \sum_{i=1}^K 1_{\{\tau=i\}} 1_{\{T_i > c_i\}} \\ &= \sum_{i=1}^K 1_{\{\{T_i > c_i\} \cap \bigcap_{j=1}^{i-1} \{T_j \leq c_j\}\}}.\end{aligned}$$

Hierbei ist  $T_i = T_i(X_1, \dots, X_i)$  eine Teststatistik basierend auf  $X_1, \dots, X_i$ .

**Beispiel 4.2** (einseitiger, gruppensequentieller Test mit Futility Grenzen). Erweiternd zu dem im vorhergehenden Beispiel beschriebenen gruppensequentiellen Test soll der nun betrachtete Test an der  $i$ -ten Auswertung mit Beibehalten der Nullhypothese stoppen, wenn  $T_i < b_i$  gilt. Es wird angenommen, dass an der Endauswertung  $c_K = b_K$  gilt. In diesem Fall ist die Stoppzeit gegeben durch  $\tau = \min\{i \in I \mid T_i > c_i \text{ oder } T_i < b_i\}$ . Einen solchen Test kann man wie folgt darstellen:

$$\begin{aligned}\psi_\tau(X) &= \sum_{i=1}^K 1_{\{\tau=i\}} 1_{\{T_i > c_i\}} \\ &= \sum_{i=1}^K 1_{\{\{T_i > c_i\} \cup \{T_i < b_i\}\} \cap \bigcap_{j=1}^{i-1} \{b_j \leq T_j \leq c_j\}} \cdot 1_{\{T_i > c_i\}}.\end{aligned}$$

Darstellungen für zweiseitige, gruppensequentielle Tests lassen sich in ähnlicher Form angeben.

### 4.1.3 Entscheidungsprozeduren mit datenabhängiger Anpassung

Eine Erweiterung der sequentiellen Prozeduren stellen Entscheidungsprozeduren dar, die eine datenabhängige Auswahl der weiteren Vorgehensweise erlauben, ohne dass die zugrunde liegende Regel vorher definiert wurde. Im

Folgenden wollen wir uns auf zweistufige, flexible Entscheidungsprozeduren beschränken, d.h. Prozeduren, die an einem Zeitpunkt eine datenabhängige Anpassung ermöglichen. Die zweistufigen sequentiellen Prozeduren, bei denen die Wahl der weiteren Vorgehensweise datenabhängig, aber nach fester Regel, erfolgt, sind somit ein Spezialfall der nun betrachteten Entscheidungsprozeduren.

**Definition 4.8** (flexibles Experiment). Sei  $(\mathfrak{X}, \mathcal{F})$  ein Stichprobenraum,  $\Theta$  ein Parameterraum und  $(P_\vartheta)_{\vartheta \in \Theta}$  eine Familie von Verteilungen auf  $(\mathfrak{X}, \mathcal{F})$ . Es sei  $\mathcal{M}$  die abzählbare Menge der möglichen Adaptionen. Zusammen mit einer zugehörigen  $\sigma$ -Algebra  $\mathcal{M}$  bildet sie den Raum der Adaptionen  $(\mathcal{M}, \mathcal{M})$ . Sei desweiteren  $(\mathcal{B}_m)_{m \in \mathcal{M}}$  eine Familie von Sub- $\sigma$ -Algebren von  $\mathcal{F}$  und seien  $\mathcal{G} \subset \mathcal{F}$  zwei weitere Sub- $\sigma$ -Algebren von  $\mathcal{F}$ , so dass  $\sigma(\bigcup_{m \in \mathcal{M}} \mathcal{B}_m)$  bedingt unabhängig von  $\mathcal{T}$  bei gegebenem  $\mathcal{G}$  ist. Dann heißt das Tupel  $((\mathfrak{X}, \mathcal{F}), \Theta, (P_\vartheta)_{\vartheta \in \Theta}, (\mathcal{M}, \mathcal{M}), (\mathcal{B}_m)_{m \in \mathcal{M}}, \mathcal{G}, \mathcal{T})$  ein *flexibles Experiment*.

*Bemerkung 4.2.* Für die Definition der bedingten Unabhängigkeit sei auf den Anhang A.1 verwiesen.

**Definition 4.9** (flexible Entscheidungsprozedur). Sei  $((\mathfrak{X}, \mathcal{F}), \Theta, (P_\vartheta)_{\vartheta \in \Theta}, (\mathcal{M}, \mathcal{M}), (\mathcal{B}_m)_{m \in \mathcal{M}}, \mathcal{G}, \mathcal{T})$  ein flexibles Experiment und  $(\mathcal{E}, \mathcal{E})$  ein Entscheidungsraum. Eine Abbildung  $\tau : \mathfrak{X} \rightarrow \mathcal{M}$ , so dass  $\tau$  messbar bezüglich  $\mathcal{T}$  ist, heißt *Adaptionsregel*. Sei  $\mathcal{T}$  eine Menge von Adaptionsregeln und sei  $\psi = (\psi_m)_{m \in \mathcal{M}}$  eine Familie von Entscheidungsregeln  $\psi_m : \mathfrak{X} \rightarrow \mathcal{E}$ , so dass  $\psi_m$  messbar bezüglich  $\mathcal{B}_m$  für alle  $m \in \mathcal{M}$  ist. Dann wird das Tupel  $(\mathcal{T}, \psi)$  als *flexible Entscheidungsprozedur* bezeichnet und  $\psi_\tau$  auch als *flexible Entscheidungsregel*.

*Bemerkung 4.3.* Aufgrund der Abzählbarkeit von  $\mathcal{M}$  gilt wieder:

$$\psi_\tau = \sum_{m \in \mathcal{M}} 1_{\{\tau=m\}} \psi_m.$$

**Definition 4.10.** Eine flexible Entscheidungsprozedur  $(\mathcal{T}, \psi)$  heißt *flexibler Niveau- $\alpha$ -Test* für  $H_0 : \vartheta \in \Theta_0$  vs.  $H_1 : \vartheta \in \Theta_1$ , wenn  $\mathbb{E}_\vartheta\{\psi_\tau(X)\} \leq \alpha$  für alle  $\vartheta \in \Theta_0$  und alle  $\tau \in \mathcal{T}$  gilt.



**Lemma 4.11.** Gegeben sei die Menge  $\mathcal{M}$  der Adaptionen, ein Stichprobenraum  $(\mathfrak{X}, \mathcal{F})$ , ein Parameterraum  $\Theta$ , eine Familie  $(P_\vartheta)_{\vartheta \in \Theta}$  von Verteilungen auf  $(\mathfrak{X}, \mathcal{F})$ , die messbaren Abbildungen  $X_1 : \mathfrak{X} \rightarrow \mathfrak{X}_1$  (die Daten der 1. Stufe),  $X_{2m} : \mathfrak{X} \rightarrow \mathfrak{X}_{2m}$  (die Daten der 2. Stufe zur Adaption  $m \in \mathcal{M}$ ), so dass  $X_1$  und  $X_{2m}$  unabhängig sind, und eine Sub- $\sigma$ -Algebra  $\mathcal{T}$  mit  $\sigma(X_1) \subset \mathcal{T} \subset \mathcal{F}$ , wobei  $\mathcal{T}$  unabhängig von  $\sigma((X_{2m})_{m \in \mathcal{M}})$  ist. Dann ist

$$((\mathfrak{X}, \mathcal{F}), \Theta, (P_\vartheta)_{\vartheta \in \Theta}, (\mathcal{M}, \mathcal{M}), (\sigma(X_1, X_{2m}))_{m \in \mathcal{M}}, \sigma(X_1), \mathcal{T})$$

ein flexibles Experiment.

*Beweis.* Gemäß der Definition eines flexiblen Experiments bleibt zu zeigen, dass  $\mathcal{T}$  bedingt unabhängig von  $\sigma(X_1, (X_{2m})_{m \in \mathcal{M}})$  bei gegebenem  $\sigma(X_1)$  ist. Sei  $T \in \mathcal{T}$ ,  $A \in \sigma(X_1)$  und  $B \in \sigma(X_{2m})_{m \in \mathcal{M}}$ , dann gilt

$$\begin{aligned} \mathbb{E}\{1_T 1_{A \cap B} | \sigma(X_1)\} &= 1_A \mathbb{E}\{1_T \mathbb{E}\{1_B | \mathcal{T}, \sigma(X_1)\} | \sigma(X_1)\} \\ &= 1_A \mathbb{E}\{1_B\} \mathbb{E}\{1_T | \sigma(X_1)\} = \mathbb{E}\{1_A 1_B | \sigma(X_1)\} \mathbb{E}\{1_T | \sigma(X_1)\} \quad \text{fs.} \end{aligned}$$

□

*Bemerkung 4.4.* Lemma 4.11 verdeutlicht, dass das Model geeignet ist, zweistufige, flexible Verfahren mit unabhängigen Stufen zu beschreiben.

*Bemerkung 4.5.* Gilt  $|\mathcal{T}| = 1$ ,  $\mathcal{G} = \mathcal{T}$  und bildet  $(\mathcal{B}_m)_{m \in \mathcal{M}}$  eine aufsteigende Familie von Sub- $\sigma$ -Algebren, so entspricht eine flexible Entscheidungsprozedur einer zweistufigen, gruppensequentiellen bzw. einer flexiblen Entscheidungsprozedur mit fester Adaptionsregel, wie sie im Abschnitt 3.3 beschrieben worden ist.

Ist hingegen  $\mathcal{T}$  die Menge aller Adaptionsregeln, d.h. die Menge aller  $\mathcal{G}$  messbaren Abbildungen  $\tau : \mathfrak{X} \rightarrow \mathcal{M}$  und  $\mathcal{G}$  eine "echte" Sub- $\sigma$ -Algebra von  $\mathcal{T}$ , so werden mit obiger Definition die "voll" flexiblen Verfahren aus Abschnitt 3.5 und 3.6 beschrieben, bei denen insbesondere die Adaptionsregel nicht bekannt ist.

**Beispiel 4.3.** Die flexible Entscheidungsregel der Designs mit bedingter Fehlerfunktion aus Abschnitt 3.5.2 lässt sich darstellen durch

$$\psi_\tau = \sum_{m \in \mathcal{M}} 1_{\{\tau=m\}} 1_{\{p_2(X_{2m}) \leq A(p_1(X_1))\}},$$

wobei  $X_1$  die Daten der 1. Stufe und  $X_{2m}$  die Daten der 2. Stufe zur Adaption  $m \in \mathcal{M}$  bezeichnet.

## 4.2 Theoretische Grundlagen der flexiblen Verfahren

Ziel dieses Abschnittes ist es, Bedingungen anzugeben, so dass die "voll" flexiblen Entscheidungsprozeduren aus Abschnitt 3.5 und 3.6 gewisse Eigenschaften besitzen wie ein Niveau- $\alpha$ -Test oder ein unverfälschter Schätzer zu sein. Ein großer Unterschied zwischen den "vorgeplant" flexiblen Designs mit fester Anpassungsregel, wie in Abschnitt 3.3, und den "voll" flexiblen Designs besteht darin, dass bei den Letzteren theoretisch jede Anpassung erlaubt sein soll und diese auch nicht in Form einer festen Regel vorgezifiziert sein muss. Das bedeutet, dass für die mögliche Anpassung nicht nur die Informationen aus dem bisherigen Studienverlauf, sondern auch weitere Informationen verwendet werden können. Die Entscheidung, welche Form der Anpassung durchgeführt wird, kann letztendlich von der Einschätzung der entscheidenden Person abhängen. Aus diesem Grunde werden wir in diesem Abschnitt für die Menge  $\mathcal{T}$  der Adaptionsregeln einer flexiblen Entscheidungsprozedur immer die Menge aller  $\mathcal{T}$  messbaren Abbildungen  $\tau : \mathfrak{X} \rightarrow \mathcal{M}$  betrachten.

**Satz 4.12.** *Gegeben seien ein flexibles Experiment  $((\mathfrak{X}, \mathcal{F}), \Theta, (P_\vartheta)_{\vartheta \in \Theta})$ ,  $(\mathcal{M}, \mathcal{M})$ ,  $(\mathcal{B}_m)_{m \in \mathcal{M}}$ ,  $\mathcal{G}, \mathcal{T}$ , die Menge aller Adaptionsregeln  $\mathcal{T}$  und eine flexible Entscheidungsprozedur  $(\mathcal{T}, \psi)$ . Sei  $A : \mathfrak{X} \rightarrow \mathbb{R}$  eine  $\mathcal{G}$  messbare Abbil-*

dung. Existiert ein  $m^* \in \mathcal{M}$ , so dass

$$\mathbb{E}\{\psi_{m^*} | \mathcal{G}\} = A \quad \text{fs},$$

dann gilt

$$\mathbb{E}\{\psi_m | \mathcal{G}\} \stackrel{\leq}{\underset{>}{\approx}} A \quad \text{fs} \quad \text{für alle } m \in \mathcal{M} \quad (4.1)$$

$$\Leftrightarrow \mathbb{E}\{\psi_\tau\} \stackrel{\leq}{\underset{>}{\approx}} \mathbb{E}\{A\} \quad \text{für alle Adaptionsregeln } \tau \in \mathcal{T}. \quad (4.2)$$

*Beweis.* "⇒" Dies folgt aufgrund der bedingten Unabhängigkeit von  $\sigma(\bigcup_{m \in \mathcal{M}} \mathcal{B}_m)$  und  $\mathcal{T}$  bei gegebenem  $\mathcal{G}$  mit Hilfe des Lemmas A.2 aus

$$\begin{aligned} \mathbb{E}\{\psi_\tau\} &= \mathbb{E}\{\mathbb{E}\{\psi_\tau | \mathcal{G}, \mathcal{T}\}\} = \sum_{m \in \mathcal{M}} \mathbb{E}\{1_{\{\tau=m\}} \mathbb{E}\{\psi_m | \mathcal{G}\}\} \\ &\stackrel{\leq}{\underset{>}{\approx}} \mathbb{E}\{A \sum_{m \in \mathcal{M}} 1_{\{\tau=m\}}\} = \mathbb{E}\{A\} \end{aligned}$$

für alle  $\tau \in \mathcal{T}$ .

"⇐" Wir wollen hier nur den Fall der Gleichheit behandeln, die beiden anderen Fälle lassen sich vollkommen analog zeigen. Angenommen, es existiert  $m \in \mathcal{M}$  mit

$$P(\mathbb{E}\{\psi_m | \mathcal{G}\} \neq A) > 0.$$

Setze  $g = \mathbb{E}\{\psi_m | \mathcal{G}\}$ ,  $M_1 = \{g > A\}$  und  $M_2 = \{g < A\}$ . Angenommen, es gilt  $P(M_1) > 0$ . Ist  $\tau = m 1_{M_1} + m^* 1_{M_2} + m^* 1_{\{g=A\}}$ , so ist  $\tau$  messbar bezüglich  $\mathcal{G} \subset \mathcal{T}$  und somit eine zulässige Adaptionsregel aus  $\mathcal{T}$ , und es gilt

$$\begin{aligned} \mathbb{E}\{\psi_\tau - A\} &= \mathbb{E}\{\mathbb{E}\{\psi_\tau | \mathcal{G}\} - A\} \\ &= \mathbb{E}\{1_{M_1} g + 1_{M_2} A - (1_{M_1} A + 1_{M_2} A)\} \\ &= \mathbb{E}\{1_{M_1} (g - A) + 1_{M_2} (A - A)\} > 0, \end{aligned}$$

ein Widerspruch. Gilt  $P(M_1) = 0$ , so folgt  $P(M_2) > 0$  und eine analoge Rechnung führt ebenfalls zum Widerspruch.

□

Aus der Gleichheit der bedingten Erwartungen (4.1) für alle  $m \in \mathcal{M}$  folgt nicht die Gleichheit der Erwartungen (4.2), wenn die bedingte Unabhängigkeit, wie sie in der Definition des flexiblen Experimentes gefordert wurde, nicht erfüllt ist. Dies zeigt das folgende Beispiel.

**Beispiel 4.4.** Es seien  $X_1, X_2$  zwei *Bernoulli*( $\frac{1}{2}$ ) verteilte Zufallsvariablen. Ferner seien  $\psi_1(X_1, X_2)$  und  $\psi_2(X_1, X_2)$  zwei Entscheidungsregeln mit

$$\psi_1(x_1, x_2) = \begin{cases} 1, & x_1 = 1, x_2 = 1 \\ 0, & \text{sonst} \end{cases}, \quad \psi_2(x_1, x_2) = \begin{cases} 1, & x_1 = 1, x_2 = 0 \\ 0, & \text{sonst} \end{cases}.$$

Die Adaptionregel sei gegeben durch  $\tau = X_2 + 1$ . Damit gilt:

$$\psi(X, \tau) = 1_{\{\tau=1\}}\psi_1(X_1, X_2) + 1_{\{\tau=2\}}\psi_2(X_1, X_2) = \psi_1(X_1, 0) + \psi_2(X_1, 1) = 0.$$

Aber:

$$\mathbb{E}\{\psi_i(X_1, X_2) | X_1 = x_1\} = \begin{cases} 0, & x_1 = 0 \\ \frac{1}{2}, & x_1 = 1 \end{cases}, \quad i = 1, 2.$$

Wenn wir das Ergebnis dieses Satzes für die Niveau- $\alpha$ -Einhaltung der flexiblen Verfahren aus Abschnitt 3.5 und 3.6 des vorherigen Kapitels für den Fall der Unabhängigkeit der Daten der beiden Stufen zusammenfassen, so erhalten wir:

**Korollar 4.13** (flexible Niveau  $\alpha$ -Tests). *Gegeben sei das flexible Experiment  $((\mathfrak{X}, \mathcal{F}), \Theta, (P_\vartheta)_{\vartheta \in \Theta}, (\mathcal{M}, \mathcal{M}), (\sigma(X_1, X_{2m}))_{m \in \mathcal{M}}, \sigma(X_1), \mathcal{T})$  mit  $X_1 : \mathfrak{X} \rightarrow \mathfrak{X}_1$  (die Daten der 1. Stufe),  $X_{2m} : \mathfrak{X} \rightarrow \mathfrak{X}_{2m}$  (die Daten der 2. Stufe zur Adaption  $m \in \mathcal{M}$ ), so dass  $X_1$  unabhängig von  $(X_{2m})_{m \in \mathcal{M}}$ , und  $\sigma(X_1) \subset \mathcal{T} \subset \mathcal{F}$  ist, sowie die Menge aller Adaptionen  $\mathcal{T}$ .*

a) (flexible Designs)

Ist  $(A_\vartheta)_{\vartheta \in \Theta_0}$  eine Familie von messbaren Abbildungen  $A_\vartheta : \mathfrak{X}_1 \rightarrow [0, 1]$

mit  $\mathbb{E}_\vartheta\{A_\vartheta(X_1)\} = \alpha$  für alle  $\vartheta \in \Theta_0$ , so gilt:

$$\begin{aligned} \mathbb{E}_\vartheta\{\psi_\tau(X, \tau)\} &\leq \alpha \quad \text{für alle } \tau \in \mathcal{T}, \vartheta \in \Theta_0 \\ \Leftrightarrow \mathbb{E}_\vartheta\{\psi_m|X_1\} &\leq A_\vartheta(X_1) \text{ fs} \quad \text{für alle } m \in \mathcal{M}, \vartheta \in \Theta_0. \end{aligned}$$

b) (Designänderungen nach dem CRP-Prinzip)

Ist  $\psi_{m_1} : \mathfrak{X} \rightarrow \mathcal{E}$  eine  $\sigma(X_1, X_{2m_1})$  messbare Abbildung und  $m_1 \in \mathcal{M}$ , so dass

$$\mathbb{E}_\vartheta\{\psi_{m_1}\} \leq \alpha \quad \text{für alle } \vartheta \in \Theta_0,$$

so gilt:

$$\begin{aligned} \mathbb{E}_\vartheta\{\psi_\tau\} &\leq \alpha \quad \text{für alle } \tau \in \mathcal{T}, \vartheta \in \Theta_0 \\ \Leftrightarrow \mathbb{E}_\vartheta\{\psi_m|X_1\} &\leq \mathbb{E}_\vartheta\{\psi_{m_1}|X_1\} + k(\vartheta) \text{ fs} \quad \text{für alle } m \in \mathcal{M}, \text{ und } \vartheta \in \Theta_0 \end{aligned}$$

mit  $k(\vartheta) = \alpha - \mathbb{E}_\vartheta\{\psi_{m_1}\}$ .

*Beweis.* Dies folgt direkt aus Satz 4.12. □

*Bemerkung 4.6.* Für die flexiblen Designs mit bedingter Fehlerfunktion wird normalerweise die bedingte Fehlerfunktion  $A : \mathfrak{X}_1 \rightarrow [0, 1]$  für ein  $\vartheta^*$  von dem Rand der Nullhypothese vorgegeben. Im Fall des einseitigen Testproblems  $H_0 : \vartheta \leq \vartheta_0$  vs.  $H_1 : \vartheta > \vartheta_0$  wäre  $\vartheta^* = \vartheta_0$ . Die Familie  $(A_\vartheta)_{\vartheta \in \Theta_0}$  kann dann gewählt werden durch  $A_\vartheta := A + k(\vartheta)$  mit  $k(\vartheta) := \alpha - \mathbb{E}_\vartheta\{A(X_1)\}$ .

*Bemerkung 4.7.* An diesem Korollar wird nochmals der Unterschied zwischen den flexiblen Verfahren gemäß Abschnitt 3.5 und den Designänderungen nach dem CRP-Prinzip deutlich. Bei Ersterem wird die bedingte Fehlerfunktion  $A(X_1)$  vorgegeben, bei Letzterem wird die bedingte Erwartung der initialen Entscheidungsregel  $\mathbb{E}\{\psi_{m_1}|X_1\}$  als bedingte Fehlerfunktion (*bedingte Ablehnungswahrscheinlichkeit*) verwendet.

**Beispiel 4.5** (CRP beim  $z$ -Test). Es seien  $X_1, \dots, X_n$  unabhängig und identisch  $N(\mu, 1)$  verteilte Zufallsvariablen. Wir wollen die Nullhypothese

$H_0 : \mu = 0$  gegen die Alternative  $H_1 : \mu > 0$  testen. Der gleichmäßig beste Niveau- $\alpha$ -Test ist der  $z$ -Test:

$$H_0 \text{ verwerfen} \Leftrightarrow T_n = \sqrt{n}\bar{X}_n > z_{1-\alpha},$$

wobei  $z_q$  das  $q$ -Quantil der Standardnormalverteilung bezeichnet.

Soll nun nach  $n_1$  Beobachtungen eine Studie geändert werden, die mit einem  $z$ -Test und einer Fallzahl  $n > n_1$  geplant wurde, so ist die bedingte Erwartung unter  $H_0$ , gegeben die ersten  $n_1$  Beobachtungen zu bestimmen. Diese ist gegeben als

$$\begin{aligned} & \mathbb{E}_{\mu=0}\{1_{\{T_n > z_{1-\alpha}\}} | x_1, \dots, x_{n_1}\} \\ &= \mathbb{P}_{\mu=0}\left(\sqrt{n}\left(\frac{n_1}{n}\bar{x}_{n_1} + \frac{n-n_1}{n}\bar{X}_{(n_1+1):n}\right) > z_{1-\alpha}\right) \\ &= \mathbb{P}_{\mu=0}\left(\frac{\sqrt{n-n_1}}{\sqrt{n}}\sqrt{n-n_1}\bar{X}_{(n_1+1):n} > z_{1-\alpha} - \frac{n_1}{\sqrt{n}}\bar{x}_{n_1}\right) \\ &= 1 - \Phi\left(\frac{\sqrt{n}}{\sqrt{n-n_1}}z_{1-\alpha} - \frac{n_1}{\sqrt{n-n_1}}\bar{x}_{n_1}\right). \end{aligned}$$

Hierbei bezeichnet  $\bar{X}_{(n_1+1):n} = \frac{1}{n-n_1} \sum_{i=n_1+1}^n X_i$  den Mittelwert aus den verbleibenden Beobachtungen. Der Verlauf der bedingten Ablehnungswahrscheinlichkeit für verschiedene Verhältnisse von  $n_1$  zu  $n$  ist in Abbildung 4.1 zu sehen.

Soll die Fallzahl in  $\tilde{n} > n_1$  geändert werden, so muss die neue Entscheidungsregel ein Test auf den verbleibenden  $\tilde{n} - n_1$  Beobachtungen zum Niveau  $\Phi\left(\frac{\sqrt{\tilde{n}}}{\sqrt{\tilde{n}-n_1}}z_{1-\alpha} - \frac{n_1}{\sqrt{\tilde{n}-n_1}}\bar{x}_{n_1}\right)$  bilden, um die Gleichheit der bedingten Erwartungen aus Korollar 4.13 zu erfüllen. Die Entscheidungsregel für die geänderte Fallzahl ist somit gegeben durch

$$\psi_{\tilde{n}}(X_1, \dots, X_{\tilde{n}}) = 1_{\left\{\frac{n_1}{\sqrt{\tilde{n}}}\bar{x}_{n_1} + \frac{\sqrt{\tilde{n}-n_1}}{\sqrt{\tilde{n}}}\sqrt{\tilde{n}-n_1}\bar{X}_{(n_1+1):\tilde{n}} > z_{1-\alpha}\right\}}.$$

*Bemerkung 4.8.* Im nächsten Kapitel werden wir sehen, dass die erhaltene, flexible Entscheidungsregel auch für die Nullhypothese  $H_0 : \mu \leq 0$  das Niveau  $\alpha$  einhält.

**Beispiel 4.6** (CRP t-Test). In diesem Beispiel wollen wir dieselben Hypothesen testen wie im vorherigen Beispiel, mit dem Unterschied, dass wir nun

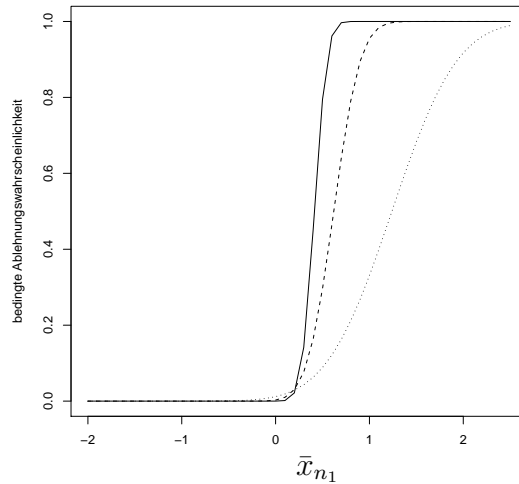


Abbildung 4.1: Bedingte Ablehnungswahrscheinlichkeit eines  $z$ -Tests, geplant mit  $n = 40$  Beobachtungen zum Niveau  $2.5\%$ , nach  $n_1 = 10, 20$  und  $30$ -Beobachtungen (gepunktete, gestrichelte und durchgezogene Linie).

annehmen wollen,  $X_1, \dots, X_n$  sind unabhängig und identisch  $N(\mu, \sigma^2)$  verteilt mit unbekannter Varianz  $\sigma^2$ . Wie bereits erwähnt ist der gleichmäßig beste Niveau- $\alpha$ -Test der  $t$ -Test:

$$H_0 \text{ verwerfen} \Leftrightarrow T_n = \sqrt{n} \frac{\bar{X}_n}{S_n} > t_{n-1; 1-\alpha}.$$

Um einen solchen  $t$ -Test nach  $n_1$ -Beobachtungen zu ändern, muss wieder die bedingte Erwartung unter  $H_0$ , gegeben die ersten  $n_1$  Beobachtungen bestimmt werden. Dazu nutzen wir die nachfolgende Zerlegung der  $t$ -Teststatistik

$$T_n(\bar{X}_1, S_1, \bar{X}_2, S_2) = \frac{\sqrt{n} \left( \frac{n_1}{n} \bar{X}_1 + \frac{n_2}{n} \bar{X}_2 \right)}{\sqrt{\frac{1}{n-1} \left( S_1^2(n_1 - 1) + S_2^2(n_2 - 1) + \frac{n_1 n_2}{n} (\bar{X}_1 - \bar{X}_2)^2 \right)}},$$

wobei wir mit  $\bar{X}_1, \bar{X}_2$  und  $S_1, S_2$  den Mittelwert bzw. Varianzschätzer aus den ersten  $n_1$  und den verbleibenden  $n_2 = n - n_1$  Beobachtungen bezeichnen.

Damit ist die bedingte Ablehnungswahrscheinlichkeit gegeben durch

$$\begin{aligned} & \mathbb{E}_{\mu=0, \sigma^2} \{1_{\{T_n > t_{n-1; 1-\alpha}\}} | x_1, \dots, x_{n_1}\} \\ &= \mathbb{P}_{\mu=0, \sigma^2} \left( \frac{\sqrt{n} \left( \frac{n_1}{n} \bar{x}_1 + \frac{n_2}{n} \bar{X}_2 \right)}{\sqrt{\frac{1}{n-1} (s_1^2 (n_1 - 1) + S_2^2 (n_2 - 1) + \frac{n_1 n_2}{n} (\bar{x}_1 - \bar{X}_2)^2)}} > t_{n-1; 1-\alpha} \right). \end{aligned}$$

Setzen wir nun

$$Z = \sqrt{n} \left( \frac{n_1}{n} \frac{\bar{x}_1}{\sigma} + \frac{n_2}{n} \frac{\bar{X}_2}{\sigma} \right)$$

und

$$V = \frac{1}{n-1} \left( (n_1 - 1) \frac{s_1^2}{\sigma^2} + (n_2 - 1) \frac{S_2^2}{\sigma^2} + \frac{n_1}{n_2} \left( \frac{n_1}{\sqrt{n}} \frac{\bar{x}_1}{\sigma} + \frac{n_2}{\sqrt{n}} \frac{\bar{X}_2}{\sigma} - \sqrt{n} \frac{\bar{x}_1}{\sigma} \right)^2 \right),$$

so ist die gemeinsame Dichte unter  $H_0 : \mu = 0$  gegeben durch

$$f_{(Z, V)}(z, v) = \varphi_{(0, n_2/n)} \left( z - \frac{n_1}{\sqrt{n}} \frac{\bar{x}_1}{\sigma} \right) f_{\frac{1}{n-1} \chi_{n_2-1}^2} \left( v - \frac{n_1 - 1}{n-1} \frac{s_1^2}{\sigma^2} - \frac{n_1 (z - \sqrt{n} \frac{\bar{x}_1}{\sigma})^2}{n_2 (n-1)} \right).$$

Hierbei ist  $\varphi(\mu, \sigma^2)$  die Dichte einer  $N(\mu, \sigma^2)$ -Verteilung und  $f_{\frac{1}{n-1} \chi_{n_2-1}^2}$  die Dichte einer  $\chi^2$  verteilten Zufallsvariable mit  $n-1$  Freiheitsgraden, welche mit dem Faktor  $\frac{1}{n-1}$  multipliziert wurde. Mit Hilfe der gemeinsamen Dichte kann die bedingte Ablehnungswahrscheinlichkeit nun mittels numerischer Integration bestimmt werden. An der Darstellung der gemeinsamen Dichte von  $Z$  und  $V$  wird deutlich, dass die bedingte Ablehnungswahrscheinlichkeit von dem unbekanntem Parameter  $\sigma$  abhängt. Einige Verläufe der bedingten Ablehnungswahrscheinlichkeit in Abhängigkeit von  $\sigma$  für festes  $\bar{x}_1$  und  $s_1^2$  sind in Abbildung 4.2 dargestellt. Diese Grafik verdeutlicht, dass der Kurvenverlauf je nach beobachtetem Wert von  $\bar{x}_1$  sehr unterschiedlich ist.

Soll die Fallzahl in  $\tilde{n} > n_1$  geändert werden, so müsste die Entscheidungsregel für die geänderte Fallzahl so konstruiert werden, dass sie denselben Kurvenverlauf besitzt, um die Gleichheit der bedingten Erwartungen aus Korollar 4.13 zu erfüllen und somit das Niveau- $\alpha$  auszuschöpfen. Dies ist nicht so einfach möglich. Um zumindest die Einhaltung des  $\alpha$ -Niveaus zu erreichen,



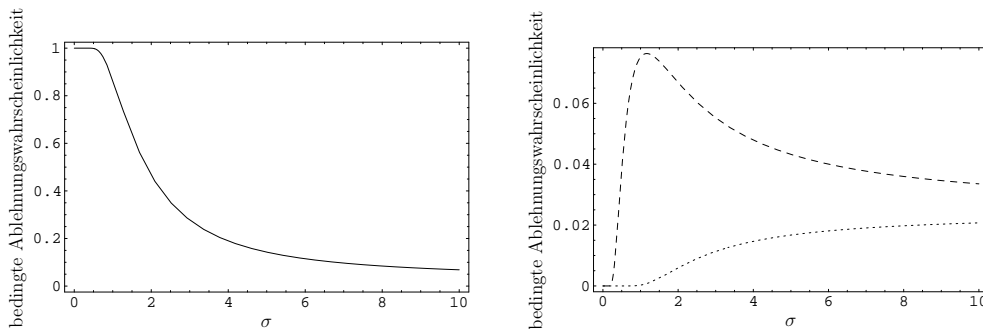


Abbildung 4.2: Bedingte Ablehnungswahrscheinlichkeit eines  $t$ -Tests, geplant mit  $n = 40$  Beobachtungen zum Niveau 2.5% nach  $n_1 = 20$ -Beobachtungen für  $s_1^2 = 1$  und  $\bar{x}_1 = 1, 0.3$  und  $-0.1$  (linker Plot: durchgezogene Linie; rechter Plot: gestrichelte und gepunktete Linie).

würde es auch ausreichen, eine Kurve zu finden, die für alle Werte von  $\sigma$  kleiner oder gleich der dargestellten Kurve ist. Ein Ansatz wäre, analog zu dem Vorgehen beim  $z$ -Test, die neue Entscheidungsregel in Form eines bedingten Tests zu einem konstanten, bedingten Niveau zu konstruieren. In den Fällen, die zu Kurvenverläufen führen, die denen der rechten Grafik aus Abbildung 4.2 entsprechen, ist dies nicht sinnvoll. In diesen Fällen müsste das bedingte Niveau gleich 0 gewählt werden, um unterhalb der Kurve zu liegen.

Für eine weitere Diskussion der Kurvenverläufe der bedingten Ablehnungswahrscheinlichkeiten des  $t$ -Tests, ihrer Eigenschaften, sowie ihrer Berechnung auch unter der Alternativhypothese sei auf Posch et al. (2004) verwiesen.

*Bemerkung 4.9.* Im nachfolgenden Kapitel werden wir eine Methode vorstellen, die es in bestimmten Verteilungssituationen ermöglicht, die bedingten Ablehnungswahrscheinlichkeiten über den gesamten Bereich eines Störparameters anzupassen. In diesem Zusammenhang werden wir nochmals das Beispiel des  $t$ -Tests verwenden. Insbesondere werden wir dort eine Methode zur Berechnung der bedingten Ablehnungswahrscheinlichkeiten angeben, die nicht auf den üblichen Statistiken  $\bar{X}$  und  $S^2$  beruht.



# 5

## Flexible Designs in Exponentialfamilien

Das Ziel dieses Kapitels ist die Entwicklung einer Methode, die es erlaubt, in Verteilungssituationen mit Störparameter, wie z.B. dem  $t$ -Test, Fallzahländerungen nach dem CRP-Prinzip vorzunehmen.

### 5.1 Modell und Eigenschaften

#### 5.1.1 Exponentialfamilien

**Definition 5.1** (Exponentialfamilie). Sei  $(\mathfrak{X}, \mathcal{F})$  ein Stichprobenraum,  $\Theta$  ein Parameterraum und  $(P_\vartheta)_{\vartheta \in \Theta}$  eine Verteilungsfamilie auf  $(\mathfrak{X}, \mathcal{F})$ . Existieren Funktionen  $C, q_1, \dots, q_k : \Theta \rightarrow \mathbb{R}$ , messbare Funktionen  $h, T_1, \dots, T_k : \mathfrak{X} \rightarrow \mathbb{R}$  und ein  $\sigma$ -finites Maß  $\mu$  auf  $(\mathfrak{X}, \mathcal{F})$  derart, dass  $P_\vartheta$  für alle  $\vartheta \in \Theta$  eine

$\mu$ -Dichte  $f_{\vartheta}$  der Form

$$f_{\vartheta}(x) = C(\vartheta) \exp \left( \sum_{i=1}^k q_i(\vartheta) T_i(x) \right) h(x), \quad (5.1)$$

besitzt, so heißt  $(P_{\vartheta})_{\vartheta \in \Theta}$  Exponentialfamilie in  $q = (q_1, \dots, q_k)$  und  $T = (T_1, \dots, T_k)$ .

**Bezeichnungen 5.2.** Die Darstellung (5.1) ist in der Regel nicht eindeutig. Ist  $k$  die kleinste natürliche Zahl, so dass  $P_{\vartheta}$  für alle  $\vartheta \in \Theta$  Dichten der Form (5.1) besitzt, so heißt die Exponentialfamilie  $k$ -parametrig. Bei geeigneter Parametrisierung der Verteilungsfamilien können wir die Dichten der Exponentialfamilie in auch schreiben in der Form

$$f_{\vartheta}(x) = C(\vartheta) \exp \left[ \sum_{i=1}^k \vartheta_i T_i(x) \right] h(x).$$

Wir sprechen dann von einer natürlichen Parametrisierung. Der zugehörige Parameterraum  $\Theta$  wird als *natürliche Parameterraum* der Exponentialfamilie bezeichnet.

*Bemerkung 5.1.* Wir wollen im Folgenden immer von dem natürlichen Parameterraum ausgehen.

Der Faktor  $h(x)$  kann durch geeignete Wahl des Maßes  $\mu$  weggelassen werden. Die Sprechweise  $(P_{\vartheta}^X)_{\vartheta \in \Theta}$  ist Exponentialfamilie in  $\vartheta$  und  $T$  bedeutet dann: Es existiert ein geeignetes Maß  $\mu$  und  $\mu$ -Dichten der Form

$$f_{\vartheta}(x) = C(\vartheta) \exp \left[ \sum_{i=1}^k \vartheta_i T_i(x) \right].$$

**Bezeichnungen 5.3.** Sei  $(P_{\theta})_{\theta \in \Theta}$  eine  $k$ -parametrig Exponentialfamilie mit  $k > 1$ . Wollen wir einen der Parameter aus dem Parametervektor  $\theta$  speziell hervorheben, so schreiben wir auch  $(P_{\theta})_{\theta \in \Theta}$  ist Exponentialfamilie in  $(\vartheta, \eta)$  und  $(T, U)$ , d.h. es existieren  $\mu$ -Dichten der Form

$$f_{\vartheta, \eta}(x) = C(\vartheta, \eta) \exp \left[ \vartheta T(x) + \sum_{j=1}^{k-1} \eta_j U_j(x) \right].$$

In diesem Fall schreiben wir anstelle von  $P_{\theta}$  auch  $P_{\vartheta, \eta}$ .

Im Folgenden stellen wir einige Eigenschaften von Exponentialfamilien zusammen, welche wir später benötigen.

**Lemma 5.4.** *Seien  $X_1 : (\mathcal{X}, \mathcal{F}) \rightarrow (\mathcal{X}_1, \mathcal{F}_1)$  und  $X_2 : (\mathcal{X}, \mathcal{F}) \rightarrow (\mathcal{X}_2, \mathcal{F}_2)$  zwei unabhängige Zufallsvariablen mit Verteilungen  $P_\vartheta^{X_i}$  auf  $(\mathcal{X}_i, \mathcal{F}_i)$ ,  $i = 1, 2$ , so dass  $(P_\vartheta^{X_i})_{\vartheta \in \Theta}$  Exponentialfamilien in  $\vartheta$  und  $T_i$  sind. Dann gilt für die gemeinsame Verteilung  $P_\vartheta^{(X_1, X_2)}$ : Die Verteilungsfamilie  $(P_\vartheta^{(X_1, X_2)})_{\vartheta \in \Theta}$  ist eine Exponentialfamilie in  $T = T_1 + T_2$  und  $\vartheta$ .*

*Beweis.* Gemäß der Voraussetzung existieren Maße  $\mu_i$ ,  $i = 1, 2$  so dass die Verteilungen  $P_\vartheta^{X_i}$   $\mu_i$ -Dichten der Form

$$f_\vartheta^{X_i}(x) = C_i(\vartheta) \exp \left[ \sum_{j=1}^k \vartheta_j T_{i,j}(x) \right] h_i(x), \quad i = 1, 2$$

besitzen. Aufgrund der Unabhängigkeit von  $X_1$  und  $X_2$  besitzt die gemeinsame Verteilung  $P_\vartheta^{(X_1, X_2)}$   $\mu_1 \otimes \mu_2$ -Dichten der Form

$$f_\vartheta^{(X_1, X_2)}(x) = C_1(\vartheta) C_2(\vartheta) \exp \left[ \sum_{j=1}^k \vartheta_j (T_{1,j}(x) + T_{2,j}(x)) \right] h_1(x) h_2(x)$$

und bildet somit eine Exponentialfamilie in  $T = T_1 + T_2$  und  $\vartheta$ .  $\square$

Nachfolgend noch einige Eigenschaften von Exponentialfamilien, welche ohne Beweise angegeben werden. Die Beweise sind an den angegebenen Literaturstellen zu finden, sowie in vielen weiteren Lehrbüchern zur mathematischen Statistik.

**Lemma 5.5** (Witting (1985), Satz 1.160). *Es sei  $(P_\vartheta^X)_{\vartheta \in \Theta}$  eine Exponentialfamilie in  $\vartheta$  und  $T$ . Dann bildet auch die Verteilung von  $T$  eine Exponentialfamilie in  $\vartheta$  und der identischen Abbildung.*

**Lemma 5.6** (Lehmann und Romano (2005), Lemma 2.7.2). *Es sei  $(P_{\vartheta, \eta}^X)_{(\vartheta, \eta) \in \Theta}$  eine Exponentialfamilie mit  $\mu$ -Dichten der Form*

$$f_{\vartheta, \eta}^X(x) = C(\vartheta, \eta) \exp \left[ \sum_{i=1}^r \vartheta_i T_i(x) + \sum_{j=1}^s \eta_j U_j(x) \right].$$

Dann existieren Maße  $\lambda_\vartheta$  und  $\nu_u$  so, dass gilt:

- (i) die Verteilung von  $U = (U_1, \dots, U_s)$  ist eine Exponentialfamilie mit  $\lambda_\vartheta$ -Dichten der Form

$$f_{\vartheta, \eta}^U(u) = C(\vartheta, \eta) \exp \left[ \sum_{j=1}^s \eta_j u_j \right].$$

- (ii) die bedingte Verteilung von  $T = (T_1, \dots, T_r)$  bei gegebenem  $U$  ist eine Exponentialfamilie mit  $\nu_u$ -Dichten der Form

$$f_{(\vartheta)}^{T|U}(t|u) = C(\vartheta) \exp \left[ \sum_{i=1}^r \vartheta_i t_i \right].$$

Aus der Definition der Exponentialfamilien und dem vorherigen Lemma folgt direkt:

**Korollar 5.7** (Suffizienz). Sei  $(P_{\vartheta, \eta}^X)_{(\vartheta, \eta) \in \Theta}$  eine Exponentialfamilie in  $(\vartheta, \eta)$  und  $(T, U)$ . Dann gilt:

- (i)  $(T, U)$  sind suffizient für  $(\vartheta, \eta)$   
(ii) Für festes  $\vartheta$  ist  $U$  suffizient für  $\eta$

**Lemma 5.8** (Vollständigkeit; Lehmann und Romano (2005), Satz 4.3.1). Es sei  $(P_{\vartheta}^X)_{\vartheta \in \Theta}$  eine Exponentialfamilie mit  $\mu$ -Dichten der Form

$$f_{\vartheta}^X(x) = C(\vartheta) \exp \left[ \sum_{i=1}^k \vartheta_i T_i(x) \right],$$

dann ist die Verteilungsfamilie  $(P_{\vartheta}^T)_{\vartheta \in \Gamma}$  von  $T = (T_1, \dots, T_k)$ , mit  $\Gamma \subset \Theta$ , vollständig, falls  $\Gamma$  ein  $k$ -dimensionales Intervall enthält.

**Lemma 5.9** (Stetigkeit und Differenzierbarkeit; Witting (1985), Satz 1.164). Sei  $(P_{\vartheta}^X)_{\vartheta \in \Theta}$  eine Exponentialfamilie in  $\vartheta$  und  $T$ . Ist  $\psi$  eine messbare Funktion auf  $(\mathcal{X}, \mathcal{F})$  mit  $\mathbb{E}_{\vartheta}\{\psi\} < \infty$  für alle  $\vartheta \in \Theta$ , so gilt: Die durch  $\beta(\vartheta) := \mathbb{E}_{\vartheta}\{\psi\}$  definierte Funktion ist stetig und beliebig oft differenzierbar.

**Lemma 5.10** (Tests in einparametrischen Exponentialfamilien; Witting (1985), Satz 2.24 und 2.70). Sei  $(P_{\vartheta}^X)_{\vartheta \in \Theta}$  eine einparametrische Exponentialfamilie in  $\vartheta$  und  $T$  und  $\alpha \in [0, 1]$ . Dann gilt:

- (i) Es existiert ein gleichmäßig bester Niveau- $\alpha$ -Test für das einseitige Testproblem  $H_0 : \vartheta \leq \vartheta_0$  gegen  $H_1 : \vartheta > \vartheta_0$ . Dieser ist gegeben in der Form

$$\psi(x) = 1_{\{T(x) > c(\alpha)\}} + \gamma(\alpha) 1_{\{T(x) = c(\alpha)\}},$$

wobei  $c(\alpha) \in \bar{\mathbb{R}}$  und  $\gamma(\alpha) \in [0, 1]$  bestimmt durch

$$\mathbb{E}_{\vartheta_0}\{\psi(X)\} = \alpha.$$

- (ii) Es existiert ein gleichmäßig bester unverfälschter Niveau- $\alpha$ -Test für das zweiseitige Testproblem  $H_0 : \vartheta = \vartheta_0$  gegen  $H_1 : \vartheta \neq \vartheta_0$ . Dieser ist gegeben in der Form

$$\psi(x) = 1_{\{T(x) < c_1(\alpha)\}} + 1_{\{T(x) > c_2(\alpha)\}} + \gamma_1(\alpha) 1_{\{T(x) = c_1(\alpha)\}} + \gamma_2(\alpha) 1_{\{T(x) = c_2(\alpha)\}},$$

wobei  $c_i(\alpha) \in \bar{\mathbb{R}}$  und  $\gamma_i(\alpha) \in [0, 1]$ ,  $i = 1, 2$  bestimmt werden durch

$$\mathbb{E}_{\vartheta_0}\{\psi(X)\} = \alpha \quad \text{und} \quad \mathbb{E}_{\vartheta_0}\{T(X)\psi(X)\} = \alpha \mathbb{E}_{\vartheta_0}\{T(X)\}.$$

**Lemma 5.11** (Tests in mehrparametrischen Exponentialfamilien; Witting (1985), Satz 3.60 und 3.62). Sei  $(P_{\vartheta, \eta}^X)_{(\vartheta, \eta) \in \Theta}$  eine  $k$ -parametrische Exponentialfamilie mit  $k > 1$  in  $(\vartheta, \eta)$  und  $(T, U)$ , und  $\alpha \in [0, 1]$ . Dann gilt:

- (i) Es existiert ein gleichmäßig bester unverfälschter Niveau- $\alpha$ -Test für das einseitige Testproblem  $H_0 : \vartheta \leq \vartheta_0$  gegen  $H_1 : \vartheta > \vartheta_0$ . Dieser ist gegeben in der Form

$$\psi(x) = 1_{\{T(x) > c(U(x); \alpha)\}} + \gamma(U(x); \alpha) 1_{\{T(x) = c(U(x); \alpha)\}},$$

wobei  $c(u; \alpha) \in \bar{\mathbb{R}}$  und  $\gamma(u; \alpha) \in [0, 1]$  bestimmt werden durch

$$\mathbb{E}_{\vartheta_0}\{\psi(X) | U(X) = u\} = \alpha.$$

(ii) Es existiert ein gleichmäßig bester unverfälschter Niveau- $\alpha$ -Test für das zweiseitige Testproblem  $H_0 : \vartheta = \vartheta_0$  gegen  $H_1 : \vartheta \neq \vartheta_0$ . Dieser ist gegeben in der Form

$$\begin{aligned} \psi(x) = & 1_{\{T(x) < c_1(U(x); \alpha)\}} + 1_{\{T(x) > c_2(U(x); \alpha)\}} \\ & + \gamma_1(U(x); \alpha) 1_{\{T(x) = c_1(U(x); \alpha)\}} + \gamma_2(U(x); \alpha) 1_{\{T(x) = c_2(U(x); \alpha)\}}, \end{aligned}$$

wobei  $c_i(u; \alpha)$  und  $\gamma_i(u; \alpha)$ ,  $i = 1, 2$  bestimmt werden durch

$$\begin{aligned} \mathbb{E}_{\vartheta_0}\{\psi(X)|U(X) = u\} &= \alpha \quad \text{und} \\ \mathbb{E}_{\vartheta_0}\{T(X)\psi(X)|U(X) = u\} &= \alpha \mathbb{E}_{\vartheta_0}\{T(X)|U(X) = u\}. \end{aligned}$$

### 5.1.2 Voraussetzungen

In diesem Abschnitt wollen wir einige Voraussetzungen formulieren, die in den nachfolgenden Abschnitten dieses Kapitels immer wieder benötigt werden.

**Voraussetzungen 5.12.** Gegeben sei die Menge der Adaptionen  $\mathcal{M} \subset \mathbb{N}$  (die möglichen Gesamtfallzahlen). Sei  $N := \max \mathcal{M}$ , falls  $|\mathcal{M}| < \infty$ , und  $N := \infty$ , falls  $|\mathcal{M}| = \infty$ . Dann definieren wir  $I \subset \mathbb{N}$  durch  $I := \{1, \dots, N\}$  falls  $|\mathcal{M}| < \infty$  und  $I := \mathbb{N}$  falls  $|\mathcal{M}| = \infty$ .

**Voraussetzungen 5.13** (und Folgerungen). Seien  $\mathcal{M}$  und  $I$  definiert gemäß Voraussetzung 5.12. Sei  $(\mathfrak{X}_n, \mathcal{F}_n)$  für alle  $n \in I$  ein Stichprobenraum. Dann bildet auch  $(\mathfrak{X}, \mathcal{F})$  gegeben durch  $(\mathfrak{X}, \mathcal{F}) = (\prod_{n \in I} \mathfrak{X}_n, \otimes_{n \in I} \mathcal{F}_n)$  einen Stichprobenraum.

**Voraussetzungen 5.14** (und Folgerungen). Seien  $\mathcal{M}$ ,  $I$ ,  $(\mathfrak{X}_n, \mathcal{F}_n)$ ,  $n \in I$  und  $(\mathfrak{X}, \mathcal{F})$  wie in Voraussetzungen 5.12 und 5.13. Sei  $(Y_n)_{n \in I}$  eine Folge von unabhängigen Zufallsvariablen  $Y_n : (\mathfrak{X}, \mathcal{F}) \rightarrow (\mathfrak{X}_n, \mathcal{F}_n)$ , sei  $X_1 = (Y_1, \dots, Y_{n_1})$  (die Daten der 1. Stufe) und  $X_{2m} = (Y_{n_1+1}, \dots, Y_m)$ ,  $m \in \mathcal{M}$  (die Daten der 2. Stufe). Dann sind  $X_1$  und  $(X_{2m})_{m \in \mathcal{M}}$  unabhängig.



**Voraussetzungen 5.15** (und Folgerungen). Seien  $(Y_n)_{n \in I}$ ,  $X_1$  und  $(X_{2m})_{m \in \mathcal{M}}$  wie in Voraussetzung 5.14 mit  $Y_n \sim P_\vartheta^{Y_n}$ . Die Verteilungsfamilien  $(P_\vartheta^{Y_n})_{\vartheta \in \Theta}$ ,  $n \in I$ , seien Exponentialfamilien in  $T_n^Y$  und  $\vartheta$ . Dann sind auch die Verteilungsfamilien  $(P_\vartheta^{X_1})_{\vartheta \in \Theta}$  von  $X_1$  und  $(P_\vartheta^{X_{2m}})_{\vartheta \in \Theta}$ ,  $m \in \mathcal{M}$ , von  $X_{2m}$  Exponentialfamilien in  $T_1 = \sum_{i=1}^{n_1} T_i^Y$  bzw.  $T_{2m} = \sum_{i=n_1+1}^m T_i^Y$  und  $\vartheta$  nach Lemma 5.4.

Wir wollen im Weiteren annehmen, dass der Parameterraum konvex ist und dass der interessierende Parameter  $\vartheta_0$  im Inneren liegt.

**Bezeichnungen 5.16.** Die entsprechenden Bildräume der Statistiken  $T_1$ ,  $T_{2m}$ ,  $U_1$  und  $U_{2m}$  seien bezeichnet durch  $\mathfrak{X}_1^T, \mathfrak{X}_{2m}^T \subset \mathbb{R}$  und  $\mathfrak{X}_1^U, \mathfrak{X}_{2m}^U \subset \mathbb{R}^{k-1}$ .

**Korollar 5.17.** Erfülle  $(X_{2m})_{m \in \mathcal{M}}$  die Voraussetzung 5.15. Dann gilt für alle  $m_1, m_2 \in \mathcal{M}$  mit  $m_1 < m_2$  und alle messbaren Abbildungen  $g : \mathfrak{X}_{2m_1} \rightarrow \mathbb{R}$  mit  $\mathbb{E}_\vartheta\{g(X_{2m_1})\} < \infty$ :

$$\mathbb{E}_\vartheta\{g(X_{2m_1}) | T_{2m_2}(X_{2m_2})\} \text{ ist unabhängig von } \vartheta.$$

*Insbesondere gilt im Falle einer  $k$ -parametrischen Exponentialfamilie*

$$\mathbb{E}_{\vartheta, \eta}\{g(X_{2m_1}) | U_{2m_2}(X_{2m_2})\} \text{ ist unabhängig von } \eta. \quad (5.2)$$

*Beweis.* Nach Voraussetzung gilt  $X_{2m_1} = (Y_{n_1+1}, \dots, Y_{m_1})$  und  $X_{2m_2} = (Y_{n_1+1}, \dots, Y_{m_1}, \dots, Y_{m_2})$ . Daher kann  $X_{2m_1}$  dargestellt werden als  $X_{2m_1} = \pi(X_{2m_2})$ , wobei  $\pi$  die Projektion auf die entsprechenden Komponenten ist. Somit folgt die Behauptung aus der Suffizienz von  $T_{2m_2}$  für  $(P_\vartheta^{X_{2m_2}})_{\vartheta \in \Theta}$  (Korollar 5.7). Aussage (5.2) folgt analog aus der Suffizienz von  $U_{2m_2}$  für  $\eta$  bei festem  $\vartheta$ .  $\square$

## 5.2 Flexible Designs in einparametrischen Exponentialfamilien

### 5.2.1 Einseitige Testprobleme

**Satz 5.18** (einseitiges Testproblem). *Gegeben sei das flexible Experiment  $((\mathfrak{X}, \mathcal{F}), \Theta, (P_\vartheta)_{\vartheta \in \Theta}, (\mathcal{M}, \mathcal{M}), (\sigma(X_1, X_{2m}))_{m \in M}, \sigma(X_1), \mathcal{T})$ , wobei  $X_1$  und  $(X_{2m})_{m \in M}$  gemäß Voraussetzung 5.12-5.15 definiert sind, so dass  $(P_\vartheta^{X_1})_{\vartheta \in \Theta}$  und  $(P_\vartheta^{X_{2m}})_{\vartheta \in \Theta}$  einparametrische Exponentialfamilien in  $\vartheta$  und  $T_1$  bzw.  $T_{2m}$  bilden. Sei  $\mathcal{T}$  die Menge aller Adaptionsregeln. Existiert eine messbare Abbildung  $A : \mathfrak{X}_1^T \rightarrow [0, 1]$  mit*

$$\begin{aligned} \mathbb{E}_{\vartheta_0} \{A(T_1(X_1))\} &= \alpha \quad \text{und} \\ A(t) &\text{ ist eine monoton wachsende Funktion in } t, \end{aligned} \quad (5.3)$$

so gilt für das einseitige Testproblem  $H_0 : \vartheta \leq \vartheta_0$  gegen  $H_1 : \vartheta > \vartheta_0$ :

- (i) *Ist  $x_1 \in \mathfrak{X}_1$  und  $m \in M$  gegeben. Dann existiert ein gleichmäßig bester Test zum bedingten Niveau  $A(T_1(x_1))$ . Dieser ist gegeben durch*

$$\begin{aligned} \psi_{1,m}^{x_1}(x_{2m}) &= \mathbb{1}_{\{T_{2m}(x_{2m}) > c_m(A(T_1(x_1)))\}} \\ &\quad + \gamma_m(A(T_1(x_1))) \mathbb{1}_{\{T_{2m}(x_{2m}) = c_m(A(T_1(x_1)))\}}, \end{aligned} \quad (5.4)$$

wobei  $c_m(A(T_1(x_1))) \in \bar{\mathbb{R}}$  und  $\gamma_m(A(T_1(x_1))) \in [0, 1]$  bestimmt werden aus

$$\mathbb{E}_{\vartheta_0} \{\psi_{1,m}^{x_1}(X_{2m})\} = A(T_1(x_1)).$$

- (ii) *Ist  $\psi_{1,m}(x_1, x_{2m}) := \psi_{1,m}^{x_1}(x_{2m})$ . Dann sind die Abbildungen  $\psi_{1,m}(X_1, X_{2m}) : \mathfrak{X} \rightarrow [0, 1]$  messbar und es gilt*

$$\mathbb{E}_{\vartheta_0} \{\psi_{1,m}(X_1, X_{2m}) | X_1\} = A(T_1(X_1)) \text{ fs.} \quad (5.5)$$

(iii) Für  $\tau \in \mathcal{T}$  ist  $\psi_{1,\tau} = \sum_{m \in \mathcal{M}} 1_{\{\tau=m\}} \psi_{1,m}(X_1, X_{2m})$ . Dann gilt:

$$\mathbb{E}_{\vartheta}\{\psi_{1,\tau}\} \leq \alpha, \quad \text{für alle } \vartheta \leq \vartheta_0$$

und

$$\mathbb{E}_{\vartheta}\{\psi_{1,\tau}\} \geq \alpha, \quad \text{für alle } \vartheta \geq \vartheta_0.$$

(iv) Sei  $(\mathcal{T}, \phi_1)$  mit  $\phi_1 = (\phi_{1,m})_{m \in \mathcal{M}}$  ein weiterer flexibler Niveau- $\alpha$ -Test. Dann gilt:

$$\mathbb{E}_{\vartheta}\{\psi_{1,\tau}\} \geq \mathbb{E}_{\vartheta}\{\phi_{1,\tau}\} \quad \text{für alle } \tau \in \mathcal{T}, \vartheta > \vartheta_0.$$

*Beweis.* (i) Nach Voraussetzung 5.15 bildet die Verteilungsfamilie von  $X_{2m}$  eine Exponentialfamilie in  $\vartheta$  und  $T_{2m}$ . Somit folgt (i) direkt aus Lemma 5.10 (i).

(ii) Nach der Konstruktion der Tests in Lemma 5.10 sind  $c_m : [0, 1] \rightarrow \bar{\mathbb{R}}$  und  $\gamma_m : [0, 1] \rightarrow [0, 1]$  messbare Abbildungen, also auch  $c_m(A(T_1(x_1)))$  und  $\gamma_m(A(T_1(x_1)))$ . Daher ist der Beweis der Messbarkeit der Menge  $\{(t_1, t_{2m}) | t_{2m} > c(A(t_1))\}$  analog zu dem Beweis des Hilfssatzes 3.61 in Witting (1985). Somit erhalten wir die Messbarkeit von  $\psi_{1,m}(X_1, X_{2m})$ . Die Gleichheit (5.5) folgt direkt aus der Unabhängigkeit von  $X_1$  und  $X_{2m}$ .

(iii) Da  $\psi_{1,m}(X_1, X_{2m})$  für gegebenes  $X_1 = x_1$  gleichmäßig bester Test zum bedingten Niveau  $A(T(x_1))$  ist, folgt für alle  $m \in \mathcal{M}$ :

$$\mathbb{E}_{\vartheta}\{\psi_{1,m}(X_1, X_{2m}) | X_1\} \leq A(T_1(X_1)) \quad \text{fs} \quad \text{für alle } \vartheta < \vartheta_0,$$

$$\mathbb{E}_{\vartheta}\{\psi_{1,m}(X_1, X_{2m}) | X_1\} \geq A(T_1(X_1)) \quad \text{fs} \quad \text{für alle } \vartheta > \vartheta_0.$$

Nach Satz 4.12 bleibt zu zeigen:

$$\mathbb{E}_{\vartheta}\{A(T_1(X_1))\} \leq \alpha \quad \text{für alle } \vartheta < \vartheta_0,$$

$$\mathbb{E}_{\vartheta}\{A(T_1(X_1))\} \geq \alpha \quad \text{für alle } \vartheta > \vartheta_0.$$

Wir werden nur die erste Ungleichung zeigen. Die zweite folgt analog. Sei  $f_{\vartheta}^{T_1}$  die  $\mu$ -Dichte von  $P_{\vartheta}^{T_1}$  nach Lemma 5.5, dann ist der Dichtequotient  $\frac{f_{\vartheta_0}^{T_1}(t)}{f_{\vartheta}^{T_1}(t)}$  monoton wachsend in  $t$  für  $\vartheta < \vartheta_0$ . Somit existiert ein  $t_0$  so, dass  $f_{\vartheta_0}^{T_1}(t) > f_{\vartheta}^{T_1}(t)$  für alle  $t > t_0$ . Sei  $a := \sup_{t \leq t_0} A(t)$  und  $b := \inf_{t > t_0} A(t)$ , dann gilt  $a \leq b$ , da  $A(t)$  monoton wachsend ist. Daher folgt:

$$\begin{aligned}
\mathbb{E}_{\vartheta_0}\{A(T_1(X_1))\} - \mathbb{E}_{\vartheta}\{A(T_1(X_1))\} &= \int A(t)(f_{\vartheta_0}^{T_1}(t) - f_{\vartheta}^{T_1}(t))d\mu(t) \\
&= \int_{\{t \leq t_0\}} A(t)(f_{\vartheta_0}^{T_1}(t) - f_{\vartheta}^{T_1}(t))d\mu(t) + \int_{\{t > t_0\}} A(t)(f_{\vartheta_0}^{T_1}(t) - f_{\vartheta}^{T_1}(t))d\mu(t) \\
&\geq a \int_{\{t \leq t_0\}} (f_{\vartheta_0}^{T_1}(t) - f_{\vartheta}^{T_1}(t))d\mu(t) + b \int_{\{t > t_0\}} (f_{\vartheta_0}^{T_1}(t) - f_{\vartheta}^{T_1}(t))d\mu(t) \\
&= (b - a) \int_{\{t > t_0\}} (f_{\vartheta_0}^{T_1}(t) - f_{\vartheta}^{T_1}(t))d\mu(t) + a \int_{\{t > t_0\}} (f_{\vartheta_0}^{T_1}(t) - f_{\vartheta}^{T_1}(t))d\mu(t) \\
&\geq 0.
\end{aligned}$$

(iv) Aus Korollar 4.13 folgt

$$\mathbb{E}_{\vartheta_0}\{\phi_{1,m}(X_1, X_{2m})|X_1\} \leq A(T_1(X_1)) \text{ für alle } m \in \mathcal{M}.$$

Da  $\psi_{1,m}(X_1, X_{2m})$  für gegebenes  $X_1 = x_1$  gleichmäßig bester Test zum bedingten Niveau  $A(T_1(x_1))$  ist, folgt

$$\mathbb{E}_{\vartheta}\{\psi_{1,m}(X_1, X_{2m})|X_1\} \geq \mathbb{E}_{\vartheta}\{\phi_{1,m}(X_1, X_{2m})|X_1\} \text{ fs für alle } m \in \mathcal{M}.$$

Somit ergibt sich

$$\begin{aligned}
\mathbb{E}_{\vartheta}\{\phi_{1,\tau}\} &= \sum_{m \in \mathcal{M}} \mathbb{E}_{\vartheta}\{1_{\{\tau=m\}}\} \mathbb{E}\{\phi_{1,m}(X_1, X_{2m})|\sigma(X_1), \mathcal{F}\} \\
&= \sum_{m \in \mathcal{M}} \mathbb{E}_{\vartheta}\{1_{\{\tau=m\}}\} \mathbb{E}\{\phi_{1,m}(X_1, X_{2m})|\sigma(X_1)\} \\
&\leq \sum_{m \in \mathcal{M}} \mathbb{E}_{\vartheta}\{1_{\{\tau=m\}}\} \mathbb{E}\{\psi_{1,m}(X_1, X_{2m})|\sigma(X_1)\} \\
&= \mathbb{E}_{\vartheta}\{\psi_{1,\tau}\}.
\end{aligned}$$

□

*Bemerkung 5.2.* Anstelle der Voraussetzung (5.3), dass  $A(t)$  monoton wachsend in  $t$  ist, hätte es auch gereicht zu fordern

$$\mathbb{E}_{\vartheta}\{A(T_1(X_1))\} \leq \alpha \text{ für alle } \vartheta < \vartheta_0 \text{ und } \mathbb{E}_{\vartheta}\{A(T_1(X_1))\} \geq \alpha \text{ für alle } \vartheta > \vartheta_0.$$

Andererseits ist die Forderung, dass  $A(t)$  monoton wachsend ist, plausibel und in der Regel auch erfüllt, wie die nachfolgenden Korollare zeigen.

**Korollar 5.19** (flexible Designs). *Gegeben sei das flexible Experiment  $((\mathfrak{X}, \mathcal{F}), \Theta, (P_{\vartheta})_{\vartheta \in \Theta}, (\mathcal{M}, \mathcal{M}), (\sigma(X_1, X_{2m}))_{m \in M}, \sigma(X_1), \mathcal{T})$  wie in Satz 5.18 und die Menge aller Adaptionsregeln  $\mathcal{T}$ . Sei  $\tilde{A} : [0, 1] \rightarrow [0, 1]$  eine bedingte Fehlerfunktion wie in Abschnitt 3.5.2. Dann ist die bedingte Ablehnungswahrscheinlichkeit bei gegebenem  $T_1(X_1) = t_1$  bestimmt durch*

$$A(t_1) = \tilde{A}(p_1(t_1)).$$

Hierbei bezeichnet  $p_1$  den  $p$ -Wert eines gleichmäßig besten Tests für das einseitige Testproblem  $H_0 : \vartheta \leq \vartheta_0$  gegen  $H_1 : \vartheta > \vartheta_0$ . Insbesondere ist  $A(t_1)$  eine monoton wachsende Funktion in  $t_1$ .

*Beweis.* Im Falle eines gleichmäßig besten Tests ist der  $p$ -Wert für gegebenes  $T_1(X_1) = t_1$  gleich

$$p_1(t_1) = P(T_1(X_1) \geq t_1).$$

Offensichtlich ist  $p_1(t_1)$  monoton fallend in  $t_1$ . Damit ergibt sich für  $t_1 < \tilde{t}_1$ ,

$$A(t_1) = \tilde{A}(p_1(t_1)) \leq \tilde{A}(p_1(\tilde{t}_1)) = A(\tilde{t}_1),$$

da  $\tilde{A}(p_1)$  nach Abschnitt 3.5.2 monoton fallend in  $p_1$  ist. □

**Korollar 5.20** (Fallzahländerungen mit dem CRP-Prinzip). *Gegeben sei das flexible Experiment  $((\mathfrak{X}, \mathcal{F}), \Theta, (P_{\vartheta})_{\vartheta \in \Theta}, (\mathcal{M}, \mathcal{M}), (\sigma(X_1, X_{2m}))_{m \in M}, \sigma(X_1), \mathcal{T})$  wie in Satz 5.18 und die Menge aller Adaptionsregeln  $\mathcal{T}$ . Sei  $\psi_{m^*}$  ein gleichmäßig bester Niveau- $\alpha$ -Test für das einseitige Testproblem  $H_0 : \vartheta \leq \vartheta_0$  gegen*

$H_1 : \vartheta > \vartheta_0$ ,  $m^* \in \mathcal{M}$ . Dann ist die bedingte Ablehnungswahrscheinlichkeit bei gegebenem  $T_1(X_1) = t_1$  gleich

$$\begin{aligned} A(t_1) = & P_{\vartheta_0}(T_{2m^*}(X_{2m^*}) > c_{m^*}(\alpha) - t_1) \\ & + \gamma_{m^*}(\alpha) P_{\vartheta_0}(T_{2m^*}(X_{2m^*}) = c_{m^*}(\alpha) - t_1). \end{aligned} \quad (5.6)$$

Insbesondere ist  $A(t_1)$  eine monoton wachsende Funktion in  $t_1$ .

*Beweis.* Da  $\psi_{m^*}$  gleichmäßig bester Niveau- $\alpha$ -Test ist, ist  $\psi_{m^*}$  gegeben durch

$$\psi_{m^*}(X_1, X_{2m}) = 1_{\{T_1(X_1) + T_{2m^*}(X_{2m^*}) > c_{m^*}(\alpha)\}} + \gamma_{m^*}(\alpha) 1_{\{T_1(X_1) + T_{2m^*}(X_{2m^*}) = c_{m^*}(\alpha)\}}.$$

Aus der Unabhängigkeit von  $X_1$  und  $X_{2m^*}$  ergibt sich somit die bedingte Ablehnungswahrscheinlichkeit (5.6).

Sei  $F_{2m^*}$  die Verteilungsfunktion von  $T_{2m^*}$ , dann gilt

$$A(t_1) = 1 - F_{2m^*}(c_{m^*}(\alpha) - t_1) + \gamma_{m^*}(\alpha) [F_{2m^*}(c_{m^*}(\alpha) - t_1) - F_{2m^*}(c_{m^*}(\alpha) - t_1 - 0)].$$

Daraus folgt

$$1 - F_{2m^*}(c_{m^*}(\alpha) - t_1 - 0) \geq A(t_1) \geq 1 - F_{2m^*}(c_{m^*}(\alpha) - t_1). \quad (5.7)$$

Sei  $t_1 < \tilde{t}_1$ , dann erhalten wir aufgrund der Monotonie der Verteilungsfunktion  $F_{2m^*}$ :

$$\begin{aligned} F_{2m^*}(c_{m^*}(\alpha) - \tilde{t}_1 - 0) & \leq F_{2m^*}(c_{m^*}(\alpha) - \tilde{t}_1) \\ & \leq F_{2m^*}(c_{m^*}(\alpha) - t_1 - 0) \leq F_{2m^*}(c_{m^*}(\alpha) - t_1). \end{aligned}$$

Hieraus folgt mit (5.7), dass  $A(t_1)$  monoton wachsend in  $t_1$  ist.  $\square$

**Korollar 5.21** (Fallzahlverlängerungen mit dem CRP-Prinzip). *Gegeben sei das flexible Experiment  $((\mathfrak{X}, \mathcal{F}), \Theta, (P_{\vartheta})_{\vartheta \in \Theta}, (\mathcal{M}, \mathcal{M}), (\sigma(X_1, X_{2m}))_{m \in \mathcal{M}}, \sigma(X_1, \mathcal{T}))$  wie in Satz 5.18 und die Menge aller Adaptionsregeln  $\mathcal{T}$ . Sei  $\psi_{m^*}$  gleichmäßig bester Niveau- $\alpha$ -Test für das einseitige Testproblem  $H_0 : \vartheta \leq \vartheta_0$  gegen  $H_1 : \vartheta > \vartheta_0$ ,  $m^* \in \mathcal{M}$  mit  $m > m^*$  für alle  $m \in \mathcal{M}$ . Dann existiert ein flexibler Niveau- $\alpha$ -Test  $(\mathcal{T}, \psi)$  zu  $\psi_{m^*}$  mit*

$$\mathbb{E}_{\vartheta}\{\psi_{\tau}\} \geq \mathbb{E}_{\vartheta}\{\psi_{m^*}(X_1, X_{2m^*})\} \text{ für alle } \vartheta \geq \vartheta_0 \text{ und alle } \tau \in \mathcal{T}. \quad (5.8)$$

*Beweis.* Nach Korollar 5.20 ist die durch  $\psi_{m^*}(X_1, X_{2m^*})$  definierte bedingte Ablehnungswahrscheinlichkeit  $A(t_1)$  monoton wachsend in  $t_1$ . Daher existiert nach Satz 5.18 ein flexibler Niveau- $\alpha$ -Test, so dass die Tests  $\psi_m$ ,  $m \in \mathcal{M}$  durch (5.4) gegeben sind.

Nach Korollar 5.17 ist

$$\tilde{\psi}_m(T_1(X_1), T_{2m}(X_{2m})) = \mathbb{E}\{\psi_{m^*}(X_1, X_{2m^*}) | T_1(X_1), T_{2m}(X_{2m})\}$$

unabhängig von  $\vartheta$ . Insbesondere gilt

$$\mathbb{E}_{\vartheta_0}\{\tilde{\psi}_m(X_1, X_{2m}) | T_1(X_1)\} = \mathbb{E}_{\vartheta_0}\{\psi_{m^*}(X_1, X_{2m^*}) | T_1(X_1)\} = A(T_1(X_1)) \text{ fs}$$

Damit ist  $\tilde{\psi}_m(X_1, X_{2m})$  ein weiterer Test, welcher das bedingte Niveau  $A(T_1(X_1))$  einhält. Da  $\psi_m(X_1, X_{2m})$  für gegebenes  $X_1 = x_1$  gleichmäßig bester Test zu dem bedingten Niveau ist, folgt für alle  $x_1 \in \mathfrak{X}_1$

$$\begin{aligned} \mathbb{E}_{\vartheta}\{\psi_m(X_1, X_{2m}) | X_1 = x_1\} &\geq \mathbb{E}_{\vartheta}\{\tilde{\psi}_m(X_1, X_{2m}) | X_1 = x_1\} \\ &= \mathbb{E}_{\vartheta}\{\psi_{m^*}(X_1, X_{2m^*}) | X_1 = x_1\}. \end{aligned}$$

Somit ergibt sich für alle  $\tau \in \mathcal{T}$  durch eine analoge Rechnung, wie im Beweis von Satz 5.18 (iv),

$$\begin{aligned} \mathbb{E}_{\vartheta}\{\psi_{\tau}\} &= \mathbb{E}_{\vartheta}\left\{\sum_{m \in \mathcal{M}} \mathbb{E}_{\vartheta}\{1_{\{\tau=m\}}\} \mathbb{E}\{\psi_{1,m}(X_1, X_{2m}) | X_1\}\right\} \\ &\geq \mathbb{E}_{\vartheta}\left\{\sum_{m \in \mathcal{M}} \mathbb{E}_{\vartheta}\{1_{\{\tau=m\}}\} \mathbb{E}\{\psi_{1,m^*}(X_1, X_{2m^*}) | X_1\}\right\} = \mathbb{E}\{\psi_{1,m^*}(X_1, X_{2m^*})\}. \end{aligned}$$

□

*Bemerkung 5.3.* Dieses Korollar zeigt, dass eine Fallzahlverlängerung nach dem CRP-Prinzip immer zu einem Powergewinn im Vergleich zu dem ursprünglichen, optimalen (gleichmäßig besten) Test führt.

## 5.2.2 Zweiseitige Testprobleme

Für ein zweiseitiges Testproblem lässt sich zu gegebener bedingter Fehlerfunktion wieder ein flexibler Niveau- $\alpha$ -Test konstruieren, der zudem noch

unverfälscht ist (analog Satz 5.18). Allerdings kann nicht mehr gezeigt werden, dass dieser ein gleichmäßig bester Test ist. Wir werden aber für den Spezialfall der Fallzahlverlängerung nach dem CRP-Prinzip auf eine weitere Art einen flexiblen unverfälschten Niveau- $\alpha$ -Test konstruieren, der eine Ungleichung wie in (5.8) erfüllt. Dieser ist in manchen Situationen dem vorherigen überlegen.

**Satz 5.22** (zweiseitiges Testproblem). *Gegeben sei das flexible Experiment  $((\mathfrak{X}, \mathcal{F}), \Theta, (P_\vartheta)_{\vartheta \in \Theta}, (\mathcal{M}, \mathcal{M}), (\sigma(X_1, X_{2m}))_{m \in M}, \sigma(X_1), \mathcal{T})$  wie in Satz 5.18 und die Menge aller Adaptionsregeln  $\mathcal{T}$ . Existiert eine messbare Abbildung  $A : \mathfrak{X}_1^T \rightarrow [0, 1]$  mit*

$$\begin{aligned} \mathbb{E}_{\vartheta_0}\{A(T_1(X_1))\} &= \alpha \quad \text{und} \\ \mathbb{E}_{\vartheta}\{A(T_1(X_1))\} &\geq \alpha \quad \text{für alle } \vartheta \neq \vartheta_0, \end{aligned} \quad (5.9)$$

so gilt für das zweiseitige Testproblem  $H_0 : \vartheta = \vartheta_0$  gegen  $H_1 : \vartheta \neq \vartheta_0$ :

(i) Sei  $x_1 \in \mathfrak{X}_1$  und  $m \in \mathcal{M}$  gegeben. Dann existiert ein gleichmäßig bester unverfälschter Test zum bedingten Niveau  $A(T_1(x_1))$ . Dieser ist gegeben durch

$$\begin{aligned} \psi_{2,m}^{x_1}(X_{2m}) &= 1_{\{T_{2m}(X_{2m}) < c_{1,m}(A(T_1(x_1)))\}} + 1_{\{T_{2m}(X_{2m}) > c_{2,m}(A(T_1(x_1)))\}} \\ &\quad + \gamma_{1,m}(A(T_1(x_1))) 1_{\{T_{2m}(X_{2m}) = c_{1,m}(A(T_1(x_1)))\}} \\ &\quad + \gamma_{2,m}(A(T_1(x_1))) 1_{\{T_{2m}(X_{2m}) = c_{2,m}(A(T_1(x_1)))\}}, \end{aligned}$$

wobei  $c_{i,m}(A(T_1(x_1))) \in \bar{\mathbb{R}}$  und  $\gamma_{i,m}(A(T_1(x_1))) \in [0, 1]$ ,  $i = 1, 2$  bestimmt werden aus

$$\begin{aligned} \mathbb{E}_{\vartheta_0}\{\psi_{2,m}^{x_1}(X_{2m})\} &= A(T_1(x_1)) \quad \text{und} \\ \mathbb{E}_{\vartheta_0}\{T_{2m}(X_{2m}) \psi_{2,m}^{x_1}(X_{2m})\} &= A(T_1(x_1)) \mathbb{E}_{\vartheta_0}\{T_{2m}(X_{2m})\}. \end{aligned}$$

(ii) Setze  $\psi_{2,m}(x_1, x_{2m}) := \psi_{2,m}^{x_1}(x_{2m})$ . Dann sind die Abbildungen  $\psi_{2,m}(X_1, X_{2m}) : \mathfrak{X} \rightarrow [0, 1]$  messbar und es gilt

$$\begin{aligned} \mathbb{E}_{\vartheta_0}\{\psi_{2,m}(X_1, X_{2m}) | X_1\} &= A(T_1(X_1)) \quad \text{fs} \quad \text{und} \\ \mathbb{E}_{\vartheta_0}\{T_{2m}(X_{2m}) \psi_{2,m}(X_1, X_{2m}) | X_1\} &= A(T_1(X_1)) \mathbb{E}_{\vartheta_0}\{T_{2m}(X_{2m}) | X_1\} \quad \text{fs}. \end{aligned}$$



(iii) Für  $\tau \in \mathcal{T}$  ist  $\psi_{2,\tau} = \sum_{m \in \mathcal{M}} 1_{\{\tau=m\}} \psi_{2,m}(X_1, X_{2m})$ . Dann gilt

$$\mathbb{E}_{\vartheta_0} \{\psi_{2,\tau}\} = \alpha$$

und

$$\mathbb{E}_{\vartheta} \{\psi_{2,\tau}\} \geq \alpha, \quad \text{für alle } \vartheta \neq \vartheta_0.$$

*Beweis.* (i) und (ii) folgen wie im Beweis von Satz 5.18 aus dem zweiseitigen Test 5.10.

(iii) Da  $\psi_{2,m}(X_1, X_{2m})$  für gegebenes  $X_1 = x_1$  gleichmäßig bester unverfälschter Test zum bedingten Niveau  $A(T(x_1))$  ist, folgt für alle  $m \in \mathcal{M}$ :

$$\mathbb{E}_{\vartheta} \{\psi_{2,m}(X_1, X_{2m}) | X_1\} \geq A(T_1(X_1)) \quad \text{für alle } \vartheta \neq \vartheta_0.$$

Hieraus folgt die Behauptung mit Hilfe von (5.9).

□

*Bemerkung 5.4.* Die in Satz 5.22 konstruierten flexiblen Niveau- $\alpha$ -Tests sind nur die besten Tests unter der Bedingung, dass die bedingte Powerfunktion, definiert als  $\mathbb{E}_{\vartheta} \{\psi_{2,m}(X_1, X_{2m}) | X_1 = x_1\}$ , für  $\vartheta = \vartheta_0$  ein Minimum besitzt. Ob dies wirklich zu guten flexiblen Tests führt ist noch nicht geklärt, zumal für die gleichmäßig besten unverfälschten Tests gilt, dass die bedingte Powerfunktion für  $\vartheta = \vartheta_0$  fast überall eine von 0 verschiedene Steigung besitzt, wie wir nachfolgend noch sehen werden.

Wenn wir das Korollar 4.13 auf eine Familie von Abbildungen  $(A_{\vartheta})_{\vartheta \in \Theta}$  und die Unverfälschtheit erweitern, dann besitzt  $A_{\vartheta}(t_1)$  als Funktion von  $\vartheta$  für festes  $t_1$  ein Maximum in  $\vartheta = \vartheta_0$ . Hierbei bedeutet die Unverfälschtheit, dass

$$\mathbb{E}_{\vartheta} \{\psi_{2,m} | T_1(X_1)\} \geq A_{\vartheta}(T_1(X_1)) \quad \text{für alle } m \in \mathcal{M} \text{ und } \vartheta \neq \vartheta_0,$$

wobei die  $A_\vartheta$  wie in Bemerkung 4.6 definiert seien. Aus der Bedingung

$$\mathbb{E}_\vartheta\{\psi_{2,m}(X_1, X_{2m})|T_1(X_1)\} \geq A_\vartheta(T_1(X_1)) \text{ f\"ur alle } \vartheta \in \Theta$$

wobei f\"ur  $\vartheta = \vartheta_0$  die Gleichheit gilt, folgt, die bedingte G\"utefunktion von  $\psi_{2,m}(X_1, X_{2m})$  muss die Funktion  $A_\vartheta(t_1)$ , bei gegebenem  $T_1(X_1) = t_1$ , an der Stelle  $\vartheta = \vartheta_0$  ber\"uhren. Daher m\"ussen beide die Steigung 0 besitzen, da f\"ur  $\vartheta = \vartheta_0$  ein Maximum von  $A_\vartheta(t_1)$  vorliegt. In diesem Fall m\"ussen die bedingten Powerfunktionen der Tests f\"ur alle  $m \in \mathcal{M}$  ein Minimum f\"ur  $\vartheta = \vartheta_0$  besitzen.

**Korollar 5.23** (Fallzahlverl\"angerungen mit dem CRP-Prinzip). *Gegeben sei das flexible Experiment  $((\mathfrak{X}, \mathcal{F}), \Theta, (P_\vartheta)_{\vartheta \in \Theta}, (\mathcal{M}, \mathcal{M}), (\sigma(X_1, X_{2m}))_{m \in \mathcal{M}}, \sigma(X_1), \mathcal{T})$  wie in Satz 5.22 und die Menge aller Adaptionsregeln  $\mathcal{T}$ . F\"ur  $m^* \in \mathcal{M}$  mit  $m > m^*$  f\"ur alle weiteren  $m \in \mathcal{M}$  sei  $\psi_{m^*}$  ein gleichm\"a\ssig bester unverf\"alschter Niveau- $\alpha$ -Test f\"ur das zweiseitige Testproblem  $H_0 : \vartheta = \vartheta_0$  gegen  $H_1 : \vartheta \neq \vartheta_0$ . Dann existiert ein flexibler Niveau- $\alpha$ -Test  $(\mathcal{T}, \psi)$  zu  $\psi_{m^*}$  mit*

$$\mathbb{E}_\vartheta\{\psi_\tau\} \geq \mathbb{E}_\vartheta\{\psi_{m^*}(X_1, X_{2m^*})\} \text{ f\"ur alle } \vartheta \geq \vartheta_0 \text{ und alle } \tau \in \mathcal{T}.$$

Dabei sind die  $\psi_m(x_1, x_{2m})$  gegeben in der Form

$$\begin{aligned} \psi_m(x_1, x_{2m}) = & 1_{\{T_{2m}(x_{2m}) < c_{1,m}(A(T_1(x_1)))\}} + 1_{\{T_{2m}(x_{2m}) > c_{2,m}(A(T_1(x_1)))\}} \\ & + \gamma_{1,m}(A(T_1(x_1)))1_{\{T_{2m}(x_{2m}) = c_{1,m}(A(T_1(x_1)))\}} \\ & + \gamma_{2,m}(A(T_1(x_1)))1_{\{T_{2m}(x_{2m}) = c_{2,m}(A(T_1(x_1)))\}}. \end{aligned}$$

F\"ur gegebenes  $X_1 = x_1$  mit  $T_1(x_1) = t_1$  werden  $c_{i,m}(A(T_1(x_1))) \in \bar{\mathbb{R}}$  und  $\gamma_{i,m}(A(T_1(x_1))) \in [0, 1]$ ,  $i = 1, 2$  bestimmt durch

$$\mathbb{E}_{\vartheta_0}\{\psi_{2,m}(x_1, X_{2m})\} = A(t_1) \quad \text{und} \quad (5.10)$$

$$\mathbb{E}_{\vartheta_0}\{T_{2m}(X_{2m}) \psi_m(x_1, X_{2m})\} = \mathbb{E}_{\vartheta_0}\{T_{2m}(X_{2m}) \psi_{m^*}(x_1, X_{2m^*})\}. \quad (5.11)$$

Hierbei ist

$$\begin{aligned} A(t_1) = & P_{\vartheta_0}(T_{2m^*}(X_{2m^*}) < c_{1,m^*}(\alpha) - t_1) + P_{\vartheta_0}(T_{2m^*}(X_{2m^*}) > c_{2,m^*}(\alpha) - t_1) \\ & + \gamma_{1,m^*}(\alpha)P_{\vartheta_0}(T_{2m^*}(X_{2m^*}) = c_{1,m^*}(\alpha) - t_1) \\ & + \gamma_{2,m^*}(\alpha)P_{\vartheta_0}(T_{2m^*}(X_{2m^*}) = c_{2,m^*}(\alpha) - t_1). \end{aligned}$$

*Beweis.* Im ersten Schritt ist für gegebenes  $x_1 \in \mathfrak{X}_1$  die Existenz der Tests  $\psi_m^{x_1}(X_{2m})$  zu zeigen, welche die Bedingung (5.10) und (5.11) erfüllen und gleichzeitig die bedingte Power maximieren. Dazu betrachten wir die Menge

$$\mathcal{A} = \{(\mathbb{E}_{\vartheta_0}\{\psi_m^{x_1}(X_{2m})\}, \mathbb{E}_{\vartheta_0}\{T_{2m}(X_{2m})\psi_m^{x_1}(X_{2m})\}) \mid \psi_m^{x_1} : \mathfrak{X}_{2m} \rightarrow [0, 1] \text{ ist messbare Abbildung}\}.$$

Nach dem verallgemeinerten Neyman-Pearson-Lemma gilt, dass die Menge  $\mathcal{A}$  konvex ist. Es bleibt zu zeigen: Der Punkt  $(A(t_1), \mathbb{E}_{\vartheta_0}\{T_{2m}(X_{2m})\psi_{m^*}(x_1, X_{2m^*})\})$  ist innerer Punkt von  $\mathcal{A}$ . Nach dem verallgemeinerten Neyman-Pearson-Lemma existieren einseitige Tests  $\psi_{m,\min}^{x_1}$  und  $\psi_{m,\max}^{x_1}$  zu dem bedingten Niveau  $A(t_1)$ , welche  $\mathbb{E}_{\vartheta_0}\{T_{2m}(X_{2m})\psi_m^{x_1}(X_{2m})\}$  minimieren bzw. maximieren. Angenommen, es ist  $\mathbb{E}_{\vartheta_0}\{T_{2m}(X_{2m})\psi_{m,\min}^{x_1}(X_{2m})\} = \mathbb{E}_{\vartheta_0}\{T_{2m}(X_{2m})\psi_{m^*}(x_1, X_{2m^*})\}$  oder  $\mathbb{E}_{\vartheta_0}\{T_{2m}(X_{2m})\psi_{m,\max}^{x_1}(X_{2m})\} = \mathbb{E}_{\vartheta_0}\{T_{2m}(X_{2m})\psi_{m^*}(x_1, X_{2m^*})\}$ . Dann setze  $c_{1,m}(x_1) = -\infty$  bzw.  $c_{2,m}(x_1) = \infty$  und das entsprechende  $\gamma_{i,m}(x_1) = 0$ .

Somit können wir nun annehmen, dass gilt

$$\begin{aligned} \mathbb{E}_{\vartheta_0}\{T_{2m}(X_{2m})\psi_{m,\min}^{x_1}(X_{2m})\} &< \mathbb{E}_{\vartheta_0}\{T_{2m}(X_{2m})\psi_{m^*}(x_1, X_{2m^*})\} \\ &< \mathbb{E}_{\vartheta_0}\{T_{2m}(X_{2m})\psi_{m,\max}^{x_1}(X_{2m})\}. \end{aligned}$$

Ferner enthält  $\mathcal{A}$  alle Punkte  $(a, a\mathbb{E}_{\vartheta_0}\{T_{2m}(X_{2m})\})$  mit  $a \in [0, 1]$ . Somit ist  $(A(t_1), \mathbb{E}_{\vartheta_0}\{T_{2m}(X_{2m})\psi_{m^*}(x_1, X_{2m^*})\})$  ein innerer Punkt von  $\mathcal{A}$  und die Existenz der  $\psi_m^{x_1}(X_{2m})$ ,  $m \in \mathcal{M}$  folgt nun analog zur Existenz zweiseitiger Tests (z.B. Lehmann und Romano, 2005, Kap.4). Insbesondere gilt für alle  $m \in \mathcal{M}$ :

$$\mathbb{E}_{\vartheta}\{\psi_m^{x_1}(X_{2m})\} \geq \mathbb{E}_{\vartheta}\{\psi_{m^*}(x_1, X_{2m^*})\}.$$

Gemäß der Konstruktion sind  $c_{i,m}(x_1)$  und  $\gamma_{i,m}(x_1)$  messbare Funktionen und daher folgt die Messbarkeit von  $\psi_m(x_1, x_{2m}) = \psi_m^{x_1}(x_{2m})$  analog zum Beweis von Satz 5.18. Somit ergibt sich durch eine analoge Rechnung, wie im Beweis

von (iv), Satz 5.18, für alle  $\tau \in \mathcal{T}$ :

$$\begin{aligned} \mathbb{E}_{\vartheta}\{\psi_{\tau}\} &= \mathbb{E}_{\vartheta}\left\{\sum_{m \in \mathcal{M}} \mathbb{E}_{\vartheta}\{1_{\{\tau=m\}} \mathbb{E}\{\psi_{1,m}(X_1, X_{2m})|X_1\}\}\right\} \\ &\geq \mathbb{E}_{\vartheta}\left\{\sum_{m \in \mathcal{M}} \mathbb{E}_{\vartheta}\{1_{\{\tau=m\}} \mathbb{E}\{\psi_{1,m^*}(X_1, X_{2m^*})|X_1\}\}\right\} = \mathbb{E}\{\psi_{1,m^*}(X_1, X_{2m^*})\} \end{aligned}$$

□

## 5.3 Flexible Designs in mehrparametrischen Exponentialfamilien

### 5.3.1 Einseitige Testprobleme

*Bemerkung 5.5.* Zur Vereinfachung der Notation wollen wir im Folgenden für bedingte Erwartungen bei gegebener Realisierung  $\mathbb{E}\{Y|X = x\}$  die Kurzschreibweise  $\mathbb{E}\{Y|x\}$  verwenden.

**Satz 5.24** (einseitiges Testproblem). *Gegeben sei das flexible Experiment  $((\mathfrak{X}, \mathcal{F}), \Theta, (P_{\vartheta})_{\vartheta \in \Theta}, (\mathcal{M}, \mathcal{M}), (\sigma(X_1, X_{2m}))_{m \in \mathcal{M}}, \sigma(X_1), \mathcal{T})$ , wobei  $X_1$  und  $(X_{2m})_{m \in \mathcal{M}}$  gemäß Voraussetzung 5.12-5.15 definiert sind, so dass  $(P_{\vartheta}^{X_1})_{\vartheta \in \Theta}$  und  $(P_{\vartheta}^{X_{2m}})_{\vartheta \in \Theta}$  eine  $k$ -parametrische Exponentialfamilie mit  $k > 1$  in  $(\vartheta, \eta)$  und  $(T_1, U_1)$  bzw.  $(T_{2m}, U_{2m})$  für alle  $m \in \mathcal{M}$  bilden. Sei  $\mathcal{T}$  die Menge aller Adaptionsregeln. Existieren für alle  $(\vartheta_0, \eta) \in \Theta$  messbare Abbildungen  $A_{\vartheta_0, \eta} : \mathfrak{X}_1^T \times \mathfrak{X}_1^U \rightarrow [0, 1]$  und für alle  $m \in \mathcal{M}$  messbare Abbildungen  $a_m : \mathfrak{X}_{2m}^U \times \mathfrak{X}_1^T \times \mathfrak{X}_1^U \rightarrow [0, 1]$ , so dass für alle  $(\vartheta_0, \eta) \in \Theta$  und  $m \in \mathcal{M}$  gilt*

$$\mathbb{E}_{\vartheta_0, \eta}\{a_m(U_{2m}(X_{2m}), T_1(X_1), U_1(X_1))|X_1\} = A_{\vartheta_0, \eta}(T_1(X_1), U_1(X_1)) \text{ f.s.} \quad (5.12)$$

und

$$\mathbb{E}_{\vartheta_0, \eta}\{A_{\vartheta_0, \eta}(T_1(X_1), U_1(X_1))\} = \alpha, \quad (5.13)$$

so gilt für das einseitige Testproblem  $H_0 : \vartheta \leq \vartheta_0$  gegen  $H_1 : \vartheta > \vartheta_0$ :

(i) Die Abbildungen  $a_m$  sind eindeutig bestimmt.

(ii) Seien  $m \in \mathcal{M}$ ,  $x_1 \in \mathfrak{X}_1$  mit  $t_1 = T_1(x_1)$  und  $u_1 = U_1(x_1)$  und  $u_{2m} \in \mathfrak{X}_{2m}^U$  gegeben, dann existiert ein gleichmäßig bester Test zum bedingten Niveau  $a_m(u_{2m}, t_1, u_1)$ . Dieser ist gegeben durch

$$\begin{aligned} \psi_{1,m}^{t_1, u_1, u_{2m}}(t_{2m}) &= \mathbb{1}_{\{t_{2m} > c_m(a_m(u_{2m}, t_1, u_1), u_{2m})\}} \\ &\quad + \gamma_m(a_m(u_{2m}, t_1, u_1), u_{2m}) \mathbb{1}_{\{t_{2m} = c_m(a_m(u_{2m}, t_1, u_1), u_{2m})\}} \end{aligned}$$

mit  $c_m(a_m(u_{2m}, t_1, u_1), u_{2m}) \in \bar{\mathbb{R}}$  und  $\gamma_m(a_m(u_{2m}, t_1, u_1), u_{2m}) \in [0, 1]$  werden bestimmt aus

$$\mathbb{E}_{\vartheta_0} \{ \psi_{1,m}^{t_1, u_1, u_{2m}}(T_{2m}(X_{2m})) | u_{2m} \} = a_m(u_{2m}, t_1, u_1).$$

(iii) Sei  $\psi_{1,m}(x_1, x_{2m}) := \tilde{\psi}_{1,m}(T_1(x_1), U_1(x_1), T_{2m}(x_{2m}), U_{2m}(x_{2m}))$  mit  $\tilde{\psi}_{1,m}(t_1, u_1, t_{2m}, u_{2m}) := \psi_{1,m}^{t_1, u_1, u_{2m}}(t_{2m})$ , dann sind die Abbildungen  $\psi_{1,m}(X_1, X_{2m}) : \mathfrak{X} \rightarrow [0, 1]$  messbar und es gilt

$$\mathbb{E}_{\vartheta_0, \eta} \{ \psi_{1,m}(X_1, X_{2m}) | X_1 \} = A_{\vartheta_0, \eta}(T_1(X_1), U_1(X_1)) \text{ fs.}$$

(iv) Existieren für alle  $(\vartheta, \eta) \in \Theta$ ,  $\vartheta \neq \vartheta_0$  messbare Abbildungen  $A_{\vartheta, \eta} : \mathfrak{X}_1^T \times \mathfrak{X}_1^U \rightarrow [0, 1]$  mit

$$\mathbb{E}_{\vartheta, \eta} \{ A_{\vartheta, \eta}(T_1(X_1), U_1(X_1)) \} \begin{cases} \leq \alpha & \text{falls } \vartheta < \vartheta_0 \\ \geq \alpha & \text{falls } \vartheta > \vartheta_0 \end{cases}, \quad (5.14)$$

so dass für alle  $m \in \mathcal{M}$  gilt

$$\begin{aligned} \mathbb{E}_{\vartheta, \eta} \{ \psi_{1,m}(X_1, X_{2m}) | X_1 \} &\leq A_{\vartheta, \eta}(T_1(X_1), U_1(X_1)) \text{ fs für alle } \vartheta < \vartheta_0 \\ \mathbb{E}_{\vartheta, \eta} \{ \psi_{1,m}(X_1, X_{2m}) | X_1 \} &\geq A_{\vartheta, \eta}(T_1(X_1), U_1(X_1)) \text{ fs für alle } \vartheta > \vartheta_0, \end{aligned} \quad (5.15)$$

dann gilt für  $\tau \in \mathcal{T}$  mit  $\psi_{1,\tau} = \sum_{m \in \mathcal{M}} \mathbb{1}_{\{\tau=m\}} \psi_{1,m}(X_1, X_{2m})$ :

$$\mathbb{E}_{\vartheta, \eta} \{ \psi_{1,\tau} \} \leq \alpha, \quad \text{für alle } (\vartheta, \eta) \in \Theta \text{ mit } \vartheta \leq \vartheta_0 \quad (5.16)$$

und

$$\mathbb{E}_{\vartheta,\eta}\{\psi_{1,\tau}\} \geq \alpha, \quad \text{für alle } (\vartheta, \eta) \in \Theta \text{ mit } \vartheta \geq \vartheta_0. \quad (5.17)$$

(v) Sei  $(\mathcal{T}, \phi_1)$  mit  $\phi_1 = (\phi_{1,m})_{m \in \mathcal{M}}$  ein weiterer flexibler Niveau- $\alpha$ -Test, der die Gleichungen (5.16) und (5.17) erfüllt, dann gilt

$$\mathbb{E}_{\vartheta}\{\psi_{1,\tau}\} \geq \mathbb{E}_{\vartheta}\{\phi_{1,\tau}\} \text{ für alle } \tau \in \mathcal{T}, (\vartheta, \eta) \in \Theta \text{ mit } \vartheta \geq \vartheta_0.$$

*Beweis.* (i) Da nach der Voraussetzung 5.15  $\Theta$  konvex ist und Punkte in  $\Theta$  mit  $\vartheta > \vartheta_0$  und  $\vartheta < \vartheta_0$  existieren, folgt, dass die Menge

$$\Gamma = \{(\vartheta, \eta) : (\vartheta, \eta) \in \Theta, \vartheta = \vartheta_0\}$$

konvex ist und ein  $k - 1$  dimensionales Intervall enthält. Daher folgt aus Lemma 5.8, dass die Verteilungsfamilie  $(\mathbb{P}_{\vartheta_0,\eta}^{U_{2m}})_{(\vartheta_0,\eta) \in \Gamma}$  von  $U_{2m}$ , vollständig ist.

Sei nun  $b_m(u_{2m}, t_1, u_1)$  eine weitere messbare Abbildung mit

$$\mathbb{E}_{\vartheta_0,\eta}\{b_m(U_{2m}(X_{2m}), T_1(X_1), U_1(X_1)) | X_1\} = A_{\vartheta_0,\eta}(T_1(X_1), U_1(X_1)) \text{ fs.}$$

Dann gilt aufgrund der Unabhängigkeit von  $X_1$  und  $X_{2m}$ :

$$\begin{aligned} 0 &= \mathbb{E}_{\vartheta_0,\eta}\{a_m(U_{2m}(X_{2m}), T_1(X_1), U_1(X_1)) | x_1\} \\ &\quad - \mathbb{E}_{\vartheta_0,\eta}\{b_m(U_{2m}(X_{2m}), T_1(X_1), U_1(X_1)) | x_1\} \\ &= \mathbb{E}_{\vartheta_0,\eta}\{a_m(U_{2m}(X_{2m}), t_1, u_1) - b_m(U_{2m}(X_{2m}), t_1, u_1)\} \end{aligned}$$

mit  $t_1 = T_1(x_1)$  und  $u_1 = u_1(x_1)$  Aufgrund der Vollständigkeit von  $(\mathbb{P}_{\vartheta_0,\eta}^{U_{2m}})_{(\vartheta_0,\eta) \in \Gamma}$  folgt hieraus

$$a_m(U_{2m}(X_{2m}), t_1, u_1) = b_m(U_{2m}(X_{2m}), t_1, u_1) \text{ fs}$$

für alle  $t_1 \in \mathfrak{X}_1^T$  und  $u_1 \in \mathfrak{X}_1^U$ .

(ii) Nach Lemma 5.6 bildet die bedingte Verteilung von  $T_{2m}$  bei gegebenem  $U_{2m}$  eine Exponentialfamilie in  $\vartheta$  und der identischen Abbildung. Somit folgt (ii) direkt aus Lemma 5.10.

- (iii) Zum Beweis der Messbarkeit ist zu zeigen:  $c_m(a, u)$  und  $\gamma_m(a, u)$  sind messbare Funktionen in  $(a, u)$ . Nach der Definition von  $c_m(a, u)$  und  $\gamma_m(a, u)$  reicht es zu zeigen, dass  $F_u^{-1}(y)$  und  $F_u^{-1}(y-0)$  messbare Funktionen von  $(y, u)$  sind. Hier soll nur die Messbarkeit von  $F_u^{-1}(y)$  gezeigt werden, da die Messbarkeit von  $F_u^{-1}(y-0)$  analog folgt. Für jedes  $c \in \mathbb{R}$  gilt:

$$\begin{aligned} \{(y, u) : F_u^{-1}(y) \leq c\} &= \{(y, u) : F_u(c) \geq y\} \\ &= \bigcup_{r \in \mathbb{Q}} \{(y, u) : F_u(c) \geq r, r \geq y\}. \end{aligned}$$

Da  $F_u(c)$  für festes  $c$  messbare Funktion von  $u$  ist, folgt die Messbarkeit von  $F_u^{-1}(y)$ . Hieraus ergibt sich die Messbarkeit von  $\psi_{1,m}(X_1, X_{2m})$  analog zum Beweis von Hilfssatz 3.61 in Witting (1985).

- (iv) Die Behauptung folgt mit Satz 4.12 aus (5.14) und den Eigenschaften 5.15 der  $A_{\vartheta, \eta}$ .
- (v) Nach Korollar 4.13 gilt:

$$\mathbb{E}_{\vartheta_0, \eta} \{\phi_{1,m}(X_1, X_{2m}) | X_1\} \leq A_{\vartheta_0, \eta}(T_1(X_1), U_1(X_1)) \text{ f\"ur alle } m \in \mathcal{M}.$$

Angenommen, es existiert ein  $m \in \mathcal{M}$  mit

$$P_{\vartheta_0, \eta}(\mathbb{E}_{\vartheta_0, \eta} \{\phi_{1,m}(X_1, X_{2m}) | X_1\} < A_{\vartheta_0, \eta}(T_1(X_1), U_1(X_1))) > 0.$$

Dann folgt

$$\mathbb{E}_{\vartheta_0, \eta} \{\phi_{1,m}(X_1, X_{2m})\} < \alpha.$$

Da die gemeinsame Verteilung von  $(X_1, X_{2m})$  nach Lemma 5.4 ebenfalls eine Exponentialfamilie in  $(\vartheta, \eta)$  und  $(T_1 + T_{2m}, U_1, U_{2m})$  bildet, folgt aus Lemma 5.9, dass

$$\beta_m(\vartheta, \eta) := \mathbb{E}_{\vartheta, \eta} \{\phi_{1,m}(X_1, X_{2m})\}$$

stetig ist. Daher existiert ein  $(\vartheta, \eta)$  mit  $\vartheta > \vartheta_0$  so, dass

$$\mathbb{E}_{\vartheta, \eta} \{\phi_{1,m}(X_1, X_{2m})\} < \alpha$$

Somit erhalten wir für  $\tau \equiv m$ :

$$\mathbb{E}_{\vartheta, \eta}\{\phi_\tau\} < \alpha.$$

Also gilt für alle  $m \in \mathcal{M}$ :

$$\mathbb{E}_{\vartheta_0, \eta}\{\phi_{1,m}(X_1, X_{2m})|X_1\} = A_{\vartheta_0, \eta}(T_1(X_1), U_1(X_1)) \text{ fs.}$$

Nach (i) gilt für alle  $m \in \mathcal{M}$ :

$$\mathbb{E}_{\vartheta_0}\{\phi_{1,m}(X_1, X_{2m})|u_{2m}, t_1, u_1\} = a_m(u_{2m}, t_1, u_1).$$

Da  $\psi_{i,m}(X_1, X_{2m})$  nach (ii) für alle  $x_1 \in \mathfrak{X}_1$  mit  $t_1 = T_1(x_1)$  und  $u_1 = U_1(x_1)$  ein gleichmäßig bester Test zum bedingten Niveau  $a_m(u_{2m}, t_1, u_1)$  ist, folgt für alle  $\vartheta > \vartheta_0$ :

$$\mathbb{E}_{\vartheta}\{\phi_{1,m}(X_1, X_{2m})|u_{2m}, t_1, u_1\} \leq \mathbb{E}_{\vartheta}\{\psi_{1,m}(X_1, X_{2m})|u_{2m}, t_1, u_1\}$$

und daher

$$\mathbb{E}_{\vartheta, \eta}\{\phi_{1,m}(X_1, X_{2m})|X_1\} \leq \mathbb{E}_{\vartheta, \eta}\{\psi_{1,m}(X_1, X_{2m})|X_1\}.$$

Hieraus folgt die Behauptung analog zum Beweis 5.18 (iv). □

**Korollar 5.25** (Fallzahlverlängerungen mit dem CRP-Prinzip). *Gegeben sei das flexible Experiment  $((\mathfrak{X}, \mathcal{F}), \Theta, (P_\vartheta)_{\vartheta \in \Theta}, (\mathcal{M}, \mathcal{M}), (\sigma(X_1, X_{2m}))_{m \in \mathcal{M}}, \sigma(X_1), \mathcal{T})$ , wie in Satz 5.24, und die Menge aller Adaptionsregeln  $\mathcal{T}$ . Für  $m^* \in \mathcal{M}$  mit  $m > m^*$  für alle weiteren  $m \in \mathcal{M}$  sei  $\psi_{m^*}$  ein gleichmäßig bester unverfälschter Niveau- $\alpha$ -Test für das einseitige Testproblem  $H_0 : \vartheta \leq \vartheta_0$  gegen  $H_1 : \vartheta > \vartheta_0$ . Bei gegebenem  $X_1 = x_1$  mit  $T_1(x_1) = t_1$  und  $U_1(x_1) = u_1$  sind die bedingten Ablehnungswahrscheinlichkeiten gegeben durch*

$$\begin{aligned} A_{\vartheta_0, \eta}(t_1, u_1) &:= \mathbb{E}_{\vartheta_0, \eta}\{\psi_{m^*}(X_1, X_{2m^*})|t_1, u_1\} \\ &= \mathbb{E}_{\vartheta_0, \eta}\{1_{\{T_{2m^*}(X_{2m^*}) > c(u_1 + U_{2m^*}(X_{2m^*}); \alpha) - t_1\}}\} \\ &\quad + \mathbb{E}_{\vartheta_0, \eta}\{\gamma(u_1 + U_{2m^*}(X_{2m^*}); \alpha) 1_{\{T_{2m^*}(X_{2m^*}) = c(u_1 + U_{2m^*}(X_{2m^*}); \alpha) - t_1\}}\}. \end{aligned} \tag{5.18}$$



Die Abbildungen  $a_m$  sind für alle  $m \in \mathcal{M}$  bestimmt durch

$$a_m(u_{2m}, t_1, u_1) := \mathbb{E}_{\vartheta_0} \{ \psi_{m^*}(X_1, X_{2m^*}) | u_{2m}, t_1, u_1 \}. \quad (5.19)$$

Insbesondere erfüllen für alle  $m \in \mathcal{M}$  und alle  $(\vartheta_0, \eta) \in \Theta$  die Abbildungen  $A_{\vartheta_0, \eta}$  und  $a_m$  die Bedingungen (5.12) und (5.13) aus Satz 5.24.

Ferner existiert ein flexibler Niveau- $\alpha$ -Test  $(\mathcal{T}, \psi)$  zu  $\psi_{m^*}$  mit

$$\mathbb{E}_{\vartheta} \{ \psi_{\tau} \} \geq \mathbb{E}_{\vartheta} \{ \psi_{m^*}(X_1, X_{2m^*}) \} \text{ für alle } \vartheta \geq \vartheta_0 \text{ und alle } \tau \in \mathcal{T}.$$

*Beweis.* Da  $\psi_{m^*}$  gleichmäßig bester unverfälschter Niveau- $\alpha$ -Test ist, hat nach Lemma 5.11  $\psi_{m^*}(X_1, X_{2m^*})$  die Form

$$\begin{aligned} \psi_{m^*}(X_1, X_{2m^*}) &= 1_{\{T_1(X_1) + T_{2m^*}(X_{2m^*}) > c_{m^*}(U_1(X_1) + U_{2m^*}(X_{2m^*}); \alpha)\}} \\ &\quad + \gamma_{m^*}(U_1(X_1) + U_{2m^*}(X_{2m^*}); \alpha) 1_{\{T_1(X_1) + T_{2m^*}(X_{2m^*}) = c_{m^*}(U_1(X_1) + U_{2m^*}(X_{2m^*}); \alpha)\}}. \end{aligned}$$

Aus der Unabhängigkeit von  $X_1$  und  $X_{2m^*}$  folgt die bedingte Ablehnungswahrscheinlichkeit (5.18).

Die in (5.19) definierten  $a_m$  hängen nach Korollar 5.17 nicht mehr von  $\eta$  ab. Die Gleichheit (5.12) folgt direkt aus den Eigenschaften bedingter Erwartungen und die Eigenschaft (5.13) ist offensichtlich, da  $\psi_{m^*}(X_1, X_{2m^*})$  ein Niveau- $\alpha$ -Test ist. Somit existieren für alle  $m \in \mathcal{M}$  die gleichmäßig besten Tests  $\psi_m(X_1, X_{2m})$  aus Satz 5.24 zu dem bedingten Niveau  $a_m(u_{2m}, t_1, u_1)$ . Mit einem ähnlichen Argument wie im Beweis zum Korollar 5.21 folgt, dass

$$\mathbb{E}_{\vartheta, \eta} \{ \psi_m(X_1, X_{2m}) | X_1 = x_1 \} \geq \mathbb{E}_{\vartheta, \eta} \{ \psi_{m^*}(X_1, X_{2m}) | X_1 = x_1 \} \text{ für alle } \vartheta > \vartheta_0$$

und

$$\mathbb{E}_{\vartheta, \eta} \{ \psi_m(X_1, X_{2m}) | x_1 \} \leq \mathbb{E}_{\vartheta, \eta} \{ \psi_{m^*}(X_1, X_{2m}) | x_1 \} \text{ für alle } \vartheta < \vartheta_0.$$

Somit erhalten wir für  $\psi_{\tau} = \sum_{m \in \mathcal{M}} 1_{\{\tau=m\}} \psi_m(X_1, X_{2m})$ :

$$\mathbb{E}_{\vartheta, \eta} \{ \psi_{\tau} \} \leq \alpha \text{ für alle } \vartheta < \vartheta_0 \text{ und}$$

$$\mathbb{E}_{\vartheta, \eta} \{ \psi_{\tau} \} \geq \mathbb{E}_{\vartheta, \eta} \{ \psi_{m^*}(X_1, X_{2m}) \}.$$

□

*Bemerkung 5.6.* Fallzahlverkürzungen mit dem CRP-Prinzip und vollständiger Anpassung der bedingten Ablehnungswahrscheinlichkeit sind in  $k$ -parametrischen Exponentialfamilien nicht immer möglich, wie die nachfolgenden Beispiele 5.4 zeigen werden.

Angenommen, für  $m \in \mathcal{M}$  mit  $m < m^*$  existiert ein Test  $\psi_m(X_1, X_{2m})$ , so dass gilt

$$A_{\vartheta_0, \eta}(t_1, u_1) = \mathbb{E}_{\vartheta_0, \eta} \{ \psi_m(X_1, X_{2m}) | t_1, u_1 \}.$$

Hierbei sei wie in Korollar 5.25  $A_{\vartheta_0, \eta}(t_1, u_1)$  wieder die bedingte Ablehnungswahrscheinlichkeit des ursprünglichen Tests  $\psi_{m^*}(X_1, X_{2m^*})$ . Das bedingte Niveau für diesen Test ist

$$a_m(u_{2m}, t_1, u_1) = \mathbb{E} \{ \psi_m(X_1, X_{2m}) | u_{2m}, t_1, u_1 \}.$$

Nach Satz 5.24 ist dies eindeutig bestimmt. Analog zu (5.19) kann für den ursprünglichen Test ein bedingtes Niveau definiert werden durch

$$\begin{aligned} \tilde{a}_{m^*}(u_{2m^*}, t_1, u_1) &:= \mathbb{E}_{\vartheta_0} \{ \psi_{m^*}(X_1, X_{2m^*}) | u_1, t_1, u_{2m^*} \} \\ &= \mathbb{E}_{\vartheta_0} \{ a_m(U_{2m}(X_{2m}), T_1(X_1), U_1(X_1)) | u_1, t_1, u_{2m^*} \}. \end{aligned}$$

Aufgrund der Eindeutigkeit (Satz 5.24) muss gelten

$$\tilde{a}_{m^*}(u_{2m^*}, t_1, u_1) = a_{m^*}(u_{2m^*}, t_1, u_1) = \mathbb{E}_{\vartheta_0} \{ \psi_{m^*}(X_1, X_{2m^*}) | u_1, t_1, u_{2m^*} \}.$$

Wegen der Unabhängigkeit von  $X_1$  und  $X_{2m}$  ergibt sich daraus

$$\mathbb{E}_{\vartheta_0} \{ a_m(U_{2m}(X_{2m}), t_1, u_1) | u_{2m^*} \} = a_{m^*}(u_{2m^*}, t_1, u_1). \quad (5.20)$$

Aus dieser Gleichung folgt, dass das bedingte Niveau  $a_m(u_{2m}, t_1, u_1)$  des verkürzten Tests eine Lösung des entsprechenden Gleichungssystems oder einer entsprechenden Integralgleichung sein muss.

### 5.3.2 Zweiseitige Testprobleme

Da im Falle  $k$ -parametrischer Exponentialfamilien die gleichen Probleme bestehen wie im einparametrischen Fall, werden wir hier auf die Formulierung

eines zu Satz 5.24 äquivalenten Satzes verzichten. In diesem Abschnitt soll nun nur noch der Spezialfall der Fallzahlverlängerung mit dem CRP-Prinzip betrachtet werden.

**Korollar 5.26** (Fallzahlverlängerungen mit dem CRP-Prinzip). *Gegeben sei das flexible Experiment  $((\mathcal{X}, \mathcal{F}), \Theta, (P_\vartheta)_{\vartheta \in \Theta}, (\mathcal{M}, \mathcal{M}), (\sigma(X_1, X_{2m}))_{m \in M}, \sigma(X_1), \mathcal{T})$  wie in Satz 5.24 und die Menge aller Adaptionsregeln  $\mathcal{T}$ . Für  $m^* \in \mathcal{M}$  mit  $m > m^*$  für alle weiteren  $m \in \mathcal{M}$  sei  $\psi_{m^*}$  ein gleichmäßig bester unverfälschter Niveau- $\alpha$ -Test für das zweiseitige Testproblem  $H_0 : \vartheta = \vartheta_0$  gegen  $H_1 : \vartheta \neq \vartheta_0$ . Dann existiert ein flexibler Niveau- $\alpha$ -Test  $(\mathcal{T}, \psi)$  zu  $\psi_{m^*}$  mit*

$$\mathbb{E}_\vartheta\{\psi_\tau\} \geq \mathbb{E}_\vartheta\{\psi_{m^*}(X_1, X_{2m^*})\} \text{ für alle } \vartheta \geq \vartheta_0 \text{ und alle } \tau \in \mathcal{T}.$$

Dabei sind die  $\psi_m(X_1, X_{2m})$  gegeben in der Form

$$\psi_m(X_1, X_{2m}) = \tilde{\psi}_m(T_1(X_1), U_1(X_1), T_{2m}(X_{2m}))$$

mit

$$\begin{aligned} \tilde{\psi}_m(t_1, u_1, t_{2m}, u_{2m}) = & 1_{\{t_{2m} < c_{1,m}(a_m(u_{2m}, t_1, u_1), u_{2m})\}} + 1_{\{t_{2m} > c_{2,m}(a_m(u_{2m}, t_1, u_1), u_{2m})\}} \\ & + \gamma_{1,m}(a_m(u_{2m}, t_1, u_1), u_{2m}) 1_{\{T_{2m} = c_{1,m}(a_m(u_{2m}, t_1, u_1), u_{2m})\}} \\ & + \gamma_{2,m}(a_m(u_{2m}, t_1, u_1), u_{2m}) 1_{\{T_{2m} = c_{2,m}(a_m(u_{2m}, t_1, u_1), u_{2m})\}}. \end{aligned}$$

Bei gegebenem  $X_1 = x_1$  mit  $T_1(x_1) = t_1$  und  $U_1(x_1) = u_1$  und  $U_{2m}(X_{2m}) = u_{2m}$  werden  $c_{i,m}(a_m(u_{2m}, t_1, u_1), u_{2m}) \in \bar{\mathbb{R}}$  und  $\gamma_{i,m}(a_m(u_{2m}, t_1, u_1), u_{2m}) \in [0, 1]$ ,  $i = 1, 2$  bestimmt durch

$$\mathbb{E}_{\vartheta_0}\{\tilde{\psi}_m(T_1(X_1), U_1(X_1), T_{2m}(X_{2m}) | u_{2m}, t_1, u_1)\} = a_m(u_{2m}, t_1, u_1) \quad (5.21)$$

und

$$\begin{aligned} & \mathbb{E}_{\vartheta_0}\{T_{2m}(X_{2m}) \tilde{\psi}_m(T_1(X_1), U_1(X_1), T_{2m}(X_{2m}) | u_{2m}, t_1, u_1)\} \\ & = \mathbb{E}_{\vartheta_0}\{T_{2m}(X_{2m}) \psi_{m^*}(x_1, X_{2m^*}) | u_{2m}, t_1, u_1\}. \end{aligned} \quad (5.22)$$

Hierbei ist

$$a_m(u_{2m}, t_1, u_1) := \mathbb{E}_{\vartheta_0}\{\psi_{m^*}(X_1, X_{2m^*}) | u_1, t_1, u_{2m}\}.$$

*Beweis.* Bei gegebenem  $X_1 = x_1$  mit  $T_1(x_1) = t_1$  und  $U_1(x_1) = u_1$  und  $U_{2m}(X_{2m}) = u_{2m}$  folgt die Existenz der Tests  $\psi_m(X_1, X_{2m})$ , welche die Bedingung (5.21) und (5.22) erfüllen und gleichzeitig die bedingte Power maximieren, analog zum Beweis von Korollar 5.25. Ebenso folgt die Messbarkeit. Insbesondere gilt für alle  $m \in \mathcal{M}$  und alle  $(\vartheta, \eta) \in \Theta$ :

$$\mathbb{E}_{\vartheta, \eta} \{ \psi_m(X_1, X_{2m}) | u_{2m}, t_1, u_1 \} \geq \mathbb{E}_{\vartheta, \eta} \{ \psi_{m^*}(X_1, X_{2m^*}) | u_{2m}, t_1, u_1 \}.$$

Hieraus folgt

$$\mathbb{E}_{\vartheta, \eta} \{ \psi_m(X_1, X_{2m}) | X_1 \} \geq \mathbb{E}_{\vartheta, \eta} \{ \psi_{m^*}(X_1, X_{2m^*}) | X_1 \}.$$

Somit ergibt sich durch eine analoge Rechnung wie im Beweis von Satz 5.18 (iv), dass für alle  $\tau \in \mathcal{T}$  gilt:

$$\begin{aligned} \mathbb{E}_{\vartheta, \eta} \{ \psi_\tau \} &= \mathbb{E}_{\vartheta, \eta} \left\{ \sum_{m \in \mathcal{M}} \mathbb{E}_{\vartheta} \{ 1_{\{\tau=m\}} \mathbb{E} \{ \psi_m(X_1, X_{2m}) | X_1 \} \} \right\} \\ &\geq \mathbb{E}_{\vartheta, \eta} \left\{ \sum_{m \in \mathcal{M}} \mathbb{E}_{\vartheta, \eta} \{ 1_{\{\tau=m\}} \mathbb{E}_{\vartheta, \eta} \{ \psi_{m^*}(X_1, X_{2m^*}) | X_1 \} \} \right\} = \mathbb{E}_{\vartheta, \eta} \{ \psi_{m^*}(X_1, X_{2m^*}) \} \end{aligned}$$

□

## 5.4 Anwendung der Methode an Beispielen

### 5.4.1 Fisher's exakter Test

Eine typische Anwendung in klinischen Studien ist der Vergleich zweier Erfolgsraten. Es seien  $Y_1, Y_2, \dots$  und  $Z_1, Z_2, \dots$  unabhängige Zufallsvariablen, die jeweils identisch Bernoulli verteilt sind mit den unbekanntem Erfolgswahrscheinlichkeiten  $p_T$  und  $p_K$ . Die gemeinsame Verteilung von  $(Y_1, \dots, Y_m)$  und

$(Z_1, \dots, Z_n)$  mit  $m, n \geq 1$  ist gegeben durch

$$\begin{aligned} P_{p_T, p_K}((Y_1, \dots, Y_m, Z_1, \dots, Z_n) = (y_1, \dots, y_m, z_1, \dots, z_n)) \\ = \left(1 + \frac{p_T}{1 - p_T}\right)^{-m} \left(1 + \frac{p_K}{1 - p_K}\right)^{-n} \\ \cdot \exp \left[ \sum_{i=1}^m y_i \log \left( \frac{p_T/(1 - p_T)}{p_K/(1 - p_K)} \right) + \left( \sum_{i=1}^m y_i + \sum_{j=1}^n z_j \right) \log (p_K/(1 - p_K)) \right] \end{aligned}$$

und somit eine 2-parametrische Exponentialfamilie in  $(\log(\frac{p_T/(1-p_T)}{p_K/(1-p_K)}), \log(p_K/(1-p_K)))$  und  $(\sum_{i=1}^m y_i, \sum_{i=1}^m y_i + \sum_{j=1}^n z_j)$ . Die zu einem Test auf Überlegenheit von T korrespondierenden Hypothesen sind gegeben durch  $H_0 : p_T \leq p_K$  gegen  $H_1 : p_T > p_K$ . Diese lassen sich umformen zu

$$H_0 : \log \left( \frac{p_T/(1 - p_T)}{p_K/(1 - p_K)} \right) \leq 0 \quad \text{vs.} \quad H_1 : \log \left( \frac{p_T/(1 - p_T)}{p_K/(1 - p_K)} \right) > 0.$$

Damit entspricht dies genau einem einseitigen Test in unserem Modell mit interessierendem Parameter  $\vartheta = \log((p_T/(1-p_T))/(p_K/(1-p_K)))$  und Störparameter  $\eta = \log(p_K/(1-p_K))$  und den Statistiken  $T(y_1, \dots, y_m, z_1, \dots, z_n) = \sum_{i=1}^m y_i$  (Anzahl der Erfolge in der Gruppe T), und  $U(y_1, \dots, y_m, z_1, \dots, z_n) = \sum_{i=1}^m y_i + \sum_{j=1}^n z_j$ . Diese wollen wir im Folgenden verkürzend mit  $T_m$  bzw.  $U_{m,n}$  bezeichnen. Der geeignete exakte Test ist Fisher's exakter Test.

Sei  $\psi_{m^*, n^*}(T_{m^*}, U_{m^*, n^*})$  ein exakter Fisher-Test zum Niveau  $\alpha$  mit den Gruppengrößen  $m^*$  und  $n^*$ . Dann ist dieser gegeben durch

$$\psi_{m^*, n^*}(T_{m^*}, U_{m^*, n^*}) = 1_{\{T_{m^*} > c(U_{m^*, n^*}; \alpha)\}} + \gamma(U_{m^*, n^*}; \alpha) 1_{\{T_{m^*} = c(U_{m^*, n^*}; \alpha)\}}.$$

Die bedingte Ablehnungswahrscheinlichkeit an einer Zwischenauswertung nach  $m_1$  und  $n_1$  Beobachtungen mit beobachteten Werten  $t_1$  und  $u_1$  der Statistiken lässt sich berechnen durch

$$\begin{aligned} A_{0, \eta}(t_1, u_1) = \sum_{u=0}^{m_2^* + n_2^*} P_{0, \eta}(U_{m_2^*} = u) \cdot [P_0(T_{m_2^*} > c(u_1 + u; \alpha) - t_1 | u) \\ + \gamma(u_1 + u; \alpha) P_0(T_{m_2^*} = c(u_1 + u; \alpha) - t_1 | u)] \end{aligned}$$

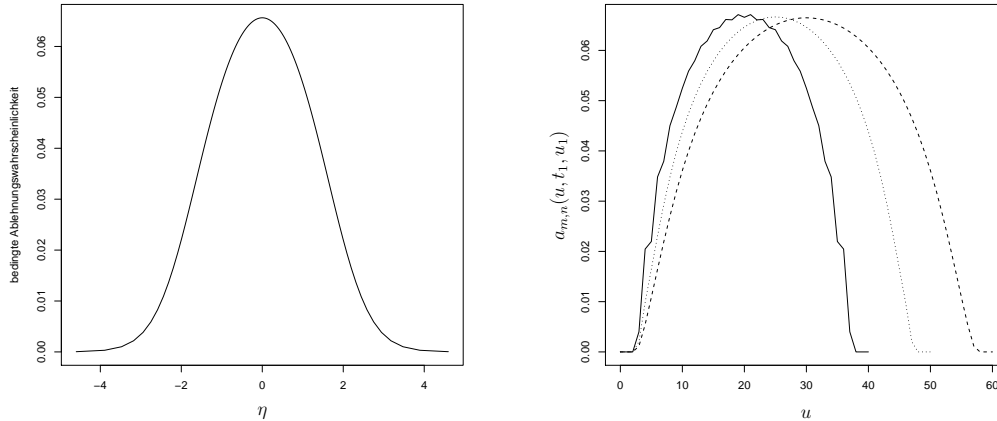


Abbildung 5.1: Bedingte Ablehnungswahrscheinlichkeit (linke Abb.) und bedingtes Niveau (rechte Abb.) eines Fisher-Tests, geplant mit  $m = n = 40$  Beobachtungen zum Niveau 2.5% nach  $m_1 = n_1 = 20$ -Beobachtungen mit  $t_1 = 12$  und  $u_1 = 20$ . Das bedingte Niveau ist sowohl für den ursprünglichen (durchgezogene Linie), als auch für verlängerte Tests (gestrichelt:  $m = n = 50$ , gepunktet:  $m = n = 45$ ) abgetragen.

mit  $m_2^* = m^* - m_1$  und  $n_2^* = n^* - n_1$ . Für  $\vartheta = 0$  ist die bedingte Verteilung von  $T$  bei gegebenem  $U$  eine Hypergeometrische-Verteilung der Form

$$P(T_m = t | U_{m,n} = u) = \frac{\binom{m}{t-y} \binom{n}{y}}{\binom{m+n}{t}}$$

und die Verteilung von  $U$  eine Binomialverteilung mit Erfolgswahrscheinlichkeit  $p_K$ . Die bedingte Ablehnungswahrscheinlichkeit als Funktion von  $\eta$  eines exakten Fisher-Test ist in der linken Grafik von Abbildung 5.1 zu sehen.

Sollen nun an dieser Zwischenauswertung die Gruppengrößen  $m^*$  und  $n^*$  vergrößert werden, so müssen wir für jeden möglichen Wert von  $U_{m_2, n_2}$  das bedingte Niveau bestimmen. Dazu sind die Funktionen  $a_{m,n}$  entsprechend Korollar 5.25 zu berechnen. Für den ursprünglichen Test, d.h.  $m = m^*$  und

$n = n^*$ , ist diese gegeben durch

$$a_{m^*,n^*}(u, t_1, u_1) = P_0(T_{2m_2^*} > c(u_1 + u; \alpha) - t_1 | u) \\ + \gamma(u_1 + u; \alpha) P_0(T_{m_2^*} = c(u_1 + u; \alpha) - t_1 | u).$$

Für alle weiteren  $m, n$  mit  $m > m^*$  und  $n > n^*$  können diese berechnet werden durch

$$a_{m,n}(u, t_1, u_1) = \sum_{\tilde{u}=0}^{m_2+n_2} a_{m^*,n^*}(\tilde{u}, t_1, u_1) \cdot P(U_{m_2^*,n_2^*} = \tilde{u} | U_{m_2,n_2} = u)$$

mit  $m_2 = m - m_1$  und  $n_2 = n - n_1$  und  $m_2^*, n_2^*$  entsprechend. Die bedingte Verteilung von  $U_{m_2^*,n_2^*}$  bei gegebenem  $U_{m_2,n_2}$  ist ebenfalls eine Hypergeometrische Verteilung:

$$P(U_{m_2^*,n_2^*} = \tilde{u} | U_{m_2,n_2} = u) = \frac{\binom{(m_2^*+n_2^*)-(m_2+n_2)}{u-\tilde{u}} \binom{m_2^*+n_2^*}{\tilde{u}}}{\binom{m_2+n_2}{u}}.$$

Die Funktionen  $a_{m,n}$  sowohl für den ursprünglichen als auch für verlängerte Tests sind zu sehen in der rechten Grafik von Abbildung 5.1.

Eine Verkürzung mit exakter Anpassung der bedingten Ablehnungswahrscheinlichkeit ist bei Fisher's exaktem Test nicht immer möglich, wie nachfolgendes Gegenbeispiel zeigt. Wir betrachten einen exakten Fisher-Test, welcher mit  $m = n = 25$  Beobachtungen pro Gruppe geplant wurde und für den eine Zwischenauswertung nach  $m_1 = n_1 = 23$  Beobachtungen erfolgte. Die beobachteten Werte der Statistiken seien  $t_1 = 12$  und  $u_1 = 20$ . Soll nun ein verkürzter Test mit  $m = n = 24$  Beobachtungen gefunden werden, so ist das Gleichungssystem gemäß (5.20) gegeben durch

$$\begin{pmatrix} 1.00000 & 0.00000 & 0.00000 \\ 0.50000 & 0.50000 & 0.00000 \\ 0.16667 & 0.66667 & 0.16667 \\ 0.00000 & 0.50000 & 0.50000 \\ 0.00000 & 0.00000 & 1.00000 \end{pmatrix} \begin{pmatrix} a_0 \\ a_1 \\ a_2 \end{pmatrix} = \begin{pmatrix} 0.00000 \\ 0.00000 \\ 0.00766 \\ 0.00000 \\ 0.00000 \end{pmatrix},$$

welches offensichtlich nicht lösbar ist.

### 5.4.2 t-Test

Es soll hier nur die Änderung eines einseitigen Einstichproben- $t$ -Tests betrachtet werden. Für die Verlängerung der weiteren  $t$ -Tests, wie z.B. des Zweistichproben- $t$ -Tests, aber auch des gruppensequentiellen- $t$ -Tests sei auf Timmesfeld et al. (2006) verwiesen. Im Beispiel 4.6 haben wir schon gesehen, dass die bedingte Ablehnungswahrscheinlichkeit von der unbekanntem Varianz abhing. Nun wollen wir zeigen, wie es uns jetzt möglich ist, die Fallzahl zu ändern und dabei die Ablehnungswahrscheinlichkeiten exakt anzupassen.

Es sei  $X_1, X_2, \dots$  unabhängig und identisch  $N(\mu, \sigma^2)$  verteilte Beobachtungen mit unbekanntem Erwartungswert  $\mu$  und unbekannter Varianz  $\sigma^2$ . Die gemeinsame Dichte von  $(X_1, \dots, X_n)$  ist gegeben durch

$$f_{\mu, \sigma}(X_1, \dots, x_n) = (2\pi\sigma^2)^{-n/2} \exp\left(-n \frac{\mu^2}{2\sigma^2}\right) \exp\left[\frac{\mu}{\sigma^2} \sum_{i=1}^n x_i - \frac{1}{2\sigma^2} \sum_{i=1}^n x_i^2\right].$$

Somit bildet die gemeinsame Verteilung eine 2-parametrische Exponentialfamilie in  $(\frac{\mu}{\sigma^2}, -\frac{1}{2\sigma^2})$  und  $(\sum_{i=1}^n x_i, \sum_{i=1}^n x_i^2)$ . Die zu testende Hypothese sei  $H_0 : \mu \leq 0$  gegen  $H_1 : \mu > 0$ , diese ist äquivalent zu  $H_0 : \frac{\mu}{\sigma^2} \leq 0$  gegen  $H_1 : \frac{\mu}{\sigma^2} > 0$ . Also lässt sich der einseitige  $t$ -Test wieder mit unserem Modell und interessierendem Parameter  $\vartheta = \frac{\mu}{\sigma^2}$ , Störparameter  $\eta = -\frac{1}{2\sigma^2}$  und den Statistiken  $V_n(x_1, \dots, x_n) = \sum_{i=1}^n x_i$  und  $U_n(x_1, \dots, x_n) = \sum_{i=1}^n x_i^2$  beschreiben. Für die Statistiken wollen wir auch hier wieder verkürzend die Bezeichnungen  $V_n$  und  $U_n$  verwenden.

Sei  $\psi_{n^*}(V_{n^*}, U_{n^*})$  ein einseitiger  $t$ -Test zum Niveau  $\alpha$  mit  $n^*$  Beobachtungen. In Form eines bedingten Tests ist dieser gegeben durch

$$\psi_{n^*}(V_{n^*}, U_{n^*}) = 1_{\{V_{n^*} \geq c\sqrt{(nU_{n^*})/(n^*-1+c^2)}\}}$$

mit  $c = t_{n^*-1; 1-\alpha}$ . Die bedingte Ablehnungswahrscheinlichkeit dieses Tests an einer Zwischenauswertung nach  $n_1$  Beobachtungen und beobachteten Werten



$v_1$  und  $u_1$  der Statistiken ist gegeben durch

$$\begin{aligned} A_{0,\eta}(v_1, u_1) &= \mathbb{E}_{0,\eta}\{\psi_{n^*}(V_n, U_n)|t_1, u_1\} \\ &= \int_0^\infty a_{n^*}(u, t_1, u_1) f_{0,\sigma^2}^{U_{n_2^*}}(u) du. \end{aligned}$$

Hierbei bezeichne  $a_{n^*}(u, t_1, u_1)$  das bedingte Niveau des Tests und  $f_{0,\sigma^2}^{U_{n_2^*}}(u)$  die Dichte von  $U_{n_2^*}$  mit  $n_2^* = n^* - n_1$ . Das bedingte Niveau des Tests ist gegeben durch

$$a_{n^*}(u, t_1, u_1) = P_0(V_{n_2^*} \geq c\sqrt{(n(u_1 + U_{n_2^*}))/n^* - 1 + c^2} - v_1 | U_{n_2^*} = u).$$

Da gilt

$$T_n = \frac{\sqrt{n-1}V_n}{\sqrt{nU_n - V_n^2}} = \frac{\sqrt{n-1}V_n/\sqrt{nU_n}}{\sqrt{1 - (V_n/\sqrt{nU_n})^2}},$$

$y/\sqrt{1-y^2}$  ist streng monoton in  $y$  für  $|y| \leq 1$ , und  $V_n/\sqrt{nU_n}$  ist unabhängig von  $U_n$  für  $\mu = 0$  (z.B. Lehmann und Romano (2005), Kapitel 5.2.), ist die bedingte Verteilung von  $V_n$  gegeben  $U_n$  für  $-\sqrt{nu_n} < x < \sqrt{nu_n}$  gegeben durch

$$\begin{aligned} P_0(V_n \leq x | u_n) &= P_0\left(\frac{\sqrt{n-1}V_n/\sqrt{nU_n}}{\sqrt{1 - (V_n/\sqrt{nU_n})^2}} \leq \frac{\sqrt{n-1}x}{\sqrt{nu_n - x^2}} | u_n\right) \\ &= F_{t_{n-1}}\left(\frac{\sqrt{n-1}x}{\sqrt{nu_n - x^2}}\right). \end{aligned}$$

Hierbei bezeichne  $F_{t_{df}}$  die Verteilungsfunktion einer  $t$ -Verteilung mit  $df$ -Freiheitsgraden. Somit erhalten wir für das bedingte Niveau des Tests  $\psi_{n^*}(V_{n^*}, U_{n^*})$

$$a_{n^*}(u; v_1, u_1) = \begin{cases} 0 & \text{falls } \bar{c}(u; v_1, u_1) \geq \sqrt{n_2^*u} \\ 1 & \text{falls } \bar{c}(u; v_1, u_1) \leq -\sqrt{n_2^*u}, \\ 1 - F_{t_{n_2^*-1}}\left(\frac{\sqrt{n_2^*-1}\bar{c}(u; v_1, u_1)}{\sqrt{n_2^*u - (\bar{c}(u; v_1, u_1))^2}}\right) & \text{sonst} \end{cases}$$

wobei  $\bar{c}(u; v_1, u_1) = c\sqrt{(n(u_1 + u))/n^* - 1 + c^2} - v_1$  ist.

Um das bedingte Niveau für einen verlängerten Test mit  $n > n^*$  zu bestimmen, müssen wir wieder die bedingte Verteilung von  $U_{n_2^*}$  bei gegebenem  $U_{n_2}$

berechnen. Es gilt

$$U_{n_2^*} \sim \sigma^2 \chi_{n_2^*}^2 \quad \text{und} \quad U_{n_2} | U_{n_2^*} \sim \sigma^2 \chi_{(n_2 - n_2^*)}^2 + U_{n_2^*}$$

Damit ergibt sich die bedingte Dichte

$$f_0^{U_{n_2^*} | U_{n_2}}(u | \tilde{u}) = \frac{\Gamma(\frac{n_2}{2})}{\Gamma(\frac{n_2^*}{2})\Gamma(\frac{n_2 - n_2^*}{2})} \frac{u^{\frac{n_2^*}{2} - 1} (\tilde{u} - u)^{\frac{n_2 - n_2^*}{2} - 1}}{\tilde{u}^{\frac{n_2}{2} - 1}}, \quad (5.23)$$

für  $u \leq \tilde{u}$  und 0 sonst. Somit können wir nun das bedingte Niveau für einen verlängerten Test gemäß Korollar 5.25 berechnen. Es gilt für alle  $n > n^*$

$$\begin{aligned} a_n(\tilde{u}; v_1, u_1) &= \mathbb{E}\{\psi_{n^*}(V_{n^*}, U_{n^*}) | \tilde{u}, t_1, u_1\} \\ &= \int_0^{\tilde{u}} a_{n^*}(u; v_1, u_1) f_0^{U_{n_2^*} | U_{n_2}}(u | \tilde{u}) du. \end{aligned}$$

In Abbildung 5.2 sind die bedingten Niveaus sowohl für den ursprünglichen, als auch für einen verlängerten  $t$ -Test zu sehen.

Im Gegensatz zu Fisher's exaktem Test ist beim  $t$ -Test auch eine Verkürzung möglich. Im Falle einer gradzahligen Verkürzung ist diese sogar explizit angebar. Sei  $n < n^*$  mit  $2m = n^* - n$ ,  $m \in \mathbb{N}$ , dann ist nach (5.20) die folgende Integralgleichung zu lösen:

$$a_{n^*}(\tilde{u}; v_1, u_1) = \int_0^{\tilde{u}} a_n(u; v_1, u_1) f_0^{U_{n_2} | U_{n_2^*}}(u | \tilde{u}) du.$$

Einsetzen der bedingten Dichte (5.23) ergibt

$$\tilde{u}^{\frac{n_2^*}{2} - 1} a_{n^*}(\tilde{u}; v_1, u_1) = \int_0^{\tilde{u}} a_n(u; v_1, u_1) \frac{\Gamma(\frac{n_2^*}{2})}{\Gamma(\frac{n_2}{2})\Gamma(m)} u^{\frac{n_2^*}{2} - 1} (\tilde{u} - u)^{m-1} du.$$

Mittels  $m$ -maligem Differenzieren erhalten wir die Lösung der Integralgleichung

$$a_n(u; v_1, u_1) = \frac{\Gamma(\frac{n_2}{2})}{\Gamma(\frac{n_2^*}{2})} \frac{1}{u^{\frac{n_2}{2} - 1}} \frac{d^m}{du^m} [u^{\frac{n_2^*}{2} - 1} a_{n^*}(u; v_1, u_1)].$$

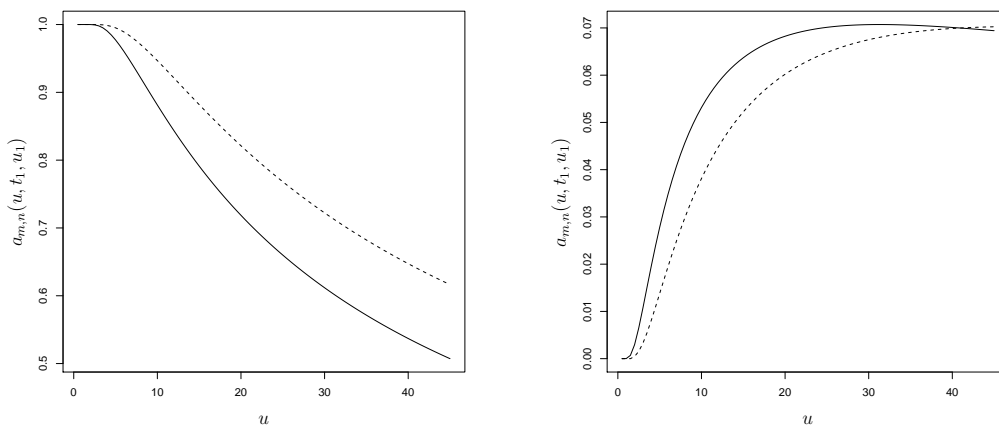


Abbildung 5.2: Bedingtes Niveau eines t-Tests, geplant mit  $n = 40$  Beobachtungen zum Niveau 2.5% nach  $n_1 = 20$ -Beobachtungen mit  $v_1 = 20$ ,  $u_1 = 59$  (linke Grafik) bzw.  $v_1 = 6$ ,  $u_1 = 22.6$  (dies entspricht  $\bar{x}_1 = 1$  bzw.  $\bar{x}_1 = 0.3$  und  $s_1^2 = 1$ , wie in Abb. 4.2). Das bedingte Niveau ist sowohl für den ursprünglichen (durchgezogene Linie), als auch für einen verlängerten Test (gestrichelt:  $n = 50$ ) dargestellt.



## 6

# Diskussion

In der vorliegenden Arbeit wurden statistische Verfahren weiterentwickelt, mit deren Hilfe während eines laufenden Experiments der Versuchsplan geändert und an die zwischenzeitlich gewonnenen Daten angepasst werden kann. Dies betrifft vor allem die Stichprobengröße, die je nach Zwischenergebnis verringert ("Fallzahlverkürzung") oder vergrößert ("Fallzahlverlängerung") werden kann. Wir verwenden als Grundlage das sogenannte CRP-Prinzip. Das CRP-Prinzip besagt, dass unter dem ursprünglichen und dem modifizierten Design die bedingte Wahrscheinlichkeit für die Ablehnung der Nullhypothese, gegeben die zum Zeitpunkt der Designänderung bereits vorliegenden Daten, unverändert bleiben soll. Dies ist eine hinreichende Bedingung für die Aufrechterhaltung des initial vorgegeben Fehlniveaus 1. Art.

Ziel der vorliegenden Arbeit war eine Erweiterung des CRP-Prinzip von Müller und Schäfer (2004) auf Tests, bei denen die bedingte Ablehnungswahrscheinlichkeit von Störparametern abhängt. Dabei tritt das Problem auf, dass die bedingte Ablehnungswahrscheinlichkeit (CRP) zusätzlich vom Störparameter abhängt, so dass eine Anpassung der CRP für jeden Wert

des unbekanntes Störparameters erreicht werden muss. Beispiele für solche Tests sind der  $t$ -Test und Fisher's exakter Test. Diese beiden Tests werden in klinischen Studien häufig verwendet. Im Kapitel 5.4 konnten wir zeigen, dass mit Hilfe der in dieser Arbeit entwickelten Theorie für beide Tests eine Fallzahlverlängerung möglich ist. Im Falle von Fisher's exakten Test ist eine Fallzahlverkürzung mit vollständiger Anpassung der bedingten Ablehnungswahrscheinlichkeiten nicht möglich. Bei dem Einstichproben- $t$ -Test kann die Fallzahl auch verkürzt werden und dabei die bedingte Ablehnungswahrscheinlichkeit für alle Werte der unbekanntes Varianz voll ausgeschöpft werden.

Die hier entwickelte Theorie geht allgemein von einer beliebigen Funktion  $A(X_1)$  aus, die die bedingte Ablehnungswahrscheinlichkeit als Funktion der bisherigen Daten  $X_1$ , darstellt. Es wurde gezeigt, dass es bei einseitigen Tests möglich ist, zu einer vorgegebenen Funktion  $A(X_1)$  gleichmäßig beste Tests zu konstruieren. Diese Funktion kann entweder zu Studienbeginn vorgegeben werden, wie es bei den flexiblen Designs der Fall ist, oder sie ergibt sich aus dem initialen Design (CRP-Prinzip). Aus Letzterem folgt, dass in jedem Studiendesign ein flexibles Design enthalten ist.

Gerade in Studien, bei denen für die Fallzahlberechnung ein unbekannter Störparameter geschätzt werden muss, stellt sich im Verlaufe häufig heraus, dass die Power zu gering ist. Deshalb ist die Erweiterung des CRP-Prinzips auf Studien mit Störparameter von großer praktischer Bedeutung. Insbesondere, da es bei einer Verlängerung der Studie nun möglich ist, das  $\alpha$ -Niveau exakt einzuhalten und die Power im Vergleich zu dem ursprünglichen Test immer zu erhöhen. Dies ist insbesondere im Falle kleiner Fallzahlen, die daher auch zu eher kleineren Fallzahlerhöhungen führen, von Bedeutung. Denn in dieser Situation führen die flexiblen Designs zu einem Powerverlust durch die Verwendung eines nicht optimalen Tests.

Zusätzlich wurde in dieser Arbeit ein Modell entwickelt (Kapitel 4), das sowohl die flexiblen Designs, als auch Designänderungen nach dem CRP-Prinzip umfasst. Mit Hilfe dieser Theorie konnte z.B. gezeigt werden, dass in Designs mit "voller" Flexibilität eine Einhaltung der bedingten Ablehnungswahrscheinlichkeit nicht nur hinreichend, sondern auch notwendig ist, um die Einhaltung des  $\alpha$ -Niveaus zu sichern. Bisher wurde dies immer als gültig angesehen, ohne dass ein Beweis erbracht wurde. Ein erster theoretischer Ansatz zur Beschreibung der flexiblen Verfahren geht auf Liu et al. (2002) zurück. Allerdings konnten mit deren Modell die Notwendigkeit der Einhaltung des bedingten Niveaus noch nicht gezeigt werden.

Das hier entwickelte Modell kann als Basis dienen, um weitere Probleme im Zusammenhang mit flexiblen Designs zu lösen. Dies sind unter anderem die Konstruktion unverfälschter Schätzer und Konfidenzintervalle zu einer festen Überdeckungswahrscheinlichkeit.





# Anhang A

## A.1 Bedingte Unabhängigkeit

Die nachfolgende Definition und das nachfolgende Lemma sind z.B. in Kaltenberg (1997, Kap. 5) zu finden und werden deshalb ohne Beweis angegeben.

**Definition A.1** (Bedingte Unabhängigkeit von  $\sigma$ -Algebren). Seien  $\mathcal{A}, \mathcal{B}, \mathcal{G} \subset \mathcal{F}$  drei Sub- $\sigma$ -Algebren von  $\mathcal{F}$ , dann heißen  $\mathcal{A}$  und  $\mathcal{B}$  bedingte unabhängig bei gegebenem  $\mathcal{G}$

$$\Leftrightarrow P(A \cap B | \mathcal{G}) = P(A | \mathcal{G})P(B | \mathcal{G}) \text{ f\"ur alle } A \in \mathcal{A}, B \in \mathcal{B}$$

**Lemma A.2.** Gegeben seien  $\mathcal{A}, \mathcal{B}, \mathcal{G} \subset \mathcal{F}$  drei Sub- $\sigma$ -Algebren von  $\mathcal{F}$ , dann gilt

$$\begin{aligned} & \mathcal{A} \text{ und } \mathcal{B} \text{ bedingte unabhängig bei gegebenem } \mathcal{G} \\ \Leftrightarrow & P(A | \mathcal{B}, \mathcal{G}) = P(A | \mathcal{G}) \text{ f\"ur alle } A \in \mathcal{A}. \end{aligned}$$



# Literaturverzeichnis

- Armitage P, McPherson C, Rowe B (1969). Repeated significance tests on accumulating data. *Journal of the Royal Statistical Society, Series A* **132**, 235–244.
- Bauer P (1989). Multistage testing with adaptive designs (with discussion). *Biometrie und Informatik in Medizin und Biologie* **20**, 130–148.
- Brannath W, Posch M, Bauer P (2002). Recursive combination tests. *Journal of the American Statistical Association* **97**, 236–244.
- Bretz F, Schmidli H, König F, Racine A, Maurer W (2006). Confirmatory seamless phase II/III clinical trials with hypotheses selection at interim: general concepts. *Biometrical Journal* **48**, 623–634.
- Denne J, Jennison C (1999). Estimating the sample size for a t-test using an internal pilot. *Statistics in Medicine* **18**, 1575–1585.
- Friede T, Kieser M (2004). Sample size recalculation for binary data in internal pilot study designs. *Pharmaceutical Statistics* **3**, 269–279.
- Friede T, Kieser M (2006). Sample size recalculation in internal pilot study designs: a review. *Biometrical Journal* **48**, 537–555.
- ICH E8 (1997). *General Considerations for Clinical Trials*. URL <http://www.ich.org/>.

- ICH E9 (1998). *Statistical Principles for Clinical Trials*. URL <http://www.ich.org/>.
- Jennison C, Turnbull B (2000). *Group Sequential Methods with Applications to Clinical Trials*. CRC Press.
- Jennison C, Turnbull BW (1991). Exact calculations for sequential t,  $\chi^2$  and F tests. *Biometrika* **78**, 133–141.
- Kallenberg O (1997). *Foundations of Modern Probability*. Springer, New York.
- Kieser M, Friede T (2000). Re-calculating the sample size in internal pilot study designs with control of the type I error rate. *Statistics in Medicine* **19**, 901–911.
- Kieser M, Friede T (2003). Simple procedures for blinded sample size adjustment that do not affect the type I error rate. *Statistics in Medicine* **22**, 3571–3581.
- König F, Bauer P, Brannath W (2006). An adaptive hierarchical test procedure for selecting safe and efficient treatments. *Biometrical Journal* **48**, 663–678.
- Lehmacher W, Wassmer G (1999). Adaptive sample size calculations in group sequential trials. *Biometrics* **55**, 1286–1290.
- Lehmann EL, Romano JP (2005). *Testing Statistical Hypotheses*. Springer Texts in Statistics. Springer, New York.
- Liu Q, Proschan M, Pledger G (2002). A unified theory of two-stage adaptive designs. *Journal of the American Statistical Association* **97**, 1034–1041.
- Müller HH, Schäfer H (2001). Adaptive group sequential designs for clinical trials: Combining the advantages of adaptive and of classical group sequential approaches. *Biometrics* **57**, 886–891.

- Müller HH, Schäfer H (2004). A general statistical principle for changing a design any time during the course of a trial. *Statistics in Medicine* **23**, 2497–2508.
- Pocock S (1983). *Clinical trials*. Wiley New York.
- Posch M, Timmesfeld N, König F, Müller HH (2004). Conditional rejection probabilities of student's t-test and design adaptations. *Biometrical Journal* **46**, 389–403.
- Proschan MA, Hunsberger SA (1995). Designed extension of studies based on conditional power. *Biometrics* **51**, 1315–1324.
- Rosenberger W (1996). New directions in adaptive designs. *Statistical Science* **11**, 137–149.
- Schäfer H, Müller HH (2001). Modification of the sample size and the schedule of interim analyses in survival trials based on data inspections. *Statistics in Medicine* **20**, 3741–3751.
- Schmitz N (1991). Sequential decision theory. *Handbook of sequential analysis*. New York: Marcel Dekker 407–28.
- Schumacher M, Schulgen G (2002). *Methodik klinischer Studien.: Methodische Grundlagen der Planung, Durchführung und Auswertung*. Springer.
- Stein C (1945). A two-sample test for a linear hypothesis whose power is independent of the variance. *Ann. Math. Stat* **16**, 243–258.
- Timmesfeld N, Schäfer H, Müller H (2006). Increasing the sample size during clinical trials with t-distributed test statistics without inflating the type I error rate. *Statistics in Medicine* **26**, 2449–2464.
- Todd S, Stallard N (2005). A new clinical trial design combining phases 2 and 3: Sequential designs with treatment selection and a change of endpoint. *Drug Information Journal* **39**, 109–118.

- Vandemeulebroecke M (2006). An investigation of two-stage tests. *Statistica Sinica* **16**, 933–951.
- Wittes J, Schabenberger O, Zucker D, Brittain E, Proschan M (1999). Internal pilot studies I: Type I error rate of the naive t-test. *Statistics in Medicine* **18**, 3481–3491.
- Witting H (1985). *Mathematische Statistik, Teil 1: Parametrische Verfahren bei festem Stichprobenumfang*. Teubner.
- Zhang W, Sargent DJ, Mandrekar S (2006). An adaptive dose-finding design incorporating both toxicity and efficacy. *Statistics in Medicine* **25**, 2365–2383.
- Zucker DM, Wittes JT, Schabenberger O, Brittain E (1999). Internal pilot studies II: Comparison of various procedures. *Statistics in Medicine* **18**, 3493–3509.

# Abbildungsverzeichnis

2.1	Power in Abhängigkeit vom Quotienten $\sigma/\sigma_0$ eines einseitigen Gausstests . . . . .	13
4.1	Bedingte Ablehnungswahrscheinlichkeit eines $z$ -Tests . . . . .	45
4.2	Bedingte Ablehnungswahrscheinlichkeit eines $t$ -Tests . . . . .	47
5.1	Bedingte Ablehnungswahrscheinlichkeit und bedingtes Niveau eines Fisher-Tests . . . . .	76
5.2	Bedingtes Niveau eines $t$ -Tests . . . . .	81





## Danksagung

Die vorliegende Arbeit wurde am Institut für medizinische Biometrie und Epidemiologie des Fachbereichs Medizin der Philipps-Universität Marburg angefertigt. Mein besonderer Dank gilt Prof. Dr. Helmut Schäfer und PD. Dr. Hans-Helge Müller für die interessante Themenstellung und die immer währende Unterstützung bei der Erstellung dieser Doktorarbeit. Insbesondere möchte ich mich für die anregenden, langen Diskussionen bedanken.

Allen derzeitigen und ehemaligen Kollegen danke ich für die gute Zusammenarbeit, die fachlichen und netten Gespräche auf dem Weg zur Mensa, welche sicherlich auch ihren Teil zu der sehr guten Arbeitsatmosphäre beitrugen. Mein besonderer Dank gilt auch Petra Fischer für das Korrekturlesen dieser Arbeit.

Danken möchte ich auch der Arbeitsgruppe um Prof. Dr. Bauer vom Institut für medizinische Statistik der Medizinischen Universität Wien, insbesondere Prof. Dr. Martin Posch und Dr. Franz König, für die gute Zusammenarbeit.

Ich möchte es auch nicht versäumen mich bei meiner Familie für die Unterstützung zu bedanken, vor allem in den letzten Wochen. Meinem Vater möchte ich danken für das Korrekturlesen dieser Arbeit. Mein ganz besonderer Dank gilt auch meinem Freund Marc Pankow für die vielen aufmunternden Worte und Allen die mich in dieser Zeit immer wieder unterstützt haben.



## Lebenslauf

9. Februar 1976	Geburt in Köln
Herbst 1980	Umzug von Köln nach Alten-Buseck
1980-1982	Besuch des Kindergartens in Alten-Buseck
1982-1986	Besuch der Grundschule in Alten-Buseck
1986-1995	Besuch der Landgraf-Ludwig-Gesamtschule in Gießen
Sommer 1995	Beendigung der Schulausbildung mit dem Erwerb der Hochschulreife an der Landgraf-Ludwig-Gesamtschule
Oktober 1995	Aufnahme des Diplomstudienganges Mathematik mit Nebenfach Informatik an der Justus-Liebig-Universität Gießen
Sommer 1997	Erwerb des Vordiploms in Mathematik
Oktober 1997	Fortführung des Mathematikstudiums im Hauptstudium
Juli / August 1998	Absolvierung eines sechs wöchigen Betriebspraktikums bei der Allianz-Leben in Stuttgart
WS 98/99 - SS 2001	Tätigkeit als studentische Hilfskraft mit der Korrektur von Hausaufgaben und der Leitung von Übungsgruppen zu den Vorlesungen: Wissenschaftliche Grundlagen des mathematischen Schulstoffes I und II, Lineare Algebra und Analytische Geometrie I und II, sowie Algebra
Frühjahr 2000	Beginn des ersten Prüfungsabschnittes zum Erwerb des Diploms in Mathematik

Herbst 2000	Beginn des zweiten Prüfungsabschnittes mit dem Verfassen der Diplomarbeit im Bereich der Stochastik
November 2001	Beendigung des Mathematikstudiums an der Justus-Liebig-Universität mit dem Erwerb des Diploms in Mathematik
15.2.2002-15.1.2007	Anstellung als wissenschaftliche Mitarbeiterin am Institut für Medizinische Biometrie und Epidemiologie im Rahmen des DFG-Projekts "Weiterentwicklung sequentieller und adaptiver statistischer Verfahren für klinisch-therapeutische und genetisch-epidemiologische Studien"
seit 15.1.2007	Anstellung als wissenschaftliche Mitarbeiterin (Landesstelle) am Institut für Medizinische Biometrie und Epidemiologie

## Erklärung

Hiermit erkläre ich, dass ich die vorliegende Dissertation selbstständig, ohne unerlaubte Hilfe verfasst und keine weiteren Hilfsmittel als die angegebenen verwendet habe.

Die Dissertation wurde in der jetzigen oder einer ähnlichen Form noch bei keiner anderen Hochschule eingereicht und hat noch keinen sonstigen Prüfungszwecken gedient.

Marburg, den 2. Januar 2008

---

Nina Timmesfeld