

# Maximal selektierte Rangstatistiken

Dissertation

zur Erlangung des akademischen Grades  
eines Doktors der Naturwissenschaften  
der Universität Dortmund

Dem Fachbereich Statistik  
der Universität Dortmund  
vorgelegt von

**Berthold Lausen**

aus Solingen

Dortmund 1990

Maximal substituierbare Katalysatoren

Thema

Die Katalysierung der Polymerisation von  
Ethen durch Metallocen-Katalysatoren  
des Typs  $Zr$ -Metallocene

Dem Fachbereich Chemie  
der Universität Freiburg  
vorgelegt von

Herrn Dr. ...

Gutachter: Prof. Dr. Martin Schumacher (Freiburg)  
Prof. Dr. Friedhelm Eicker

Tag der mündlichen Prüfung: 19. 2. 1990

## Inhaltsverzeichnis

	Vorwort	4
<b>1</b>	<b>Einleitung</b>	<b>6</b>
1.1	Problemstellung	6
1.2	Verwandte Probleme und Strategien	9
1.3	Überblick über die Ergebnisse der Arbeit	12
<b>2</b>	<b>Maximal selektierte Rangstatistiken</b>	<b>14</b>
2.1	Nullhypothese	14
2.2	Prüfpunktmodell	15
2.3	Lineare Rangstatistiken	17
2.4	Maximal selektierte Statistiken	18
2.5	Prüfpunktschätzung	21
2.6	Bindungen	21
2.7	Zensierung	22
<b>3</b>	<b>Asymptotische Verteilung unter der Nullhypothese</b>	<b>25</b>
3.1	Invarianzprinzip bei austauschbaren Zufallsvariablen	25
3.2	Invarianzprinzip maximal selektierter Statistiken	26
3.3	Approximation der asymptotischen Verteilung	28
<b>4</b>	<b>Zusammenhang zu Strukturbruchmodellen</b>	<b>30</b>
4.1	Abruptes Strukturbruchmodell	30
4.2	Strukturbruchstatistiken	32
<b>5</b>	<b>Exakte Verteilung und Approximationsmöglichkeiten</b>	<b>36</b>
5.1	Exakte Verteilung maximal selektierter Statistiken	36
5.2	Monte-Carlo Ergebnisse	37
5.3	Bonferroni-Ungleichung und eine Verbesserung	46
5.4	Vergleich verschiedener Approximationsmöglichkeiten	48
<b>6</b>	<b>Endliche Prüfpunktselektion</b>	<b>51</b>
6.1	Vorgegebene mögliche Prüfpunkte	51
6.2	Diskrete Einflußgröße	52
6.3	Diskrete Selektion	53

7	<b>Asymptotische relative Effizienz</b>	54
7.1	Exakte Bahadur-Effizienz	54
7.2	Approximative Bahadur-Effizienz	68
7.3	Diskussion anderer Optimalitätskonzepte	73
8	<b>Monte-Carlo Ergebnisse bei verschiedenen Alternativen</b>	76
8.1	Prüfpunktalternative	77
8.2	Lineare Abhängigkeit	83
8.3	Toxikologisches Schwellenwertmodell	89
9	<b>Diskussion des Prüfpunktschätzers</b>	96
9.1	Konsistenz des Prüfpunktschätzers	96
9.2	Monte-Carlo Ergebnisse	99
10	<b>Überlegungen zur Modellüberprüfung</b>	110
10.1	Abhängigkeit von der Einflußgröße	110
10.2	Zur Modellverifikation	111
11	<b>Anwendungen in der klinischen Forschung</b>	112
11.1	Studie zur antiarrhythmischen Therapie	112
11.2	Erlangener Brustkrebsstudie	115
12	<b>Diskussion der Ergebnisse der Arbeit und Ausblick</b>	123
	<b>Literaturverzeichnis</b>	127

## Vorwort

Die Problemstellung der vorliegenden Arbeit wurde während meiner Zeit als wissenschaftlicher Mitarbeiter am Institut für Medizinische Biometrie und Medizinische Informatik der Universität Freiburg von Herrn Prof. Dr. Martin Schumacher aufgeworfen. Eine vorläufige teilweise Lösung des Problems wird in dem Preprint Lausen und Schumacher (1988) gegeben.

Bei Herrn Prof. Dr. Martin Schumacher möchte ich mich für die wissenschaftliche Betreuung der Arbeit herzlich bedanken. Des weiteren möchte ich mich bedanken bei den Mitarbeitern des Instituts für Medizinische Biometrie und Medizinische Informatik, insbesondere für die Bereitschaft, Anwendungsmöglichkeiten in der klinischen Diagnostik zu diskutieren. Hierbei möchte ich Herrn Dipl.-Stat. Wilhelm Sauerbrei und Herrn Dipl.-Stat. Werner Vach hervorheben. Herrn Prof. Dr. Rudolf Lerche, Mathematische Fakultät der Universität Freiburg, und Herrn Prof. Dr. Friedhelm Eicker gilt mein Dank für wertvolle Hinweise. Herrn Priv. Doz. Dr. S. Hohnloser, Abteilung Innere Medizin III (Kardiologie) des Universitätsklinikums Freiburg, und Herrn Dr. W. Jäger, Universitäts-Frauenklinik Erlangen, danke ich für die Erlaubnis, ihre Daten exemplarisch analysieren zu dürfen. Bei Herrn Prof. Dr. Friedhelm Eicker, Fachbereich Statistik der Universität Dortmund, Herrn Dr. Theo Wember, Forschungsinstitut für Kinderernährung, Dortmund, und Herrn Dr. Hubertus Rochel, Forschungsinstitut für Kinderernährung, Dortmund, möchte ich mich für institutionelle Unterstützung bedanken.

Die Monte-Carlo Berechnungen der Arbeit wurden mittels eines portablen Fortran 77 Programms erstellt. Interessierten stelle ich den Quell-Code gerne zur Verfügung.

Dortmund, im Dezember 1989, Berthold Lausen

Hans Jonas:

"Niemals darf Existenz oder Wesen des Menschen im Ganzen zum Einsatz in den Wetten des Handelns gemacht werden. Daraus folgt ohne weiteres, daß hier bloße *Möglichkeiten* der bezeichneten Ordnung als unannehmbare Risiken anzusehen sind, die keine gegenüberstehenden Möglichkeiten annehmbarer machen."

Aus: Jonas, H. (1979), *Das Prinzip Verantwortung - Versuch einer Ethik für die technologische Zivilisation*, Insel Verlag: Frankfurt. Zitiert nach der Lizenz Ausgabe des Suhrkamp Taschenbuch Verlags, 1984, p.81.

## 1. Einleitung

Die Auswertung einer klinischen Studie am Institut für Medizinische Biometrie und Medizinische Informatik der Universität Freiburg führte zu der in der vorliegenden Arbeit behandelten Problemstellung (Abschnitt 1.1). Anhand von maximal selektierten Statistiken soll der prädiktive bzw. diagnostische Wert einer quantitativen Einflußgröße für eine beliebige Zielgröße analysiert werden. Es handelt sich hierbei um ein für verschiedene Anwendungsgebiete relevantes Problem, das sich insbesondere in der klinischen Forschung oft stellt. Im Abschnitt 1.2 werden Spezialfälle maximal selektierter Statistiken diskutiert. In diesem Zusammenhang gehen wir auf die Arbeiten von Gail und Green (1976) und Miller und Siegmund (1982) ein, die als Vorläufer unserer eigenen Untersuchungen anzusehen sind. Einen Überblick über die Ergebnisse der Arbeit enthält Abschnitt 1.3.

### 1.1. Problemstellung

Bei der Auswertung einer Studie zur nichtinvasiven Kontrolle der Wirksamkeit einer antiarrhythmischen Therapie vor Patienten mit malignen Tachyarrhythmien betrachteten Hohnloser u.a. (1987) (vgl. auch Hohnloser 1987) den prädiktiven Wert der linksventrikulären Ejektionsfraktion (EF). Als Zielgröße verwendeten sie die rezidivfreie Überlebenszeit nach der Klinikentlassung, wobei die Diagnose eines Arrhythmie rezidives oder eines plötzlichen Herztodes als terminierendes Ereignis für die Zeitdauer verwendet wurde. Falls die Beobachtung eines Patienten nach der Klinikentlassung aus anderen Gründen beendet wurde, wurde die beobachtete ereignisfreie Zeit als zensierte Beobachtung gewertet. Die Studie umfaßt 94 Patienten, und der Beobachtungszeitraum schwankt zwischen einem Monat und 39 Monaten. Hohnloser u.a. (1987) dichotomisierten die Ejektionsfraktion bei 35% EF. Hierdurch wurden die Beobachtungen in zwei Gruppen aufgeteilt. Der durchgeführte zweiseitige Log-Rank-Test legte nahe, daß die

beobachtete ereignisfreie Zeit von der EF abhängt ( $P=0.003$ ). Damit konnten Patienten mit einer EF kleiner oder gleich 35% als Patienten mit 'erhöhtem Risiko' eines Arrhythmie-rezidives oder eines plötzlichen Herztodes charakterisiert werden. Die von Hohnloser u.a. (1987) verwendete Dichotomisierung ist ein sinnvolles Vorgehen, da sie die klinische Interpretation 'erhöhtes Risiko' bzw. 'normales Risiko' ermöglicht, ohne eine Modellierung des unbekannt funktionalen Zusammenhangs zwischen der hypothetischen Einflußgröße, EF, und der betrachteten Zielgröße, ereignisfreie Zeit, vornehmen zu müssen. Die Wahl des Aufteilungspunktes - im weiteren als *Prüfpunkt* bezeichnet - wird in der Arbeit von Hohnloser u.a. (1987) nicht näher begründet. Deshalb ergibt sich als naheliegende Strategie, als Prüfpunkt den Wert der EF auszuwählen, der den Unterschied bzgl. der rezidivfreien Überlebenszeit zwischen der Gruppe mit 'erhöhtem Risiko' und der Gruppe mit 'normalem Risiko' maximiert. Des weiteren ist es manchmal sinnvoll anzunehmen, daß den Beobachtungen zwei relativ homogene Gruppen zugrunde liegen, welche sich bzgl. einer Charakteristik der Zielgröße, z.B. erwartete ereignisfreie Zeit, unterscheiden. Der Unterschied zwischen den Gruppen hängt ab von einer Einflußgröße, z.B. der EF. Bei dieser Modellierung kann man die den Beobachtungen zugrundeliegenden Gruppen anhand der Einflußgröße schätzen; d.h. die Einflußgröße ermöglicht eine Klassifikation der Beobachtungen in zwei Gruppen, die durch eine univariate Klassifikationslinie, den Prüfpunkt, gegeben sind.

Bei der Analyse des prädiktiven Wertes der EF betrachten wir die absolute standardisierte Log-Rank-Statistik als Prozeß in Abhängigkeit des hypothetischen Prüfpunktes in der EF (Abb.1.1.1). Die maximal selektierte absolute standardisierte Log-Rank-Statistik liefert den Prüfpunkt 39% EF. Der maximal selektierte Gruppenunterschied wird von den Kaplan-Meier-Schätzern der Verteilung der ereignisfreien Zeit in den Gruppen veranschaulicht (Abb.1.1.2). Offensichtlich würde die Verwendung der Verteilung der Log-Rank-Statistik bei festem Prüfpunkt zu einem zu kleinen  $P$ -Wert führen; d.h. der  $P$ -Wert würde unterschätzt.

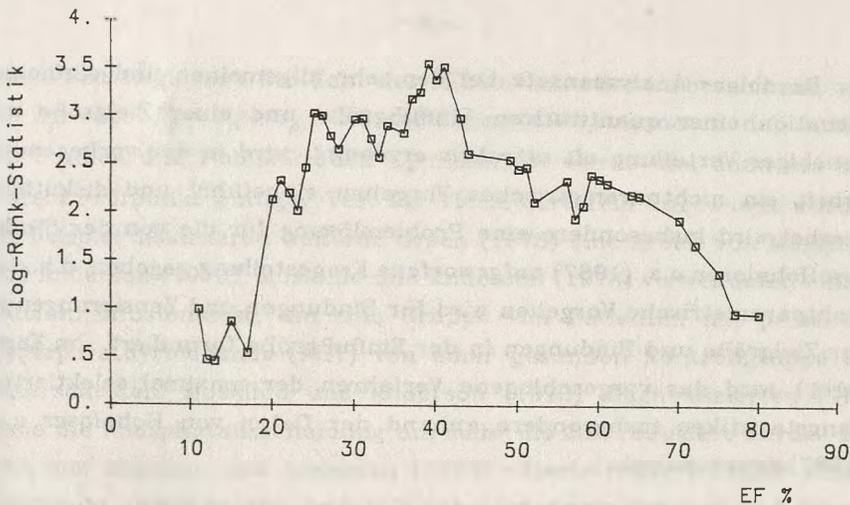


Abb. 1.1.1.: Die absolute standardisierte Log-Rank-Statistik als Prozeß in Abhängigkeit der möglichen Prüfpunkte in der ventrikulären Ejektionsfraktion (EF).

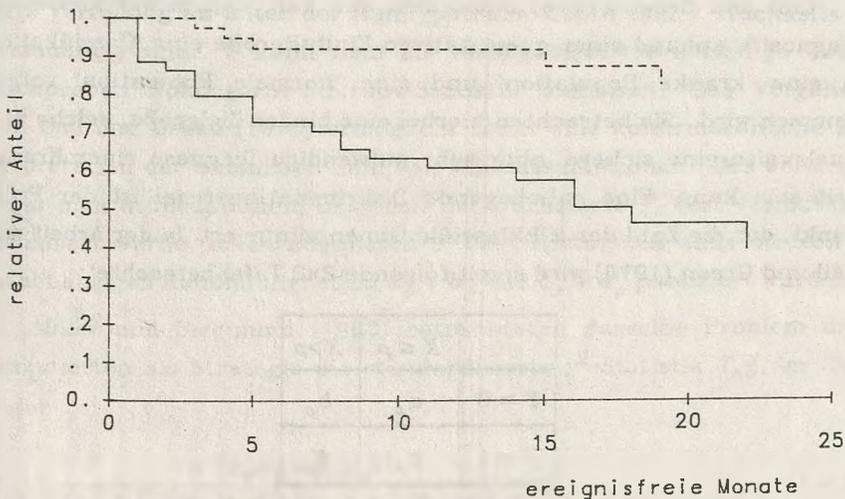


Abb. 1.1.2.: Der prädiktive Wert des maximal selektierten Prüfpunktes 39% EF wird veranschaulicht von den Kaplan-Meier Schätzern für die Patienten mit  $EF \leq 39\%$  (durchgezogene Linie) und  $EF > 39\%$  (gestrichelte Linie).

Da obiger Analyseansatz bei der sehr allgemeinen und einfachen Situation einer quantitativen Einflußgröße und einer Zielgröße mit beliebiger Verteilung oft attraktiv erscheint, wird in der vorliegenden Arbeit ein nichtparametrisches Vorgehen eingeführt und diskutiert. Hierbei wird insbesondere eine Problemlösung für die von der Studie von Hohnloser u.a. (1987) aufgeworfene Fragestellung gegeben; d.h. das nichtparametrische Vorgehen wird für Bindungen und Zensierungen in der Zielgröße und Bindungen in der Einflußgröße formuliert. Im Kapitel 11 wird das vorgeschlagene Verfahren der maximal selektierten Rangstatistiken insbesondere anhand der Daten von Hohnloser u.a. (1987) demonstriert.

## 1.2. Verwandte Probleme und Strategien

Gail und Green (1976) behandeln den Fall, daß in der klinischen Diagnostik anhand einer quantitativen Einflußgröße eine Klassifikation in eine 'kranke Population' und eine 'normale Population' vorgenommen wird. Sie betrachten hierbei eine binäre Zielgröße, welche beispielsweise eine sichere, aber sehr aufwendige Diagnose einer Krankheit sein kann. Eine naheliegende Diskriminationsregel ist der Prüfpunkt, der die Zahl der Mißklassifikationen minimiert. In der Arbeit von Gail und Green (1976) wird somit folgende  $2 \times 2$  Tafel betrachtet:

	$X \leq \rho$	$X > \rho$
$Y = 0$	$a_\rho$	$b_\rho$
$Y = 1$	$c_\rho$	$d_\rho$

hierbei bezeichnet  $X$  die quantitative Einflußgröße,  $Y$  die Zielgröße und  $a_\rho, b_\rho, c_\rho, d_\rho$  die von dem betrachteten Prüfpunkt abhängenden Zellhäufigkeiten. Hierbei entspräche  $Y = 1$  dem Vorliegen der Krankheit und  $X > \rho$  der Zuordnung zur kranken Population; d.h. der Diagnose:

'Krankheit liegt vor'. Die Zahl der Mißklassifikationen bezeichnen wir mit  $L_\rho$ ; wobei  $L_\rho = c_\rho + b_\rho$ . Gail und Green (1976) vermuten, daß in vielen Studien und Publikationen  $L_\rho$  minimiert wurde und dann ein bei festem Prüfpunkt gültiger Test für Vierfeldertafeln angewandt wurde. Als Beispiel diskutieren Gail und Green (1976) eine Arbeit von Moskilde und Anderson (1973). Moskilde und Anderson (1973) verwendeten einen Calcium Infusionstest, um eine Gruppe von Patienten mit primären Hyperparathyroidismus (PHT) von einer gesunden Kontrollgruppe zu unterscheiden. Moskilde und Anderson (1973) diagnostizierten PHT, wenn die Phosphatausscheidung um mehr als 20% reduziert wurde. Da eine von Moskilde und Anderson (1973) zitierte frühere Studie einen Prüfpunkt zwischen 30% und 40% nahelegt, vermuten Gail und Green (1976), daß der Prüfpunkt 20% gewählt wurde, um  $L_\rho$  zu minimieren. Gail und Green (1976) geben für die minimale gewichtete Zahl von Mißklassifikationen,  $T = \min_\rho (w c_\rho + (1 - w) b_\rho)$ ,  $0 < w < 1$ , die exakte Verteilung an unter der Nullhypothese, daß  $X$  und  $Y$  stochastisch unabhängig sind.  $T$  kann man als verallgemeinerte einseitige Zweistichproben Kolmogorov-Smirnov-Statistik auffassen. Das Vorgehen von Gail und Green (1976) ermöglicht somit eine konfirmatorische Interpretation der minimalen Zahl von Mißklassifikationen. Des weiteren kann man in Situationen, in denen der publizierte  $P$ -Wert vermutlich minimiert wurde, einen adäquaten  $P$ -Wert bestimmen, falls die von  $\rho$  unabhängigen Randhäufigkeiten  $a_\rho + b_\rho$  und  $c_\rho + d_\rho$  publiziert wurden.

Miller und Siegmund (1982) betrachteten dasselbe Problem und diskutierten als Strategie die standardisierte  $\chi^2$ -Statistik  $T_{n,\rho}^{\chi^2}$  der  $2 \times 2$  Tafel:

$$T_{n,\rho}^{\chi^2} = \frac{n (a_\rho d_\rho - b_\rho c_\rho)^2}{(a_\rho + b_\rho) (c_\rho + d_\rho) (a_\rho + c_\rho) (b_\rho + d_\rho)}, \quad (1.2.1)$$

wobei  $n = a_\rho + b_\rho + c_\rho + d_\rho$ . Miller und Siegmund (1982) zeigen, daß die Wurzel von  $T_{n,\rho}^{\chi^2}$  unter der Nullhypothese für  $a_\rho + b_\rho \rightarrow \infty$  und  $c_\rho + d_\rho \rightarrow \infty$  nach Verteilung gegen den Absolutbetrag einer standardisierten Brownschen Brücke konvergiert. Für die asymptotische

Verteilung der maximal selektierten  $\chi^2$ -Statistik geben Miller und Siegmund (1982) eine einfache Approximationsformel an (vgl. Abschnitt 3.3). Das Maximum wird hierbei gebildet bzgl. eines gegebenen Intervalls von möglichen Prüfpunkten  $\rho$  mit  $0 < \varepsilon_1 \leq F_X(\rho) \leq \varepsilon_2 < 1$ .  $F_X$  bezeichnet die Verteilungsfunktion von  $X$ . In der Tabelle 1.2.1 (Miller und Siegmund 1982, Tabelle 1) werden approximative asymptotische Niveaus der maximal selektierten  $\chi^2$ -Statistik bzgl. der bei festem Prüfpunkt gültigen  $\chi^2$  Verteilung angegeben. Dies veranschaulicht die Größenordnung beim Unterschätzen des  $P$ -Wertes bzw. bei der Niveauverletzung der Testvorschrift.

$\alpha$	$\varepsilon$				
	1/3	1/4	1/5	1/10	1/100
.10	.40	.49	.55		
.05	.24	.31	.35	.49	
.01	.07	.09	.11	.16	
.001	.010	.014	.017	.025	.051

Tabelle 1.2.1: Approximative asymptotische Wahrscheinlichkeiten unter der Nullhypothese, daß das maximal selektierte  $\chi^2$  das  $1 - \alpha$  Quantil der  $\chi^2$  Verteilung überschreitet, wobei das Maximum bzgl. dem Intervall  $\varepsilon \leq F_X(\rho) \leq 1 - \varepsilon$  gebildet wird.

In einer neueren Arbeit dichotomisiert Jespersen (1986) eine stetige Kovariable im Coxschen Regressionsmodell (Cox 1972). Jespersen (1986) zeigt, daß die maximal selektierte standardisierte Scorestatistik unter gewissen Regularitätsbedingungen dieselbe asymptotische Verteilung hat wie die maximal selektierte  $\chi^2$ -Statistik von Miller und Siegmund (1982).

Diese Arbeiten, die jeweils nur Spezialfälle von maximal selektierten Teststatistiken behandeln, lassen vermuten, daß ähnliche Ergebnisse auch für andere Teststatistiken zu erhalten sind. Dies ist umso

wichtiger, als die maximale Selektion von Teststatistiken in der explorativen Datenanalyse sehr häufig explizit oder implizit verwendet wird (siehe beispielsweise Breiman u.a. 1984 oder Abel, Berger und Wiebelt 1984). Die vorliegende Arbeit ermöglicht die Berechnung adäquater  $P$ -Werte bzw. eine konfirmatorische Interpretation der maximalen Selektion einer Teststatistik bzgl. einer möglichen Einflußgröße.

### 1.3. Überblick über die Ergebnisse der Arbeit

Im Kapitel 2 führen wir das Prüfpunktmodell und die Methode der maximal selektierten Rangstatistiken ein. Hierbei formulieren wir die Methode für eine quantitative mögliche Einflußgröße und eine beliebige Zielgröße. Insbesondere erlauben wir Bindungen in der Zielgröße und die Situation, daß die beobachtbare Zielgröße zensiert ist. Die asymptotische Verteilung unter der Nullhypothese leiten wir anhand eines sehr allgemeinen Invarianzprinzips her (Kapitel 3). Die asymptotischen Resultate von Miller und Siegmund (1982) und Jespersen (1986) ergeben sich als Spezialfälle. Der Zusammenhang mit Strukturbruchmodellen wird im Kapitel 4 diskutiert. Unter der Nullhypothese des Prüfpunktmodells ist die Verteilung der maximal selektierten Rangstatistik und die Verteilung der entsprechenden Rangstatistik in einem Strukturbruchmodell identisch. Die in der vorliegenden Arbeit eingeführte Strategie der maximal selektierten Rangstatistiken kann man als eine Verallgemeinerung nichtparametrischer Strategien bei einfachen Strukturbruchmodellen ansehen. Insbesondere die Verwendung des bei Bindungen und zensierten Zielgrößen gültigen Invarianzprinzips (Kapitel 3) ist im Kontext von Strukturbruchmodellen neu. Die exakte Verteilung unter der Nullhypothese und verschiedene Approximationsmöglichkeiten werden diskutiert und verglichen (Kapitel 5). Monte-Carlo Ergebnisse legen nahe, daß die Verwendung der asymptotischen Verteilung einer maximal selektierten Rangstatistik bei moderatem Stichprobenumfang in der Regel zu einem konservativen Test führt.

Approximative Ungleichungen vom Bonferroni-Typ ermöglichen in Abhängigkeit von der Größe des Intervalls von möglichen Prüfpunkten und des Stichprobenumfangs eine Verbesserung. Das Problem einer endlichen Anzahl der interessierenden Prüfpunkte behandeln wir im Kapitel 6. Die Konsistenz und die asymptotische relative Effizienz wird unter bestimmten Bedingungen im Kapitel 7 betrachtet. Wir verwenden und modifizieren insbesondere Ergebnisse über die exakte Bahadur-Effizienz nichtparametrischer Strukturbruchstatistiken von Praagman (1988). Da man für maximal selektierte Rangstatistiken bei größerem Stichprobenumfang in der Regel die asymptotische Verteilung verwenden wird, leiten wir auch die approximative Bahadur-Effizienz her. Des weiteren werden Effizienzvergleiche durchgeführt. Das Verhalten maximal selektierter Rangstatistiken bei endlichen Stichprobenumfang unter verschiedenen Alternativmodellen wird im Kapitel 8 anhand von Monte-Carlo Ergebnissen aufgezeigt und mit alternativen Methoden verglichen. Die Eigenschaften der Prüfpunktschätzung besprechen wir anhand theoretischer Überlegungen und von Monte-Carlo Ergebnissen (Kapitel 9). Wir diskutieren hierbei die Konsistenz des Prüfpunktschätzers. Im Kapitel 10 beenden wir die Einführung und Diskussion der Methode der maximal selektierten Rangstatistiken mit Überlegungen und Vorschlägen zur Modellüberprüfung. Im Kapitel 11 wird die eingeführte Methodik auf Beispiele aus der klinischen Forschung angewendet. Abschließend werden die Ergebnisse der vorliegenden Arbeit diskutiert, und wir gehen auf naheliegende Anwendungsmöglichkeiten und Weiterentwicklungen ein (Kapitel 12).

## 2. Maximal selektierte Rangstatistiken

Zur Analyse eines Prüfpunktmodells bei bivariaten Daten schlagen wir in dieser Arbeit die nichtparametrische Methode der maximal selektierten Rangstatistiken vor. Die von maximal selektierten Rangstatistiken implizierte Hypothese wird im Abschnitt 2.1 diskutiert, und eine naheliegende Formulierung einer Alternative, das Prüfpunktmodell, wird im Abschnitt 2.2 erläutert. Im Abschnitt 2.3 führen wir eine Terminologie für lineare Rangstatistiken ein. Danach definieren wir die Methode der maximal selektierten Statistiken (Abschnitt 2.4) und den hierdurch gegebenen Prüfpunktschätzer (Abschnitt 2.5). Das Vorgehen bei Bindungen in der Zielgröße erläutern wir im Abschnitt 2.6. Im Abschnitt 2.7 betrachten wir maximal selektierte Rangstatistiken bei zensierten Daten.

### 2.1. Nullhypothese

Wir betrachten die allgemeine Situation einer Zielgröße  $Y$  und einer quantitativen Einflußgröße  $X$ .  $(X_1, Y_1), \dots, (X_n, Y_n)$  sind unabhängig und identisch verteilte (u.i.v.) bivariate Zufallsvektoren. Wir nehmen an, daß die marginalen Verteilungsfunktionen von  $X$  und  $Y$  stetig sind. Auf diese technische Annahme wird später verzichtet (siehe Abschnitt 2.6, 2.7 und 6.2). Im folgenden wollen wir den prädiktiven Wert der quantitativen Einflußgröße  $X$  für die Zielgröße  $Y$  beurteilen. Deshalb nehmen wir an, daß die Einflußgröße es ermöglicht, die bivariate Stichprobe in zwei Gruppen von Beobachtungen aufzuteilen: Alle Beobachtungen mit Werten der Einflußgröße kleiner oder gleich einem bestimmten Prüfpunkt  $\mu$  definieren die eine Gruppe, die Beobachtungen mit Werten der Einflußgröße größer als der Prüfpunkt  $\mu$  definieren die andere Gruppe. Da der Prüfpunkt unbekannt sein soll, ergibt sich als interessierende Nullhypothese  $H_0$ :

$$H_0: P(Y \leq y | X \leq \rho) = P(Y \leq y | X > \rho) \quad (2.1.1)$$

für alle  $y$  und  $\rho \in \mathbb{R}$ . Die Nullhypothese  $H_0$  ist äquivalent zu der Hypothese, daß  $X$  und  $Y$  stochastisch unabhängig sind. Die Äquivalenz zeigen wir für den nicht entarteten Fall  $0 < P(X \leq \rho) < 1$ . Für alle  $y, \rho \in \mathbb{R}$  mit  $0 < P(X \leq \rho) < 1$  erhalten wir:

$$P(Y \leq y | X \leq \rho) = P(Y \leq y | X > \rho)$$

$$\Leftrightarrow P(X > \rho) P(Y \leq y, X \leq \rho) = P(X \leq \rho) P(Y \leq y, X > \rho)$$

$$\Leftrightarrow (1 - P(X \leq \rho)) P(Y \leq y, X \leq \rho)$$

$$= P(X \leq \rho) [P(Y \leq y) - P(Y \leq y, X \leq \rho)]$$

$$\Leftrightarrow P(Y \leq y, X \leq \rho) = P(X \leq \rho) P(Y \leq y).$$

Die Formulierung (2.1.1) der Nullhypothese suggeriert, daß  $Y$  die Zielgröße und  $X$  die Einflußgröße ist. Im nächsten Abschnitt wird dies durch die im Prüfpunktmodell vorliegende Abhängigkeitsstruktur konkretisiert.

## 2.2. Prüfpunktmodell

Allgemein gehen wir davon aus, daß der unbekannte Prüfpunkt  $\mu \in \mathbb{R}$  die bivariaten Beobachtungen in zwei Gruppen aufteilt. Die Gruppen sind jeweils u.i.v., aber die Verteilungen in den zwei Gruppen sind unter der Alternative des Prüfpunktmodells nicht identisch. Unter der Alternative fordern wir somit, daß für einen unbekanntes, aber festen Prüfpunkt,  $\mu \in \mathbb{R}$  mindestens ein Wert  $y \in \mathbb{R}$  existiert mit

$$P(Y \leq y | X \leq \mu) \neq P(Y \leq y | X > \mu). \quad (2.2.1)$$

Also fordern wir für den unbekanntes, aber festen Prüfpunkt  $\mu \in \mathbb{R}$ :

$$0 < P(X \leq \mu) < 1. \quad (2.2.2)$$

In der vorliegenden Arbeit betrachten wir im allgemeinen die speziellere Formulierung des Prüfpunktmodells (2.2.1) als *Lokationsmodell*.

Hierbei kann man folgende einseitige *Lokationsalternative* formulieren:

$H_a^>$ : Es existiert ein unbekannter Prüfpunkt  $\mu \in \mathbb{R}$ , so daß für alle  $y \in \mathbb{R}$  gilt:  $P(Y \leq y | X \leq \mu) \geq P(Y \leq y | X > \mu)$ , des weiteren existiert mindestens ein Wert  $y_0 \in \mathbb{R}$  mit  $P(Y \leq y_0 | X \leq \mu) > P(Y \leq y_0 | X > \mu)$ .

Die einseitige Alternative  $H_a^<$  wird analog definiert. Eine zweiseitige Alternative  $H_a$  definieren wir folgendermaßen:

$H_a: H_a^<$  oder  $H_a^>$  gilt.

Spezialfälle der Lokationsalternativen sind *Translationsalternativen*:

$H_a^{\vartheta > 0}$ : Es existiert ein unbekannter Prüfpunkt  $\mu \in \mathbb{R}$  und ein  $\vartheta > 0$ , so daß für alle  $y \in \mathbb{R}$  gilt:  $P(Y \leq y | X \leq \mu) = P(Y - \vartheta \leq y | X > \mu)$ .

$H_a^{\vartheta < 0}$  wird analog definiert und eine zweiseitige Alternative  $H_a^{\vartheta}$  analog zu  $H_a$ .

Das durch die Translationsalternativen gegebene Prüfpunktmodell kann man als einen Spezialfall des Zweigruppenmischungsmodells der Clusteranalyse auffassen:

$$f_Y(y) = P(X \leq \mu) f_{Y|X \leq \mu}(y) + P(X > \mu) f_{Y|X > \mu}(y - \vartheta),$$

wobei  $f_Z$  die Wahrscheinlichkeitsdichtefunktion einer Zufallsvariablen  $Z$  bezeichnet. Beispielsweise diskutiert Bock (1985) Mischungsmodelle und Tests, die in der Clusteranalyse gebräuchlich sind.

Die Spezifikation der Alternative im Prüfpunktmodell sollte von dem zugrundeliegenden Anwendungsproblem abhängen. Um beispielsweise die klinische Relevanz der Gruppen sicherzustellen, ist oft eine Einschränkung des unbekanntes Prüfpunkts sinnvoll:

$$0 < \varepsilon_1 \leq F_X(\mu) \leq \varepsilon_2 < 1,$$

wobei  $F_X$  die Verteilungsfunktion von  $X$  bezeichnet. Die in dem jeweiligen Anwendungsproblem festgelegten Parameter  $\varepsilon_1$  und  $\varepsilon_2$  determinieren die minimal bzw. maximal möglichen Gruppenanteile.

### 2.3. Lineare Rangstatistiken

Lineare Rangstatistiken erlauben eine nichtparametrische Analyse von mindestens ordinalskalierten Daten. Der Rang  $R_{ni}$  einer Beobachtung  $Y_i$  wird hierbei folgendermaßen definiert:

$$R_{ni} = \sum_{j=1}^n 1_{\{Y_j \leq Y_i\}}.$$

Da wir bislang annehmen, daß die Verteilungsfunktion von  $Y$  stetig ist, ist der Rangvektor  $(R_{n1}, \dots, R_{nn})$  mit Wahrscheinlichkeit eins (m.W.1) eine Permutation  $\pi_n$  der Zahlen  $(1, \dots, n)$ . Des weiteren falls  $Y_1, \dots, Y_n$  u.i.v. sind - dies gilt unter der Nullhypothese -, dann ist der Rangvektor auf der Menge aller Permutationen der Zahlen  $(1, \dots, n)$   $\Pi_n$  gleichverteilt; d.h. für  $\pi_n \in \Pi_n$  gilt:

$$P((R_{n1}, \dots, R_{nn}) = \pi_n) = 1/n!.$$

Eine *lineare Rangstatistik* definieren wir folgendermaßen (vgl. Hajek und Sidak 1967, p.61):

$$T_n = \sum_{i=1}^n c_n(i) a_n(R_{ni}), \quad (2.3.1)$$

wobei  $(a_n(1), \dots, a_n(n)), (c_n(1), \dots, c_n(n)) \in \mathbb{R}^n$  gegebene Vektoren sind. Die über die Ränge indizierten Konstanten  $a_{ni} = a_n(i)$  werden oft als *Scores* bezeichnet und die Gewichte  $c_{ni} = c_n(i)$  als *Regressoren*. Unter  $H_0$  gilt für den Erwartungswert und die Varianz der linearen Rangstatistik  $T_n$ :

$$E(T_n) = n \bar{a}_n \bar{c}_n, \quad (2.3.2)$$

$$\text{Var}(T_n) = 1/(n-1) \left( \sum_{i=1}^n (a_{ni} - \bar{a}_n)^2 \right) \left( \sum_{i=1}^n (c_{ni} - \bar{c}_n)^2 \right),$$

wobei  $\bar{a}_n = 1/n \sum a_{ni}$  und  $\bar{c}_n = 1/n \sum c_{ni}$ . Des weiteren seien  $a_n$  und  $c_n$  nicht konstant; d.h.  $\sum (a_{ni} - \bar{a}_n)^2 > 0$  und  $\sum (c_{ni} - \bar{c}_n)^2 > 0$ . Im

allgemeinen werden wir  $T_n$  in der standardisierten Form  $(T_n - E(T_n)) / (\text{Var}(T_n))^{1/2}$  betrachten. Unter gewissen Regularitätsbedingungen ist diese standardisierte Form asymptotisch normalverteilt (Hajek und Sidak 1967, p.163).

Da die Hypothese  $H_0$  äquivalent ist mit der Hypothese der stochastischen Unabhängigkeit von  $X$  und  $Y$ , ist es naheliegend, einen bekannten nichtparametrischen Test auf stochastische Unabhängigkeit zu verwenden. Deshalb vergleichen wir in der vorliegenden Arbeit Tests vom Spearman-Typ (2.3.3) mit der Methode der maximal selektierten Rangstatistiken. Ledwina (1986) und Praagman (1988, 1989) führten diesen Vergleich anhand der exakten Bahadur-Effizienz bei verwandten Problemstellungen durch (siehe auch Abschnitt 7.1). Eine *Rangstatistik vom Spearman-Typ* definieren wir folgendermaßen:

$$S_n = \sum_{i=1}^n a_n(R_{ni}) d_n(Q_{ni}), \quad (2.3.3)$$

wobei  $(d_{n1}, \dots, d_{nn}), (a_{n1}, \dots, a_{nn}) \in \mathbb{R}^n$  und  $(Q_{n1}, \dots, Q_{nn})$  den Rangvektor von  $X_1, \dots, X_n$  bezeichnet.

## 2.4. Maximal selektierte Statistiken

Bei der Beurteilung des prädiktiven Wertes der Einflußgröße  $X$  für die Zielgröße  $Y$  untersuchen wir im folgenden den Fall, daß die Beobachtungen gemäß einem unbekanntem Prüfpunkt  $\mu$  in zwei Gruppen aufgeteilt sind (vgl. Abschnitt 2.1 und 2.2). Ein naheliegendes Vorgehen ist somit die Berechnung des Gruppenunterschiedes in Abhängigkeit von dem variablen Prüfpunkt. Einen nichtparametrischen Gruppenvergleich ermöglicht eine spezielle lineare Rangstatistik, die *Zweistichproben-Rangstatistik*:

$$T_{n\rho} = \sum_{\{i: X_i \leq \rho\}} a_n(R_{ni}) = \sum_{i=1}^n c_\rho(X_i) a_n(R_{ni}), \quad (2.4.1)$$

wobei  $c_\rho(X_i) = 1_{\{X_i \leq \rho\}}$  und der Scorevektor nicht konstant ist (vgl. Abschnitt 2.3). In vielen Anwendungsfällen kann man den Prüfpunkt auf einen Bereich bzgl. der Einflußgröße einschränken. Deshalb gehen wir davon aus, daß zwei Konstanten  $\varepsilon_1$  und  $\varepsilon_2$  mit  $0 < \varepsilon_1 < \varepsilon_2 < 1$  vorgegeben sind. Der Bereich der möglichen Prüfpunkte  $\rho$  wird durch diese Konstanten bestimmt:

$$0 < \varepsilon_1 \leq F_X(\rho) \leq \varepsilon_2 < 1, \quad (2.4.2)$$

bzw.  $x_1 \leq \rho \leq x_2$ , wobei  $x_i = F_X^{-1}(\varepsilon_i)$ ;  $i=1,2$ . Mit  $F_X^{-1}$  bezeichnen wir die inverse Verteilungsfunktion der Zufallsvariablen  $X$ , die folgendermaßen definiert ist:  $F_X^{-1}(t) = \min\{x: F_X(x) \geq t\}$ . Da die Verteilung von  $X$  im allgemeinen unbekannt ist, verwenden wir für  $F_X$  die empirische Verteilungsfunktion  $F_{nX}$ ; wobei  $F_{nX}(x) = \sum 1_{\{X_i \leq x\}} / n$ . Damit betrachten wir als Bereich der möglichen Prüfpunkte:

$$0 < \varepsilon_1 \leq F_{nX}(\rho) \leq \varepsilon_2 < 1, \quad (2.4.3)$$

bzw.  $x_{n1} \leq \rho \leq x_{n2}$ , wobei  $x_{ni} = F_{nX}^{-1}(\varepsilon_i)$ ,  $i=1,2$ . Ein naheliegendes Vorgehen ist die Auswahl der Statistik  $T_{n\rho}$  mit  $\rho \in [x_{n1}, x_{n2}]$ , die den Gruppenunterschied maximiert. Das heißt, es wird die maximale Rangstatistik selektiert. Hierbei erscheint es sinnvoll, die Selektion bzgl. der absoluten standardisierten Rangstatistik durchzuführen. Als *maximal selektierte Rangstatistik*  $M_n(\varepsilon_1, \varepsilon_2)$  definieren wir:

$$M_n(\varepsilon_1, \varepsilon_2) = \max_{x_{n1} \leq \rho \leq x_{n2}} \frac{|T_{n\rho} - E(T_{n\rho})|}{(\text{Var}(T_{n\rho}))^{1/2}}. \quad (2.4.4)$$

Für spezielle Scores betrachtet Ledwina (1986)  $M_n(\varepsilon, 1 - \varepsilon)$  als einen nichtparametrischen Test auf positive quadratische Abhängigkeit. Falls die Scores monoton sind, d.h. falls gilt  $a_i \leq a_{i+1}$ ;  $i=1, \dots, n-1$  bzw.  $a_i \geq a_{i+1}$ ;  $i=1, \dots, n-1$ , dann liefert  $M_n(\varepsilon_1, \varepsilon_2)$  einen naheliegenden zweiseitigen Test auf einen Lageunterschied im Prüfpunktmodell,  $H_0$  vs.  $H_a$  (vgl. Abschnitt 2.2). In der vorliegenden Arbeit betrachten wir im allgemeinen das zweiseitige Testproblem. Eine naheliegende Statistik

für die einseitigen Testprobleme  $H_0$  vs.  $H_a >$  bzw.  $H_0$  vs.  $H_a <$  ist bei entsprechenden monotonen Scores die maximal selektierte standardisierte Rangstatistik; d.h.  $(T_{n\rho} - E(T_{n\rho})) / (\text{Var}(T_{n\rho}))^{1/2}$  wird maximiert.

Die für Rangstatistiken beschriebene Vorgehensweise kann man natürlich auch auf andere Statistiken anwenden. Beispielsweise haben Miller und Siegmund (1982) bei einer binären Zielgröße die maximal selektierte  $\chi^2$ -Statistik betrachtet (vgl. Kapitel 1), die man auch als extremen Fall einer Rangstatistik bei Bindungen auffassen kann (vgl. Abschnitt 2.6). Die *Zweistichproben-t-Statistik*  $T_{n\rho}^t$  (2.4.5) ist eine parametrische Alternative zur Zweistichproben-Rangstatistik:

$$T_{n\rho}^t = (n_{1\rho} n_{2\rho} / n)^{1/2} (\bar{Y}_{1\rho} - \bar{Y}_{2\rho}) / s_\rho, \quad (2.4.5)$$

wobei  $s_\rho^2 = 1 / (n - 2) (\sum_{\{i: X_i \leq \rho\}} (Y_i - \bar{Y}_{1\rho})^2 + \sum_{\{i: X_i > \rho\}} (Y_i - \bar{Y}_{2\rho})^2)$ ,

$n_{1\rho} = n F_{nX}(\rho)$ ,  $n_{2\rho} = n (1 - F_{nX}(\rho))$ ,  $\bar{Y}_{1\rho} = 1 / n_{1\rho} \sum_{\{i: X_i \leq \rho\}} Y_i$  und  $\bar{Y}_{2\rho} = 1 / n_{2\rho} \sum_{\{i: X_i > \rho\}} Y_i$ . Falls die Varianz  $\sigma^2$  von  $Y$  bekannt ist, erhalten wir die *Zweistichproben-Gauß-Statistik*, indem wir  $s_\rho^2$  durch die Varianz,  $\sigma^2$ , ersetzen.

Die Zweistichproben-Rangstatistik (2.4.1) kann man als Rangstatistik vom Spearman-Typ auffassen, wobei  $d_n$  allerdings von  $F_{nX}$  abhängt. Falls gilt  $d_{ni} = c_{ni}$  für  $i=1, \dots, n$ , dann ist die Verteilung der Rangstatistik vom Spearman-Typ  $S_n$  und der linearen Rangstatistik  $T_n$  unter  $H_0$  identisch. Denn

$$\begin{aligned} P(S_n \leq s) &= P\left(\sum_{i=1}^n d_{ni} a_n(R_n \circ Q_n^{-1}(i)) \leq s\right) \\ &= P\left(\sum_{i=1}^n c_{ni} a_n(R_{ni}) \leq s\right) = P(T_n \leq s), \end{aligned}$$

da für alle  $\pi_n \in \Pi_n$  gilt  $P(R_n \circ Q_n^{-1} = \pi_n) = P(R_n = \pi_n)$  (vgl. z.B. Hajek und Sidak, 1967, p.63).  $Q_n^{-1}(i)$  wird auch als *Antirang* bezeichnet.

## 2.5. Prüfpunktschätzung

Naheliegende nichtparametrische Schätzer für den unbekanntem Prüfpunkt  $\mu \in \mathbb{R}$  sind die Prüfpunkte, die die Rangstatistik maximieren. Somit liefert das Verfahren der maximal selektierten Rangstatistiken einerseits einen Test und andererseits eine Prüfpunktschätzung. Der Prüfpunkt, der die Rangstatistik maximiert, ist nicht eindeutig, da  $|T_{n\rho}|$  konstant ist für  $\mu \in [F_{n\bar{X}}^{-1}(k/n), F_{n\bar{X}}^{-1}((k+1)/n)]$ , wobei  $k \in \{1, \dots, n-1\}$ . Des weiteren kann das Maximum von verschiedenen  $k \in \{1, \dots, n-1\}$  angenommen werden. Die durch das Verfahren der maximal selektierten Rangstatistiken gegebene Menge von Prüfpunkten ist:

$$\{\rho: \rho \in [x_{n1}, x_{n2}], M_n(\varepsilon_1, \varepsilon_2) = |T_{n\rho} - E(T_{n\rho})| / (\text{Var}(T_{n\rho}))^{\frac{1}{2}}\}.$$

Wir definieren das Minimum der Menge dieser Prüfpunktschätzer als den *Prüfpunktschätzer*  $\hat{\mu}$ :

$$\hat{\mu} = \min \{\rho: \rho \in [x_{n1}, x_{n2}], M_n(\varepsilon_1, \varepsilon_2) = \frac{|T_{n\rho} - E(T_{n\rho})|}{(\text{Var}(T_{n\rho}))^{\frac{1}{2}}}\}. \quad (2.5.1)$$

Andere naheliegende Möglichkeiten, den Prüfpunktschätzer eindeutig zu definieren, sind Median und Maximum.

## 2.6. Bindungen

In diesem Abschnitt machen wir keine Annahme über die marginale Verteilung der Zielgröße  $Y$ . Damit verallgemeinern wir die Methode der maximal selektierten Rangstatistiken für nicht stetige und insbesondere diskrete Verteilungen. In beiden Fällen können mit positiver Wahrscheinlichkeit identische Realisationen der Zielgröße  $Y$  auftreten. Identische Realisationen verursachen das Problem der sogenannten *Bindungen*, welches sich folgendermaßen beschreiben läßt:  $(Y_{(1)}, \dots, Y_{(n)})$  bezeichnet den Vektor der Ordnungsstatistiken von  $(Y_1, \dots, Y_n)$ , und es gilt,

$$Y_{(1)} = \dots = Y_{(\tau_1)} < Y_{(\tau_1+1)} = \dots = Y_{(\tau_1+\tau_2)}$$

$$\dots < Y_{(K_{g-1}+1)} = \dots = Y_{(K_g)}.$$

hierbei bezeichnet  $\tau = (\tau_1, \dots, \tau_g)$  den Vektor der *Bindungslängen* und  $K = (K_1, \dots, K_g)$  den Vektor der kumulativen Bindungslänge;  $K_j = \sum_{i=1}^j \tau_i$ .  $K, g$  und  $\tau$  sind Zufallsvariablen. Wir lösen das Problem der Bindungen wie üblich, indem wir die Standardmethode der *Durchschnittsscores* verwenden. Den Vektor der Durchschnittsscores  $(a_n(1, \tau), \dots, a_n(n, \tau))$  kann man folgendermaßen definieren:

Für  $i$  mit  $K_{m-1} < i \leq K_m$

$$a_n(i, \tau) = \frac{1}{\tau_m} \sum_{k=K_{m-1}+1}^{K_m} a_n'(k), \quad (2.6.1)$$

wobei  $a_n' \in \mathbb{R}^n$ . Eine ausführliche Darstellung der Güteeigenschaften linearer Rangstatistiken unter Bindungen gibt beispielsweise Behnen (1974). Die Methode der maximal selektierten Rangstatistiken verallgemeinern wir, indem wir bei der Standardisierung der Teststatistik die unter dem Durchschnittscorevektor  $a_n$  bedingten Momente verwenden:

$$M_n(\varepsilon_1, \varepsilon_2) = \max_{x_{n1} \leq \rho \leq x_{n2}} \frac{|T_{n\rho} - E(T_{n\rho} | a_n)|}{(\text{Var}(T_{n\rho} | a_n))^{1/2}}. \quad (2.6.2)$$

## 2.7. Zensierung

Im folgenden verallgemeinern wir die Methode der maximal selektierten Rangstatistiken für zensierte Zielgrößen. Man spricht allgemein von zensierten Zielgrößen, wenn die interessierende Zielgröße nicht vollständig beobachtbar ist und ein Zensierungsmechanismus existiert, der die Beobachtung beenden kann. Wir betrachten also

$(X_1, Y_1, \delta_1), \dots, (X_n, Y_n, \delta_n)$ , wobei  $\delta_i$  der Zensierungsindikator und  $Y_i$  die beobachtbare Zielgröße ist, z.B. die beobachtbare Überlebenszeit eines Probanden.  $\delta_i = 1$  bedeutet, daß die beobachtete Zielgröße nicht zensiert ist und damit vollständig beobachtet wurde. Wir beschränken uns auf die sogenannte *zufällige Zensierung* (random censorship model, z.B. Miller, 1981, p.5). Als Nullhypothese  $H_0$  erhalten wir:

$H_0 : (Y, \delta)$  und  $X$  sind stochastisch unabhängig.

Als nichtparametrische Zweistichproben-Statistik betrachten wir die im Bereich der Analyse von Überlebenszeiten üblicherweise verwendeten Statistiken vom *Tarone-Ware-Typ* (2.7.1) (siehe z.B. Miller, 1981, pp.103-105).

$$T_{n\rho} = \sum_{i=1}^n \delta_{D_{ni}} w_i (c_\rho(X_{D_{ni}}) - \bar{c}_{\rho ni}), \quad (2.7.1)$$

wobei  $\bar{c}_{\rho ni} = \sum_{j=i}^n c_\rho(X_{D_{nj}}) / (n - i + 1)$ ,  $(w_1, \dots, w_n) \in \mathbb{R}^n$  ein gegebener Gewichtsvektor ist,  $D_n = (D_{n1}, \dots, D_{nn}) = R_n^{-1}$  und  $c_\rho(x) = 1_{\{x \leq \rho\}}$ . Die *Log-Rank-Statistik* (Savage-Scores) erhält man für  $w_i = 1$  und die *Gehan-Statistik* (Wilcoxon-Scores) für  $w_i = n - i + 1$ .

Man kann die Statistik vom Tarone-Ware-Typ als lineare Rangstatistik auffassen, deren Scores von dem bzgl.  $Y$  geordneten Vektor  $\delta_{()} = (\delta_{D_{n1}}, \dots, \delta_{D_{nn}})$  abhängen (z.B. Mehrotra, Michalek und Mihalka 1982). Denn es gilt:

$$\begin{aligned} T_{n\rho} &= \sum_{i=1}^n \delta_{D_{ni}} w_i (c_\rho(X_{D_{ni}}) - \bar{c}_{\rho ni}) & (2.7.2) \\ &= \sum_{i=1}^n \delta_{D_{ni}} w_i (c_\rho(X_{D_{ni}}) - \sum_{j=i}^n c_\rho(X_{D_{nj}}) / (n - i + 1)) \\ &= \sum_{i=1}^n c_\rho(X_{D_{ni}}) (\delta_{D_{ni}} w_i - \sum_{j=1}^i \delta_{D_{nj}} w_j / (n - j + 1)) \\ &= \sum_{i=1}^n c_\rho(X_i) a_n(R_{ni}, \delta_{()}), \end{aligned}$$

wobei

$$a_n(i, \delta_{(j)}) = \delta_{D_{ni}} w_i - \sum_{j=1}^i \delta_{D_{nj}} w_j / (n - j + 1).$$

Im Falle von *Bindungen* in der Zielgröße  $Y$  hängen die Scores auch von den Bindungslängen ab. Die identischen Beobachtungen erhalten bis auf den Zensierungsindikator identische Scores, wobei der Nenner  $(n - j + 1)$ , die Anzahl der Beobachtungen unter Risiko ohne Bindungen, durch die Anzahl der Beobachtungen unter Risiko bei Bindungen ersetzt wird. Des weiteren werden die Gewichte  $w_j$  bzw.  $w_i$  mit dem Rang von  $Y_{(j)}$  bzw.  $Y_{(i)}$  indiziert. Also verwenden wir wie üblich keine Durchschnittsscores bei zensierten Zielgrößen (z.B. Jones und Crowley 1989, Formel 7) und können somit die Interpretation als *verallgemeinerte Mantel-Haenzel-Statistik* aufrechterhalten. Die Teststatistik (2.7.1) standardisieren wir, indem wir, wie bei Bindungen, die unter dem Scorevektor  $a_n$  bedingten Momente verwenden, und wir erhalten somit analog zu (2.6.2) die maximal selektierte Tarone-Ware-Typ Statistik.

### 3. Asymptotische Verteilung unter der Nullhypothese

Im folgenden leiten wir die asymptotische Verteilung maximal selektierter Rangstatistiken unter der Nullhypothese her. Das verwendete Invarianzprinzip gilt auch bei Bindungen und bei Zensierung für die in den Abschnitten 2.6 und 2.7 eingeführten Verallgemeinerungen. Im Abschnitt 3.1 führen wir das benötigte Invarianzprinzip der austauschbaren Zufallsvariablen von Billingsley (1968, pp. 212-214) ein, und im Abschnitt 3.2 wenden wir das Invarianzprinzip auf die im Kapitel 2 eingeführte Methodik an. Eine einfache Approximationsformel von Jennen (1985) ermöglicht die Berechnung von approximativen asymptotischen Quantilen (Abschnitt 3.3).

#### 3.1. Invarianzprinzip bei austauschbaren Zufallsvariablen

Billingsley (1968, pp.212-214) leitet für austauschbare Zufallsvariablen ein Invarianzprinzip her. Hierbei definiert Billingsley *austauschbare Zufallsvariablen* folgendermaßen: Für jedes  $n \in \mathbb{N}$  sind  $\xi_{n1}, \dots, \xi_{nn}$  austauschbare Zufallsvariablen, falls die gemeinsame Verteilung permutationsinvariant ist; d.h. falls für alle  $\pi_n \in \Pi_n$  und  $x \in \mathbb{R}^n$  gilt,

$$P(\xi_{n1} \leq x_1, \dots, \xi_{nn} \leq x_n) = P(\xi_n(\pi_{n1}) \leq x_1, \dots, \xi_n(\pi_{nn}) \leq x_n).$$

Des weiteren fordert Billingsley (1968) folgende Regularitätsbedingungen:

$$\sum_{i=1}^n \xi_{ni} \xrightarrow{P} 0, \quad \sum_{i=1}^n (\xi_{ni})^2 \xrightarrow{P} 1, \quad \max_{1 \leq i \leq n} |\xi_{ni}| \xrightarrow{P} 0 \text{ für } n \rightarrow \infty, \quad (3.1.1)$$

wobei  $\xrightarrow{P}$  die Konvergenz nach Wahrscheinlichkeit bezeichnet.

*Satz 3.1.1. (Billingsley 1968, Theorem 24.2):*

Falls  $\xi_{n1}, \dots, \xi_{nn}$  austauschbare Zufallsvariablen sind und falls (3.1.1) gilt, dann gilt:

$\{B_n(t); 0 \leq t \leq 1\} \xrightarrow{d} \{B_0(t); 0 \leq t \leq 1\}$  für  $n \rightarrow \infty$ ,

wobei  $B_n(t) = \sum_{i=1}^{[nt]} \xi_{ni}$ ,  $\sum_1^0 = 0$ ,  $[x] = \max\{n \in \mathbb{N} : n \leq |x|\}$   $\text{sign}(x)$ ,  $B_0$

die Brownsche Brücke bezeichnet und  $\xrightarrow{d}$  die Konvergenz nach Verteilung bezeichnet.

*Korollar 3.1.2. (Chernoff und Teicher, 1958):*

Unter den Voraussetzungen von Satz 3.1.1 und  $0 < t_0 < 1$  gilt:  $B_n(t_0) \xrightarrow{d} N(0, t_0(1-t_0))$  für  $n \rightarrow \infty$ , wobei  $N(0, t_0(1-t_0))$  die Normalverteilung mit Erwartungswert 0 und Varianz  $t_0(1-t_0)$  bezeichnet.

### 3.2. Invarianzprinzip maximal selektierter Statistiken

Um das Invarianzprinzip aus Satz 3.1.1 auf die im Kapitel 2 eingeführten maximal selektierten Statistiken anwenden zu können, ist zu zeigen, daß die verwendeten Scores austauschbare Zufallsvariablen sind. Hierzu betrachten wir die Scores in Abhängigkeit der Antiränge von  $X$ , d.h.

$$(a_n(R_n \circ Q_n^{-1}(1)), \dots, a_n(R_n \circ Q_n^{-1}(n))) . \quad (3.2.1)$$

Hierbei hängt  $a_n$  bei Bindungen auch von  $\tau$  (Abschnitt 2.6), bei zensierten Daten von  $D_n$  und  $\delta$  ab (Abschnitt 2.7). Unter der Nullhypothese sind  $X$  und  $Y$  bzw.  $X$  und  $(Y, \delta)$  stochastisch unabhängig, und somit ist die gemeinsame Verteilung von (3.2.1) permutationsinvariant, und die Zufallsvariablen (3.2.1) sind somit austauschbar. Da wir die standardisierte Form der jeweiligen Teststatistik betrachten, fordern wir anstelle von Bedingung (3.1.1):

$$\max_{1 \leq i \leq n} \frac{(a_{ni} - \bar{a}_n)^2}{\sum_{i=1}^n (a_{ni} - \bar{a}_n)^2} \xrightarrow{P} 0 \text{ für } n \rightarrow \infty . \quad (3.2.2)$$

Diese Bedingung ist hinreichend für (3.1.1), falls

$$\xi_{ni} = (a_{ni} - \bar{a}_n) / (\sum (a_{ni} - \bar{a}_n)^2)^{1/2}.$$

Falls (3.2.2) erfüllt ist, gilt somit m.W.1:

$$\{B_n(t); 0 \leq t \leq 1\} \xrightarrow{d} \{B_0(t); 0 \leq t \leq 1\} \text{ für } n \rightarrow \infty,$$

$$\text{wobei } B_n(t) = \sum_{i=1}^{[nt]} (a_n(R_n \circ Q_n^{-1}(i)) - \bar{a}_n) / (\sum_{i=1}^n (a_{ni} - \bar{a}_n)^2)^{1/2}.$$

Die Aussage gilt m.W.1, da  $Q_n^{-1}$  m.W.1 eine Permutation ist, und mit  $t = F_{nX}(\rho)$  gilt dann:

$$B_n(t) = B_{n\rho}, \text{ wobei } B_{n\rho} = (T_{n\rho} - nF_{nX}(\rho)\bar{a}_n) / (\sum (a_{ni} - \bar{a}_n)^2)^{1/2}.$$

Die Konsistenz der empirischen Verteilungsfunktion  $F_{nX}$  sichert die Konsistenz der empirischen Quantile. Damit wird die Selektion asymptotisch bzgl. dem  $\varepsilon_1$ - und  $\varepsilon_2$ -Quantil der Verteilung der Einflußgröße  $X$  durchgeführt.

$D[0,1]$  sei der Raum der rechtsseitig stetigen, reellwertigen Funktionen mit linksseitigen Grenzwert auf dem Intervall  $[0,1]$  versehen mit der Supremums-Metrik. Des weiteren sei  $0 < \varepsilon_1 < \varepsilon_2 < 1$  und  $F_{\varepsilon_1\varepsilon_2} : D[0,1] \rightarrow \mathcal{R}$  mit

$$F_{\varepsilon_1\varepsilon_2}(f) = \sup_{\varepsilon_1 \leq t \leq \varepsilon_2} \frac{|f(t)|}{(t(1-t))^{1/2}}.$$

Für stetige Funktionale  $F : D[0,1] \rightarrow \mathcal{R}$  gilt allgemein (z.B. Billingsley, 1968, p.29), daß aus der Konvergenz des Prozesses die Konvergenz des Funktionals folgt;

d.h. aus  $\{B_n(t)\} \xrightarrow{d} \{B_0(t)\}$  folgt  $F(\{B_n(t)\}) \xrightarrow{d} F(\{B_0(t)\})$ .  $F_{\varepsilon_1\varepsilon_2}$  ist ein stetiges Funktional auf  $D[0,1]$ , und für die Varianz von  $T_{n\rho}$  gilt:

$$\text{Var}(T_{n\rho}) = n / (n-1) F_{nX}(\rho) (1 - F_{nX}(\rho)) \sum (a_{ni} - \bar{a}_n)^2. \text{ Damit läßt sich } M_n(\varepsilon_1, \varepsilon_2) \text{ als Funktional } F_{\varepsilon_1\varepsilon_2} \text{ von } B_n(t) \text{ darstellen. Also gilt für die}$$

asymptotische Verteilung maximal selektierter Statistiken folgender Satz:

Satz 3.2.1.:

$M_n(\varepsilon_1, \varepsilon_2)$  sei gemäß Kapitel 2 definiert und (3.2.2) sei erfüllt, dann gilt unter  $H_0$ :

$$M_n(\varepsilon_1, \varepsilon_2) \xrightarrow{d} \sup_{\varepsilon_1 \leq t \leq \varepsilon_2} \frac{|B_0(t)|}{(t(1-t))^{1/2}} \text{ für } n \rightarrow \infty \text{ m.W.1.}$$

Miller und Siegmund (1982) erhielten für die Wurzel der maximal selektierten  $\chi^2$ -Statistik dasselbe asymptotische Resultat. Die Ergebnisse von Miller und Siegmund (1982) für binäre Zielgrößen und von Jespersen (1986) für zensierte Zielgrößen erhält man als Spezialfälle von Satz 3.2.1. Satz 3.2.1 läßt sich auch auf eine maximal selektierte  $t$ - bzw. Gauß-Statistik anwenden, falls das zweite Moment der Zielgröße  $Y$  endlich ist, wobei  $a_{ni} = Y_i$ . Dieser Spezialfall läßt sich allerdings auch durch eine direkte Anwendung von Donskers Theorem (z.B. Billingsley 1968, p.68) zeigen.

### 3.3. Approximation der asymptotischen Verteilung

Miller und Siegmund (1982) geben für die asymptotische Verteilung folgende Approximation (Jennen 1985, Sek. 5E) an:

$$P \left[ \sup_{t \in [\varepsilon_1, \varepsilon_2]} \frac{|B_0(t)|}{(t(1-t))^{1/2}} \geq b \right] = \frac{4}{b} \varphi(b) + \varphi(b) \left( b - \frac{1}{b} \right) \log \left( \frac{\varepsilon_2(1-\varepsilon_1)}{(1-\varepsilon_2)\varepsilon_1} \right) + o\left(\frac{\varphi(b)}{b}\right) \quad (3.3.1)$$

für  $0 < \varepsilon_1 < \varepsilon_2 < 1$  und  $b \rightarrow \infty$ , wobei  $\varphi(b)$  die Lebesque-Dichte der Standardnormalverteilung bezeichnet; also  $\varphi(b) = 1/(2\pi)^{1/2} \exp(-\frac{1}{2}b^2)$ . Die Formel (3.3.1) ermöglicht eine Approximation der asymptotischen Quantile, und somit können wir eine approximative Testvorschrift zum Niveau  $\alpha$  für die maximal selektierte Statistik  $M_n(\varepsilon_1, \varepsilon_2)$  angeben. Tabelle 3.3.1 erlaubt den Vergleich einiger anhand von Formel (3.3.1) berechneter Quantile mit den Quantilen des Absolutbetrages einer

Standardnormalverteilung (siehe auch Tabelle 1.2.1).

$\alpha$	Quantile	asymptotische Quantile für			
	$ N(0,1) $	(0.25,0.75)	(0.4,0.6)	(0.1,0.9)	(0.4,0.9)
0.100	1.645	2.539	2.263	2.784	2.597
0.090	1.695	2.585	2.310	2.828	2.642
0.080	1.751	2.636	2.362	2.875	2.693
0.070	1.812	2.693	2.420	2.927	2.748
0.060	1.881	2.756	2.485	2.986	2.811
0.050	1.960	2.829	2.560	3.054	2.883
0.040	2.054	2.915	2.649	3.135	2.968
0.030	2.170	3.022	2.760	3.235	3.073
0.025	2.241	3.088	2.828	3.297	3.138
0.020	2.326	3.166	2.910	3.371	3.215
0.010	2.576	3.395	3.149	3.588	3.442

Tabelle 3.3.1.: Quantile der asymptotischen Verteilung in Abhängigkeit von  $(\varepsilon_1, \varepsilon_2)$  und die Quantile des Absolutbetrages der Standardnormalverteilung,  $|N(0,1)|$ .

#### 4. Zusammenhang zu Strukturbruchmodellen

Im Abschnitt 4.1 diskutieren wir den Zusammenhang zwischen dem Prüfpunktmodell und einem abrupten Strukturbruchmodell. Im Gegensatz zu den sequentiellen Strukturbruchmodellen kann man das Prüfpunktmodell als eine Verallgemeinerung eines nichtsequentiellen einfachen Strukturbruchmodells ansehen. Man geht hierbei davon aus, daß eine geordnete Folge von Beobachtungen einer Zielgröße gegeben ist. Normalerweise wird die Ordnung bei Strukturbruchmodellen direkt aus dem Beobachtungsmechanismus abgeleitet; beispielsweise die Ordnung der Beobachtungszeitpunkte. Beim Prüfpunktmodell kann man die Beobachtungen bzgl. der hypothetischen Einflußgröße ordnen und anhand maximal selektierter Statistiken den Einfluß dieser Ordnung überprüfen. Man erhält somit ein einfaches Strukturbruchproblem. In den letzten Jahren sind viele Arbeiten erschienen, die Strukturbruchmodelle und mehrphasige Regressionsmodelle behandeln (Abschnitt 4.2). Zusammenhänge, Analogien und Unterschiede zu diesen Arbeiten werden im folgenden aufgezeigt.

##### 4.1. Abruptes Strukturbruchmodell

Die Analyse von abrupten Strukturbruchmodellen geht auf die Arbeiten von Page (1954, 1955) zurück. Page (1955) geht davon aus, daß eine Folge von stochastisch unabhängigen Zufallsvariablen  $Z_1, \dots, Z_n$  vorliegt, die Verteilung der Zufallsvariablen  $Z_1, \dots, Z_\nu$  identisch ist, die Verteilung der Zufallsvariablen  $Z_{\nu+1}, \dots, Z_n$  ebenfalls identisch ist und die Verteilung der Zufallsvariablen  $Z_\nu$  und  $Z_{\nu+1}$  unterschiedlich sein kann. Des weiteren sei der *Wechselpunkt* bzw. *Strukturbruchpunkt*  $\nu$  mit  $1 \leq \nu < n$  unbekannt. Das Prüfpunktmodell (Abschnitt 2.2) kann man als verallgemeinertes abruptes Strukturbruchmodell auffassen. Hierbei sei  $(X_1, Y_1), \dots, (X_n, Y_n)$  eine bivariate Stichprobe gemäß Kapitel 2 und  $(D_{n1}, \dots, D_{nn})$  bezeichnet den Vektor der Antiränge von  $X$ . Da  $D_n$  m.W.1 eine Permutation der Zahlen  $(1, \dots, n)$

liefert, kann man  $Y$  m.W.1 anhand von  $D_n$  eindeutig ordnen. Die bzgl. der Einflußgröße  $X$  geordnete Folge der Zielgrößen bezeichnen wir mit  $Y'$ ; wobei  $Y'_i = Y_{D_{ni}}$ . Somit läßt sich das Prüfpunktmodell m.W.1 in ein abruptes Strukturbruchmodell überführen. Wir vernachlässigen hierbei allerdings Information über die Verteilung von  $X$ , und der unbekannt Wechsellpunkt  $\nu'$  wird zu einer Zufallsvariablen, wobei  $\nu' = nF_{nX}(\mu)$ . Der Wechsellpunkt  $\nu'$  modelliert somit die Anzahl der Beobachtungen in der Gruppe mit Werten der Einflußgröße kleiner oder gleich dem Prüfpunkt  $\mu$ . Der hier beschriebene Zusammenhang erlaubt einerseits eine teilweise Übertragung von Ergebnissen aus Arbeiten über abrupte Strukturbruchmodelle (vgl. Abschnitt 5.1 und 7.1), und andererseits sind einige Ergebnisse unserer Arbeit auf den Spezialfall abruptes Strukturbruchmodell anwendbar. Dies gilt insbesondere für die Verteilung der maximal selektierten Rangstatistiken unter der Nullhypothese (siehe Abschnitt 3.2, 5.1 und 5.2) und für die diskutierten Approximationsmöglichkeiten (Abschnitt 5.3, 5.4 und 6.3), da die Verteilung unter der Nullhypothese bei beiden Modellierungen m.W.1 identisch ist. Ein wesentlicher Unterschied des in der vorliegenden Arbeit betrachteten Prüfpunktmodells gegenüber einem abrupten Strukturbruchmodell ist der Klassifikationsaspekt der Modellierung. Hierdurch ist die Einführung von den minimalen Gruppenanteilen  $\varepsilon_1$  und  $\varepsilon_2$  naheliegend und beispielsweise für die diagnostische Interpretation einer Einflußgröße sinnvoll. Dagegen kann man den durch  $\varepsilon_1$  und  $\varepsilon_2$  gegebenen Selektionsbereich bei Strukturbruchproblemen verwenden, um als asymptotische Verteilung die standardisierte Brownsche Brücke zu erhalten (James, James und Siegmund 1987). Dies gilt auch für die von Ledwina (1986) betrachtete maximal selektierte Rangstatistik als nichtparametrischer Test auf positive quadratische Abhängigkeit. Des weiteren ist bei der Analyse des Prüfpunktmodells der Einfluß der Verteilung der möglichen Einflußgröße von Bedeutung (vgl. Abschnitte 9.1 und 10.1).

## 4.2. Strukturbruchstatistiken

Ausgehend von den Arbeiten von Page (1954,1955) sind vielfältige Arbeiten über Strukturbruchmodelle erschienen. Dementsprechend fand die Thematik in den letzten Jahren auch in Lehrbüchern Beachtung. Beispiele hierfür sind Krämer und Sonnberger (1986), Schulze (1987) und Shorack und Wellner (1986). Krämer und Sonnberger (1986) und Schulze (1987) diskutieren parametrische Strategien zur Analyse von Strukturbrüchen im linearen Modell. Aspekte des durch lineare Rangstatistiken gegebenen empirischen Prozesses werden bei Shorack und Wellner (1986) behandelt. Siegmund (1986, 1988) bespricht Strukturbruchstatistiken im Zusammenhang mit parametrischen sequentiellen Strategien und gibt eine Übersicht über asymptotische Ergebnisse. Des weiteren kann man die Arbeit von Wolfe und Schechtmann (1984) erwähnen, wo erste nichtparametrische Prozeduren zur Analyse des abrupten Strukturbruchproblems zusammengefaßt werden. Eine Bibliographie zu Strukturbruchproblemen geben Shaban (1980) und Hackl und Westlund (1989). Hackl (1989) und Lombard (1989) bieten eine aktuelle Übersicht und Einführung zu Strukturbruchproblemen in der Ökonometrie bzw. zum abrupten Strukturbruchproblem.

Im folgenden geben wir eine Übersicht über Arbeiten zu Strukturbruchproblemen und diskutieren Ergebnisse bei der Nullhypothese, daß kein Wechsellpunkt existiert. Wir erwähnen insbesondere Ergebnisse, die eine maximierte Statistik verwenden, da hierdurch in der Regel auch ein Wechsellpunkt geschätzt wird und insofern der Methode der maximal selektierten Rangstatistiken beim Prüfpunktmodell entspricht. Quandt (1958, 1960) analysiert einen Likelihoodquotiententest und verwirft aufgrund von Monte-Carlo Ergebnissen eine  $\chi^2$ -Approximation für diesen Test. Eine detaillierte Argumentation bzgl. der nicht statthaften  $\chi^2$ -Approximation findet sich bei Feder (1975). Hinkley (1970) leitet Ergebnisse bei existierendem Wechsellpunkt her, d.h. Aussagen über den Wechsellpunktschätzer unter der Alternative. MacNeill (1974) behandelt

Invarianzprinzipien von parametrischen Strukturbruchstatistiken sowohl unter der Nullhypothese, kein Wechsellpunkt, als auch unter Alternativen, mindestens ein Wechsellpunkt. Hawkins (1977) und Worsley (1979) leiten eine rekursive Integraldarstellung für die exakte Verteilung unter der Nullhypothese her. Sie betrachten das klassische Zweistichproben-Normalverteilungsmodell mit bekannter und unbekannter Varianz. Für dasselbe Problem geben James, James und Siegmund (1987, 1988) Approximationsformeln an. Eine andere Approximation der exakten Verteilung von Strukturbruchstatistiken liefert die verbesserte Bonferroni-Ungleichung von Hunter (1976) (siehe auch Worsley 1982). Worsley (1982, Bsp. 3.2) diskutiert die Verwendung bei Strukturbruchproblemen als einen Spezialfall des allgemeinen Vorschlags. Das Verschwinden des Wechsellpunktes unter der Hypothese verursacht die Probleme bei der Verwendung der üblichen Maximum-Likelihood Asymptotik. Diese Problematik beim Hypothesen Testen, wenn ein Parameter nur unter der Alternative existiert, behandelt Davies (1977, 1987). Davies (1977) approximiert die Verteilung von Teststatistiken in solchen Situationen durch einen stetig differenzierbaren Gaußprozeß. In dieser Situation schlägt Davies (1987) die Verwendung einer Schätzung der Kovarianzfunktion des Gaußprozesses in seiner Approximationsformel vor. Worsley (1983) betrachtet die Verteilung des Likelihoodquotiententests in der zweiphasigen Regression und demonstriert insbesondere die Verwendung einer Approximation der Verteilung unter der Nullhypothese mittels einer verbesserten Bonferroni-Ungleichung (Worsley 1982). Freeman (1986) schlägt eine Strategie unter Benutzung eines Anpassungstests vor und führt Vergleiche zu dem Maximum-Likelihood Vorgehen bei Sen und Srivastava (1975) durch.

Sen und Srivastava (1975) schlagen auch maximale Rangstatistiken als Strukturbruchstatistiken vor. Insbesondere diskutieren sie den maximalen standardisierten Median- und Wilcoxon-Test. Pettitt (1979) verwendet nichtstandardisierte Rangstatistiken und bespricht den Zusammenhang zum Kolmogorov-Smirnov-Test (vgl. Abschnitt 1.2).

Schechtman (1982) vergleicht beide Strategien anhand von Monte-Carlo Ergebnissen und der exakten Permutationsverteilung für  $n \leq 9$ . Weitere neuere Arbeiten zu nichtparametrischen Tests beim abrupten Strukturbruchmodell sind Lombard (1987) und Csörgö und Horvath (1987, 1988). Bhattacharyya und Johnson (1968) analysieren lineare Rangstatistiken (Tests auf Trend) bei einem abrupten Strukturbruchmodell und leiten Optimalitätsaussagen in Abhängigkeit einer apriori Verteilung des Wechsellpunkts her. Praagman (1988, 1989) vergleicht lineare Rangstatistiken und maximale Rangstatistiken anhand ihrer exakten Bahadur-Effizienz (siehe auch Abschnitt 7.1).

Darkhovskh (1976) zeigt, daß die Wilcoxon-Statistik unter allgemeinen Bedingungen eine konsistente Wechsellpunktschätzung ermöglicht. Pettitt (1980) diskutiert die Wechsellpunktschätzung mittels maximaler Rangstatistiken anhand von Monte-Carlo Ergebnissen. Carlstein (1988) leitet unter äußerst allgemeinen Bedingungen einen konsistenten Wechsellpunktschätzer her (vgl. Dümbgen 1989). Carlsteins Strategie beruht auf einer maximalen Statistik vom Anderson-Darling Typ zwischen den durch den Wechsellpunkt determinierten Teilfolgen. Weitere neuere Arbeiten über die nichtparametrische Wechsellpunktschätzung sind Miao und Zhao (1988) und Scariano und Watkins (1988). Miao und Zhao (1988) behandeln die gemeinsame konsistente Schätzung von mehreren Wechsellpunkten.

In der vorliegenden Arbeit verwenden und modifizieren wir die Ergebnisse von Praagman (1988) bei der Berechnung der exakten Bahadur-Effizienz (Abschnitt 7.1), und wir verallgemeinern die Konsistenzaussage von Darkhovskh (1976) (Abschnitt 9.1). Die von Lombard (1987) und von Pettitt (1979) betrachtete asymptotische Brownsche-Brücke unter der Nullhypothese erhalten wir im Kapitel 3 als Spezialfall. Das bei Bindungen und Zensierung gültige Invarianzprinzip ist auch für Strukturbruchprobleme neu. Zusammenfassend betonen wir, daß die in der vorliegenden Arbeit eingeführte Methode der maximal selektierten Rangstatistiken keine triviale Verallgemeinerung eines nichtparametrischen Vorgehens bei Strukturbruch-

problemen ist. Die Abhängigkeit von der Verteilung der Einflußgröße, die Interpretation des Prüfpunktmodells und die Festlegung der Selektionsgrenzen verdeutlichen diese Feststellung (vgl. Abschnitt 4.1).

## 5. Exakte Verteilung und Approximationsmöglichkeiten

Die in Kapitel 3 unter der Nullhypothese hergeleitete asymptotische Verteilung liefert eine Möglichkeit, einen maximal selektierten Rangtest durchzuführen. Hierbei stellt sich das Problem, wie gut die exakte Verteilung bei festem Stichprobenumfang durch die asymptotische approximiert wird. Im Abschnitt 5.1 diskutieren wir das Problem der Berechnung und der Approximation der exakten Verteilung. Des weiteren approximieren wir die exakte Verteilung für verschiedene Scorefunktionen anhand einer Monte-Carlo Studie und vergleichen sie mit der asymptotischen Verteilung (Abschnitt 5.2). Ungleichungen vom Bonferroni-Typ liefern eine andere Möglichkeit, das Niveau eines maximal selektierten Tests abzuschätzen (Abschnitt 5.3). Im Abschnitt 5.4 werden die diskutierten Approximationsmöglichkeiten verglichen.

### 5.1. Exakte Verteilung maximal selektierter Statistiken

Die exakte Verteilung der maximal selektierten Rangstatistiken kann anhand der Permutationsverteilung des Scorevektors berechnet werden. Die  $n!$  möglichen Permutationen ergeben maximal  $n! / ((n_1 - 1)! (n - n_2)!)$  relevante Kombinationen, wobei  $n_1 = F_{nX}(x_{n_1}) n$ ,  $n_2 = F_{nX}(x_{n_2}) n$  und  $x_{n_1}, x_{n_2}$  bezeichnen die  $\varepsilon_1$ -,  $\varepsilon_2$ -Quantile bzgl.  $F_{nX}$  (vgl. Kapitel 2). Allerdings ist die direkte Berechnung über die möglichen Permutationen des Scorevektors nur für kleine Stichprobenumfänge praktikabel, die für die Anwendung des Prüfpunktmodells irrelevant sind.

Die exakte Verteilung der maximal selektierten Gauß-Statistik (vgl. Abschnitt 2.4) könnte man als Approximationsmöglichkeit betrachten. Für diese parametrische Situation leitet Hawkins (1977) eine rekursive numerische Berechnungsmöglichkeit her und bestimmt so exakte Quantile bis zu einem Stichprobenumfang von  $n = 50$ . Er betrachtet ein abruptes Strukturbruchmodell und die dadurch gegebene Folge von Gauß-Statistiken. Sein Resultat gilt allgemein für die Verteilung des

maximalen Absolutbetrages einer Folge von Teststatistiken  $\{T_1, \dots, T_l\}$ , falls  $\{T_i; 1 \leq i \leq l\}$  ein Markoffscher standardisierter Gaußprozeß ist. Worsley (1979) leitet für  $t$ -Statistiken ein hierzu analoges Verfahren her. Die von James, James und Siegmund (1987) vorgeschlagene Approximation zweiter Ordnung liefert in beiden Fällen eine numerisch einfachere Lösung. Für die exakte Verteilung der maximal selektierten Gauß-Statistik kann man folgende Approximationsformel in Abhängigkeit des Stichprobenumfangs  $n$  verwenden (James u.a. 1987; Formel 18,17):

$$P(M_n(\varepsilon_1, \varepsilon_2) \geq b) \approx \quad (5.1.1)$$

$$2(1 - \Phi(b) + b \varphi(b)) \int_{b(1/n_2 - 1/n)^{1/2}}^{b(1/n_1 - 1/n)^{1/2}} \frac{1}{x} \exp(-0.583(x + \frac{b^2}{nx})) dx,$$

wobei  $\Phi$  die Verteilungsfunktion und  $\varphi$  die Lebesgue-Dichte der Standardnormalverteilung bezeichnet, des weiteren  $n_1 = F_{nX}(x_{n1})n$  und  $n_2 = F_{nX}(x_{n2})n$ . James u.a. (1987) und Siegmund (1985) empfehlen obige Approximation, falls die Integrationsgrenzen innerhalb des Intervalls  $[0,2]$  liegen.

## 5.2. Monte-Carlo Ergebnisse

Anhand von Monte-Carlo Berechnungen approximieren wir die exakte Verteilung maximal selektierter Rangstatistiken bei endlichen Stichprobenumfängen. Die exakte Verteilung hängt von der Wahl des Scorevektors ab, deshalb führten wir Monte-Carlo Berechnungen für verschiedene gebräuchliche Scorevektoren durch: Wilcoxon-Scores,  $a_n(i) = i$ ; Median-Scores,  $a_n(i) = 1_{\{i \leq n/2\}}$ ; Log-Rank-Scores,  $a_n(i) = 1 - \sum_{j=1}^i 1/(n-j+1)$  und van-der-Waerden-Scores,  $a_n(i) = \Phi^{-1}(i/(n+1))$ , wobei  $\Phi$  die Verteilungsfunktion der Standardnormalverteilung bezeichnet. Die Median-Score-Statistik

dichotomisiert die Zielgröße  $Y$ , und die betrachtete Teststatistik ist somit äquivalent zu der Wurzel der maximal selektierten  $\chi^2$ -Statistik von Miller und Siegmund (1982) bei  $a_\rho + b_\rho = [n/2]$  und  $c_\rho + d_\rho = n - [n/2]$  (vgl. Abschnitt 1.2). Wir können die maximal selektierte  $\chi^2$ -Statistik als eine maximal selektierte Quantil-Score-Statistik auffassen und Monte-Carlo Ergebnisse für die maximal selektierte Median-Score-Statistik mit Monte-Carlo Ergebnissen von Halpern (1982) für die maximal selektierte  $\chi^2$ -Statistik vergleichen. Des Weiteren führten wir Monte-Carlo Berechnungen für die maximal selektierte Zweistichproben- $t$ -Statistik und Gauß-Statistik durch.

Die Verteilung der maximal selektierten Rangstatistik ist unter der Nullhypothese invariant bzgl. stetiger Verteilungsfunktionen von  $X$  und  $Y$  und insofern nichtparametrisch. Bei unseren Monte-Carlo Berechnungen wählten wir für die Einflußgröße  $X$  die Gleichverteilung auf dem Intervall  $[0,1]$  und für die Zielgröße  $Y$  die Standardnormalverteilung. Gleichverteilte Pseudo-Zufallszahlen berechneten wir mit dem Algorithmus von Wichmann und Hill (1985) und Normalverteilungsquantile mit dem Algorithmus von Beasley und Springer (1985). Letzteren Algorithmus verwendeten wir auch für die Transformation der gleichverteilten Pseudo-Zufallszahlen in standardnormalverteilte Pseudo-Zufallszahlen. Wir berechneten 10 000 Wiederholungen für jede betrachtete Parameterkombination, wobei  $n = 10, 20, 30, 50, 100, 200$  und  $(\varepsilon_1, \varepsilon_2) = (0.25, 0.75), (0.4, 0.6), (0.1, 0.9)$  und  $(0.4, 0.9)$ .

Ein oberer Test bzgl.  $M_n(\varepsilon_1, \varepsilon_2)$  liefert eine Testvorschrift für das zweiseitige Testproblem (vgl. Abschnitt 2.2); d.h. zur Konstruktion eines exakten Tests ist die exakte Verteilung der Teststatistik erst ab dem 90% Quantil relevant. Deshalb beschränken wir uns im folgenden auf den Teil der jeweiligen Verteilung zwischen dem 90% und 100% Quantil. Unter einem  $K\%$  Quantil verstehen wir den Wert der inversen Verteilungsfunktion für  $K/100$ . Für festes  $x_0 \in \mathbb{R}$  ist  $p_m(x_0) = \#\{M_n(\varepsilon_1, \varepsilon_2) \leq x_0\} / m$  binomialverteilt; damit kann man die Genauigkeit der Monte-Carlo Berechnung von der Verteilungsfunktion der maximal selektierten Rangstatistik  $M_n(\varepsilon_1, \varepsilon_2)$ ,  $P(M_n(\varepsilon_1, \varepsilon_2) \leq x_0)$ ,

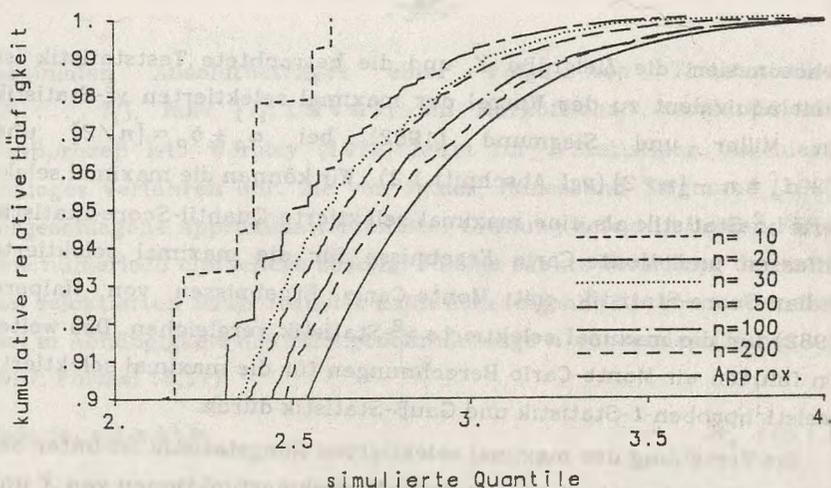


Abb. 5.2.1.: Simulierter und approximativer oberer Teil der Verteilungsfunktion der maximal selektierten Wilcoxon-Statistik unter der Nullhypothese bei  $(\varepsilon_1, \varepsilon_2) = (0.25, 0.75)$ . Die durchgezogene Linie veranschaulicht die approximative asymptotische Verteilung.

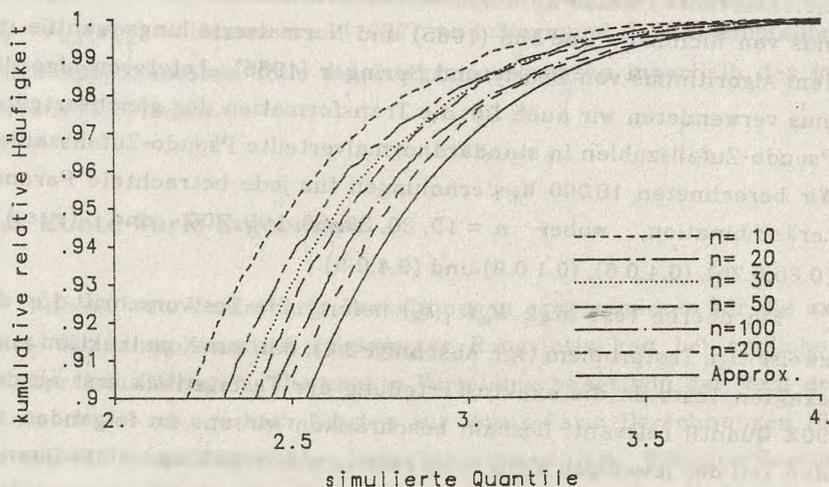


Abb. 5.2.2.: Simulierter und approximativer oberer Teil der Verteilungsfunktion der maximal selektierten Gauß-Statistik unter der Nullhypothese bei  $(\varepsilon_1, \varepsilon_2) = (0.25, 0.75)$ . Die durchgezogene Linie veranschaulicht die approximative asymptotische Verteilung.

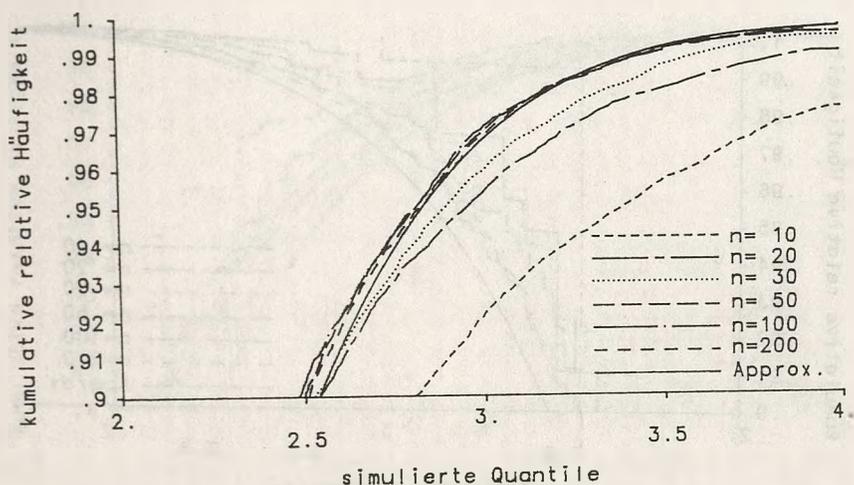


Abb. 5.2.3.: Simulierter und approximater oberer Teil der Verteilungsfunktion der maximal selektierten  $t$ -Statistik unter der Nullhypothese bei  $(\varepsilon_1, \varepsilon_2) = (0.25, 0.75)$ . Die durchgezogene Linie veranschaulicht die approximative asymptotische Verteilung.

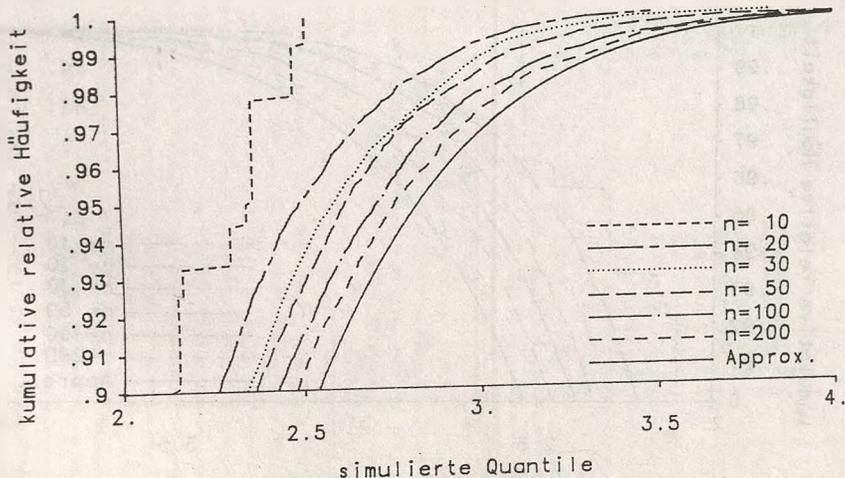


Abb. 5.2.4.: Simulierter und approximater oberer Teil der Verteilungsfunktion der maximal selektierten van-der-Waerden-Statistik unter der Nullhypothese bei  $(\varepsilon_1, \varepsilon_2) = (0.25, 0.75)$ . Die durchgezogene Linie veranschaulicht die approximative asymptotische Verteilung.

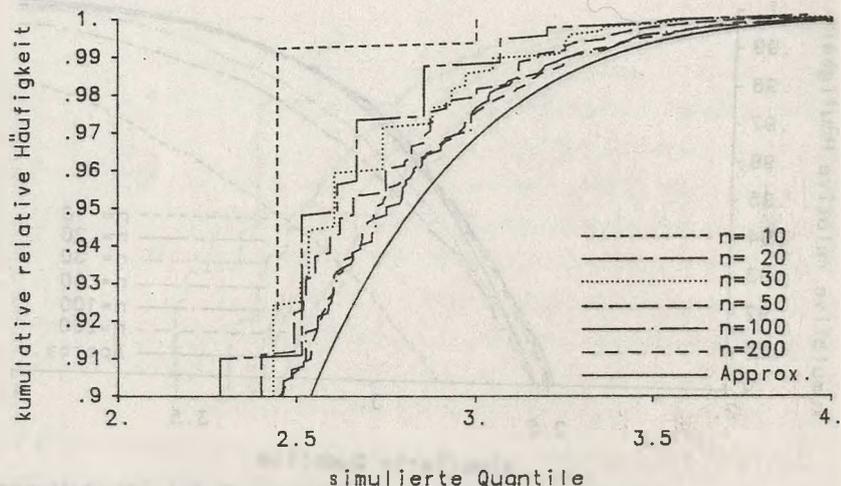


Abb. 5.2.5.: Simulierter und approximativer oberer Teil der Verteilungsfunktion der maximal selektierten Median-Statistik unter der Nullhypothese bei  $(\varepsilon_1, \varepsilon_2) = (0.25, 0.75)$ . Die durchgezogene Linie veranschaulicht die approximative asymptotische Verteilung.

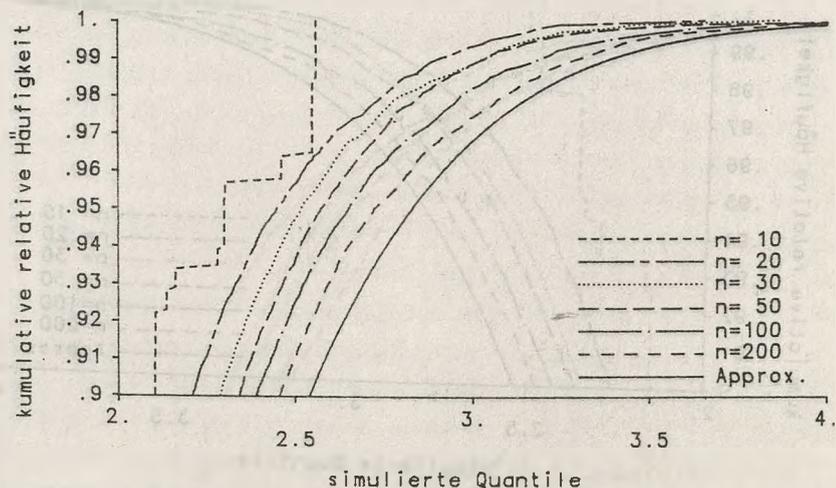


Abb. 5.2.6.: Simulierter und approximativer oberer Teil der Verteilungsfunktion der maximal selektierten Log-Rank-Statistik unter der Nullhypothese bei  $(\varepsilon_1, \varepsilon_2) = (0.25, 0.75)$ . Die durchgezogene Linie veranschaulicht die approximative asymptotische Verteilung.

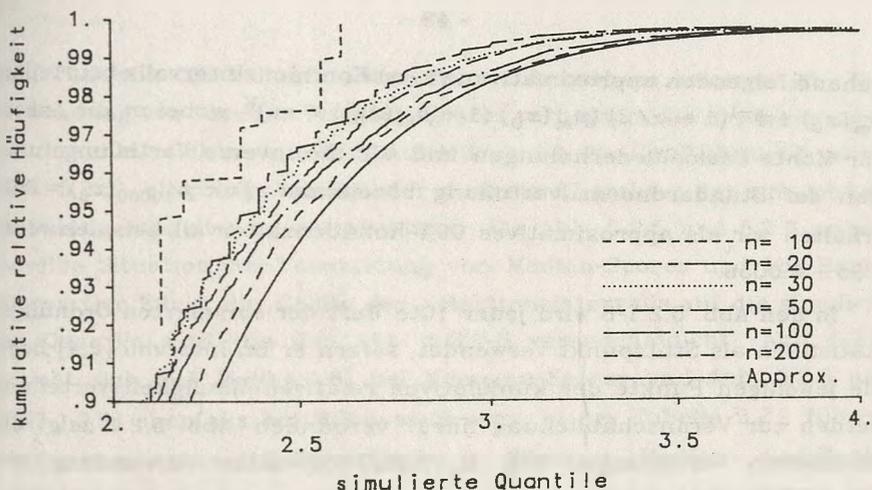


Abb. 5.2.7.: Simulierter und approximativer oberer Teil der Verteilungsfunktion der maximal selektierten Wilcoxon-Statistik unter der Nullhypothese bei  $(\epsilon_1, \epsilon_2) = (0.4, 0.6)$ . Die durchgezogene Linie veranschaulicht die approximative asymptotische Verteilung.

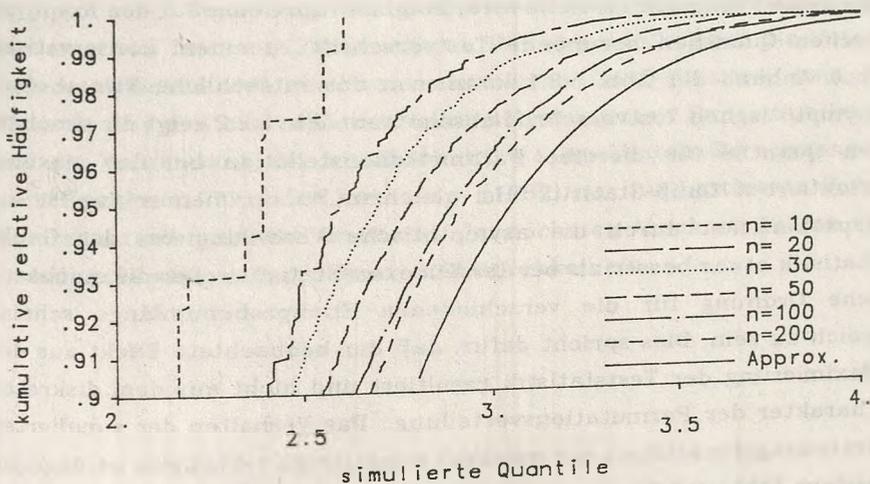


Abb. 5.2.8.: Simulierter und approximativer oberer Teil der Verteilungsfunktion der maximal selektierten Wilcoxon-Statistik unter der Nullhypothese bei  $(\epsilon_1, \epsilon_2) = (0.1, 0.9)$ . Die durchgezogene Linie veranschaulicht die approximative asymptotische Verteilung.

anhand folgenden approximativen  $(1-\alpha)$  Konfidenzintervalls beurteilen:  $p_m(x_0) \pm \Phi^{-1}(1 - \alpha/2) (p_m(x_0) (1 - p_m(x_0)) / m)^{1/2}$ , wobei  $m$  die Anzahl der Monte-Carlo Wiederholungen und  $\Phi^{-1}$  die inverse Verteilungsfunktion der Standardnormalverteilung bezeichnet. Für  $p_{10000}(x_0) = 0.95$  erhalten wir als approximatives 99%-Konfidenzintervall beispielsweise:  $0.95 \pm 0.0056$ .

In den Abb. 5.2.1-8 wird jeder 10te Wert der simulierten Ordnungsstatistiken als Stützpunkt verwendet, sofern er im Intervall  $[2,4]$  liegt. Die jeweiligen Punkte der kumulativen relativen Häufigkeitsverteilung werden zur Veranschaulichung linear verbunden. Abb. 5.2.1 zeigt die simulierten Verteilungen von  $M_n(0.25,0.75)$  unter Verwendung von Wilcoxon-Scores bei den betrachteten Stichprobenumfängen und die approximative asymptotische Verteilung gemäß Formel (3.3.1). Abb. 5.2.1 legt nahe, daß die asymptotische Verteilung stochastisch größer ist als die simulierten Verteilungen, wobei die asymptotische Verteilung nur relativ langsam erreicht wird. Folglich führt eine auf den asymptotischen Quantilen beruhende Testvorschrift zu einem konservativen Test. Anhand der Abb. 5.2.1 kann man das tatsächliche Niveau einer asymptotischen Testvorschrift abschätzen. Abb. 5.2.2 zeigt die simulierten Quantile für dieselbe Parameterkonstellation bei der maximal selektierten Gauß-Statistik. Bei gleichem Stichprobenumfang ist die Approximation durch die asymptotische Verteilung bei der Gauß-Statistik etwas besser als bei der Wilcoxon-Statistik, aber die stochastische Ordnung für die verschiedenen Stichprobenumfänge scheint gleich zu sein. Dies spricht dafür, daß der beobachtete Effekt aus der Maximierung der Teststatistik resultiert und nicht aus dem diskreten Charakter der Permutationsverteilung. Das Verhalten der simulierten Verteilungsfunktionen der maximal selektierten  $t$ -Statistik ist dagegen anders (Abb. 5.2.3). Für  $n = 10, 20, 30$  ist die simulierte Verteilungsfunktion stochastisch größer als die asymptotische Verteilung, und ab  $n = 50$  ist die Approximation der simulierten Verteilung durch die asymptotische Verteilung gut. Diese Unterschiede sind als Effekt der Varianzschätzung plausibel, da bei den Rangstatistiken und bei der

Gauß-Statistik die Varianz unter der Nullhypothese bekannt ist. Die Verwendung von van-der-Waerden-Scores verbessert die Approximation der maximal selektierten Rangstatistik durch die Gauß-Statistik nicht (Abb. 5.2.4), allerdings scheint der diskrete Charakter der simulierten Verteilung schneller zu verschwinden. Die Abb. 5.2.5 und 5.2.6 zeigen dieselbe Situation bei Verwendung von Median-Scores und Log-Rank-Scores. Der Effekt der Größe des Selektionsintervalls auf die simulierten Quantile wird von den Abb. 5.2.7-8 veranschaulicht. Abb. 5.2.7 bezieht sich auf  $M_n(0.4,0.6)$  bei Wilcoxon-Scores und Abb. 5.2.8 auf  $M_n(0.1,0.9)$  ebenfalls bei Wilcoxon-Scores. In der Tabelle 5.2.1 führen wir die simulierten 95% Quantile der  $t$ -, Wilcoxon-, Median-, Log-Rank-Statistik und die Approximation der 95%-Quantile nach James u.a. (1987; Formel 18,17) auf.

Zusammenfassend kann man feststellen, daß die Approximation der exakten Verteilung maximal selektierter Rangstatistiken durch die asymptotische Verteilung bei moderaten Stichprobenumfängen zu konservativen Tests zu führen scheint. Des weiteren belegen die Unterschiede zwischen der maximal selektierten Gauß-Statistik bzw.  $t$ -Statistik und den betrachteten maximal selektierten Rangstatistiken, daß die Verwendung der im Abschnitt 5.1 erwähnten Berechnungen nach Hawkins (1977) oder James u.a. (1987) nur für bestimmte Parametersituationen sinnvoll erscheint. Im nächsten Abschnitt führen wir deshalb eine weitere Approximationsmöglichkeit ein.

$(\varepsilon_1, \varepsilon_2)$	$n$	$t$	M	W	LR	JJS
(.25,.75)	10	3.39	2.45	2.39	2.30	2.49
(.25,.75)	20	2.95	2.62	2.50	2.45	2.58
(.25,.75)	30	2.90	2.61	2.60	2.52	2.62
(.25,.75)	50	2.81	2.66	2.64	2.57	2.66
(.25,.75)	100	2.81	2.75	2.73	2.65	2.71
(.25,.75)	200	2.82	2.72	2.77	2.72	2.74
(.25,.75)	$\infty$	2.83	2.83	2.83	2.83	2.83
(.4,.6)	10	2.87	2.45	2.19	2.04	2.25
(.4,.6)	20	2.70	2.28	2.39	2.26	2.30
(.4,.6)	30	2.59	2.51	2.41	2.30	2.33
(.4,.6)	50	2.54	2.29	2.43	2.36	2.39
(.4,.6)	100	2.52	2.44	2.46	2.41	2.41
(.4,.6)	200	2.54	2.55	2.50	2.49	2.43
(.4,.6)	$\infty$	2.56	2.56	2.56	2.56	2.56
(.1,.9)	10	3.70	2.45	2.39	2.55	2.66
(.1,.9)	20	3.20	2.67	2.59	2.70	2.77
(.1,.9)	30	3.06	2.69	2.70	2.76	2.82
(.1,.9)	50	2.98	2.82	2.78	2.82	2.87
(.1,.9)	100	3.00	2.88	2.90	2.91	2.93
(.1,.9)	200	2.98	2.94	2.93	2.94	2.97
(.1,.9)	$\infty$	3.05	3.05	3.05	3.05	3.05
(.4,.9)	10	3.34	2.45	2.35	2.29	2.51
(.4,.9)	20	2.98	2.52	2.47	2.50	2.61
(.4,.9)	30	2.92	2.54	2.58	2.62	2.65
(.4,.9)	50	2.86	2.66	2.66	2.67	2.70
(.4,.9)	100	2.87	2.79	2.78	2.74	2.75
(.4,.9)	200	2.88	2.82	2.82	2.81	2.79
(.4,.9)	$\infty$	2.88	2.88	2.88	2.88	2.88

Tabelle 5.2.1: Simulierte und approximierte 95% Quantile unter der Nullhypothese. Hierbei bezeichnet  $t$  die  $t$ -Statistik, M die Median-Statistik, W die Wilcoxon-Statistik, LR die Log-Rank-Statistik und JJS das anhand (5.1.1) approximierte 95% Quantil der Gauß-Statistik nach James u.a. (1987). Für  $n = \infty$  wurden die 95% Quantile mittels der Approximation der asymptotischen Verteilung (3.3.1) berechnet.

### 5.3. Bonferroni-Ungleichung und eine Verbesserung

Die Bonferroni-Ungleichungen (5.3.1) liefern eine allgemeine und in der Literatur oft verwandte Möglichkeit, das Niveau eines multiplen Tests nach oben bzw. nach unten abzuschätzen (z.B. Hawkins 1977; Worsley 1979; Krämer und Sonnberger 1986, p.147; Bolviken 1988). Ein multipler Test sei durch die Folge der Einzeltests  $A_1, \dots, A_l$  gegeben, wobei  $A_i = \{ |T_i| > b \}$  für  $i = 1, \dots, l$ . Wenn man die Verteilung der maximalen Teststatistik betrachtet, d.h.  $P(\max |T_i| > b)$ , dann ergeben die Bonferroni-Ungleichungen erster bzw. zweiter Ordnung eine einfache Abschätzung nach oben bzw. nach unten:

$$\sum_{i=1}^l P(A_i) - \sum_{i=1}^{l-1} \sum_{j=i+1}^l P(A_i \cap A_j) \leq P\left(\bigcup_{i=1}^l A_i\right) \leq \sum_{i=1}^l P(A_i). \quad (5.3.1)$$

Die obere Abschätzung liefert einen konservativen Test und kann einfach anhand der Verteilungen der einzelnen Tests berechnet werden. Im allgemeinen ist die untere Abschätzung in (5.3.1) eine bessere Näherung, wobei allerdings die bivariaten Verteilungen benötigt werden. Die asymptotischen univariaten bzw. bivariaten Verteilungen der Teststatistiken ergeben eine naheliegende Approximationsmöglichkeit der Ungleichungen (5.3.1). Worsley (1982, Korollar 1) leitet folgende verbesserte obere Abschätzung her (Hunter 1976):

$$P\left(\bigcup_{i=1}^l A_i\right) \leq \sum_{i=1}^l P(A_i) - \sum_{i=1}^{l-1} P(A_i \cap A_{i+1}). \quad (5.3.2)$$

Beweis nach Worsley (1982):

$$\begin{aligned} P\left(\bigcup_{i=1}^l A_i\right) &\leq P\left(A_1 \cup \left(\bigcup_{i=1}^{l-1} \{A_{i+1} - A_i\}\right)\right) \\ (5.3.1) \quad &\leq P(A_1) + \sum_{i=1}^{l-1} P(A_{i+1} \cap A_i^c) \\ &\leq \sum_{i=1}^l P(A_i) - \sum_{i=1}^{l-1} P(A_i \cap A_{i+1}). \end{aligned}$$

Worsley (1982, Theorem 1) betrachtet die Ereignisse  $A_i$  als Knoten eines Graphen  $G$ , wobei zwei Knoten  $A_i, A_j$  nur dann durch eine Kante  $e_{ij}$  verbunden sind, falls die beiden Ereignisse nicht disjunkt sind. Worsley zeigt für einen beliebigen Teilgraphen  $H$  von  $G$ , daß gilt:

$$P\left(\bigcup_{i=1}^l A_i\right) \leq \sum_{i=1}^l P(A_i) - \sum_{\{i,j:e_{ij} \in H\}} P(A_i \cap A_j),$$

genau dann, wenn  $H$  ein Baum ist.

Man erhält die beste obere Schranke nach (5.3.2) für den Baum  $H$  der  $\sum P(A_i \cap A_j)$  maximiert. Wenn  $T_1, \dots, T_l$  eine bezüglich  $X$  geordnete Folge der in Kapitel 2 eingeführten Rangstatistiken ist, dann ist (5.3.2) mit  $H = \{e_{12}, e_{23}, \dots, e_{l-1,l}\}$  bzgl. der asymptotischen Normalverteilung unter den zulässigen Teilgraphen optimal (vgl. Worsley 1982, Bsp. 3.2).

Des weiteren ist die asymptotische  $l$ -variate Normalverteilung von  $T_1, \dots, T_l$  bekannt, und die asymptotische bivariate Normalverteilung von  $T_i$  und  $T_j$  wird durch die Korrelation  $\text{Korr}(T_i, T_j)$  determiniert. Wenn  $T_1, \dots, T_n$  die Folge von Zweistichproben-Gauß-Statistiken bezeichnet (vgl. Abschnitt 2.4), dann gilt für  $i < j$ :  $\text{Korr}(T_i, T_j) = (i(n-j)/((n-i)j))^{1/2}$  (Worsley 1983; Formel 6.3, bei  $p=1$ ). Worsley (1983) leitet eine einfache Approximationsformel für korrelierte chiquadratverteilte Teststatistiken her. Für die hier betrachteten Zweistichprobentests erhält man mittels der von Worsley (1983; Formel 6.5) benutzten Reihenentwicklung:

$$P(A_i \cap A_j) =$$

$$1 - F_1(b^2) - (2/\pi) f_2(b^2) [t_{ij} - (b^2/4 - 1)(t_{ij}^3/6] + O(t_{ij}^5),$$

wobei  $i \neq j$ ,  $t_{ij} = (1 - \text{Korr}(T_i, T_j))^2)^{1/2}$ ,  $b > 0$ ,  $F_1$  die Verteilungsfunktion der  $\chi_1^2$  Verteilung und  $f_2$  die Dichte der  $\chi_2^2$  Verteilung bezeichnet. Somit kann man die Abschätzung (5.3.2) für die maximal selektierte Gaußstatistik in einfacher Weise approximieren. Die Verwendung bei Rangstatistiken läßt sich asymptotisch rechtfertigen (vgl. auch Kapitel 6).

#### 5.4. Vergleich verschiedener Approximationsmöglichkeiten

In diesem Abschnitt wollen wir die eingeführten Approximationsmöglichkeiten der Verteilung maximal selektierter Rangstatistiken bei moderatem Stichprobenumfang vergleichen. In den Abb. 5.4.1-4 werden die simulierten Verteilungen der maximal selektierten Gauß-Statistik und der Wilcoxon-Statistik für das Selektionsintervall  $(\epsilon_1, \epsilon_2) = (0.25, 0.75)$  zusammen mit drei Approximationsmöglichkeiten dargestellt. Die verbesserte Bonferroni-Ungleichung ist anhand einer Reihenentwicklung nach Worsley (1983) gemäß Abschnitt 5.3 berechnet worden. Die Verteilungsfunktion der Standardnormalverteilung (fester Prüfpunkt) ist als Vergleichsgröße ebenfalls dargestellt. Die vier Abbildungen beziehen sich auf die Stichprobenumfänge  $n = 10, 20, 30, 50$ . Die verbesserte Bonferroni-Ungleichung verbessert die Approximation gegenüber der asymptotischen Verteilung für  $n = 10, 20, 30$  erheblich. Für  $n = 50$  ist die verbesserte Bonferroni-Ungleichung noch etwas weniger konservativ als die Approximation durch die asymptotische Verteilung. Die aufwendigeren Berechnungen nach Hawkins (1977) bzw. James u.a. (1987) erscheinen ab etwa  $n = 50$  als mögliche alternative Approximation. Da diese Berechnungen jedoch relativ aufwendig sind und für die Rangstatistiken nur approximativ gelten, verzichten wir auf eine umfassende Analyse derselben (vgl. Tabelle 5.2.1). Die beiden Möglichkeiten, approximative asymptotische Verteilung und verbesserte Bonferroni-Ungleichung, lassen beide eine einfache Berechnung von kritischen Werten zu und scheinen beide zu eher konservativen Tests zu führen.

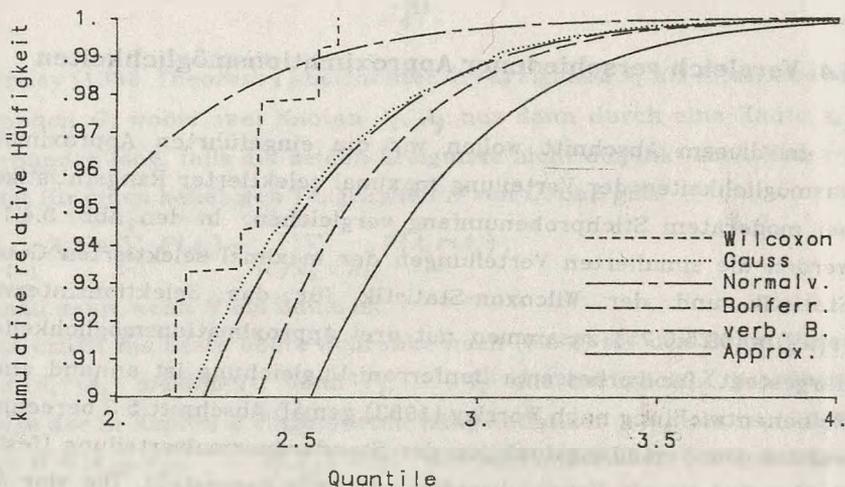


Abb. 5.4.1.: Vergleich der Approximationsmöglichkeiten der exakten Verteilung der maximal selektierten Wilcoxon-Statistik und der maximal selektierten Gauß-Statistik für  $n = 10$  und  $(\epsilon_1, \epsilon_2) = (0.25, 0.75)$ .

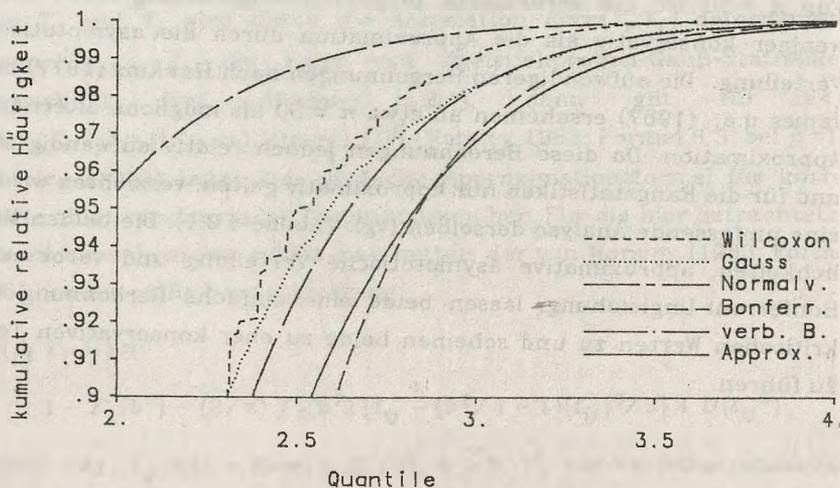


Abb. 5.4.2.: Vergleich der Approximationsmöglichkeiten der exakten Verteilung der maximal selektierten Wilcoxon-Statistik und der maximal selektierten Gauß-Statistik für  $n = 20$  und  $(\epsilon_1, \epsilon_2) = (0.25, 0.75)$ .

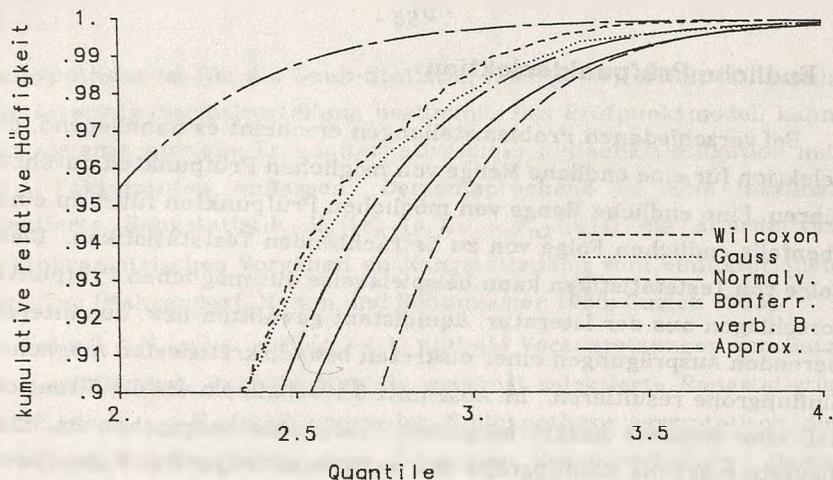


Abb. 5.4.3.: Vergleich der Approximationsmöglichkeiten der exakten Verteilung der maximal selektierten Wilcoxon-Statistik und der maximal selektierten Gauß-Statistik für  $n = 30$  und  $(\varepsilon_1, \varepsilon_2) = (0.25, 0.75)$ .

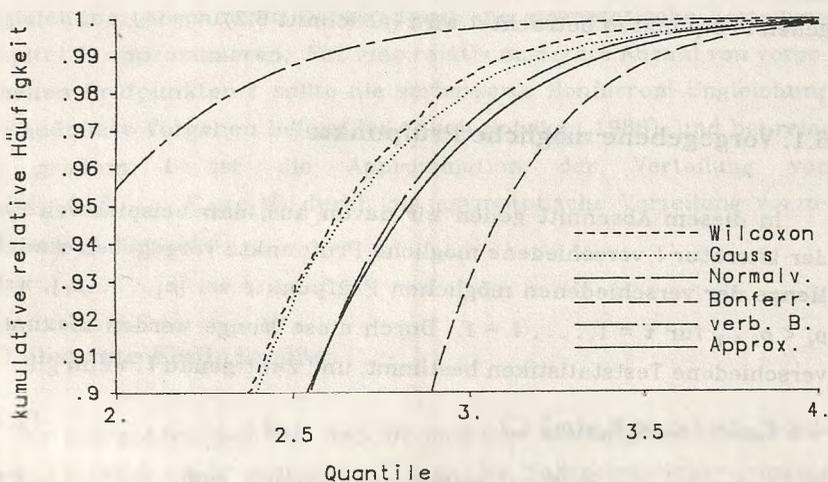


Abb. 5.4.4.: Vergleich der Approximationsmöglichkeiten der exakten Verteilung der maximal selektierten Wilcoxon-Statistik und der maximal selektierten Gauß-Statistik für  $n = 50$  und  $(\varepsilon_1, \varepsilon_2) = (0.25, 0.75)$ .

## 6. Endliche Prüfpunktselektion

Bei verschiedenen Problemstellungen erscheint es naheliegend, die Selektion für eine endliche Menge von möglichen Prüfpunkten durchzuführen. Eine endliche Menge von möglichen Prüfpunkten führt zu einer ebenfalls endlichen Folge von zu betrachtenden Teststatistiken. Diese Folge von Teststatistiken kann beispielsweise aus möglichen Prüfpunkt-vorschlägen aus der Literatur, äquidistant gewählten bzw. aus interessierenden Ausprägungen einer diskreten bzw. diskretisierten möglichen Einflußgröße resultieren. Im Abschnitt 6.1 behandeln wir die Situation, daß eine endliche Anzahl möglicher Prüfpunkte vorgegeben ist. Eine diskrete mögliche Einflußgröße mit endlichem Träger der Wahrscheinlichkeitsmasse führt zu einer endlichen Anzahl von relevanten möglichen Prüfpunkten (Abschnitt 6.2). Des weiteren kann man die Methode der maximal selektierten Rangstatistiken modifizieren, indem die Selektion endlich definiert wird; d.h. nur eine endliche Anzahl von möglichen Prüfpunkten betrachtet wird (Abschnitt 6.3).

### 6.1. Vorgegebene mögliche Prüfpunkte

In diesem Abschnitt gehen wir davon aus, daß beispielsweise aus der Literatur  $l$  verschiedene mögliche Prüfpunkte vorgegeben sind. Die Menge der verschiedenen möglichen Prüfpunkte sei  $\{\rho_1, \dots, \rho_l\}$ , wobei  $\rho_i < \rho_{i+1}$  für  $i = 1, \dots, l-1$ . Durch diese Menge werden maximal  $l$  verschiedene Teststatistiken bestimmt, und zwar genau  $l$ , wenn gilt:

$$0 < F_{nX}(\rho_1) < \dots < F_{nX}(\rho_l) < 1. \quad (6.1.1)$$

Wir benennen die maximal selektierte Statistik bzgl.  $\{\rho_1, \dots, \rho_l\}$  mit  $M_n(\{F_{nX}(\rho_1), \dots, F_{nX}(\rho_l)\})$  (vgl. Formel 6.1.2).

$$M_n(\{F_{nX}(\rho_1), \dots, F_{nX}(\rho_l)\}) = \max_{\rho \in \{\rho_1, \dots, \rho_l\}} \frac{|T_{n\rho} - E(T_{n\rho})|}{(\text{Var}(T_{n\rho}))^{1/2}}. \quad (6.1.2)$$

Die exakte Verteilung von  $M_n(\{F_{nX}(\rho_1), \dots, F_{nX}(\rho_l)\})$  unter der

Nullhypothese ist für die Gauß-Statistik, falls (6.1.1) erfüllt ist, durch eine  $l$ -variante Normalverteilung bestimmt. Das Prüfpunktmodell kann man als eine spezielle Trendalternative einer Einfachklassifikation mit  $l + 1$  Faktorstufen auffassen. Dementsprechend ist eine maximal selektierte Rangstatistik  $M_n(\{F_{nX}(\rho_1), \dots, F_{nX}(\rho_l)\})$  ein alternatives nichtparametrisches Vorgehen zu Rangstatistiken vom einfaktoriellen Page-Typ (Wahrendorf, Mahon und Schumacher 1985; Cuzick 1985).

Falls  $0 < F_X(\rho_1) < \dots < F_X(\rho_l) < 1$  und die Voraussetzungen von Satz 3.2.1 erfüllt sind, dann besitzt die maximal selektierte Rangstatistik  $M_n(\{F_{nX}(\rho_1), \dots, F_{nX}(\rho_l)\})$  unter der Nullhypothese asymptotisch die Verteilung des Maximums einer  $l$ -variaten Normalverteilung. Damit liefert die Arbeit von Hawkins (1977) eine numerisch aufwendige Approximationsmöglichkeit der Verteilung für festes  $l$ . In Abhängigkeit der Anzahl der vorgegebenen Prüfpunkte kann man die Verteilung von  $M_n(\{F_{nX}(\rho_1), \dots, F_{nX}(\rho_l)\})$  einfach durch die verbesserte Bonferroni-Ungleichung (Abschnitt 5.3) bzw. durch die asymptotische Verteilung (Kapitel 3) approximieren. Für eine relativ moderate Anzahl von vorgegebenen Prüfpunkten  $l$  sollte die verbesserte Bonferroni-Ungleichung das adäquate Vorgehen liefern (vgl. auch Bolviken 1988), und bei relativ großem  $l$  ist die Approximation der Verteilung von  $M_n(\{F_{nX}(\rho_1), \dots, F_{nX}(\rho_l)\})$  durch die asymptotische Verteilung vorzuziehen (vgl. Kapitel 5).

## 6.2. Diskrete Einflußgröße

Wir betrachten den Fall, daß die mögliche Einflußgröße  $X$  eine diskrete Verteilung mit endlichem Träger der Wahrscheinlichkeitsmasse hat. Das heißt, wir gehen davon aus, daß die Anzahl der beobachtbaren Prüfpunkte m.W.1 endlich ist. Hierdurch erhalten wir mit positiver Wahrscheinlichkeit Bindungen in der möglichen Einflußgröße  $X$ . Es ergeben sich somit die gleichen Approximationsmöglichkeiten wie im Abschnitt 6.1. Unter der Nullhypothese kann man das Problem einer

diskreten Einflußgröße auf das Problem von vorgegebenen möglichen Prüfpunkten zurückführen. Die von  $\varepsilon_1$  und  $\varepsilon_2$  abhängige Menge der beobachteten möglichen Prüfpunkte ist:

$$\{X_i : \varepsilon_1 \leq F_{nX}(X_i) \leq \varepsilon_2, i = 1, \dots, n\}.$$

### 6.3. Diskrete Selektion

Im Kapitel 5 beschreiben wir Approximationsmöglichkeiten der (exakten) Verteilung maximal selektierter Statistiken unter der Nullhypothese. Da die Güte einer Approximation im Fall der endlichen Selektion einfacher zu analysieren ist, erscheint es naheliegend, die Selektion endlich zu definieren. Das heißt, wir führen die Selektion für eine vorgegebene Menge von Quantilen bzgl.  $\{\varepsilon_1, \dots, \varepsilon_l\}$  durch. Die Verteilung der maximal selektierten Gauß-Statistiken ist dann beispielsweise bei  $\{\varepsilon_1, \dots, \varepsilon_6\} = \{0.25, 0.35, 0.45, 0.55, 0.65, 0.75\}$  für  $n = 10K, K \in \mathbb{N}$  identisch, und somit werden mögliche Approximationen in Abb. 5.4.1 dargestellt und verglichen. Da die Güte der Approximation einer diskret maximal selektierten Rangstatistik durch die entsprechende diskret maximal selektierte Gaußstatistik mit dem Verhältnis von  $n/l$  wächst, wird das Approximationsproblem durch die diskrete Selektion bei Rangstatistiken reduziert. Ein Nachteil dieses Vorgehens ist allerdings, daß der Parameterraum des Prüfpunktmodells durch die Diskretisierung eingeschränkt wird.

## 7. Asymptotische relative Effizienz

Das Konzept der asymptotisch relativen Effizienz liefert eine Vergleichsmöglichkeit für konkurrierende Testvorschriften. Rothe (1980, 1981) zeigt auf, daß man die verschiedenen Effizienzkonzepte als Grenzübergänge des Quotienten von Stichprobenumfängen auffassen kann. Hierbei wird der Stichprobenumfang betrachtet, der erforderlich ist, um bei vorgegebenem Niveau und fester Alternative eine bestimmte Schärfe zu erreichen. Im Abschnitt 7.1 betrachten wir die exakte Bahadur-Effizienz (z.B. Bahadur 1967, 1971; Rothe 1980) bzgl. maximal selektierter Rangtests und bzgl. Rangtests vom Spearman-Typ in Analogie zu Praagman (1988, 1989) (vgl. auch Ledwina 1986), wobei sich Optimalitätsaussagen im Sinne der exakten Bahadur-Effizienz lokal, d.h. nahe der Nullhypothese, präzisieren lassen (Praagman 1988, Sec. 7; Ledwina 1986; Bajorski 1987). Im Abschnitt 7.2 bestimmen wir die approximative Bahadur-Effizienz, und im Abschnitt 7.3 diskutieren wir die Möglichkeit, andere Optimalitätskonzepte anzuwenden.

### 7.1. Exakte Bahadur-Effizienz

In diesem Abschnitt nehmen wir an, daß die Verteilungsfunktion von  $(X, Y)$  stetig ist. Praagman (1988) betrachtet die exakte Bahadur-Effizienz unter einem abrupten Strukturbruchmodell für maximal selektierte Rangstatistiken mit  $(\varepsilon_1, \varepsilon_2) = [0, 1]$ . Aufgrund dieser Unterschiede müssen wir die Ergebnisse und Beweise von Praagman (1988) ausführlich diskutieren und teilweise modifizieren. Wir berechnen den *Bahadur-Slope* anhand folgenden Satzes (vgl. Rothe 1980, Paragraph 3):

*Satz 7.1.1. (Bahadur 1967):*

$\{T_n\}_{n \in \mathbb{N}}$  sei eine Folge von Teststatistiken bzgl. des Testproblems  $H_0: \vartheta \in \Theta_0$  vs.  $H_a: \vartheta \in \Theta - \Theta_0 = \Theta_1$ . Falls  $b(\vartheta) \in \mathbb{R}$  mit  $\vartheta \in \Theta_1$  existiert und eine nicht negative und für  $t = b(\vartheta)$  stetige Funktion  $h(t)$  existiert, so daß:

$$\lim_{n \rightarrow \infty} n^{-1} T_n = b(\vartheta) \quad P_{\vartheta}\text{-f.s., } \vartheta \in \Theta_1, \quad (7.1.1)$$

und

$$\lim_{n \rightarrow \infty} -n^{-1} \log[\sup\{P_{\vartheta}(T_n \geq nt); \vartheta \in \Theta_0\}] = h(t), \quad (7.1.2)$$

für alle  $t \in \mathbb{R}$ , dann ist  $2h(b(\vartheta))$  der Bahadur-Slope von  $\{T_n\}$  in  $\vartheta$ .

Um den Bahadur-Slope mittels Satz 7.1.1 bestimmen zu können, müssen wir die Konsistenzaussage (7.1.1) und die Aussage über die "exponentielle Konvergenzgeschwindigkeit" (7.1.2) herleiten. Wir passen unsere Notation und das Prüfpunktmodell teilweise an die Arbeit von Praagman (1988) an. Danach zeigen wir analog zu Praagman (1988) (7.1.2) und (7.1.1). Diese Aussagen lassen einen speziellen Effizienzvergleich von maximal selektierten Rangtests und Rangtests vom Spearman-Typ in Analogie zu dem Vorgehen von Praagman (1988) zu. Bei dem speziellen Vergleich beweisen wir, daß der maximal selektierte Rangtest mindestens so gut ist im Sinne der Bahadur-Effizienz wie der analoge Rangtest vom Spearman-Typ, falls der wahre Prüfpunkt zwischen dem  $\varepsilon_1$ - und  $\varepsilon_2$ -Quantil liegt.

In diesem Abschnitt behandeln wir das in Kapitel 2 eingeführte Prüfpunktmodell und das durch  $H_a^<$  gegebene Testproblem in folgender Formulierung:  $H(x, y)$  bezeichnet die gemeinsame Verteilungsfunktion von  $(X_i, Y_i)$ , und wir betrachten das Testproblem:

$$H_0 : H(x, y) = F_X(x) F_Y(y) \text{ vs.}$$

$$H_a^< : H(x, y) = F_X(x) F_Y(y) \quad \text{falls } x \leq \mu$$

$$H(x, y) = F_X(\mu) F_Y(y) + (F_X(x) - F_X(\mu)) G_Y(y) \quad \text{sonst,}$$

wobei  $H: \mathbb{R} \times \mathbb{R} \rightarrow [0, 1]$ , und  $F_Y < G_Y$ .  $F_X$ ,  $F_Y$ ,  $G_Y$  und  $\mu$  sind unbekannt; wobei  $0 < F_X(\mu) < 1$ . Anzumerken ist, daß die Modellierung von Praagman (1988),  $F_X(x) = x$  für  $0 \leq x \leq 1$ , unter  $H_0$  keine Einschränkung obiger Modellierung bedeutet. Praagman (1988) verwendet die gleichverteilte Zufallsvariable  $X$  nur aus beweistechnischen Gründen.

Als nächstes betrachten wir maximal selektierte Rangstatistiken und Rangstatistiken vom Spearmanotyp in Abhängigkeit von *empirischen Verteilungsfunktionen* (Praagman 1988, Sektion 2). Hierzu führen wir folgende von den Scores  $a_n(1), \dots, a_n(n)$  abhängende Funktion ein:  $\varphi_n : [0,1] \rightarrow \mathbb{R}$  mit  $\varphi_n(u) = a_n(i)$  für  $u \in ((i-1)/n, i/n]$  und  $\varphi_n(0) = 0$ ,  $i = 1, \dots, n$ .  $\psi_n$  und  $\gamma_n$  werden analog in Abhängigkeit von den Konstanten  $d_n(1), \dots, d_n(n)$  und  $c_n(1), \dots, c_n(n)$  definiert. In der Regel konvergieren  $\varphi_n$ ,  $\psi_n$  und  $\gamma_n$  gegen Funktionen  $\varphi$ ,  $\psi$  und  $\gamma$ , und man kann  $a_n$ ,  $d_n$  und  $c_n$  in Abhängigkeit von  $\varphi$ ,  $\psi$  und  $\gamma$  definieren. Man spricht dann von den *Scoreerzeugenden-Funktionen*  $\varphi$ ,  $\psi$  und  $\gamma$ .  $H_n(x, y) = \sum 1_{\{X_i \leq x, Y_i \leq y\}}/n$  ist die empirische Verteilungsfunktion bzgl.  $(X_i, Y_i)$ ,  $i = 1, \dots, n$  und  $H_{n,X}(x) = H_n(x, \infty)$  und  $H_{n,Y}(y) = H_n(\infty, y)$  sind die marginalen empirischen Verteilungsfunktionen. Dann gilt für die Statistik vom Spearmanotyp m.W.1:

$$S_n = n \int_0^1 \int_0^1 J_n(u, v) dH_n(H_{n,X}^{-1}(u), H_{n,Y}^{-1}(v)), \quad (7.1.3)$$

wobei

$$J_n(u, v) = \varphi_n(u) \psi_n(v), \quad (7.1.4)$$

und für die maximal selektierte Rangstatistik  $M_n(\varepsilon_1, \varepsilon_2)$  gilt m.W.1:

$$M_n(\varepsilon_1, \varepsilon_2) = \sup_{\varepsilon_1 \leq \rho \leq \varepsilon_2} n \int_0^1 \int_0^1 J_{n,\rho}(u, v) dH_n(H_{n,X}^{-1}(u), H_{n,Y}^{-1}(v)), \quad (7.1.5)$$

wobei

$$J_{n,\rho}(u, v) = \gamma_n(\rho) \varphi_n(u) 1_{[0,\rho]}(v). \quad (7.1.6)$$

Dagegen betrachtet Praagman (1988, Formel 2.4) als 'max-type' Statistik das Supremum über  $0 \leq \rho \leq 1$ .

### *Wahrscheinlichkeit großer Abweichungen*

Die Darstellung der Teststatistik in Abhängigkeit von empirischen Verteilungsfunktionen ermöglicht es, die *Wahrscheinlichkeit großer*

*Abweichungen* (7.1.2) anhand eines Ergebnisses von Groeneboom, Oosterhoff und Ruymgaart (1979) zu bestimmen. Hierbei sei  $H_\nu$  die Verteilungsfunktion von  $(X_i, Y_i)$ ,  $i=1,2,\dots$  und  $D$  der Raum der zweidimensionalen Verteilungsfunktionen  $H(x,y)$  versehen mit der Topologie  $\tau$  der Konvergenz auf allen Borel-Mengen. Für  $H \in D$  bezeichnet  $K(H, H_\nu)$  die *Kullback-Leibler Information* von  $H$  bzgl.  $H_\nu$  (siehe z.B. Shorack und Wellner 1986, p.789-792).

Des weiteren sei  $T: D \rightarrow \overline{\mathbb{R}}$ ,  $\Omega_{T,t} = \{H \in D \mid T(H) \geq t\}$ ,  $t \in \mathbb{R}$ , und  $K(\Omega_{T,t}, H_\nu) = \inf\{K(H, H_\nu) \mid H \in \Omega_{T,t}\}$ .

*Satz 7.1.2* (Groeneboom, Oosterhoff und Ruymgaart 1979, Theorem 3.2):  $T: D \rightarrow \overline{\mathbb{R}}$  sei  $\tau$ -stetig für jedes  $H$  mit  $K(H, H_\nu) < \infty$ , die Funktion  $t \rightarrow K(\Omega_{T,t}, H_\nu)$ ,  $t \in \mathbb{R}$ , sei rechtsseitig stetig für  $t = \tau$  und für  $\{u_n\}_{n \in \mathbb{N}}$ ,  $u_n \in \mathbb{R}$  gelte:  $\lim_{n \rightarrow \infty} u_n = 0$ . Dann gilt:

$$\lim_{n \rightarrow \infty} -n^{-1} \log[P(T(H_n) \geq \tau + u_n)] = K(\Omega_{T,\tau}, H_\nu).$$

Zur Berechnung der Wahrscheinlichkeit großer Abweichungen verwenden wir weitere Bezeichnungen und Definitionen:  $H(x,y) \in D$ ,  $H_X(x)$  und  $H_Y(y)$  bezeichnen die Verteilungsfunktionen der Randverteilungen,  $\overline{H}(u,v) = H(H_X^{-1}(u), H_Y^{-1}(v))$ , und für festes  $m \in \mathbb{N}$ ,  $M_{Jm, \varepsilon_1, \varepsilon_2}: D \rightarrow \overline{\mathbb{R}}$ .

$$M_{Jm, \varepsilon_1, \varepsilon_2}(H) = \sup_{\varepsilon_1 \leq \rho \leq \varepsilon_2} \iint J_{m,\rho}(u,v) d\overline{H}(u,v). \text{ Des weiteren sei}$$

$$\Omega = \{H \in D \mid h \text{ Lebesgue-Dichte von } H \text{ mit } \int h(u,v) dv = \int h(u,v) du = 1\}$$

$$\text{und für alle } J: [0,1] \times [0,1] \rightarrow \mathbb{R} \text{ mit } |\iint J(u,v) du dv| < \infty \text{ sei}$$

$$\Omega(t, J) = \{H \in \Omega \mid \iint J(u,v) dH(u,v) \geq t\},$$

$$I(t, J) = \inf\{\iint h \log(h) du dv \mid H \in \Omega(t, J)\},$$

$$t^+(J) = \sup\{\iint J(u,v) dH(u,v) \mid H \in \Omega\}.$$

*Lemma 7.1.3:*

$M_{Jm, \varepsilon_1, \varepsilon_2}: D \rightarrow \overline{\mathbb{R}}$  ist  $\tau$ -stetig für jede stetige Funktion  $H$  und  $0 < \varepsilon_1 < \varepsilon_2 < 1$ .

Beweis: Der Beweis ist analog zu Praagman (1988, Lemma 3.5), da  $[\varepsilon_1, \varepsilon_2]$  kompakt ist.

Lemma 7.1.4:

Sei  $J_\rho: [0,1] \times [0,1] \rightarrow \mathbb{R}$ ,  $|\iint J_\rho(u,v) du dv| < \infty$ , für alle  $\rho \in [\varepsilon_1, \varepsilon_2]$ .

(i) Dann ist das  $\inf I(t, J_\rho)$  nicht negativ, monoton wachsend und stetig in  $t$  für  $0 < t < \sup t^+(J_\rho)$ .

(ii) Des weiteren sei  $\tilde{J}_\rho: [0,1] \times [0,1] \rightarrow \mathbb{R}$ ,  $|\iint \tilde{J}_\rho(u,v) du dv| < \infty$  für alle  $\rho \in [\varepsilon_1, \varepsilon_2]$ , falls eine streng monoton wachsende, stetige Abbildung  $q$  von  $[\varepsilon_1, \varepsilon_2]$  auf  $[\varepsilon_1, \varepsilon_2]$  existiert mit

$$\sup_{H \in \Omega} \sup_{\rho \in [\varepsilon_1, \varepsilon_2]} |\iint (\tilde{J}_q(\rho)(u,v) - J_\rho(u,v)) h(u,v) du dv| < \delta,$$

dann gilt für alle  $t$  mit  $0 < t - \delta < t + \delta < \sup t^+(J_\rho)$ ,

$$\inf_{\rho} I(t - \delta, J_\rho) \leq \inf_{\rho} I(t, \tilde{J}_\rho) \leq \inf_{\rho} I(t + \delta, J_\rho) \text{ und}$$

$$\sup_{\rho} t^+(J_\rho) < \sup_{\rho} t^+(\tilde{J}_\rho) + \delta.$$

(iii)  $H_\nu$  sei die Gleichverteilung auf dem Einheitsquadrat. Dann gilt für  $0 < t < \sup t^+(J_\rho)$ ,  $K(\Omega_{M_{J_m \varepsilon_1 \varepsilon_2, t}}, H_\nu) = \inf_{\rho} I(t, J_{m, \rho})$ .

Beweis: Für festes  $\rho$  liefert Lemma 3.3 von Praagman (1988) bzw. Woodworth (1970) obige Aussagen. Damit ist  $I(t, J_\rho)$  nicht negativ, monoton wachsend und stetig in  $t$  für  $0 < t < t^+(J_\rho)$ , also gilt (i).

Des weiteren gilt  $I(t - \delta, J_\rho) \leq I(t, \tilde{J}_q(\rho)) \leq I(t + \delta, J_\rho)$  und  $t^+(J_\rho) < t^+(\tilde{J}_q(\rho)) + \delta$ , und somit gilt (ii).

Nach Lemma 3.3 (iii) von Praagman (1988) gilt

$$K(\Omega_{M_{J_m \varepsilon \varepsilon, t}}, H_\nu) = I(t, J_{m, \varepsilon}).$$

$$\text{mit } \{H \in D \mid M_{J_m \varepsilon_1 \varepsilon_2}(H) \geq t\} = \bigcup_{\rho = \varepsilon_1}^{\varepsilon_2} \{H \in D \mid J_{m, \rho} d\bar{H} \geq t\}$$

folgt, daß

$$\inf \{ \iint h \log h \mid H \in D, M_{J_m \varepsilon_1 \varepsilon_2}(H) \geq t \} = \inf_{\rho} I(t, J_{m, \rho}),$$

und damit (iii).

Satz 7.1.5:

$(X_i, Y_i)$ ,  $i=1, \dots, n$  sei u.i.v. mit der stetigen Verteilungsfunktion

$H_\nu(x, y) = F_X(x) F_Y(y)$ ;  $x, y \in \mathbb{R}$ .  $M_n(\varepsilon_1, \varepsilon_2)$  sei gemäß (7.1.5) und (7.1.6). Für die Funktion  $\varphi: [0, 1] \rightarrow \mathbb{R}$  gelte:  $\int \varphi(u) du = 0$ ,  $\int (\varphi(u))^2 du < \infty$  und  $\int (\varphi_n(u) - \varphi(u))^2 du \rightarrow 0$  für  $n \rightarrow \infty$ ; und für die Funktion  $\gamma: [\varepsilon_1, \varepsilon_2] \rightarrow \mathbb{R}^+$  gelte:  $\gamma_n$  konvergiert gegen  $\gamma$  in der Skorohod-Topologie auf  $[\varepsilon_1, \varepsilon_2]$ , und  $\gamma_n$  ist beschränkt. Für  $\{u_n\}_{n \in \mathbb{N}}$ ,  $u_n \in \mathbb{R}$  gelte  $\lim_{n \rightarrow \infty} u_n = 0$ , dann gilt für alle  $t$  mit  $0 < t < \sup_{\rho} t^+(J_\rho)$ :

$$\lim_{n \rightarrow \infty} -n^{-1} \log(P(n^{-1} M_n(\varepsilon_1, \varepsilon_2) \geq t + u_n)) = \inf_{\varepsilon_1 \leq \rho \leq \varepsilon_2} I(t, J_\rho),$$

wobei  $J_\rho(u, v) = \varphi(u) \gamma(\rho) 1_{[0, \rho]}(v)$ .

Bemerkung: Skorohod-Topologie und Skorohod-Konvergenz werden z.B. von Billingsley (1968, p.111-112) diskutiert.

Beweis: Da die Verteilung von  $M_n(\varepsilon_1, \varepsilon_2)$  unter der Nullhypothese  $H_0$  von  $F_X$  und  $F_Y$  unabhängig ist, können wir o.B.d.A. annehmen, daß  $H_\nu$  die Gleichverteilung auf dem Einheitsquadrat sei. Wähle  $\delta > 0$ , so daß  $0 < t - 2\delta < t + 2\delta < \sup_{\rho} t^+(J_\rho)$ .

Als erstes zeigen wir, daß eine streng monoton wachsende stetige Abbildung  $q_n: [\varepsilon_1, \varepsilon_2] \rightarrow [\varepsilon_1, \varepsilon_2]$  existiert, so daß  $q_n(\rho)$  gleichförmig in  $\rho$  gegen  $\rho$  konvergiert und daß gilt:

$$\sup_{\rho \in [\varepsilon_1, \varepsilon_2]} \left| \int \int (J_{n, q_n(\rho)}(u, v) - J_\rho(u, v)) h(u, v) du dv \right| \rightarrow 0. \quad (7.1.7)$$

$$\begin{aligned} & \sup_{\rho \in [\varepsilon_1, \varepsilon_2]} \left| \int \int (J_{n, q_n(\rho)}(u, v) - J_\rho(u, v)) h(u, v) du dv \right| \\ & \leq \sup_{\rho} \left| \int \int (\gamma_n(q_n(\rho)) \varphi_n(u) 1_{[0, q_n(\rho)]}(v) - \gamma_n(q_n(\rho)) \varphi(u) 1_{[0, q_n(\rho)]}(v)) \right. \\ & \quad \left. + \gamma_n(q_n(\rho)) \varphi(u) 1_{[0, q_n(\rho)]}(v) - \gamma(\rho) \varphi(u) 1_{[0, \rho]}(v) \right) h du dv \left| \right. \\ & \leq \sup_{\rho} \left| \int \int \gamma_n(q_n(\rho)) 1_{[0, q_n(\rho)]}(v) (\varphi_n(u) - \varphi(u)) h du dv \right. \\ & \quad \left. + \int \int (\gamma_n(q_n(\rho)) 1_{[0, q_n(\rho)]}(v) - \gamma(\rho) 1_{[0, \rho]}(v)) \varphi(u) h du dv \right| \\ & \leq \sup_{\rho} \gamma_n(q_n(\rho)) \int |\varphi_n(u) - \varphi(u)| du \end{aligned}$$

$$+ \sup_{\rho} \left( \int (\gamma_n(q_n(\rho))1_{[0,q_n(\rho)]}(v) - \gamma(\rho)1_{[0,\rho]}(v))^2 dv \right)^{\frac{1}{2}} \\ \times \left( \int (\varphi(u))^2 du \right)^{\frac{1}{2}},$$

da  $Kov(\gamma, \varphi) \leq (\text{Var}(\gamma) \text{Var}(\varphi))^{\frac{1}{2}}$ ; wegen  $\int (\varphi_n(u) - \varphi(u))^2 du \rightarrow 0$ , der Skorohod-Konvergenz von  $\gamma_n \rightarrow \gamma$  auf  $[\varepsilon_1, \varepsilon_2]$  und der Unabhängigkeit obiger Abschätzung von  $H \in \Omega$  existiert ein  $m \in \mathbb{N}$ , so daß für  $n \geq m$  gilt:

$$\sup_{H \in \Omega} \sup_{\rho} \left| \int \int (J_{n,q_n(\rho)}(u,v) - J_{\rho}(u,v)) h(u,v) du dv \right| < \delta/2 \quad (7.1.8)$$

und

$$\sup_{H \in \Omega} \sup_{\rho} \left| \int \int (J_{m,q_m(\rho)}(u,v) - J_{n,q_n(\rho)}(u,v)) h(u,v) du dv \right| < \delta/2. \quad (7.1.9)$$

Um diese Ungleichungen auf die maximal selektierte Rangstatistik in Abhängigkeit der empirischen Verteilung anwenden zu können, definieren wir als empirische Lebesgue-Dichte  $\bar{H}_n(u,v) = n$  für  $(u,v) \in ((R_i - 1)/n, R_i/n] \times ((Q_i - 1)/n, Q_i/n]$  und 0 sonst, wobei  $Q_1, \dots, Q_n$  die Ränge von  $X_1, \dots, X_n$  bezeichnen. Nun gilt für großes  $n \in \mathbb{N}$  m.W. 1:

$$|M_n(\varepsilon_1, \varepsilon_2) - M_{Jm\varepsilon_1\varepsilon_2}(H_n)| < \delta. \quad (7.1.10)$$

Denn

$$\begin{aligned} & |M_n(\varepsilon_1, \varepsilon_2) - M_{Jm\varepsilon_1\varepsilon_2}(H_n)| \\ &= \left| \sup_{\rho} \int \int J_{n,\rho}(u,v) d\bar{H}_n(u,v) - \sup_{\rho} \int \int J_{m,\rho}(u,v) d\bar{H}_n(u,v) \right| \\ &\leq \sup_{\rho} \left| \int \int (J_{n,q_n(\rho)}(u,v) - J_{m,q_m(\rho)}(u,v)) d\bar{H}_n(u,v) \right| \\ &\leq \sup_{\rho} \sum_{i=1}^n |J_{n,q_n(\rho)}(R_i/n, Q_i/n) - J_{m,q_m(\rho)}(R_i/n, Q_i/n)| / n \\ &\leq \sup_{\rho} \int \int |J_{n,q_n(\rho)}(u,v) - J_{m,q_m(\rho)}(u,v)| \bar{H}_n(u,v) du dv \\ &\quad + \frac{2m}{n} \sup_{\rho} \max \left| J_{m,\rho}\left(\frac{k}{m}, \frac{l}{m}\right) - J_{m,\rho}\left(\frac{k'}{m}, \frac{l'}{m}\right) \right|, \end{aligned}$$

wobei das Maximum bzgl.  $1 \leq k, k', l, l' \leq m$ ,  $|k - k'|, |l - l'| \leq 1$  ist. Für großes  $n \in \mathbb{N}$  ist der zweite Summand kleiner  $\delta/2$ , und der erste Summand ist gemäß (7.1.9) kleiner  $\delta/2$ , also gilt (7.1.10).

$M_{J_m, \varepsilon_1, \varepsilon_2}$  ist  $\tau$ -stetig (Lemma 7.1.3), d.h. um Satz 7.1.2 auf  $M_{J_m, \varepsilon_1, \varepsilon_2}$  anwenden zu können, bleibt zu zeigen, daß die Abbildung  $t \rightarrow \inf \{ \int \int h \log h \mid H \in D, M_{J_m, \varepsilon_1, \varepsilon_2}(H) \geq t \}$  rechtsseitig stetig ist. Mit Lemma 7.1.4 (iii) folgt die benötigte Stetigkeit aus Lemma 7.1.4 (i). Deshalb gilt gemäß Satz 7.1.2

$$\begin{aligned} \inf_{\rho} I(t - \delta, J_{m, \rho}) &\leq \liminf_{n \rightarrow \infty} \inf - n^{-1} \log (P(M_n(\varepsilon_1, \varepsilon_2) \geq n(t + u_n))) \\ &\leq \limsup_{n \rightarrow \infty} - n^{-1} \log (P(M_n(\varepsilon_1, \varepsilon_2) \geq n(t + u_n))) \\ &\leq \inf_{\rho} I(t + \delta, J_{m, \rho}). \end{aligned}$$

Des weiteren gilt mit (7.1.8) und mit Lemma 7.1.4 (ii),

$$\inf_{\rho} I(t - 2\delta, J_{\rho}) \leq \inf_{\rho} I(t - \delta, J_{m, \rho})$$

und

$$\inf_{\rho} I(t + \delta, J_{m, \rho}) \leq \inf_{\rho} I(t + 2\delta, J_{\rho}),$$

und aus der Stetigkeit von  $\inf_{\rho} I(t, J_{\rho})$  (Lemma 7.1.4 (i)) folgt die Aussage des Satzes.

Zur Vereinfachung der Aussage von Satz 7.1.5 können wir folgendes Korollar formulieren.

*Korollar 7.1.6:*

$$\inf_{\rho} I(t, J_{\rho}) = \inf \{ \int \int h \log h \mid H \in \Omega^*(t, J_{\rho}), \rho \in [\varepsilon_1, \varepsilon_2] \},$$

wobei  $\Omega^*(t, J_{\rho}) = \{ H \in \Omega \mid h(u, v) = f(u) 1_{[0, \rho]}(v) + g(u) 1_{(\rho, 1]}(v),$   
und  $\int \int J_{\rho} h \, du \, dv \geq t \}$ .

Beweis: Wähle ein beliebiges  $H \in \Omega$  mit  $\int \int J_{\rho}(u, v) \, dH(u, v) \geq t$  für ein  $\rho \in [\varepsilon_1, \varepsilon_2]$ . Des weiteren sei  $f(u) = \rho^{-1} \int_0^{\rho} h(u, v) \, dv,$

$g(u) = (1 - \rho)^{-1} \int_0^1 h(u, v) dv$  und  $h_\rho(u, v) = f(u)$  für  $0 \leq v \leq \rho$  und  $h_\rho(u, v) = g(u)$  für  $\rho < v \leq 1$ . Nun gilt nach Praagman (1988, p.208)

$$\iint J_\rho(u, v) h_\rho(u, v) du dv = \iint J_\rho(u, v) h(u, v) du dv \quad (7.1.11)$$

und

$$\iint h(u, v) \log(h(u, v)) du dv \quad (7.1.12)$$

$$\geq \iint h_\rho(u, v) \log(h_\rho(u, v)) du dv .$$

Also gilt:

$$\inf_\rho I(t, J_\rho)$$

$$= \inf_\rho \inf \{ \iint h \log h \mid H \in \Omega, \iint J_\rho(u, v) dH(u, v) \geq t \}$$

(7.1.11)

$$= \inf_\rho \inf \{ \iint h \log h \mid H \in \Omega, \iint J_\rho(u, v) h_\rho(u, v) du dv \geq t \}$$

(7.1.12)

$$= \inf_\rho \inf \{ \iint h \log h \mid H \in \Omega^*(t, J_\rho) \}$$

$$= \inf \{ \iint h \log h \mid H \in \Omega^*(t, J_\rho), \rho \in [\varepsilon_1, \varepsilon_2] \} .$$

Für die Wahrscheinlichkeit großer Abweichungen der Statistik vom Spearmanotyp gilt:

*Satz 7.1.7 (Praagman 1988, Theorem 3.4):*

$(X_i, Y_i), i=1, \dots, n$  sei u.i.v. mit der stetigen Verteilungsfunktion  $H_v(x, y) = F_X(x) F_Y(y)$ ;  $x, y \in \mathbb{R}$ .  $S_n$  sei gemäß (7.1.3) und (7.1.4), und eine Funktion  $J : [0, 1] \times [0, 1] \rightarrow \mathbb{R}$  existiert, so daß

$$\sup_{H \in \Omega} \left| \iint (J_n(u, v) - J(u, v)) h(u, v) du dv \right| \rightarrow 0 . \quad (7.1.13)$$

Für  $\{u_n\}_{n \in \mathbb{N}}, u_n \in \mathbb{R}$  gelte  $\lim_{n \rightarrow \infty} u_n = 0$ , dann gilt für alle  $t$  mit  $\iint J(u, v) du dv < t < t^+(J)$ :

$$\lim_{n \rightarrow \infty} -n^{-1} \log(P(n^{-1} S_n \geq t + u_n)) = I(t, J) ,$$

wobei  $J(u, v) = \varphi(u) \psi(v)$ .

Bemerkung: Eine hinreichende Bedingung für (7.1.13) ist, daß  $\varphi_n \rightarrow \varphi$  und  $\psi_n \rightarrow \psi$  auf  $[0, 1]$  im quadratischen Mittel konvergieren (Woodsworth 1970, p.257).

Als eine andere hinreichende Bedingung führt Praagman (1988, p.204) auf, daß eine der Scoreerzeugenden-Funktionen beschränkt sei und die andere in  $L_1$  konvergiert.  $\varphi$  sei beschränkt und  $\|\varphi - \varphi_n\| \rightarrow 0$  für  $n \rightarrow \infty$ , des weiteren konvergiere  $\psi_n \rightarrow \psi$  bzgl.  $L_1$ , also  $\int |\psi_n(v) - \psi(v)| dv \rightarrow 0$ , dann ist (7.1.13) erfüllt. Denn

$$\begin{aligned} & \left| \int \int (J_n(u, v) - J(u, v)) h(u, v) dudv \right| \\ & \leq \left| \int \int (\varphi_n(u) \psi_n(v) - \varphi(u) \psi(v)) h(u, v) dudv \right| \\ & \leq \left| \int \int (\varphi_n(u) \psi_n(v) - \varphi_n(u) \psi(v) \right. \\ & \quad \left. + \varphi_n(u) \psi(v) - \varphi(u) \psi(v)) h(u, v) dudv \right| \\ & \leq \left| \int \int \varphi_n(u) (\psi_n(v) - \psi(v)) h(u, v) dudv \right. \\ & \quad \left. + \int \int (\varphi_n(u) - \varphi(u)) \psi(v) h(u, v) dudv \right| \\ & \leq C_\varphi \int |\psi_n(v) - \psi(v)| dv + C_{\varphi_n} \int |\psi(v)| dv, \end{aligned}$$

wobei  $C_\varphi = \sup_u |\varphi_n(u)|$  und  $C_{\varphi_n} = \|\varphi_n - \varphi\|$ . Da  $C_\varphi$  und  $\int |\psi(v)| dv$  konstant sind und  $\int |\psi_n(v) - \psi(v)| dv$  und  $C_{\varphi_n}$  gegen 0 konvergieren, ist (7.1.13) erfüllt.

### Konsistenz

Um den Bahadur-Slope anhand Satz 7.1.1 berechnen zu können, benötigen wir noch die jeweiligen Konsistenzaussagen. Für die Rangstatistik vom Spearman-Typ gilt für beliebige bivariate Verteilungen:

*Satz 7.1.8 (Praagman 1988, Theorem 4.1):*

$H$  sei die Verteilungsfunktion von  $(X_i, Y_i)$  und  $S_n$  sei gemäß (7.1.3) und (7.1.4). Falls  $J_n, J: [0, 1] \times [0, 1] \rightarrow \mathbb{R}$  (7.1.13) erfüllen, dann gilt:

$$\lim_{n \rightarrow \infty} n^{-1} S_n = \iint J(u, v) d\bar{H}(u, v) \quad P_H\text{-f.s.}$$

Für das durch  $H_0$  und  $H_a$  gegebene Prüfpunktmodell erhält man mit  $\bar{F}_{Y\mu}(u) = F_Y(H_Y^{-1}(u))$  und  $\bar{G}_{Y\mu}(u) = G_Y(H_Y^{-1}(u))$ :

*Korollar 7.1.9 (Praagman 1988, Corollary 4.2):*

Für eine durch  $F_X$ ,  $F_Y$ ,  $G_Y$  und  $\mu$  gegebene stetige Verteilungsfunktion  $H(u, v)$  gilt mit  $J(u, v) = \varphi(u) \psi(v)$ :

$$\begin{aligned} \lim_{n \rightarrow \infty} n^{-1} S_n &= \int_0^{F_X(\mu)} \psi(v) dv \int_0^1 \varphi(u) d\bar{F}_{Y\mu}(u) \\ &+ \int_{F_X(\mu)}^1 \psi(v) dv \int_0^1 \varphi(u) d\bar{G}_{Y\mu}(u) \quad P_H\text{-f.s.} \end{aligned}$$

Für die maximal selektierte Rangstatistik gilt entsprechend für beliebige bivariate Verteilungen:

*Satz 7.1.10:*

$H$  sei die Verteilungsfunktion von  $(X_i, Y_i)$  und  $M_n(\varepsilon_1, \varepsilon_2)$  sei gemäß (7.1.5) und (7.1.6). Falls  $J_{n, \rho}, J_\rho: [0, 1] \times [0, 1] \rightarrow \mathbb{R}$  (7.1.7) erfüllen, dann gilt:

$$\lim_{n \rightarrow \infty} n^{-1} M_n(\varepsilon_1, \varepsilon_2) = \sup_{\rho \in [\varepsilon_1, \varepsilon_2]} \iint J_\rho(u, v) d\bar{H}(u, v) \quad P_H\text{-f.s.}$$

Beweis: Da  $|\iint J_{n, \rho_n(\rho)}(u, v) d\bar{H}(u, v) - \iint J_\rho(u, v) d\bar{H}(u, v)| \rightarrow 0$  gleichmäßig in  $\rho \in [\varepsilon_1, \varepsilon_2]$ , ansonsten analog zu Praagman (1988, Beweis Theorem 4.3).

Unter dem Prüfpunktmodell definieren wir

$$\begin{aligned} b_\mu(\rho) &= \rho \int \varphi(u) d\bar{F}_{Y\mu}(u) \quad \text{für } \rho \in [0, F_X(\mu)] \text{ und} \\ b_\mu(\rho) &= F_X(\mu) \int \varphi(u) d\bar{F}_{Y\mu}(u) + (\rho - F_X(\mu)) \int \varphi(u) d\bar{G}_{Y\mu}(u) \quad \text{für} \\ &\rho \in (F_X(\mu), 1]. \end{aligned}$$

Dann gilt:

*Korollar 7.1.11:*

Für eine durch  $F_X$ ,  $F_Y$ ,  $G_Y$  und  $\mu$  gegebene stetige Verteilungsfunktion  $H(u, v)$  gilt:

$$\lim_{n \rightarrow \infty} n^{-1} M_n(\varepsilon_1, \varepsilon_2) = \sup_{\rho \in [\varepsilon_1, \varepsilon_2]} \gamma(\rho) b_{\mu}(\rho) \quad P_H\text{-f.s.}$$

### Exakte Bahadur-Slope

Als nächstes zeigen wir, daß bei einer bestimmten Wahl der Funktion  $\gamma$  in Abhängigkeit der Funktion  $\psi$  der Bahadur-Slope der maximal selektierten Rangstatistik größer oder gleich als derjenige der Rangstatistik vom Spearman-Typ ist, falls für den Prüfpunktparameter gilt  $0 < \varepsilon_1 \leq F_X(\mu) \leq \varepsilon_2 < 1$ . Die Abhängigkeit der Scoreerzeugenden-Funktionen  $\gamma$  und  $\psi$  ist definiert durch:

$$\Gamma(\rho) = \int_0^{\rho} \psi(s) ds \quad \text{und} \quad \gamma : [\varepsilon_1, \varepsilon_2] \rightarrow \mathbb{R} \quad (7.1.14)$$

mit  $\gamma(\rho) = (\Gamma(\rho) - \rho \Gamma(1)) / (\rho(1 - \rho))$ .

$\Gamma_n$  und  $\gamma_n$  werden analog in Abhängigkeit von  $\psi_n$  definiert.

#### Satz 7.1.12:

$(X_i, Y_i)$ ,  $i=1,2,\dots$  sei gemäß der durch  $F_X$ ,  $F_Y$ ,  $G_Y$  und  $\mu$  gegebenen Verteilungsfunktion  $H(u, v)$  verteilt, und es gelte  $F_Y < G_Y$   $0 < \varepsilon_1 \leq F_X(\mu) \leq \varepsilon_2 < 1$ .  $\{S_n\}$  sei eine durch (7.1.3) und (7.1.4) gegebene Folge von Teststatistiken, wobei  $\int \varphi_n(u) du = 0$  und  $\psi_n$  monoton fallend sei.  $\{M_n(\varepsilon_1, \varepsilon_2)\}$  sei eine durch (7.1.5) und (7.1.6) definierte Folge von Teststatistiken, wobei die Score-Funktionen  $\varphi_n$  für  $S_n$  und  $M_n$  identisch seien und die Gewichtsfunktion  $\gamma_n$  gemäß (7.1.14) definiert sei. Des weiteren sei  $\varphi : (0, 1) \rightarrow \mathbb{R}$  monoton fallend und nicht konstant. Es gelte:  $\int (\varphi_n(u) - \varphi(u))^2 du \rightarrow 0$  und  $\int (\psi_n(v) - \psi(v))^2 dv \rightarrow 0$ , wobei  $\psi : [0, 1] \rightarrow \mathbb{R}^+$ . Dann ist der exakte Bahadur-Slope von  $\{S_n\}$  kleiner oder gleich als der exakte Bahadur-Slope von  $\{M_n(\varepsilon_1, \varepsilon_2)\}$ .

Zur Vereinfachung des Beweises von Satz 7.1.12 formulieren wir:

#### Lemma 7.1.13:

Die Voraussetzungen von Satz 7.1.12 seien erfüllt, dann gilt:

$$\lim_{n \rightarrow \infty} n^{-1} S_n = \lim_{n \rightarrow \infty} n^{-1} M_n(\varepsilon_1, \varepsilon_2) = \gamma(F_X(\mu)) b_{\mu}(F_X(\mu)) \quad P_H\text{-f.s.}$$

Beweis: Praagman (1988, Lemma 5.2).

Beweis Satz 7.1.12: Nach Lemma 7.1.13 sind die f.s. Grenzwerte identisch, folglich unterscheiden sich die Bahadur-Slopes von  $\{M_n(\varepsilon_1, \varepsilon_2)\}$  und  $\{S_n\}$  nur, wenn sich die Wahrscheinlichkeiten großer Abweichungen unterscheiden. Für  $h(u, v) = f(u) 1_{[0, \rho]} + g(u) 1_{(\rho, 1]}$  gemäß Korollar 7.1.6 gilt (Praagman 1988, Beweis von Theorem 5.1)

$$\begin{aligned} \int \int J_\rho(u, v) h(u, v) du dv &= \int \int \varphi(u) \psi(u) h(u, v) du dv \\ &= \int \int J(u, v) h(u, v) du dv . \end{aligned}$$

Also gilt  $\{\Omega^*(t, J_\rho), \rho \in [\varepsilon_1, \varepsilon_2]\} \subset \Omega(t, J)$ , und damit ist die Wahrscheinlichkeit großer Abweichungen von  $S_n$  kleiner oder gleich als die von  $M_n(\varepsilon_1, \varepsilon_2)$ .

*Lemma 7.1.14:*

$\varphi$  sei beschränkt,  $\|\varphi_n - \varphi\| \rightarrow 0$  und für  $\psi_n, \psi$  fordern wir nur die Konvergenz bzgl.  $L_1$ , ansonsten gelten die Voraussetzungen von Satz 7.1.12. Dann ist der exakte Bahadur-Slope von  $\{S_n\}$  kleiner oder gleich als der exakte Bahadur-Slope von  $\{M_n(\varepsilon_1, \varepsilon_2)\}$ .

Beweis: Die Voraussetzungen von Satz 7.1.7 sind erfüllt (siehe Bemerkung zu Satz 7.1.7).  $\gamma(\rho)$  ist nach (7.1.14) gleichmäßig stetig, und  $\psi_n$  konvergiert in  $L_1$  gegen  $\psi$ . Damit konvergiert  $\gamma_n$  gegen  $\gamma$  in der Skorohod-Topologie auf  $[\varepsilon_1, \varepsilon_2]$  (z.B. Billingsley 1968, p.112), und  $\gamma_n$  ist beschränkt. Also sind auch die Voraussetzungen von Satz 7.1.5 erfüllt, und ansonsten ist der Beweis analog zu Satz 7.1.12 bzw. Praagman (1988).

*Korollar 7.1.15:*

Falls gilt  $0 < \varepsilon_1 \leq \varepsilon'_1 \leq F_X(\mu) \leq \varepsilon'_2 \leq \varepsilon_2 < 1$  und die Voraussetzungen von Satz 7.1.12 bzw. Lemma 7.1.14 erfüllt sind, dann ist der exakte Bahadur-Slope von  $M_n(\varepsilon'_1, \varepsilon'_2)$  größer oder gleich als derjenige von  $M_n(\varepsilon_1, \varepsilon_2)$ .

Beweis:  $\inf_{\rho \in [\varepsilon_1, \varepsilon_2]} I(t, J_\rho) \leq \inf_{\rho \in [\varepsilon'_1, \varepsilon'_2]} I(t, J_\rho)$ .

Falls  $M_n(\varepsilon_1, \varepsilon_2)$  der standardisierte einseitige Zweistichproben-Wilcoxon-Test ist, dann ergeben sich folgende Scoreerzeugende-Funktionen:

$$\begin{aligned} \varphi(u) &= \sqrt{12}(0.5 - u), & \gamma(\rho) &= (\rho(1 - \rho))^{-\frac{1}{2}} & \text{und} \\ \psi(v) &= 0.5(1 - 2v)(v(1 - v))^{-\frac{1}{2}}. \end{aligned}$$

$\varphi$  und  $\psi$  erfüllen die Voraussetzungen von Lemma 7.1.14.

Damit konnten wir zeigen, daß maximal selektierte Rangstatistiken unter schwachen Regularitätsbedingungen zu konsistenten Tests führen (Korollar 7.1.11). Des weiteren kann man die jeweiligen Scorefunktionen so wählen, daß die exakte Bahadur-Effizienz der maximal selektierten Rangstatistik gegenüber der entsprechend gewählten Statistik vom Spearman-Typ größer oder gleich eins ist (Satz 7.1.12, Lemma 7.1.14).

## 7.2. Approximative Bahadur-Effizienz

Die Berechnung der Wahrscheinlichkeit großer Abweichungen der exakten Verteilung ist oft aufwendig; deshalb ist es naheliegend, die approximative Bahadur-Effizienz zu betrachten, da hierbei nur eine Aussage über die Wahrscheinlichkeit großer Abweichungen der asymptotischen Verteilung benötigt wird. Des weiteren spricht für die approximative Bahadur-Effizienz, daß man oft eine Testvorschrift bzgl. der asymptotischen Verteilung verwendet (vgl. Rothe 1980, Paragraph 4). Wir berechnen den *approximativen Bahadur-Slope* anhand folgenden Satzes:

*Satz 7.2.1. (Bahadur 1960):*

$\{T_n\}_{n \in \mathbb{N}}$  sei eine Folge von Teststatistiken bzgl. des Testproblems  $H_0: \vartheta \in \Theta_0 = \{\vartheta_0\}$  vs.  $H_a: \vartheta \in \Theta - \Theta_0 = \Theta_1$ . Des weiteren gelte:

Es existiert eine Zufallsvariable  $Z$  mit stetiger Verteilungsfunktion  $F_Z$  derart, daß

$$T_n \stackrel{d}{\rightarrow} Z \quad \text{für } n \rightarrow \infty \quad \text{und } \vartheta = \vartheta_0, \quad (7.2.1)$$

$$\lim_{n \rightarrow \infty} n^{-1/2} T_n = b(\vartheta) \quad P_{\vartheta}\text{-f.s., } \vartheta \in \Theta_1, \quad (7.2.2)$$

$$-\log(1 - F_Z(t)) = a t^2 (1 + o(1)) \quad \text{für } t \rightarrow \infty. \quad (7.2.3)$$

Dann ist  $a (b(\vartheta))^2$  der approximative Bahadur-Slope von  $\{T_n\}$  in  $\vartheta$ .

Man kann zeigen, daß der Quotient der approximativen Bahadur-Slopes zweier Folgen von Teststatistiken gemäß Satz 7.2.1 die *approximative Bahadur-Effizienz* ist (z.B. Bahadur 1960; Rothe 1980). Somit benötigen wir zur Berechnung die asymptotische Verteilung (7.2.1), die Konstante  $a$  aus der Aussage über die Wahrscheinlichkeit großer Abweichungen der asymptotischen Verteilung (7.2.3) und die Konsistenzaussage (7.2.2). Falls  $Z$  standardnormal verteilt oder zentralchiquadrat verteilt ist, ist (7.2.3) für  $a = 1/2$  erfüllt (Bahadur 1960). Für die asymptotische Verteilung der maximal selektierten Rangstatistik verwenden wir ein allgemeines Resultat von Marcus und Shepp (1972).

*Lemma 7.2.2. (Marcus und Shepp 1972):*

$\{Z(s), 0 \leq s \leq 1\}$  sei ein Gaußprozeß mit stetigen Pfaden, dann gilt:

$$\lim_{t \rightarrow \infty} 1/t^2 \log (P(\sup_{0 \leq s \leq 1} |Z(s)| \geq t)) = -(2\sigma^2)^{-1}, \quad (7.2.4)$$

wobei  $\sigma^2 = \sup_{0 \leq s \leq 1} \text{Var}(Z(s))$ .

Folglich gilt für das Supremum des Absolutbetrages der standardisierten Brownschen Brücke aus Satz 3.2.1 ebenfalls  $\alpha = \frac{1}{2}$ . Die approximative Bahadur-Effizienz ist damit unter den Voraussetzungen von Satz 7.1.12 für  $\{S_n\}$  und  $\{M_n(\varepsilon_1, \varepsilon_2)\}$  gleich eins. Des weiteren gilt für  $\{M_n(\varepsilon_1, \varepsilon_2)\}$  und  $\{M_n(\varepsilon'_1, \varepsilon'_2)\}$  analog zu Korollar 7.1.14, daß die approximative Bahadur-Effizienz eins ist; d.h. die Größe des Selektionsintervalls hat keinen Einfluß auf die approximativen Bahadur-Slopes, sofern der Prüfpunkt  $\mu$  im Intervall liegt. Die maximal selektierte Rangstatistik  $\{M_n(\varepsilon_1, \varepsilon_2)\}$  hat damit auch die approximative Bahadur-Effizienz eins zu der Zweistichproben-Rangstatistik bei bekanntem Prüfpunkt  $\mu$ ,  $\{M_n(F_{nX}(\mu), F_{nX}(\mu))\}$ , falls  $\varepsilon_1 \leq F_X(\mu) \leq \varepsilon_2$  gilt. Um den Einfluß der Größe des Selektionsintervalls aufzuzeigen, ist das Konzept der approximativen Bahadur-Effizienz somit nicht geeignet.

#### Gleichverteilung

Im folgenden berechnen wir die approximativen Bahadur-Slopes für  $0 < \varepsilon_1 \leq \mu \leq \varepsilon_2 < 1$ ,  $F_X(x) = x \mathbb{1}_{[0,1]}(x)$ ,  $F_Y(y) = y \mathbb{1}_{[0,1]}(y)$  und  $G_Y(y) = (y - \vartheta) \mathbb{1}_{[\vartheta, 1 + \vartheta]}(y)$ . Als erstes betrachten wir den maximal selektierten Wilcoxon-Test, also  $\varphi(u) = \sqrt{12} (0.5 - u)$  und  $\gamma(\rho) = (\rho (1 - \rho))^{-\frac{1}{2}}$ . Nach Lemma 7.1.13 ist  $\gamma(\mu) b_\mu(\mu)$  der  $P_H$ -f.s. Grenzwert (7.2.2). Wobei

$$\begin{aligned} b_\mu(\mu) &= \mu \int \varphi(u) d\overline{F}_{Y\mu}(u) \\ &= \mu \int \varphi(u) dF_Y(H_Y^{-1}(u)) \\ &= \mu \int \varphi(H_Y(u)) dF_Y(u), \end{aligned}$$

mit  $H_Y(y) = \mu y \mathbb{1}_{[0,1]}(y) + (1-\mu)(y-\vartheta) \mathbb{1}_{[\vartheta, 1+\vartheta]}(y)$

$$+ \mu 1_{(1, 1+\vartheta]}(y) + 1_{(1+\vartheta, -\infty)}(y).$$

Somit erhalten wir als approximativen Bahadur-Slope:

$$\begin{aligned} & 6\mu^2 (0.5 - 0.5\mu - (1 - \mu) (0.5 - \vartheta + 0.5\vartheta^2))^2 (\gamma(\mu))^2 \\ & = 6\mu (1 - \mu) \vartheta^2 (1 - 0.5\vartheta)^2, \end{aligned}$$

für  $0 < \vartheta \leq 1$ .

Die Scoreerzeugende-Funktion der maximal selektierten Median-Statistik ist:  $\varphi(u) = 1_{\{u \leq 0.5\}} - 1_{\{u > 0.5\}}$ . Wir erhalten als approximativen Bahadur-Slope:

$$\begin{aligned} & (\gamma(\mu))^2 0.5\mu^2 (\min\{0.5/\mu, \vartheta\} \\ & + 1_{\{\mu\vartheta < 0.5\}} (\min\{1, 0.5 + \vartheta - \mu\vartheta\} - \vartheta) \\ & - 1_{\{0.5 < \mu\vartheta\}} (\vartheta - 0.5/\mu) \\ & - 1_{\{\max\{\vartheta, 0.5 + \vartheta - \mu\vartheta\} < 1\}} (1 - \max\{\vartheta, 0.5 + \vartheta - \mu\vartheta\}))^2 \end{aligned}$$

für  $0 < \vartheta \leq 1$ ;  $0 < \mu < 1$ . Der approximative Bahadur-Slope der maximal selektierten  $t$ -Statistik ist unter Verwendung einer zu Lemma 7.1.13 analogen Argumentation bzw. Praagman (1989b; Theorem 10.18):

$$6\vartheta^2 \mu (1 - \mu), \text{ für } \vartheta > 0.$$

Die Scoreerzeugenden Funktionen des standardisierten Spearman-Rangtests sind  $\varphi(u) = \sqrt{12} (0.5 - u)$  und  $\psi(u) = \sqrt{12} (0.5 - u)$ . Nach Korollar 7.1.9 gilt  $P_H$ -f.s.:

$$\begin{aligned} \lim_{n \rightarrow \infty} n^{-1} S_n &= \int_0^\mu \psi(v) dv \int_0^1 \varphi(u) d\bar{F}_{Y\mu}(u) \\ &+ \int_\mu^1 \psi(v) dv \int_0^1 \varphi(u) d\bar{G}_{Y\mu}(u) \\ &= 6\mu (1 - \mu) \vartheta (1 - \vartheta/2), \end{aligned}$$

also erhalten wir als approximativen Bahadur-Slope:

$$18 \mu^2 (1 - \mu)^2 \vartheta^2 (1 - \vartheta/2)^2.$$

Die approximativen Bahadur-Effizienzen werden in den Abb.7.2.1-2 dargestellt. Die Abb.7.2.1 zeigt die Abhängigkeit von  $\vartheta$  und die Abb.7.2.2 von  $\mu$ . Die Knickpunkte der Bahadur-Effizienz bzgl. des maximal selektierten Median-Tests resultieren aus dem nicht stetigen Charakter der Mischung der beiden Gleichverteilungen und der nicht streng monotonen Scoreerzeugenden-Funktion des Median-Tests. Denn beispielsweise für  $\mu \neq 0.5$  und  $\vartheta = 1.0$  werden die beiden f.s. getrennten Gleichverteilungen durch die Scoreerzeugende-Funktion des Median-Tests nicht vollständig getrennt, da die Scoreerzeugende-Funktion nur zwei Werte annimmt. Die approximative Bahadur-Effizienz vom maximal selektierten Wilcoxon-Test zum maximal selektierten  $t$ -Test ist unabhängig von  $\mu$ , und sie ist für  $\vartheta \rightarrow 0$  gleich eins. Dieses Resultat entspricht der Pitman-Effizienz der jeweiligen Zweistichproben-Tests (Randles und Wolfe 1979, pp. 163-171). Der maximal selektierte Wilcoxon-Test ist dem Spearman-Test im Sinne der approximativen Bahadur-Effizienz in dem betrachteten Modell überlegen, und die approximative Bahadur-Effizienz ist unabhängig von  $\vartheta$ . Für kleine  $\vartheta$  ist der maximal selektierte Wilcoxon-Test dem maximal selektierten Median-Test im Sinne der approximativen Bahadur-Effizienz ebenfalls überlegen.

#### *Doppelexponentialverteilung*

Die Wahrscheinlichkeitsdichten der standardisierten Doppelexponentialverteilung sind:

$$f_Y(y) = 2^{-\frac{1}{2}} \exp(-2^{\frac{1}{2}} |y|) \text{ bzw. } g_Y(y) = 2^{-\frac{1}{2}} \exp(-2^{\frac{1}{2}} |y - \vartheta|),$$

wobei  $\vartheta > 0$ . Des weiteren gilt für die Verteilungsfunktion der Randverteilung von  $Y$ :

$$\begin{aligned} H_Y(y) = & \mu (1_{(-\infty, 0]}(y) \frac{1}{2} \exp(2^{\frac{1}{2}} y) + 1_{(0, \infty)}(y) (1 - \frac{1}{2} \exp(-2^{\frac{1}{2}} y))) \\ & + (1 - \mu) (1_{(-\infty, \vartheta]}(y) \frac{1}{2} \exp(2^{\frac{1}{2}} (y - \vartheta)) \\ & + 1_{(\vartheta, \infty)}(y) (1 - \frac{1}{2} \exp(-2^{\frac{1}{2}} (y - \vartheta)))) . \end{aligned}$$

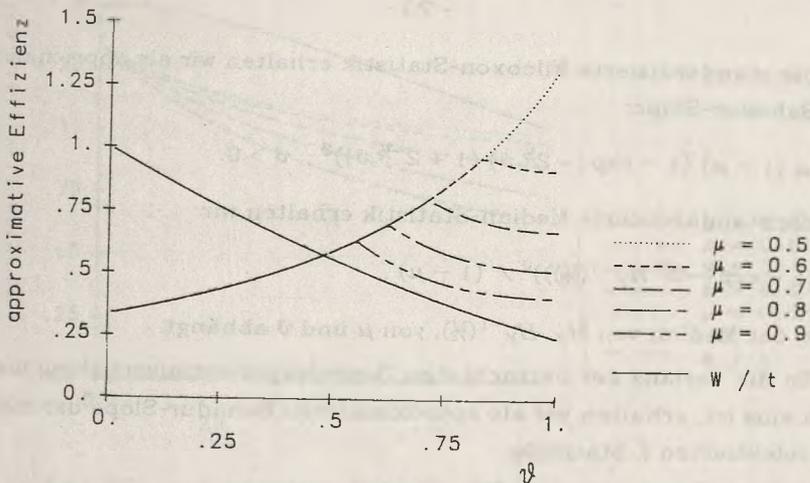


Abb. 7.2.1.: Die approximative Bahadur-Effizienz bei der Gleichverteilung in Abhängigkeit des Lokationsunterschiedes  $\vartheta$ : maximal selektierter Median-Test zum maximal selektierten Wilcoxon-Test bei verschiedenen Prüfpunkten  $\mu$  und maximal selektierter Wilcoxon-Test zum maximal selektierten  $t$ -Test ( $W/t$ ).

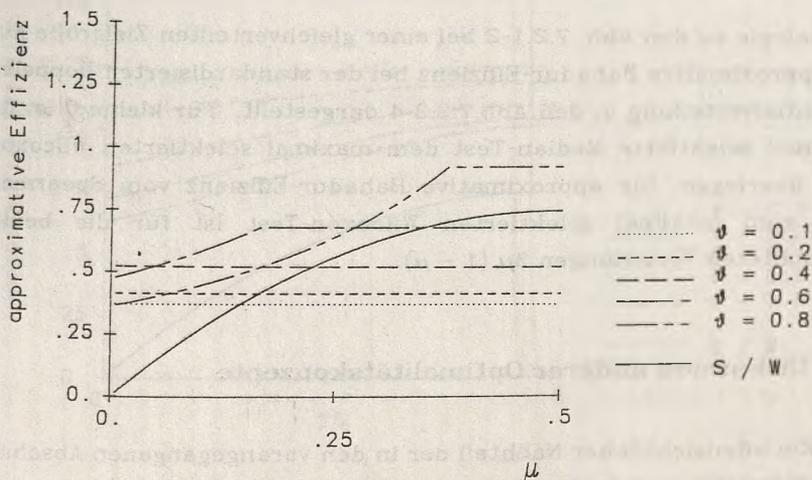


Abb. 7.2.2.: Die approximative Bahadur-Effizienz bei der Gleichverteilung in Abhängigkeit des Prüfpunkts  $\mu$ : maximal selektierter Median-Test zum maximal selektierten Wilcoxon-Test bei verschiedenen Lokationsunterschieden  $\vartheta$  und Spearman-Test zum maximal selektierten Wilcoxon-Test ( $S/W$ ).

Für die standardisierte Wilcoxon-Statistik erhalten wir als approximativen Bahadur-Slope:

$$3/2 \mu (1 - \mu) (1 - \exp(-2^{1/2} \vartheta) (1 + 2^{-1/2} \vartheta))^2, \vartheta > 0.$$

Für die standardisierte Median-Statistik erhalten wir:

$$1/2 \mu (1 - \exp(-2^{1/2} H_Y^{-1}(1/2)))^2 / (1 - \mu),$$

wobei der Median von  $H_Y$ ,  $H_Y^{-1}(1/2)$ , von  $\mu$  und  $\vartheta$  abhängt.

Da die Varianz der betrachteten Doppelexponentialverteilung identisch eins ist, erhalten wir als approximativen Bahadur-Slope der maximal selektierten  $t$ -Statistik:

$$1/2 \vartheta^2 \mu (1 - \mu), \text{ für } \vartheta > 0.$$

Für den standardisierten Spearman-Rangtest erhält man:

$$9/2 \mu^2 (1 - \mu)^2 (1 - \exp(-2^{1/2} \vartheta) (1 + 2^{-1/2} \vartheta))^2, \vartheta > 0.$$

In Analogie zu den Abb. 7.2.1-2 bei einer gleichverteilten Zielgröße wird die approximative Bahadur-Effizienz bei der standardisierten Doppelexponentialverteilung in den Abb. 7.2.3-4 dargestellt. Für kleine  $\vartheta$  ist der maximal selektierte Median-Test dem maximal selektierten Wilcoxon-Test überlegen. Die approximative Bahadur-Effizienz vom Spearman-Test zum maximal selektierten Wilcoxon-Test ist für die beiden betrachteten Verteilungen  $3\mu(1 - \mu)$ .

### 7.3. Diskussion anderer Optimalitätskonzepte

Ein offensichtlicher Nachteil der in den vorangegangenen Abschnitten diskutierten Bahadur-Effizienz ist, daß die Größe des Selektionsintervalls auf die approximative Bahadur-Effizienz keinen Einfluß hat, bzw. bei der exakten Bahadur-Effizienz kein Einfluß nachgewiesen wurde (vgl. Korollar 7.1.15). Es ist zu vermuten, daß die Betrachtung der asymptotisch relativen Pitman-Effizienz diesen Nachteil beseitigt.

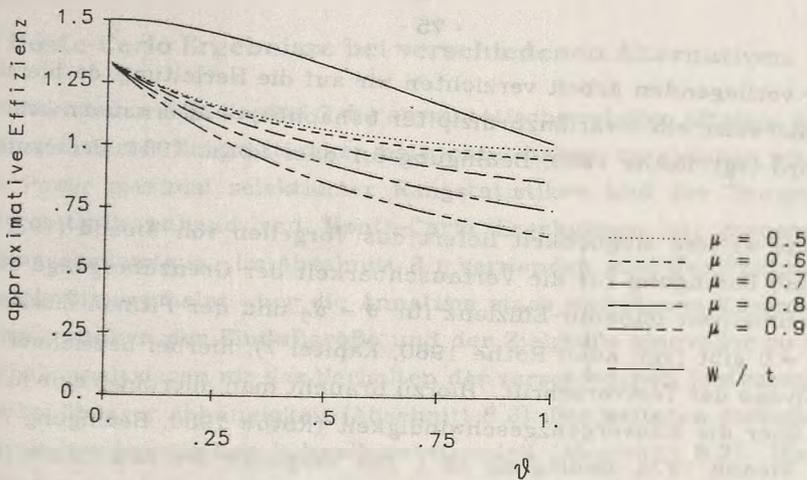


Abb. 7.2.3.: Die approximative Bahadur-Effizienz bei der Doppel exponentialverteilung in Abhängigkeit des Lokationsunterschiedes  $\vartheta$ : maximal selektierter Median-Test zum maximal selektierten Wilcoxon-Test bei verschiedenen Prüfpunkten  $\mu$  und maximal selektierter Wilcoxon-Test zum maximal selektierten  $t$ -Test ( $W/t$ ).

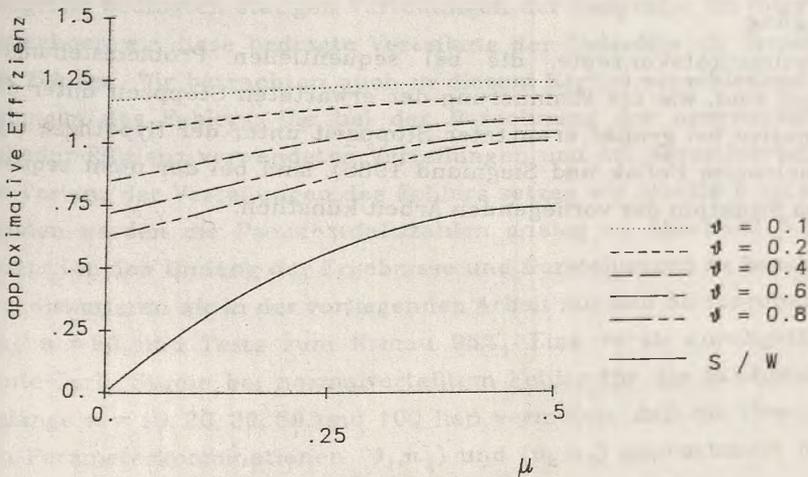


Abb. 7.2.4.: Die approximative Bahadur-Effizienz bei der Doppel exponentialverteilung in Abhängigkeit des Prüfpunktes  $\mu$ : maximal selektierter Median-Test zum maximal selektierten Wilcoxon-Test bei verschiedenen Lokationsunterschieden  $\vartheta$  und Spearman-Test zum maximal selektierten Wilcoxon-Test ( $S/W$ ).

In der vorliegenden Arbeit verzichten wir auf die Herleitung, da hierzu normalerweise ein Invarianzprinzip für benachbarte Alternativen benötigt wird (vgl. Rothe 1980, Bedingung 6.1 oder Rothe 1981, Bedingung A).

Eine weitere Möglichkeit liefert das Vorgehen von Wieand (1976), der eine Bedingung für die Vertauschbarkeit der Grenzübergänge der approximativen Bahadur-Effizienz für  $\vartheta \rightarrow \vartheta_0$  und der Pitman-Effizienz für  $\alpha \rightarrow 0$  gibt (vgl. auch Rothe 1980, Kapitel 7); hierbei bezeichnet  $\alpha$  das Niveau der Testvorschrift. Hierzu braucht man allerdings eine Aussage über die Konvergenzgeschwindigkeit (Rothe 1980, Bedingung 7.2 bzw. Wieand 1976, Bedingung III\*). Der Vergleich der Abb.7.2.1-4 für  $\vartheta \rightarrow 0$  mit den Werten der Pitman-Effizienz aus Randles und Wolfe (1979, Tabellen 5.4.7,11) ergibt keine Unterschiede. Für eine nichtstandardisierte maximal selektierte Rangstatistik bzgl.  $(\varepsilon_1, \varepsilon_2) = (0,1)$  bei positiver quadratischer Abhängigkeit überprüft Ledwina (1986b) Wieands Bedingung.

Optimalitätskonzepte, die bei sequentiellen Problemstellungen sinnvoll sind, wie die Minimierung der erwarteten Stoppzeit unter der Alternative bei großer erwarteter Stoppzeit unter der Hypothese (vgl. beispielsweise Pollak und Siegmund 1985), sind bei der nicht sequentiellen Situation der vorliegenden Arbeit künstlich.

## 8. Monte-Carlo Ergebnisse bei verschiedenen Alternativen

Nachdem wir im Kapitel 7 die asymptotische relative Effizienz maximal selektierter Rangstatistiken betrachtet haben, vergleichen wir nun die Power maximal selektierter Rangstatistiken und der Spearman-Rangstatistik anhand von Monte-Carlo Ergebnissen bei moderatem Stichprobenumfang. Im Abschnitt 8.1 verwenden wird das Prüfpunktmodell. Oft erscheint aber die Annahme eines monotonen Zusammenhangs zwischen der Einflußgröße und der Zielgröße sinnvoller zu sein; deshalb analysieren wir das Verhalten der verschiedenen Testvorschriften bei linearer Abhängigkeit (Abschnitt 8.2). Des weiteren diskutieren wir ein toxikologisches Schwellenwertmodell (Abschnitt 8.3). Hierbei schlagen wir insbesondere eine nichtparametrische Analyse des *no observed effect level* (NOEL) vor.

Unter der Alternative ist die Verteilung der maximal selektierten Rangstatistik in der Regel nicht invariant unter den bzgl. der Einflußgröße bedingten stetigen Verteilungen der Zielgröße. Im folgenden bezeichnen wir diese bedingte Verteilung der Zielgröße als *Verteilung des Fehlers*. Wir betrachten auch in diesem Kapitel verschiedene Verteilungen des Fehlers: Die bei der Berechnung der approximativen Bahadur-Effizienz verwendeten Verteilungen und die Normalverteilung. Die Varianz der Verteilungen des Fehlers setzen wir identisch eins, ansonsten werden die Pseudozufallszahlen analog zu Abschnitt 5.2 erzeugt. Um den Umfang der Ergebnisse und Darstellungen zu beschränken, diskutieren wir in der vorliegenden Arbeit nur den Stichprobenumfang  $n = 50$  und Tests zum Niveau 95%. Eine vorab durchgeführte Monte-Carlo Studie bei normalverteiltem Fehler für die Stichprobenumfänge  $n = 10, 20, 30, 50$  und  $100$  ließ vermuten, daß die Power bei den Parameterkombinationen  $(\vartheta_1, n_1)$  und  $(\vartheta_2, n_2)$  approximativ übereinstimmt, falls gilt:  $\vartheta_1 / \sqrt{n_1} = \vartheta_2 / \sqrt{n_2}$ . Die Abhängigkeit der Verteilung unter der Nullhypothese vom Stichprobenumfang wird im Kapitel 5 diskutiert. Die Hypothese wird verworfen, wenn die betrachtete Statistik größer als die Monte-Carlo Approximation des exakten 95%-Quantils ist. Wir verzichten auf die Konstruktion eines randomisierten

Tests zum exakten Niveau 5%. Unter der simulierten Power verstehen wir die relative Häufigkeit von verworfenen Hypothesen. Wie im Abschnitt 5.2 wurden 10 000 Wiederholungen für die Monte-Carlo Berechnungen verwendet. Als Selektionsintervall analysierten wir  $(\varepsilon_1, \varepsilon_2) = (0.1, 0.9), (0.2, 0.8), (0.3, 0.7), (0.4, 0.6)$ . Im Zusammenhang mit einer Betrachtung von Strukturbruchstatistiken haben beispielsweise Hackl und Katzenbeisser (1989) für verschiedene parametrische und nichtparametrische Statistiken eine Monte-Carlo Studie durchgeführt. Hackl und Katzenbeisser (1989) betrachten allerdings nicht standardisierte Statistiken, und insofern sind ihre Ergebnisse nur teilweise mit den Ergebnissen der vorliegenden Arbeit vergleichbar (vgl. auch Lombard 1989; p.13).

### 8.1. Prüfpunktalternative

In diesem Abschnitt betrachten wir die durch das Prüfpunktmodell (Abschnitt 2.2) gegebene Alternative:

$$Y = 1_{\{X \leq \mu\}} \vartheta + Z, \quad (8.1.1)$$

wobei  $Z$  einen von  $X$  stochastisch unabhängigen Fehler bzw. eine Störgröße bezeichnet. Analog zu den Monte-Carlo Berechnungen der Verteilung unter der Nullhypothese im Abschnitt 5.2 simulierten wir die Power für die vier verschiedenen maximal selektierten Tests und den Spearman-Rangtest. Wir betrachten 10 Werte für den Lageparameter,  $\vartheta = [8^{1/2} i] / 10, i=1, \dots, 10$ , und 5 Werte für den Prüfpunkt,  $\mu = 0.5, 0.6, 0.7, 0.8, 0.9$ .

#### *Gleichverteilung*

In den Abb.8.1.1-4 werden die fünf Tests verglichen bei einem gleichverteilten Fehler. Der maximal selektierte Median-Test ist für alle betrachteten Parameterkombinationen schlechter als die anderen Tests. Der maximal selektierte Wilcoxon-Test und der  $t$ -Test sind besser

als der Spearman-Test; allerdings ist der Unterschied für  $(\varepsilon_1, \varepsilon_2) = (0.2, 0.8)$  und  $\mu = 0.5$  sehr gering.

Die Abb. 8.1.5-6 lassen den Einfluß der Größe des Selektionsintervalls auf die Power der maximal selektierten Wilcoxon- bzw. Median-Statistik als relativ gering erscheinen. Dagegen ist der Einfluß des Prüfpunktparameters  $\mu$  vergleichsweise groß (Abb.8.1.7-8). Hierbei muß man allerdings berücksichtigen, daß die Einflußgröße gleichverteilt ist und die Selektion bzgl. der empirischen Quantile definiert ist. Denn dadurch hängt die Wahrscheinlichkeit, daß der Prüfpunkt außerhalb der Realisation des Selektionsintervalls liegt, von der Wahl der Parameter  $\mu$ ,  $\varepsilon_1$  und  $\varepsilon_2$  ab.

#### *Doppelexponentialverteilung*

Die Abb.8.1.9-10 ermöglichen einen Vergleich der simulierten Power der fünf Tests bei einem doppelexponentialverteilten Fehler (analog zu Abb.8.1.1,4). Der maximal selektierte  $t$ -Test und der maximal selektierte Log-Rank-Test sind für die betrachteten Parameterkombinationen vergleichsweise schlecht; dies gilt auch für die nicht dargestellten Parameterkombinationen aus Abb.8.1.2-3. Allerdings verbessert sich die Situation für den maximal selektierten  $t$ -Test für größere Werte des Lokationsunterschiedes  $\vartheta$  und für Werte des Prüfpunkts  $\mu$ , die in größerer Entfernung zum Median der Einflußgröße liegen. Für den maximal selektierten Median-Test erhält man unter der Doppelexponentialverteilung ein besseres Resultat, was aufgrund der Optimalitätseigenschaften der Median-Scores im Zweistichprobenfall (z.B. Hajek und Sidak 1967) zu erwarten ist. Für kleine Werte von  $\vartheta$ ,  $\mu = 0.5$  und  $(\varepsilon_1, \varepsilon_2) = (0.4, 0.6)$  ist die simulierte Power des maximal selektierten Median-Tests deutlich größer als die der anderen Tests.

Analoge Abbildungen zu Abb.8.1.5-8 lassen ebenfalls nur einen sehr geringen Einfluß der Größe des Selektionsintervalls erkennen, und dagegen ist der Einfluß des wahren Prüfpunkts  $\mu$  in Abhängigkeit des verwendeten Selektionsintervalls relativ groß.

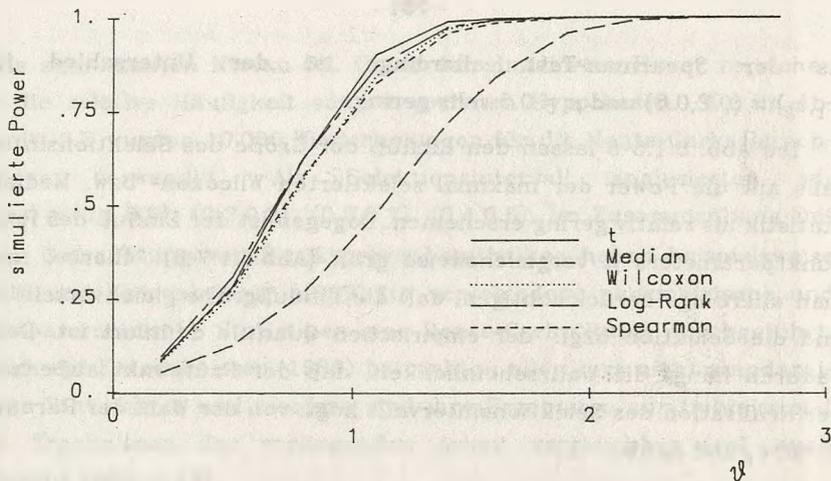


Abb. 8.1.1.: Die simulierte Power bei gleichverteiltem Fehler,  $(\varepsilon_1, \varepsilon_2) = (0.2, 0.8)$  und  $\mu = 0.5$  in Abhängigkeit des Lokationsunterschiedes  $\psi$ .

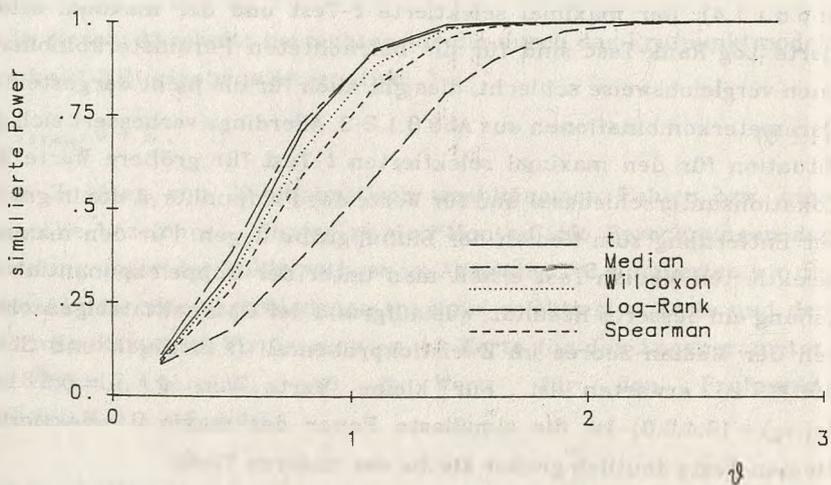


Abb. 8.1.2.: Die simulierte Power bei gleichverteiltem Fehler,  $(\varepsilon_1, \varepsilon_2) = (0.4, 0.6)$  und  $\mu = 0.5$  in Abhängigkeit des Lokationsunterschiedes  $\psi$ .

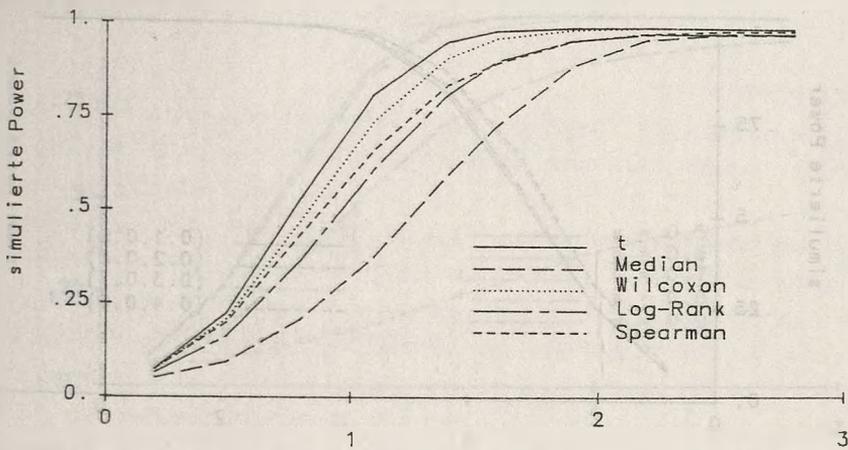


Abb. 8.1.3.: Die simulierte Power bei gleichverteiltem Fehler,  $(\varepsilon_1, \varepsilon_2) = (0.2, 0.8)$  und  $\mu = 0.7$  in Abhängigkeit des Lokationsunterschiedes  $\vartheta$ .

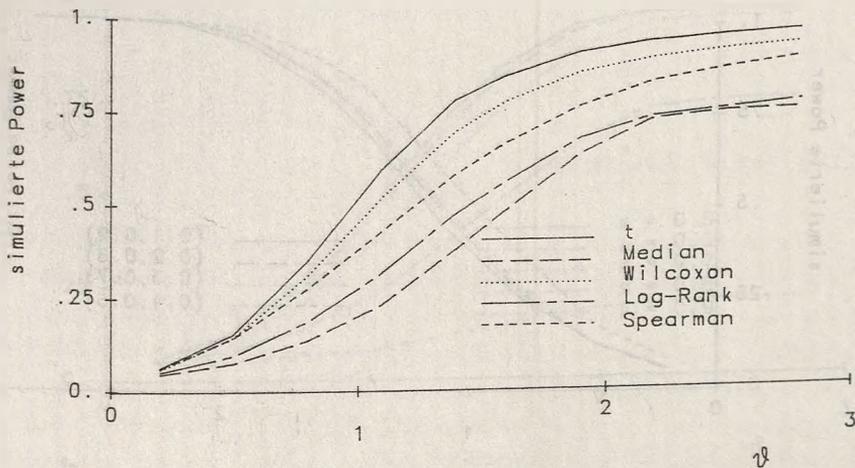


Abb. 8.1.4.: Die simulierte Power bei gleichverteiltem Fehler,  $(\varepsilon_1, \varepsilon_2) = (0.2, 0.8)$  und  $\mu = 0.8$  in Abhängigkeit des Lokationsunterschiedes  $\vartheta$ .

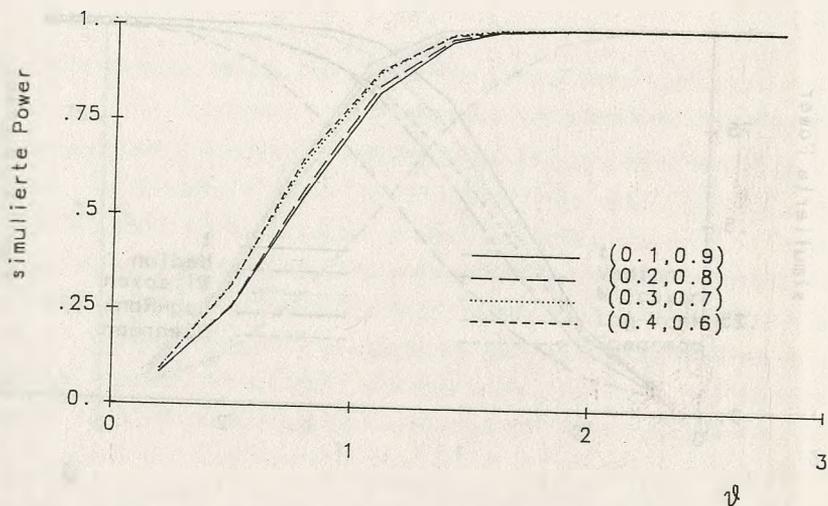


Abb. 8.1.5.: Die simulierte Power bei gleichverteiltem Fehler des maximal selektierten Wilcoxon-Tests für verschiedene Selektionsintervalle bei  $\mu = 0.5$  in Abhängigkeit des Lokationsunterschiedes  $\vartheta$ .

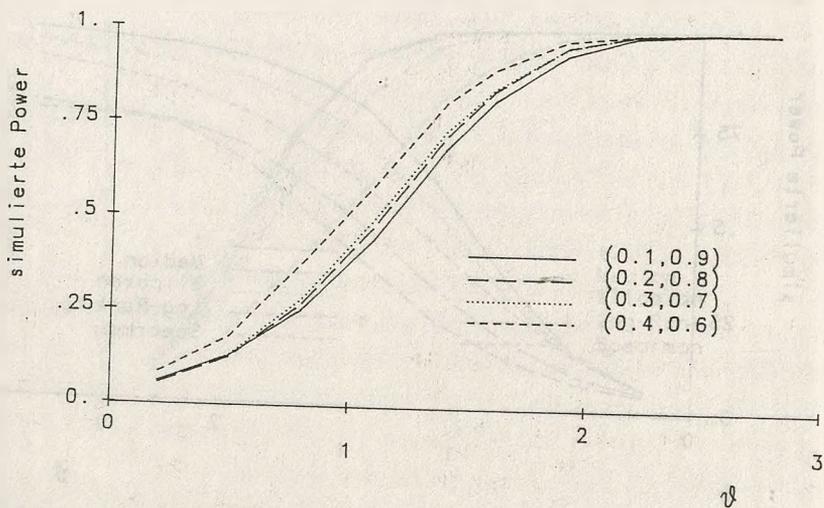


Abb. 8.1.6.: Die simulierte Power bei gleichverteiltem Fehler des maximal selektierten Median-Tests für verschiedene Selektionsintervalle bei  $\mu = 0.5$  in Abhängigkeit des Lokationsunterschiedes  $\vartheta$ .

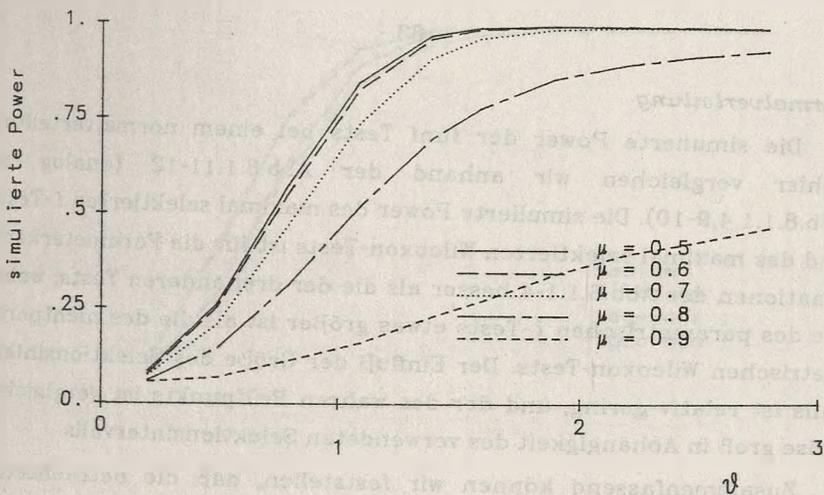


Abb. 8.1.7.: Die simulierte Power bei gleichverteiltem Fehler des maximal selektierten Wilcoxon-Tests für verschiedene Prüfpunkte bei  $(\varepsilon_1, \varepsilon_2) = (0.2, 0.8)$  in Abhängigkeit des Lokationsunterschiedes  $\vartheta$ .

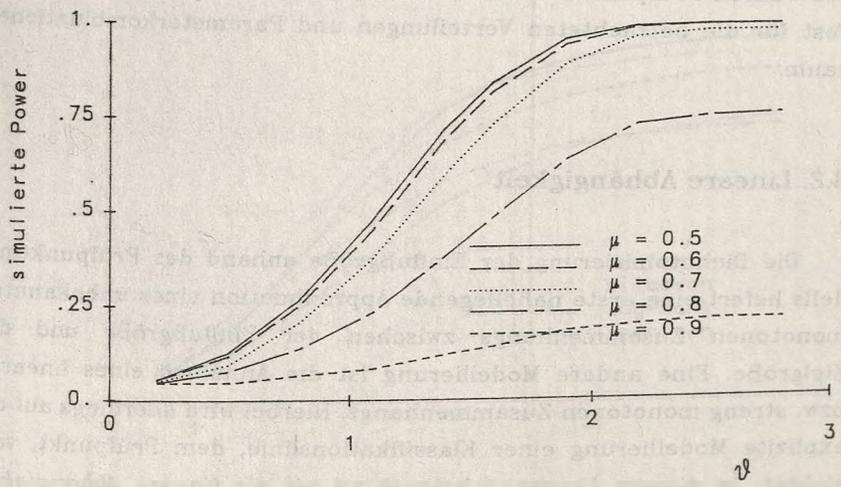


Abb. 8.1.8.: Die simulierte Power bei gleichverteiltem Fehler des maximal selektierten Median-Tests für verschiedene Prüfpunkte bei  $(\varepsilon_1, \varepsilon_2) = (0.2, 0.8)$  in Abhängigkeit des Lokationsunterschiedes  $\vartheta$ .

### *Normalverteilung*

Die simulierte Power der fünf Tests bei einem normalverteilten Fehler vergleichen wir anhand der Abb.8.1.11-12 (analog zu Abb.8.1.1,4,9-10). Die simulierte Power des maximal selektierten  $t$ -Tests und des maximal selektierten Wilcoxon-Tests ist für die Parameterkombinationen der Abb.8.1.1-4 besser als die der drei anderen Tests, wobei die des parametrischen  $t$ -Tests etwas größer ist als die des nichtparametrischen Wilcoxon-Tests. Der Einfluß der Größe des Selektionsintervalls ist relativ gering, und der des wahren Prüfpunkts ist vergleichsweise groß in Abhängigkeit des verwendeten Selektionsintervalls.

Zusammenfassend können wir feststellen, daß die betrachteten Verteilungen bei den analysierten Parameterkombinationen das Verhältnis der simulierten Power der Tests untereinander verändern. Allerdings unterscheidet sich der Einfluß des Selektionsintervalls bzw. der des wahren Prüfpunkts bei einem fest gewählten maximal selektierten Test für die betrachteten Verteilungen und Parameterkombinationen kaum.

## 8.2. Lineare Abhängigkeit

Die Dichotomisierung der Einflußgröße anhand des Prüfpunktmodells liefert eine erste naheliegende Approximation eines unbekanntem monotonen Zusammenhangs zwischen der Einflußgröße und der Zielgröße. Eine andere Modellierung ist die Annahme eines linearen bzw. streng monotonen Zusammenhangs. Hierbei wird allerdings auf die explizite Modellierung einer Klassifikationslinie, dem Prüfpunkt, verzichtet. In diesem Abschnitt betrachten wir als *lineare Abhängigkeit* zwischen der Zielgröße  $Y$  und der Einflußgröße  $X$ :

$$Y = \vartheta X + Z, \quad (8.2.1)$$

wobei  $Z$  einen von  $X$  stochastisch unabhängigen Fehler bzw. eine Störgröße und  $\vartheta$  den *Steigungsparameter* bezeichnet. Unter diesem

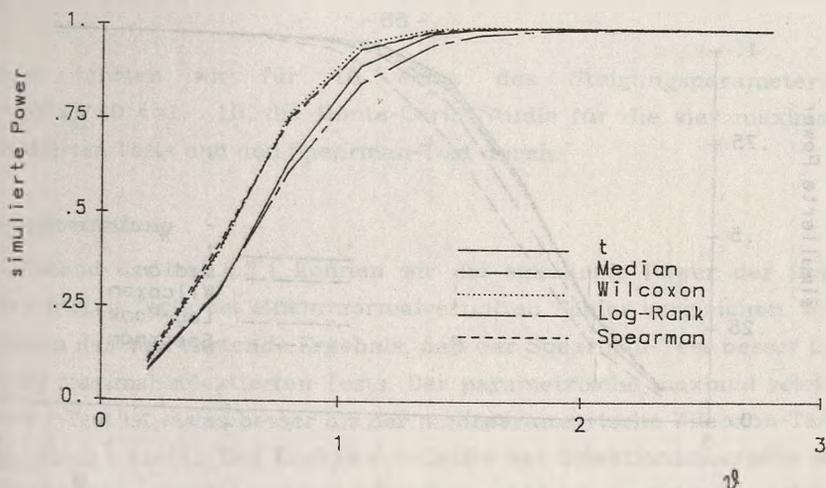


Abb. 8.1.9.: Die simulierte Power bei doppel exponentialverteiltem Fehler,  $(\varepsilon_1, \varepsilon_2) = (0.2, 0.8)$  und  $\mu = 0.5$  in Abhängigkeit des Lokationsunterschiedes  $v$ .

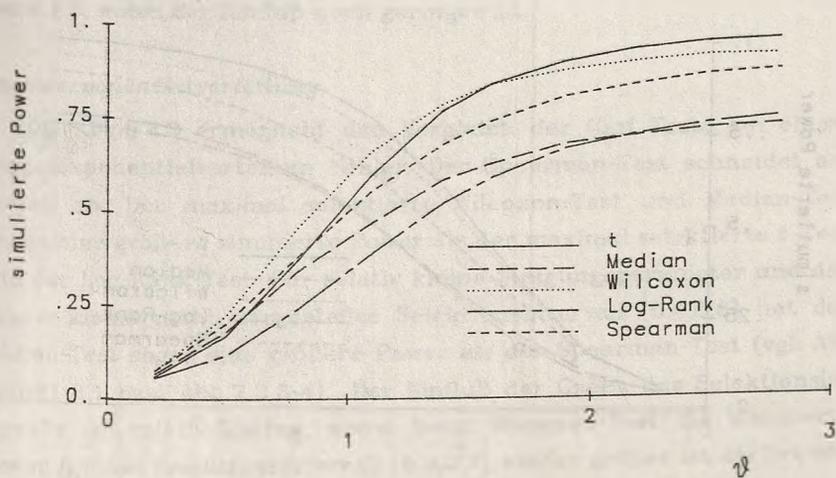


Abb. 8.1.10.: Die simulierte Power bei doppel exponentialverteiltem Fehler,  $(\varepsilon_1, \varepsilon_2) = (0.2, 0.8)$  und  $\mu = 0.8$  in Abhängigkeit des Lokationsunterschiedes  $v$ .

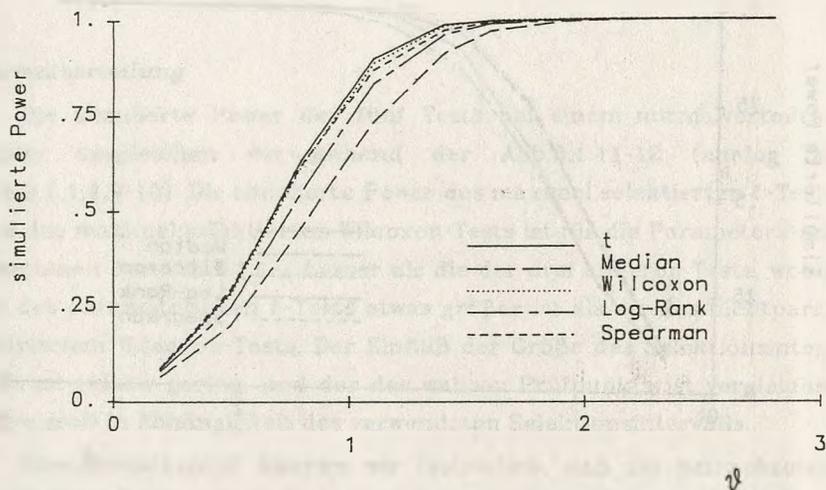


Abb. 8.1.11.: Die simulierte Power bei normalverteiltem Fehler,  $(\varepsilon_1, \varepsilon_2) = (0.2, 0.8)$  und  $\mu = 0.5$  in Abhängigkeit des Lokationsunterschiedes  $\delta$ .

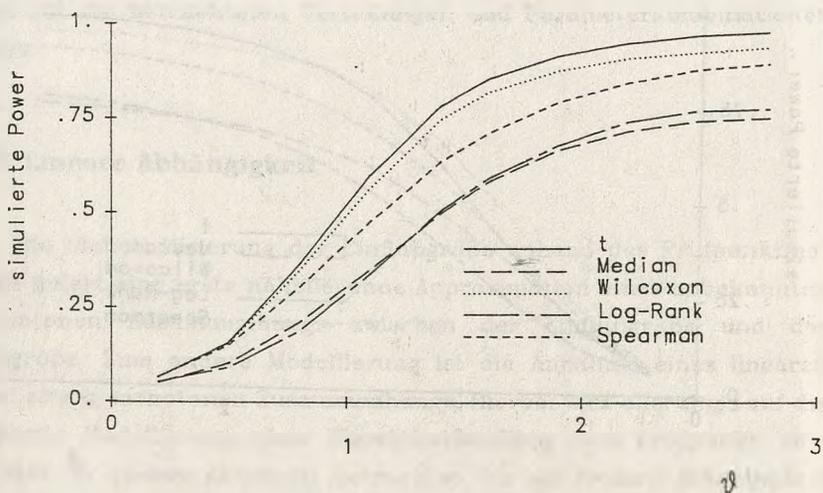


Abb. 8.1.12.: Die simulierte Power bei normalverteiltem Fehler,  $(\varepsilon_1, \varepsilon_2) = (0.2, 0.8)$  und  $\mu = 0.8$  in Abhängigkeit des Lokationsunterschiedes  $\delta$ .

Modell führten wir für 10 Werte des Steigungsparameters,  $\vartheta = [8^{\frac{1}{2}} i] / 10, i=1, \dots, 10$ , die Monte-Carlo Studie für die vier maximal selektierten Tests und den Spearman-Test durch.

#### *Normalverteilung*

Anhand der Abb.8.2.1 können wir die simulierte Power der fünf betrachteten Tests bei einem normalverteilten Fehler vergleichen. Wir erhalten das zuerwartende Ergebnis, daß der Spearman-Test besser ist als die maximal selektierten Tests. Der parametrische maximal selektierte  $t$ -Test ist etwas besser als der nichtparametrische Wilcoxon-Test (vgl. Abb.8.1.11-12). Der Einfluß der Größe des Selektionsintervalls ist relativ gering, allerdings ist der Zusammenhang beim maximal selektierten Wilcoxon-Test nicht monoton (Abb. 8.2.1). Die simulierte Power ist für  $(\varepsilon_1, \varepsilon_2) = (0.3, 0.7)$  am größten. Dagegen entspricht das nicht dargestellte Verhalten des maximal selektierten Median-Tests in etwa Abb.8.1.6, wobei der Einfluß noch geringer ist.

#### *Doppelexponentialverteilung*

Die Abb.8.2.3 ermöglicht den Vergleich der fünf Tests bei einem doppelexponentialverteilten Fehler. Der Spearman-Test schneidet am besten ab. Der maximal selektierte Wilcoxon-Test und Median-Test haben eine größere simulierte Power als der maximal selektierte  $t$ -Test und der Log-Rank-Test. Für relativ kleine Steigungsparameter und das relativ kleine nicht dargestellte Selektionsintervall  $(0.4, 0.6)$  hat der Median-Test sogar eine größere Power als der Spearman-Test (vgl. Abschnitt 8.1 bzw. Abb.7.2.3-4). Der Einfluß der Größe des Selektionsintervalls ist relativ gering, wobei beim Wilcoxon-Test die simulierte Power für das Selektionsintervall  $(0.3, 0.7)$  wieder größer ist als bei den drei anderen betrachteten Intervallgrößen.

#### *Gleichverteilung*

In der Abb.8.2.4 werden die fünf betrachteten Tests anhand der simulierten Power bei gleichverteiltem Fehler verglichen. Der

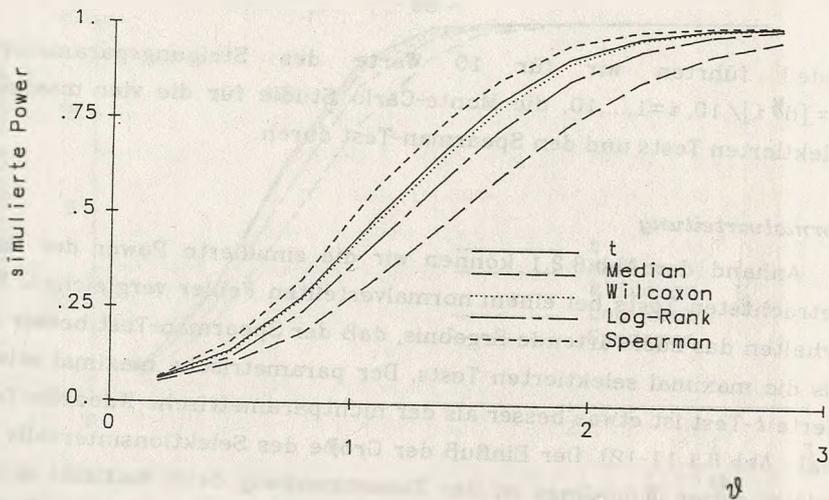


Abb. 8.2.1.: Die simulierte Power bei normalverteiltem Fehler und  $(\varepsilon_1, \varepsilon_2) = (0.2, 0.8)$  in Abhängigkeit des Steigungsparameters  $\vartheta$ .

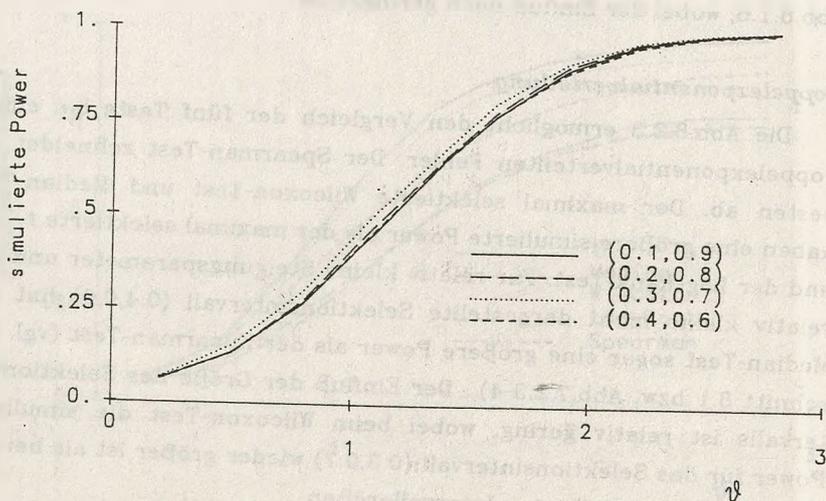


Abb. 8.2.2.: Die simulierte Power bei normalverteiltem Fehler des maximal selektierten Wilcoxon-Tests für verschiedene Selektionsintervalle in Abhängigkeit des Steigungsparameters  $\vartheta$ .

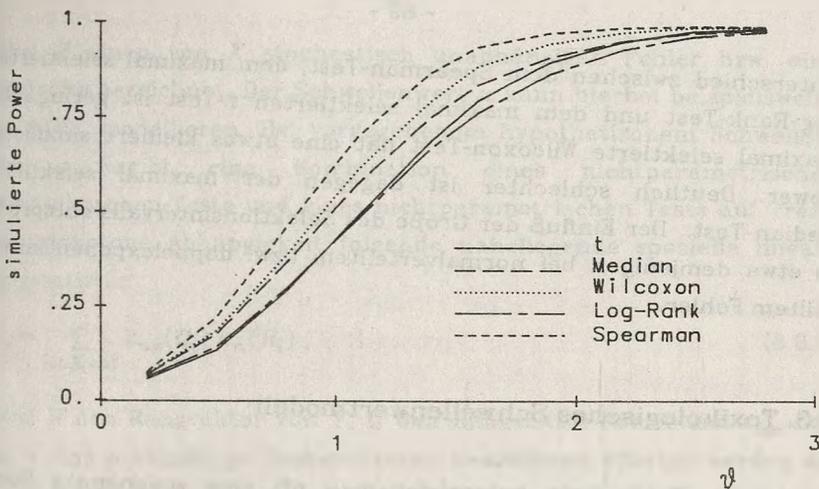


Abb. 8.2.3.: Die simulierte Power bei doppel exponentialverteiltem Fehler und  $(\varepsilon_1, \varepsilon_2) = (0.2, 0.8)$  in Abhängigkeit des Steigungsparameters  $\vartheta$ .

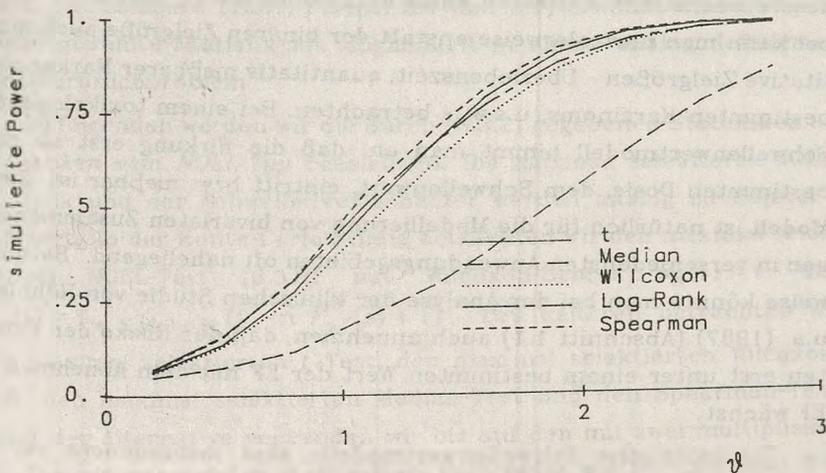


Abb. 8.2.4.: Die simulierte Power bei gleichverteiltem Fehler und  $(\varepsilon_1, \varepsilon_2) = (0.2, 0.8)$  in Abhängigkeit des Steigungsparameters  $\vartheta$ .

Unterschied zwischen dem Spearman-Test, dem maximal selektierten Log-Rank-Test und dem maximal selektierten  $t$ -Test ist gering. Der maximal selektierte Wilcoxon-Test hat eine etwas kleinere simulierte Power. Deutlich schlechter ist dagegen der maximal selektierte Median-Test. Der Einfluß der Größe des Selektionsintervalls entspricht in etwa demjenigen bei normalverteiltem bzw. doppel exponentialverteiltem Fehler.

### 8.3. Toxikologisches Schwellenwertmodell

in der Toxikologie betrachtet man oft eine sogenannte Dosis-Wirkungs-Beziehung. Hierbei wird in der Regel die Einflußgröße - Dosis eines Medikamentes, Belastung durch einen Schadstoff, u.s.w. - experimentell variiert. Die Wirkung wird mittels einer binären Zielgröße - Tod des Versuchstiers, Auftreten eines Karzinoms, u.s.w. - analysiert. Hierbei kann man normalerweise anstatt der binären Zielgröße auch quantitative Zielgrößen - Überlebenszeit, quantitativ meßbarer Marker eines bestimmten Karzinoms, u.s.w. - betrachten. Bei einem toxikologischen Schwellenwertmodell nimmt man an, daß die Wirkung erst ab einer bestimmten Dosis, dem Schwellenwert, eintritt bzw. meßbar ist. Dieses Modell ist natürlich für die Modellierung von bivariaten Zusammenhängen in verschiedensten Anwendungsgebieten oft naheliegend. Beispielsweise könnte man bei der Analyse der klinischen Studie von Hohnloser u.a. (1987) (Abschnitt 1.1) auch annehmen, daß das Risiko der Patienten erst unter einem bestimmten Wert der EF mit dem Abnehmen der EF wächst.

Toxikologische Schwellenwertmodelle sind insbesondere zur Analyse von binären Zielgrößen gebräuchlich. Beispielsweise diskutiert Cox (1987) hierzu verschiedene Modellierungen. In der vorliegenden Arbeit betrachten wir eine beliebig verteilte Zielgröße und als toxikologisches Schwellenwertmodell folgenden Zusammenhang:

$$Y = 1_{\{X > \mu\}} \vartheta (X - \mu) + Z, \quad (8.3.1)$$

wobei  $Z$  einen von  $X$  stochastisch unabhängigen Fehler bzw. eine Störgröße bezeichnet. Der Schwellenwert  $\mu$  kann hierbei beispielsweise den NOEL modellieren. Bei vorgegebenem hypothetischem Schwellenwert  $\rho$  ergibt eine Kombination eines nichtparametrischen Zweistichproben-Tests und eines nichtparametrischen Tests auf Trend bzw. monotone Abhängigkeit folgende naheliegende spezielle lineare Rangstatistik:

$$T_{n\rho} = \sum_{\{i: X_i \geq \rho\}} c_{n\rho}(Q_i) a_n(R_i), \quad (8.3.2)$$

wobei  $R$  den Rangvektor von  $Y$ ,  $Q$  den Rangvektor von  $X$  und  $c_{n\rho}$ ,  $a_n$  von  $n$  und  $\rho$  abhängige Scorevektoren bezeichnen. Hierbei werden die Beobachtungen mit Werten der Einflußgröße kleiner als  $\rho$  mit dem Regressor identisch Null gewichtet, und die Beobachtungen mit Werten der Einflußgröße größer oder gleich  $\rho$  werden mit monotonen und nichtnegativen Regressoren gewichtet; d.h.  $0 \leq c_{n\rho}(i-1) \leq c_{n\rho}(i)$  für  $i = 2, \dots, n$ . Lombard (1987, Formel 2.4 und 5.1) erwähnt eine entsprechend gewählte Statistik als Möglichkeit in einem analog definierten Strukturbruchproblem.

Im folgenden werden wir die durch (8.3.2) gegebenen Statistiken als *Statistiken vom NOEL-Typ* bezeichnen. Die maximal selektierte NOEL-Statistik und der Schwellenwertschätzer werden analog zu Kapitel 2 definiert. In der Monte-Carlo Studie betrachten wir den maximal selektierten NOEL-Test (8.3.2) mit Wilcoxon-Scores  $a_n(i) = i$  und  $c_{n\rho}(i) = 1_{\{n F_{nX}(\rho) \leq i\}} (i - n F_{nX}(\rho) + 1)$ . Des weiteren betrachten wir den maximal selektierten  $t$ -Test, den maximal selektierten Wilcoxon-Test, den maximal selektierten Median-Test und den Spearman-Test. Unter der Alternative verwenden wir bis auf den mit zwei multiplizierten Steigungsparameter  $\vartheta$  dieselben Parameter wie im Abschnitt 8.1. Als nächstes diskutieren wir die Verteilung der maximal selektierten NOEL-Statistik unter der Nullhypothese anhand von einer heuristischen Herleitung der asymptotischen Verteilung und von Monte-Carlo Approximationen der exakten Verteilung (vgl. Abschnitt 5.2).

*Verteilung der NOEL-Statistik unter der Nullhypothese*

Da die maximal selektierte NOEL-Statistik unter Berücksichtigung der Thematik der vorliegenden Arbeit nur am Rande eingeführt und diskutiert werden kann, beschränken wir uns auf folgende heuristische Motivation der asymptotischen Verteilung unter der Nullhypothese. Die Statistiken vom NOEL-Typ sind bzgl. eines festen hypothetischen Schwellenwerts einfache lineare Rangstatistiken und haben somit unter den üblichen Regularitätsbedingungen an die Scores und Regressoren eine asymptotische Normalverteilung (vgl. Kapitel 2). Damit ist die asymptotische endlich dimensionale Randverteilung des durch  $\{T_{n\rho}; 0 \leq F_X(\rho) \leq 1\}$  gegebenen Prozesses eine multivariate Normalverteilung. Bei obiger Wahl von  $c_{n\rho}$  erhalten wir als Scoreerzeugende-Funktion für  $c_{n\rho}$ :  $c_\rho(v) = 1_{\{\rho \leq v\}}(v - \rho)$ , wobei  $\rho, v \in [0,1]$ . Damit kann man die asymptotische Verteilung von (8.3.2) auch als das Integral eines mit  $c_\rho$  gewichteten Pfades einer Brownschen Brücke auffassen. Folglich vermuten wir, daß der maximal selektierte NOEL-Test gegen folgendes maximal selektiertes absolutes standardisiertes Integral der Brownschen-Brücke konvergiert:

$$M(\varepsilon_1, \varepsilon_2) = \sup_{\varepsilon_1 \leq \rho \leq \varepsilon_2} \frac{1}{\rho} \frac{|\int_\rho^1 (t - \rho) B_0(t) dt|}{(\text{Var}(\int_\rho^1 (t - \rho) B_0(t) dt))^{1/2}}, \quad (8.3.3)$$

wobei  $\{B_0(t); 0 \leq t \leq 1\}$  die Brownsche Brücke bezeichnet. Die vermutete Grenzverteilung (8.3.3) ist das Supremum eines f.s. stetig differenzierbaren Gaußprozesses. Deshalb können wir die Verteilung von (8.3.3) anhand der folgenden Formel nach Davies (1977,1987) approximieren:

$$P(M(\varepsilon_1, \varepsilon_2) \leq b) \geq \quad (8.3.4)$$

$$1 - 2\Phi(-b) - \varphi(b) \log(\varepsilon_2(1 - \varepsilon_1) / ((1 - \varepsilon_2)\varepsilon_1)) / (2\pi)^{1/2},$$

hierbei vermuten wir, daß der  $\log(\varepsilon_2(1 - \varepsilon_1) / ((1 - \varepsilon_2)\varepsilon_1))$  das

Integral über die zweite Ableitung der Kovarianzfunktion ist (vgl. Davies 1987, Formel 2.1). In der Abb. 8.3.1 werden die verschiedenen simulierten Verteilungsfunktionen für  $(\varepsilon_1, \varepsilon_2) = (0.1, 0.9)$  gemeinsam mit der Approximation aus Abschnitt 3.3 nach Jennen (1985) dargestellt. Die Approximation der NOEL-Statistik (8.3.4) nach Davies (1987) ist verglichen mit der Approximation nach Jennen (1985) für die maximal selektierten Rangstatistiken aus Abschnitt 5.2 relativ gut. Für ein kleineres Selektionsintervall wird die Approximation noch besser (Abb. 8.3.2;  $(\varepsilon_1, \varepsilon_2) = (0.3, 0.7)$ ), und dieses Verhalten entspricht dem der maximal selektierten Rangstatistiken (vgl. Abschnitt 5.2).

#### *Verhalten unter der Alternative*

Die Abb. 8.3.3-4 ermöglichen einen Vergleich der simulierten Power der 5 analysierten Statistiken bei normalverteiltem Fehler für  $(\varepsilon_1, \varepsilon_2) = (0.2, 0.8)$ ,  $(0.4, 0.6)$  und  $\mu = 0.5$ . Die maximal selektierte NOEL-Statistik hat die größte Power. Die Median-Statistik ist vergleichsweise schlecht. Die maximal selektierte Wilcoxon-Statistik und  $t$ -Statistik haben für das größere Selektionsintervall eine größere Power. Dies wird teilweise durch das Verhalten des Prüfpunktschätzers erklärt (vgl. Abschnitt 9.2). Für den nicht anhand besonderer Abbildungen dargestellten Einfluß des Selektionsintervalls gilt analog zu Abschnitt 8.1, daß er relativ gering ist. Dagegen ist der Einfluß des Schwellenwertes  $\mu$  relativ groß. Der Einfluß der verschiedenen Verteilungen des Fehlers - Normalverteilung, Gleichverteilung und Doppelsexponentialverteilung - ist relativ gering (vgl. Abb. 8.3.3, 5-6 für  $(\varepsilon_1, \varepsilon_2) = (0.2, 0.8)$  und  $\mu = 0.5$ ).

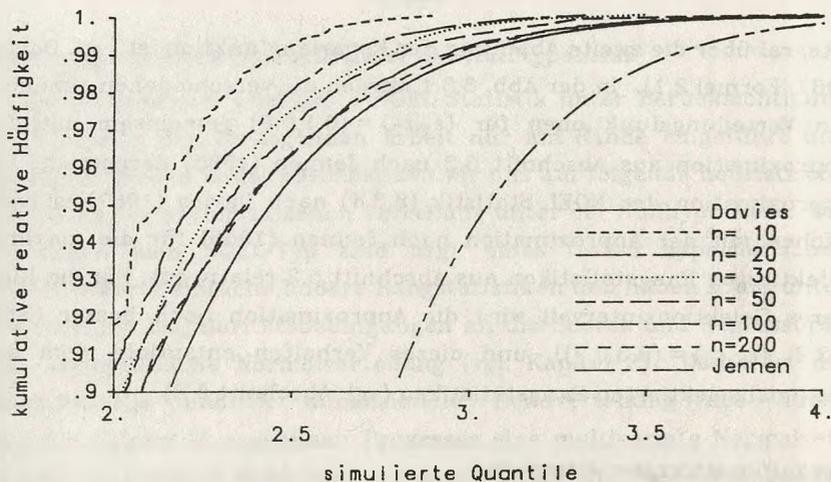


Abb. 8.3.1.: Simulierter und approximativer oberer Teil der Verteilungsfunktion der maximal selektierten NOEL-Statistik unter der Nullhypothese bei  $(\epsilon_1, \epsilon_2) = (0.1, 0.9)$ . Die durchgezogene Linie veranschaulicht die vermutete approximative asymptotische Verteilung.

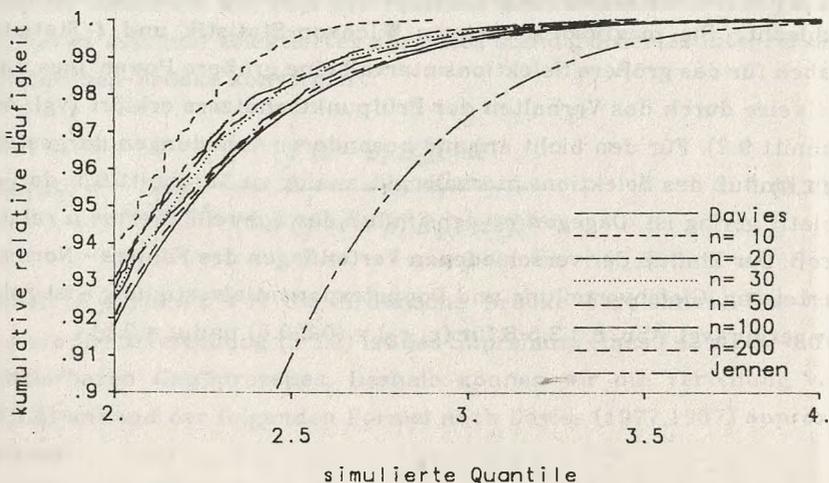


Abb. 8.3.2.: Simulierter und approximativer oberer Teil der Verteilungsfunktion der maximal selektierten NOEL-Statistik unter der Nullhypothese bei  $(\epsilon_1, \epsilon_2) = (0.3, 0.7)$ . Die durchgezogene Linie veranschaulicht die vermutete approximative asymptotische Verteilung.

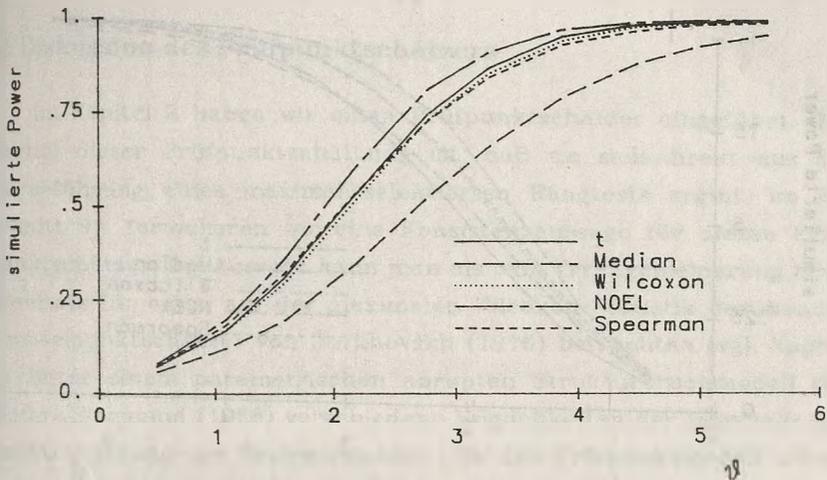


Abb. 8.3.3.: Die simulierte Power bei normalverteiltem Fehler,  $(\varepsilon_1, \varepsilon_2) = (0.2, 0.8)$  und dem Schwellenwert  $\mu = 0.5$  in Abhängigkeit des Steigungsparameters  $\vartheta$ .

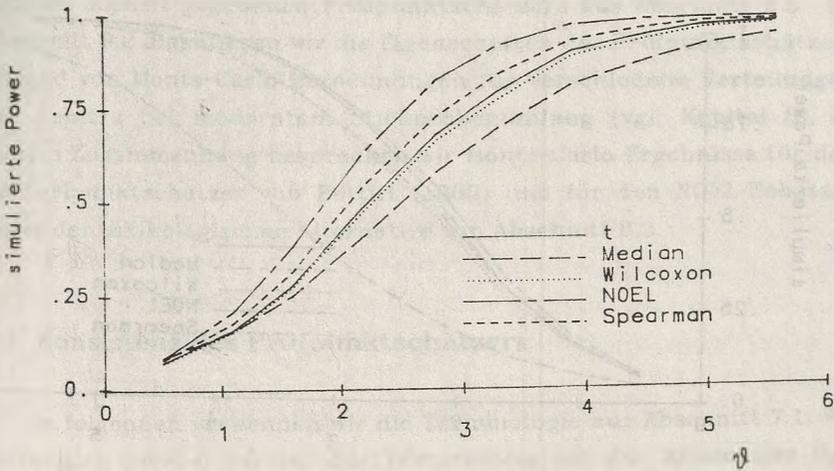


Abb. 8.3.4.: Die simulierte Power bei normalverteiltem Fehler,  $(\varepsilon_1, \varepsilon_2) = (0.4, 0.6)$  und dem Schwellenwert  $\mu = 0.5$  in Abhängigkeit des Steigungsparameters  $\vartheta$ .

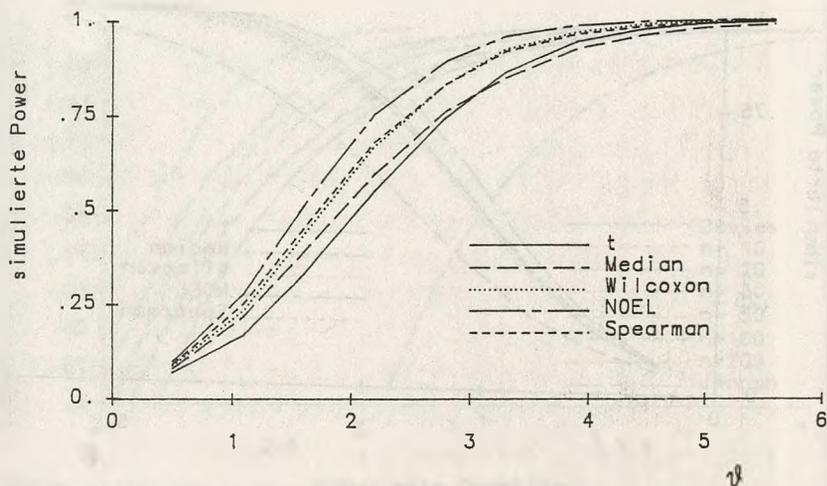


Abb. 8.3.5.: Die simulierte Power bei doppel exponentialverteiltem Fehler,  $(\varepsilon_1, \varepsilon_2) = (0.2, 0.8)$  und dem Schwellenwert  $\mu = 0.5$  in Abhängigkeit des Steigungsparameters  $\psi$ .

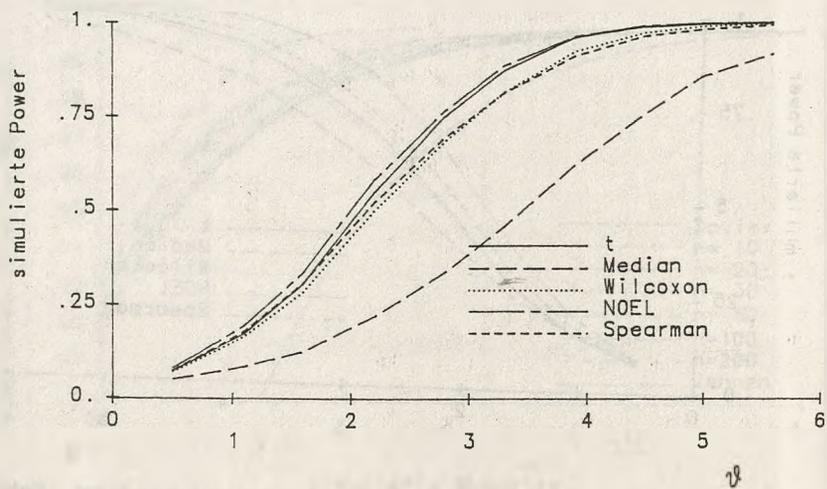


Abb. 8.3.6.: Die simulierte Power bei gleichverteiltem Fehler,  $(\varepsilon_1, \varepsilon_2) = (0.2, 0.8)$  und dem Schwellenwert  $\mu = 0.5$  in Abhängigkeit des Steigungsparameters  $\psi$ .

## 9. Diskussion des Prüfpunktschätzers

Im Kapitel 2 haben wir einen Prüfpunktschätzer eingeführt. Ein Vorteil dieser Prüfpunktschätzung ist, daß sie sich direkt aus der Durchführung eines maximal selektierten Rangtests ergibt. Im Abschnitt 9.1 formulieren wir eine Konsistenzaussage für diesen Prüfpunktschätzer. Die Aussage kann man als eine Verallgemeinerung eines Resultats für einen auf der maximalen Wilcoxon-Statistik beruhenden Wechsellpunktschätzer von Darkhovskh (1976) betrachten (vgl. Kapitel 4). Unter einem parametrischen abrupten Strukturbruchmodell diskutiert Siegmund (1988) verschiedene Möglichkeiten der Intervall- und Punktschätzung des Wechsellpunktes. Da das Prüfpunktmodell oft als eine stark vereinfachte Approximation des möglichen funktionalen Zusammenhangs zwischen Zielgröße und Einflußgröße anzusehen ist und dies ebenso für den Parameter Prüfpunkt gilt, beschränken wir uns auf die Diskussion des von der Methode der maximal selektierten Rangstatistiken implizit gegebenen Prüfpunktschätzers aus Abschnitt 2.5. Im Abschnitt 9.2 diskutieren wir die Eigenschaften des Prüfpunktschätzers anhand von Monte-Carlo Berechnungen für verschiedene Verteilungen des Fehlers bei moderatem Stichprobenumfang (vgl. Kapitel 8). In diesem Zusammenhang besprechen wir Monte-Carlo Ergebnisse für den Wechsellpunktschätzer von Pettitt (1980) und für den NOEL-Schätzer unter der toxikologischen Alternative von Abschnitt 8.3.

### 9.1. Konsistenz des Prüfpunktschätzers

Im folgenden verwenden wir die Terminologie aus Abschnitt 7.1. Wir betrachten hierbei vorerst das Testproblem mit der Alternative  $H_a >$  (Kapitel 2). Unter Verwendung von Lemma 7.1.13 erhalten wir folgende Konsistenzaussage:

*Lemma 9.1.1.:*

$H(u, v)$  sei eine Verteilungsfunktion gemäß Korollar 7.1.11,  $M_n(\varepsilon_1, \varepsilon_2)$  erfülle die Voraussetzungen von Satz 7.1.12 und  $0 < \varepsilon_1 \leq F_X(\mu) \leq \varepsilon_2 < 1$ .

Dann gilt für alle  $\delta > 0$ :

$$P_H(|F_X(\hat{\mu}) - F_X(\mu)| > \delta) \rightarrow 0 \text{ für } n \rightarrow \infty,$$

falls  $\gamma(F_X(\mu)) b_\mu(F_X(\mu)) > \gamma(\rho) b_\mu(\rho)$  für  $\rho \in [\varepsilon_1, \varepsilon_2]$  und  $\rho \neq F_X(\mu)$ .

Beweis:  $\kappa : [\varepsilon_1, \varepsilon_2] \rightarrow \mathbb{R}$ , wobei  $\kappa(\rho) = \gamma(\rho) b_\mu(\rho)$ . Da die Funktion  $\kappa$  für  $\mu$  ihr isoliertes globales Maximum annimmt, existiert für alle  $\delta > 0$  ein  $\eta > 0$ , so daß  $\kappa(\rho) - \kappa(\rho') > \varepsilon$  für  $\rho' \in \{\rho' : |\rho' - \rho| > \delta\}$ .

$T_{n\rho} = \iint J_{n\rho}(u, v) du dv$  konvergiert im Sinne der Supremumsnorm  $P_H$ -f.s. gegen  $\kappa(\rho)$  (Praagman 1988, p.210), somit existiert ein  $n_0(\delta) \in \mathbb{N}$  und für alle  $n \geq n_0(\delta)$  gilt:

$$P_H(|T_{nF_X(\hat{\mu})} - \kappa(F_X(\hat{\mu}))| > \eta/2) = P_H(|M_n(\varepsilon_1, \varepsilon_2) - \kappa(F_X(\hat{\mu}))| > \eta/2) < \delta$$

und  $P_H(|T_{nF_X(\hat{\mu})} - \kappa(F_X(\mu))| > \eta/2) < \delta$  (Lemma 7.1.13), da nach der

Definition des Prüfpunktschätzers gilt:  $T_{nF_X(\hat{\mu})} = M_n(\varepsilon_1, \varepsilon_2)$  (Abschnitt

2.5). Also gilt:  $P_H(|\kappa(F_X(\mu)) - \kappa(F_X(\hat{\mu}))| > \eta) < \delta$  und damit

$$P_H(|F_X(\hat{\mu}) - F_X(\mu)| > \delta) < \delta.$$

Darkhovskh (1976) zeigt für die maximale standardisierte Wilcoxon-Statistik in einem abrupten Strukturbruchmodell die Konsistenz des Wechsellpunktschätzers und zeigt auch, daß die Konvergenzrate  $O(1/n)$  ist. Unter Verwendung des Lemma 9.1.1 beweisen wir für die Lokationsalternative  $H_a^{\vartheta > 0}$ , daß der auf einer maximal selektierten Rangstatistik beruhende Prüfpunktschätzer konsistent ist, sofern die Scoreerzeugende Funktion  $\varphi$  schwache Regularitätsbedingungen erfüllt und der Prüfpunkt gemäß folgender Definition identifizierbar ist.

*Definition 9.1.2.:*

Der Prüfpunkt  $\mu$  ist *identifizierbar*, genau dann wenn ein  $\delta_0 > 0$  existiert, so daß  $F_X(\mu')$  für  $\mu' \in [\mu - \delta_0, \mu + \delta_0]$  bijektiv ist.

*Satz 9.1.3.:*

Die bivariate Verteilungsfunktion  $H$  ist durch die Lokationsalternative  $H_a^{\vartheta > 0}$  bzw. die Verteilungsfunktionen  $F_X$ ,  $F_Y$  und  $G_Y$  und dem Prüfpunkt  $F_X(\mu) \in [\varepsilon_1, \varepsilon_2]$  gegeben.  $\mu$  sei identifizierbar.  $M_n(\varepsilon_1, \varepsilon_2)$  sei gemäß (7.1.5) und (7.1.6). Für die Funktionen  $\varphi$  und  $\psi$  gelte:  $\int \varphi(u) du = 0$ ,  $0 < \int (\varphi(u))^2 du < \infty$ ,  $\int (\varphi(u) - \varphi_n(u))^2 du \rightarrow 0$  für  $n \rightarrow \infty$ ,  $\varphi$  sei monoton

und  $\gamma(\rho) = (\rho(1-\rho))^{-\frac{1}{2}}$  für  $\rho \in [\varepsilon_1, \varepsilon_2]$ . Dann gilt für alle  $\delta > 0$ :

$$P_H(|\hat{\mu} - \mu| > \delta) \rightarrow 0 \text{ für } n \rightarrow \infty. \quad (9.1.1)$$

Beweis: O.B.d.A. sei  $\varphi$  monoton steigend. Die Voraussetzungen von Lemma 9.1.1 werden erfüllt, somit ist zu zeigen, daß:

$$\gamma(F_X(\mu)) b_{\mu}(F_X(\mu)) > \gamma(\rho) b_{\mu}(\rho) \text{ für } \rho \in [\varepsilon_1, \varepsilon_2] \text{ und } \rho \neq F_X(\mu). \quad (9.1.2)$$

Da  $\varphi$  monoton steigend ist gilt für die Lokationsalternative:

$$\begin{aligned} \int \varphi(H_Y(u)) dG_Y(u) &< 0 < \int \varphi(H_Y(u)) dF_Y(u) \text{ und} \\ F_X(\mu) \int \varphi(H_Y(u)) dF_Y(u) + (1 - F_X(\mu)) \int \varphi(H_Y(u)) dG_Y(u) \\ &= \int \varphi(H_Y(u)) d(F_X(\mu) F_Y(u) + (1 - F_X(\mu)) G_Y(u)) = 0, \end{aligned}$$

da  $H_Y = F_X(\mu) F_Y + (1 - F_X(\mu)) G_Y$ .

Also erhalten wir  $\gamma(\rho) b_{\mu}(\rho) = (\rho(1-\rho))^{-\frac{1}{2}} \rho \int \varphi(H_Y(u)) dF_Y(u)$  für  $\varepsilon_1 \leq \rho \leq F_X(\mu)$  und  $\gamma(\rho) b_{\mu}(\rho) = (\rho(1-\rho))^{-\frac{1}{2}} (\rho-1) \int \varphi(H_Y(u)) dG_Y(u)$  für  $F_X(\mu) < \rho \leq \varepsilon_2$ .  $\gamma(\rho) b_{\mu}(\rho)$  ist stetig in  $F_X(\mu)$  und für die Ableitung gilt:

$$(d/d\rho) \gamma(\rho) b_{\mu}(\rho) = 1/(2(1-\rho)(\rho(1-\rho))^{\frac{1}{2}}) \int \varphi(H_Y(u)) dF_Y(u) > 0$$

für  $\varepsilon_1 \leq \rho < F_X(\mu)$  und

$$(d/d\rho) \gamma(\rho) b_{\mu}(\rho) = 1/(2\rho(\rho(1-\rho))^{\frac{1}{2}}) \int \varphi(H_Y(u)) dG_Y(u) < 0$$

für  $F_X(\mu) < \rho \leq \varepsilon_2$ .

Aus der lokalen Bijektivität von  $F_X$  und der Konsistenz der empirischen Verteilungsfunktion folgt; für alle  $\delta$  mit  $\delta_0 > \delta > 0$  existiert ein  $n_0(\delta) \in \mathbb{N}$ , so daß für  $n \geq n_0(\delta)$  gilt:

$$\begin{aligned} P_H(|\hat{\mu} - \mu| > \delta) &< P_H(|F_X(\hat{\mu}) - F_X(\mu)| > \eta) \\ &< P_H(\|F_{nX} - F_X\| > \eta/2, |F_{nX}(\hat{\mu}) - F_X(\mu)| > \eta/2) \\ &< \delta, \end{aligned}$$

wobei  $\eta = \min\{F_X(\mu + \delta) - F_X(\mu), F_X(\mu) - F_X(\mu - \delta)\}$ .

Folglich gilt (9.1.2).

**Korollar 9.1.4.:**

Die Konsistenzaussage von Satz 9.1.3 gilt für die Alternativen des

zweiseitigen Testproblems  $H_a^{\vartheta \neq 0}$ , sofern  $M_n(\varepsilon_1, \varepsilon_2)$  die zweiseitige Teststatistik aus Kapitel 2 ist.

Beweis: Für  $H_a^{\vartheta > 0}$  gilt  $\gamma(\rho) b_\mu(\rho) > 0$  für  $\rho \in (0, 1)$  und für  $H_a^{\vartheta < 0}$  gilt  $\gamma(\rho) b_\mu(\rho) < 0$  für  $\rho \in (0, 1)$ , also gilt:

$$\gamma(F_X(\mu)) |b_\mu(F_X(\mu))| > \gamma(\rho) |b_\mu(\rho)| \text{ für } \rho \in [\varepsilon_1, \varepsilon_2] \text{ und } \rho \neq F_X(\mu).$$

*Korollar 9.1.5.:*

Die Konsistenzaussage von Satz 9.1.3 gilt für die Alternative  $H_a^>$ , falls  $\varphi$  streng monoton ist.

Beweis: O.B.d.A. sei  $\varphi$  streng monoton steigend. Unter  $H_a^>$  gilt:  $F_Y < H_Y < G_Y$ . Da  $\varphi$  streng monoton steigend ist, folgt somit:

$$\int \varphi(H_Y(u)) dG_Y(u) < 0 < \int \varphi(H_Y(u)) dF_Y(u).$$

Die Voraussetzungen von Satz 9.1.3 werden beispielsweise von der Median-Statistik und der Wilcoxon-Statistik erfüllt.

## 9.2. Monte-Carlo Ergebnisse

Nachdem wir im vorangegangenen Abschnitt eine asymptotische Konsistenzaussage hergeleitet haben, betrachten wir nun das Verhalten des Prüfpunktschätzers bei moderatem Stichprobenumfang für verschiedene Alternativen. Hierzu verwenden wir die für Kapitel 8 durchgeführten Monte-Carlo Berechnungen und beschränken uns somit auf den Stichprobenumfang  $n = 50$  und auf eine auf dem Intervall  $[0, 1]$  gleichverteilte Einflußgröße  $X$ .

### *Prüfpunktalternative*

Die Abb.9.2.1 veranschaulicht die simulierten Verteilungsfunktionen des durch die maximal selektierte Wilcoxon-Statistik gegebenen Prüfpunktschätzers bei normalverteiletem Fehler für verschiedene Lageparameter  $\vartheta$ . Der Prüfpunktparameter ist hierbei 0.5, und der Prüfpunktschätzer tendiert zu einem leichten Unterschätzen. Dieses

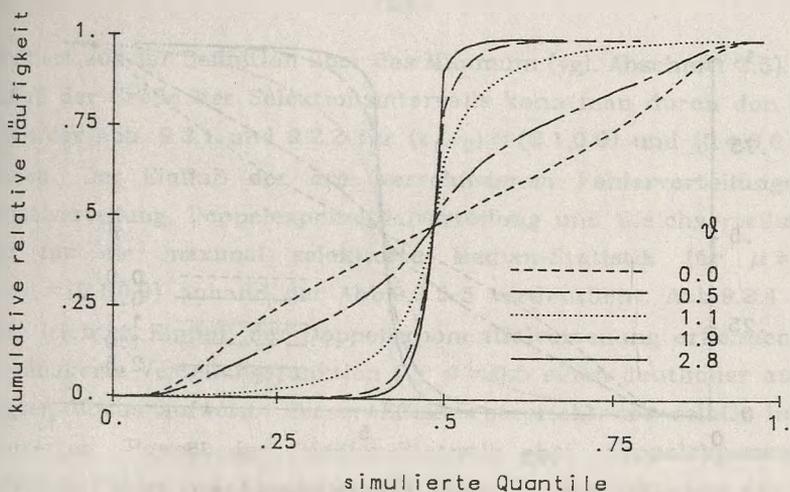


Abb. 9.2.1.: Simulierte Verteilungsfunktionen bei normalverteiltem Fehler des durch die maximal selektierte Wilcoxon-Statistik gegebenen Prüfpunktschätzers für verschiedene Lageparameter  $\vartheta$  bei  $(\varepsilon_1, \varepsilon_2) = (0.1, 0.9)$  und  $\mu = 0.5$ .

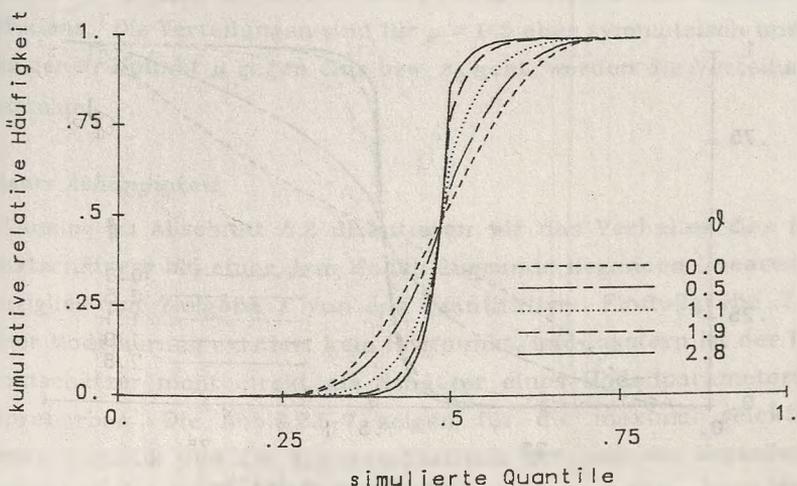


Abb. 9.2.2.: Simulierte Verteilungsfunktionen bei normalverteiltem Fehler des durch die maximal selektierte Wilcoxon-Statistik gegebenen Prüfpunktschätzers für verschiedene Lageparameter  $\vartheta$  bei  $(\varepsilon_1, \varepsilon_2) = (0.1, 0.9)$  und  $\mu = 0.5$ .

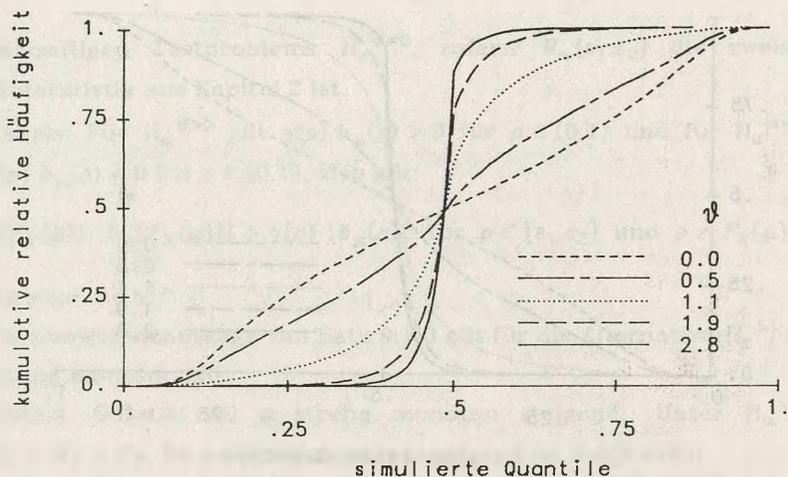


Abb. 9.2.3.: Simulierte Verteilungsfunktionen bei normalverteiltem Fehler des durch die maximal selektierte Median-Statistik gegebenen Prüfpunktschätzers für verschiedene Lageparameter  $\vartheta$  bei  $(\varepsilon_1, \varepsilon_2) = (0.1, 0.9)$  und  $\mu = 0.5$ .

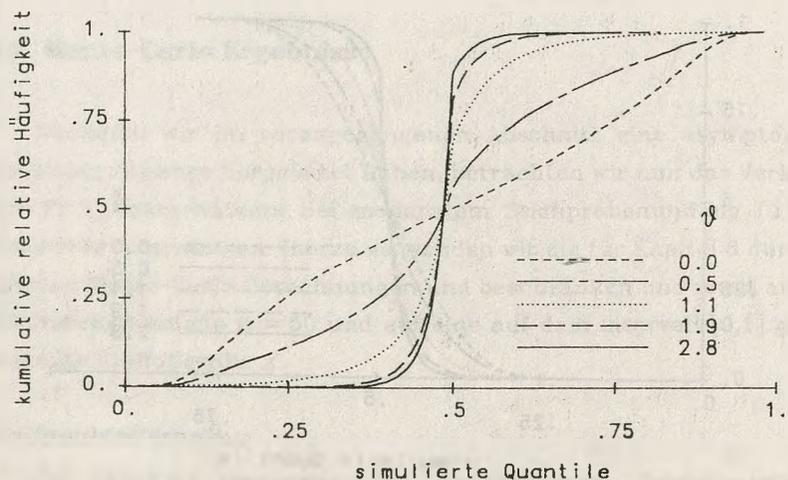


Abb. 9.2.4.: Simulierte Verteilungsfunktionen bei doppel exponentialverteiltem Fehler des durch die maximal selektierte Median-Statistik gegebenen Prüfpunktschätzers für verschiedene Lageparameter  $\vartheta$  bei  $(\varepsilon_1, \varepsilon_2) = (0.1, 0.9)$  und  $\mu = 0.5$ .

resultiert aus der Definition über das Minimum (vgl. Abschnitt 2.5). Der Einfluß der Größe des Selektionsintervalls kann man durch den Vergleich der Abb. 9.2.1 und 9.2.2 für  $(\varepsilon_1, \varepsilon_2) = (0.1, 0.9)$  und  $(0.4, 0.6)$  erkennen. Der Einfluß der drei verschiedenen Fehlerverteilungen - Normalverteilung, Doppel exponentialverteilung und Gleichverteilung - wird für die maximal selektierte Median-Statistik für  $\mu = 0.5$ ,  $(\varepsilon_1, \varepsilon_2) = (0.1, 0.9)$  anhand der Abb. 9.2.3-5 verdeutlicht. Abb. 9.2.4 läßt einen leichten Einfluß der Doppel exponentialverteilung erkennen, da die simulierte Verteilungsfunktion für  $\vartheta = 0.5$  einen deutlicher ausgeprägten Modus aufweist. Dieser Effekt entspricht der relativ hohen simulierten Power der Median-Statistik bei doppel exponentialverteiltem Fehler (vgl. Abschnitt 8.1). Unter Berücksichtigung der hier nicht dargestellten Parameterkombinationen erhalten wir damit für die maximal selektierte Wilcoxon-Statistik ein ähnliches Resultat wie Pettitt (1980) für den Wechselpunktschätzer: "*The distribution of the Mann-Whitney type estimate remains fairly constant over various distributions.*" Die Verteilungen sind für  $\mu = 0.5$  eher symmetrisch und, indem der Prüfpunkt  $\mu$  gegen eins bzw.  $\varepsilon_2$  geht, werden die Verteilungen linksschief.

### *Lineare Abhängigkeit*

Analog zu Abschnitt 8.2 diskutieren wir das Verhalten des Prüfpunktschätzers bei einer dem Modell zugrunde liegenden linearen Abhängigkeit der Zielgröße  $Y$  von der quantitativen Einflußgröße  $X$ . Bei dieser Modellierung existiert kein Prüfpunkt, und insofern ist der Prüfpunktschätzer nicht direkt als Schätzer eines Modellparameters interpretierbar. Die Abb. 9.2.6-7 zeigen für die maximal selektierte Median-Statistik und die Wilcoxon-Statistik auf, daß die Verteilungen des Prüfpunktschätzers für größere Steigungsparameter einen Modalwert bei etwa 0.5 haben. Allerdings ist die Abhängigkeit vom Steigungsparameter  $\vartheta$  gegenüber der für das Prüfpunktmodell beschriebenen Abhängigkeit vom Lageparameter  $\vartheta$  vergleichsweise gering. Abb. 9.2.8 veranschaulicht, daß diese Abhängigkeit mit der Größe des Selektions-

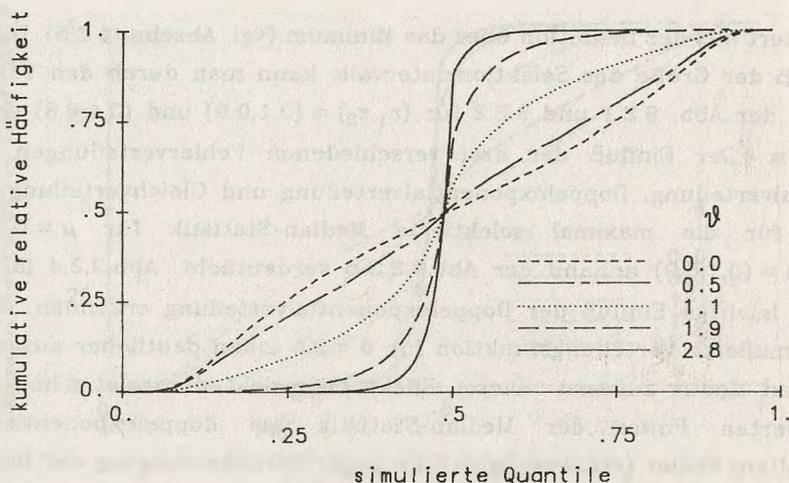


Abb. 9.2.5.: Simulierte Verteilungsfunktionen bei gleichverteiltem Fehler des durch die maximal selektierte Median-Statistik gegebenen Prüfpunktschätzers für verschiedene Lageparameter  $\vartheta$  bei  $(\varepsilon_1, \varepsilon_2) = (0.1, 0.9)$  und  $\mu = 0.5$ .

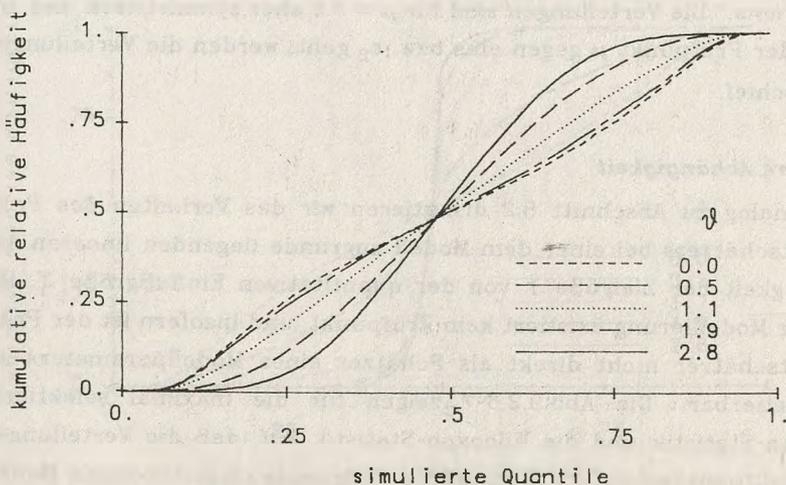


Abb. 9.2.6.: Simulierte Verteilungsfunktionen bei normalverteiltem Fehler des durch die maximal selektierte Median-Statistik gegebenen Prüfpunktschätzers für verschiedene Steigungsparameter  $\vartheta$  bei  $(\varepsilon_1, \varepsilon_2) = (0.1, 0.9)$ .

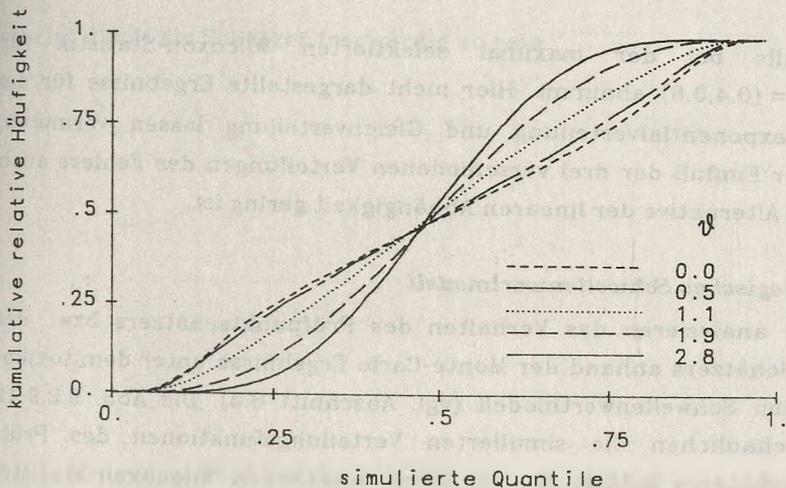


Abb. 9.2.7.: Simulierte Verteilungsfunktionen bei normalverteiltem Fehler des durch die maximal selektierte Wilcoxon-Statistik gegebenen Prüfpunktschätzers für verschiedene Steigungsparameter  $\psi$  bei  $(\epsilon_1, \epsilon_2) = (0.1, 0.9)$ .

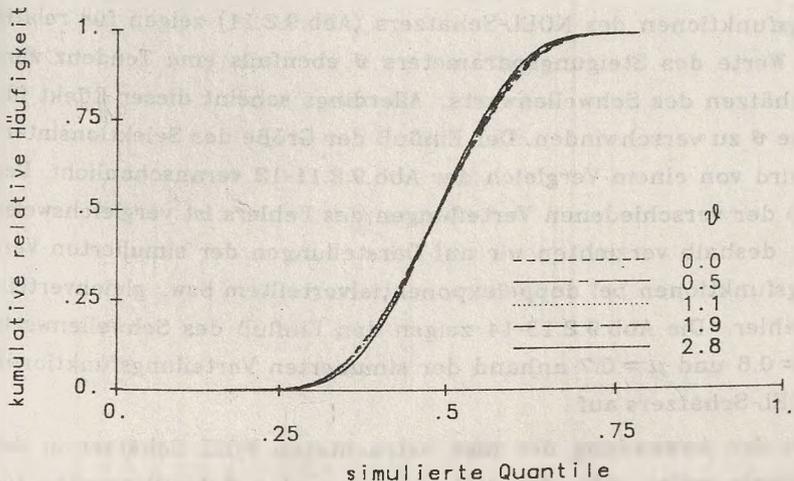


Abb. 9.2.8.: Simulierte Verteilungsfunktionen bei normalverteiltem Fehler des durch die maximal selektierte Wilcoxon-Statistik gegebenen Prüfpunktschätzers für verschiedene Steigungsparameter  $\psi$  bei  $(\epsilon_1, \epsilon_2) = (0.4, 0.6)$ .

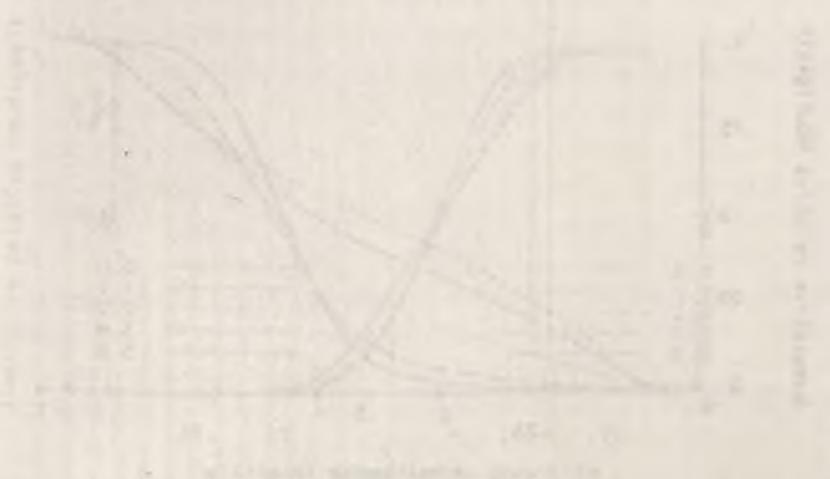
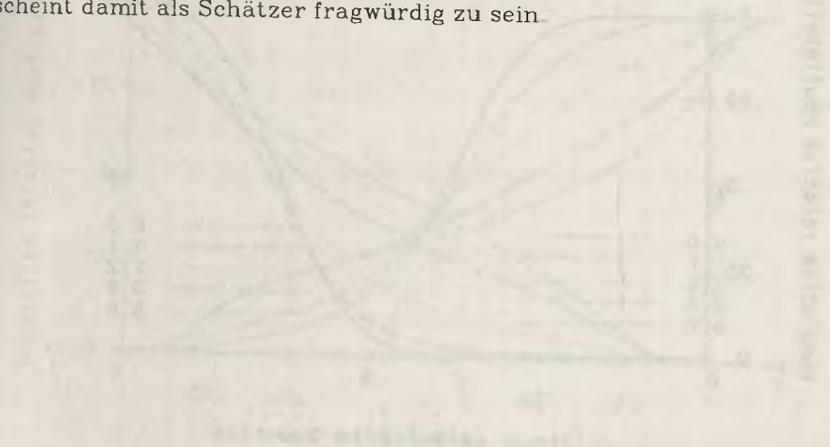
intervalls bei der maximal selektierten Wilcoxon-Statistik für  $(\varepsilon_1, \varepsilon_2) = (0.4, 0.6)$  abnimmt. Hier nicht dargestellte Ergebnisse für die Doppel exponentialverteilung und Gleichverteilung lassen vermuten, daß der Einfluß der drei verschiedenen Verteilungen des Fehlers auch für die Alternative der linearen Abhängigkeit gering ist.

#### *Toxikologisches Schwellenwertmodell*

Wir analysieren das Verhalten des Prüfpunktschätzers bzw. des NOEL-Schätzers anhand der Monte-Carlo Ergebnisse unter dem toxikologischen Schwellenwertmodell (vgl. Abschnitt 8.3). Die Abb. 9.2.9-10 veranschaulichen die simulierten Verteilungsfunktionen des Prüfpunktschätzers mittels der maximal selektierten Wilcoxon-Statistik bzw. Median-Statistik bei normalverteiltem Fehler,  $(\varepsilon_1, \varepsilon_2) = (0.1, 0.9)$  und  $\mu = 0.5$ . Der Prüfpunktschätzer tendiert dazu, den Schwellenwert zu überschätzen; dies erklärt die größere simulierte Power bei einem größeren Selektionsintervall (vgl. Abschnitt 8.3). Die simulierten Verteilungsfunktionen des NOEL-Schätzers (Abb. 9.2.11) zeigen für relativ kleine Werte des Steigungsparameters  $\vartheta$  ebenfalls eine Tendenz zum Überschätzen des Schwellenwerts. Allerdings scheint dieser Effekt für größere  $\vartheta$  zu verschwinden. Der Einfluß der Größe des Selektionsintervalls wird von einem Vergleich der Abb. 9.2.11-12 veranschaulicht. Der Einfluß der verschiedenen Verteilungen des Fehlers ist vergleichsweise gering, deshalb verzichten wir auf Darstellungen der simulierten Verteilungsfunktionen bei doppel exponentialverteiltem bzw. gleichverteiltem Fehler. Die Abb. 9.2.13-14 zeigen den Einfluß des Schwellenwerts für  $\mu = 0.6$  und  $\mu = 0.7$  anhand der simulierten Verteilungsfunktionen des NOEL-Schätzers auf.

Bei der Anwendung der hier betrachteten NOEL-Schätzer in der Toxikologie sollte man das Unterschätzen des Schwellenwertes für kleine  $\vartheta$  und die offensichtlich langsame Konvergenzgeschwindigkeit des NOEL-Schätzers gegen  $\mu$  für wachsende  $\vartheta$  berücksichtigen. Der durch die maximal selektierte Rangstatistik gegebene Prüfpunktschätzer tendiert zum deutlichen Überschätzen des Schwellenwertes und

scheint damit als Schätzer fragwürdig zu sein.



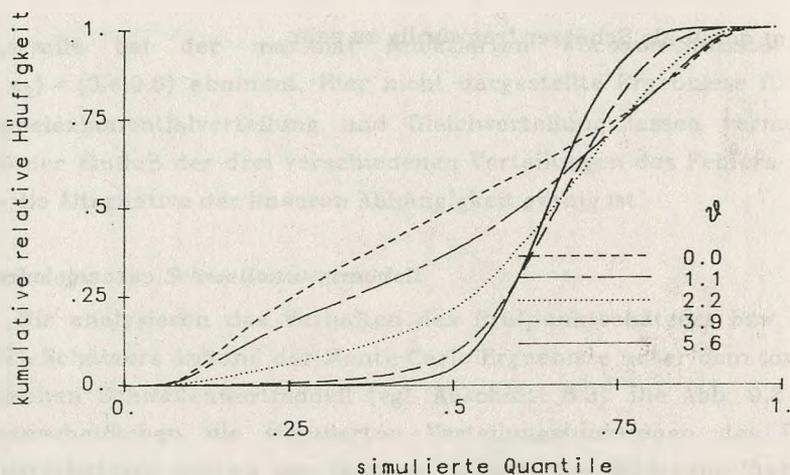


Abb. 9.2.9.: Simulierte Verteilungsfunktionen bei normalverteiltem Fehler des durch die maximal selektierte Wilcoxon-Statistik gegebenen Prüfpunktschätzers für verschiedene Steigungsparameter  $\vartheta$  bei  $(\varepsilon_1, \varepsilon_2) = (0.1, 0.9)$  und dem Schwellenwert  $\mu = 0.5$ .

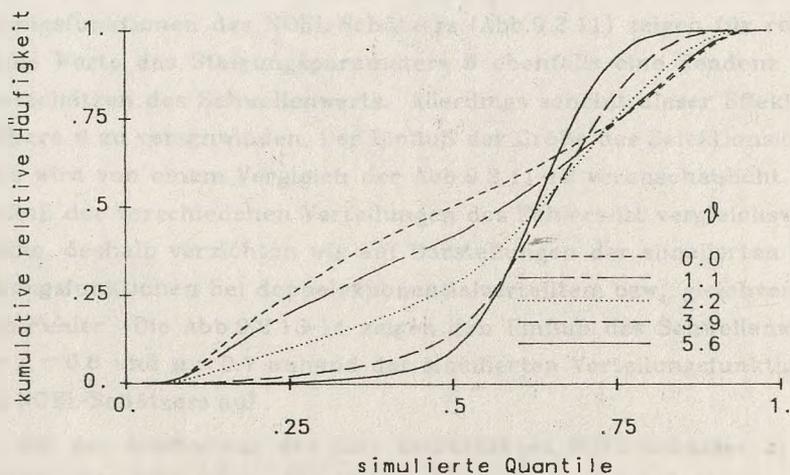


Abb. 9.2.10.: Simulierte Verteilungsfunktionen bei normalverteiltem Fehler des durch die maximal selektierte Median-Statistik gegebenen Prüfpunktschätzers für verschiedene Steigungsparameter  $\vartheta$  bei  $(\varepsilon_1, \varepsilon_2) = (0.1, 0.9)$  und dem Schwellenwert  $\mu = 0.5$ .

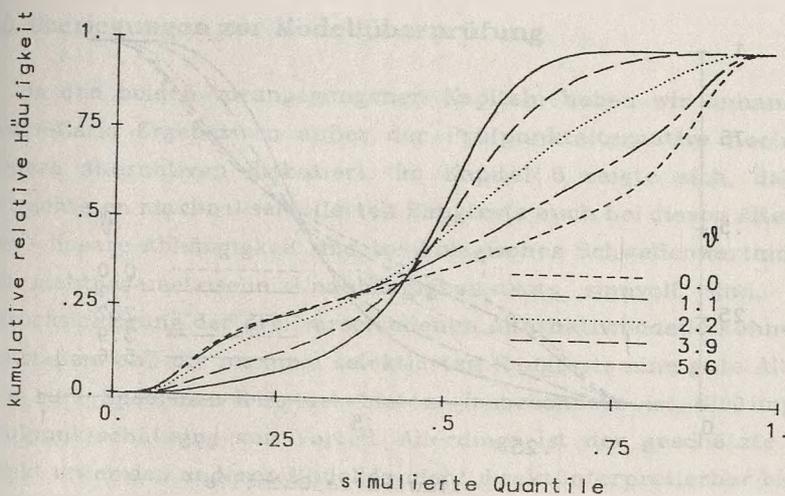


Abb. 9.2.11.: Simulierte Verteilungsfunktionen bei normalverteiltem Fehler des durch die maximal selektierte NOEL-Statistik gegebenen NOEL-Schätzers für verschiedene Steigungsparameter  $\vartheta$  bei  $(\epsilon_1, \epsilon_2) = (0.1, 0.9)$  und dem Schwellenwert  $\mu = 0.5$ .

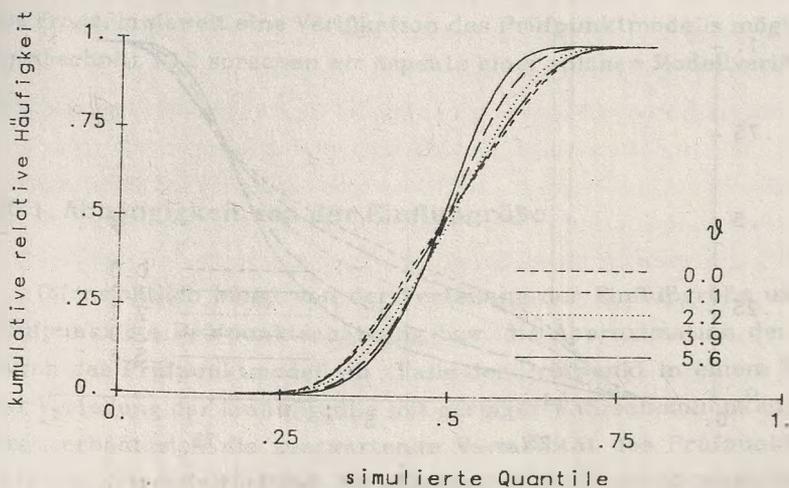


Abb. 9.2.12.: Simulierte Verteilungsfunktionen bei normalverteiltem Fehler des durch die maximal selektierte NOEL-Statistik gegebenen NOEL-Schätzers für verschiedene Steigungsparameter  $\vartheta$  bei  $(\epsilon_1, \epsilon_2) = (0.4, 0.6)$  und dem Schwellenwert  $\mu = 0.5$ .

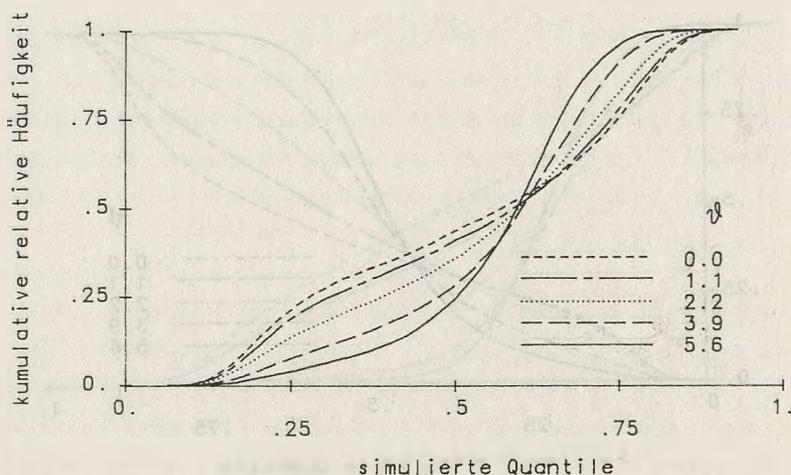


Abb. 9.2.13.: Simulierte Verteilungsfunktionen bei normalverteiltem Fehler des durch die maximal selektierte NOEL-Statistik gegebenen NOEL-Schätzers für verschiedene Steigungsparameter  $\vartheta$  bei  $(\varepsilon_1, \varepsilon_2) = (0.2, 0.8)$  und dem Schwellenwert  $\mu = 0.6$ .

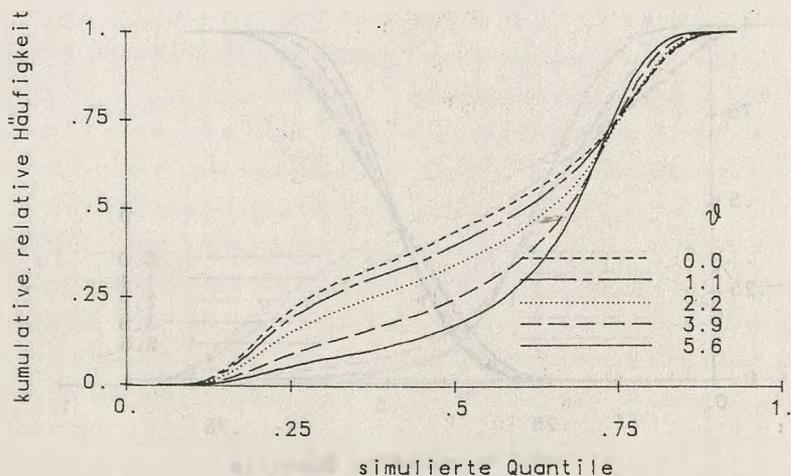


Abb. 9.2.14.: Simulierte Verteilungsfunktionen bei normalverteiltem Fehler des durch die maximal selektierte NOEL-Statistik gegebenen NOEL-Schätzers für verschiedene Steigungsparameter  $\vartheta$  bei  $(\varepsilon_1, \varepsilon_2) = (0.2, 0.8)$  und dem Schwellenwert  $\mu = 0.7$ .

## 10. Überlegungen zur Modellüberprüfung

In den beiden vorangegangenen Kapiteln haben wir anhand von Monte-Carlo Ergebnissen außer der Prüfpunktalternative noch zwei weitere Alternativen diskutiert. Im Kapitel 8 zeigte sich, daß die betrachteten maximal selektierten Rangtests auch bei diesen Alternativen - lineare Abhängigkeit und toxikologisches Schwellenwertmodell - als nichtparametrische Unabhängigkeitstests sinnvoll sind. Unter Berücksichtigung der drei verschiedenen Alternativmodelle können wir feststellen, daß die maximal selektierten Rangtests eine gute Alternative zum Spearman-Rangtest bieten. Insbesondere ist die implizite Prüfpunktschätzung von Vorteil. Allerdings ist der geschätzte Prüfpunkt unter den anderen Modellen nicht direkt interpretierbar bzw. als Parameter ein Artefakt der Methode (vgl. Abschnitt 9.2). Deshalb erscheint eine Bewertung des geschätzten Prüfpunkts bzw. des verwendeten Prüfpunktmodells geboten. Hierbei ist die Verteilung der Einflußgröße zuberücksichtigend (Abschnitt 10.1). Des weiteren stellt sich die Frage, inwieweit eine Verifikation des Prüfpunktmodells möglich ist. Im Abschnitt 10.2 sprechen wir Aspekte einer solchen Modellverifikation an.

### 10.1. Abhängigkeit von der Einflußgröße

Offensichtlich hängt von der Verteilung der Einflußgröße und dem Prüfpunkt die Prüfpunktschätzung bzw. die Approximation der Daten durch das Prüfpunktmodell ab. Falls der Prüfpunkt in einem Bereich der Verteilung der Einflußgröße mit geringer Wahrscheinlichkeitsmasse liegt, erhöht sich die zuerwartende Variabilität des Prüfpunktschätzers. Im Extremfall ist der Parameter Prüfpunkt nicht identifizierbar (vgl. Definition 9.1.2 und Satz 9.1.3). Falls der Einflußgröße eine ausgeprägte bimodale Verteilung zugrunde liegt, werden beispielsweise auch Beobachtungen bei einem linearen Zusammenhang zwischen der Einflußgröße und der Zielgröße größtenteils durch das Prüfpunktmodell

gut approximiert. Diese Problematik ergänzt die triviale Erkenntnis, daß man funktionale Zusammenhänge nur in Bereichen, in denen ausreichend viele Beobachtungen vorliegen, überprüfen kann. Bei der Interpretation des Prüfpunktschätzers sind diese trivialen Zusammenhänge allerdings von besonderer Bedeutung.

## 10.2. Zur Modellverifikation

In diesem Abschnitt gehen wir auf vorläufige Vorschläge einer Verifikation des Prüfpunktmodells ein. Der Pfad des durch die verwendete Statistik gegebenen Prozesses (siehe Abb.1.1.1) beinhaltet Informationen über die Güte der Approximation der Daten durch das Prüfpunktmodell. Ein Pfad, der genau einen ausgeprägten Extremwert hat, scheint für das Prüfpunktmodell zu sprechen; dagegen stellen mehrere ausgeprägte Extremwerte gleicher Größenordnung die Prüfpunktalternative in Frage. Dasselbe gilt, wenn man von keinem ausgeprägten Extremwert sprechen kann.

Die Stabilität des Prüfpunktschätzers ergibt eine weitere Möglichkeit, das Prüfpunktmodell zu evaluieren. Die Bootstrapverteilung des verwendeten Prüfpunktschätzers kann man für eine Stabilitätsbeurteilung verwenden. Hierbei sollte man allerdings die Ergebnisse vom Abschnitt 9.2 beachten, denn beispielsweise hat die Verteilungsfunktion des Prüfpunktschätzers unter einer linearen Abhängigkeit auch einen deutlichen Extremwert. Die Verwendung einer modifizierten - evt. nicht standardisierten; d.h. anders skalierten - Teststatistik kann diesen Nachteil beheben. Als Konstruktionsprinzip kann man hierbei die  $P_H-f$ s. Grenzfunktion der Teststatistik (vgl. Lemma 9.1.1) benutzen. Diese Grenzfunktion sollte beispielsweise bei vorliegender linearer Abhängigkeit keinen Extremwert oder Extremwerte am Rand des Selektionsintervalls aufweisen.

## 11. Anwendungen in der klinischen Forschung

In den vorangegangenen Kapiteln haben wir die nichtparametrische Methode der maximal selektierten Rangstatistiken eingeführt und ihre Eigenschaften ausführlich diskutiert. Es soll nun die Relevanz und Anwendbarkeit dieser Methodik an Beispielen aus der klinischen Forschung aufgezeigt werden. Im Abschnitt 11.1 besprechen wir die in der Einleitung (Abschnitt 1.1) vorgestellte klinische Studie zur antiarrhythmischen Therapie von Hohnloser u.a. (1987). Bei einer Brustkrebsstudie der Universitäts-Frauenklinik Erlangen wurden unter anderem verschiedene Hormon-Rezeptoren und Tumor-Marker analysiert. Wir betrachten hierbei insbesondere die Abhängigkeit der Rezeptoren vom Alter der Patientinnen (Abschnitt 11.2).

### 11.1. Studie zur antiarrhythmischen Therapie

Im Abschnitt 1.1 haben wir die Problemstellung der vorliegenden Arbeit anhand der klinischen Studie von Hohnloser u.a. (1987) erläutert. Hierbei beschränkten wir uns auf die Einflußgröße linksventrikuläre Ejektionsfraktion (EF) und die Zielgröße rezidivfreie Überlebenszeit nach der Klinikentlassung, wobei die Diagnose eines Arrhythmierzidives oder eines plötzlichen Herztodes als terminierendes Ereignis für die Zeitdauer verwendet wurde. Nach dieser Definition liegt die rezidivfreie Überlebenszeit von 66 Patienten als zensierte Beobachtung vor. Hohnloser u.a. (1987) untersuchten nicht nur den diagnostischen Wert der EF, sondern auch den diagnostischen Wert des Anteils wirkender Medikamente (WM). Hierbei wurden zwischen drei und 10 verschiedene Medikamente zur Suppression von Arrhythmien ausprobiert. Eine detaillierte Beschreibung dieses Experiments und der Definition von WM findet sich bei Hohnloser u.a. (1987). Im folgenden beurteilen wir den diagnostischen Wert der EF und der WM mittels maximal selektierter Log-Rank-Statistiken.

Abb.11.1.1 veranschaulicht die empirische Verteilung der EF, welche bei der Verwendung des Prüfpunktmodells von besonderer Bedeutung ist (vgl. Abschnitt 10.1). Die empirische Verteilung der EF ist zwischen 10% und 70% EF relativ gleichförmig und weist keinen ausgeprägten Modalwert auf. Allerdings treten Bindungen auf. Die größte Bindungslänge beträgt sieben bei 24% EF. Grundsätzlich könnten wir somit das Vorgehen der endlichen Prüfpunktselektion bei diskreter Einflußgröße wählen (Abschnitt 6.2). Da 45 verschiedene Ausprägungen der EF vorliegen, entscheiden wir uns dennoch für die Approximation anhand der asymptotischen Verteilung bei stetiger Einflußgröße (vgl. Abschnitt 5.4 und 6.1).

Unter der Annahme, daß die Anteile der Gruppe mit 'erhöhtem Risiko' bzw. der mit 'normalem Risiko' mindestens 10% betragen sollten, führen wir die Selektion zwischen dem 10% und 90% Quantil der EF durch. Wir erhalten für die maximal selektierte Log-Rank-Statistik einen approximativen  $P$ -Wert von etwa 0.01, und die implizit gegebene Prüfpunktschätzung ergibt 39% EF (vgl. Abb.1.1.1). Der Unterschied zwischen den beiden durch den geschätzten Prüfpunkt gegebenen Gruppen wird durch die Kaplan-Meier Schätzer veranschaulicht (Abb.1.1.2).

Zur Beurteilung der klinischen Bedeutung des maximal selektierten Prüfpunkts 39% EF müssen wir das Prüfpunktmodell diskutieren (vgl. Abschnitt 10.2). Der durch die Log-Rank-Statistik gegebene Pfad (Abb.1.1.1) weist keinen ausgeprägten Extremwert auf, deshalb interpretieren wir das Prüfpunktmodell nur als eine erste Approximation eines funktionalen Zusammenhangs der EF und der rezidivfreien Überlebenszeit. Eine explorative Analyse der Abhängigkeit der Zielgröße von möglichen Einflußgrößen (vgl. Hohnloser u.a. 1987) kann man mittels komplexerer Regressionsmodelle bzw. anderer Regressionsfunktionen durchführen. Hierbei bietet sich beispielsweise das Coxsche Regressionsmodell an. Zusammenfassend können wir feststellen, daß der Prüfpunkt 39% EF den Unterschied zwischen den Gruppen im Sinne der standardisierten Log-Rank-Statistik maximiert und dieser

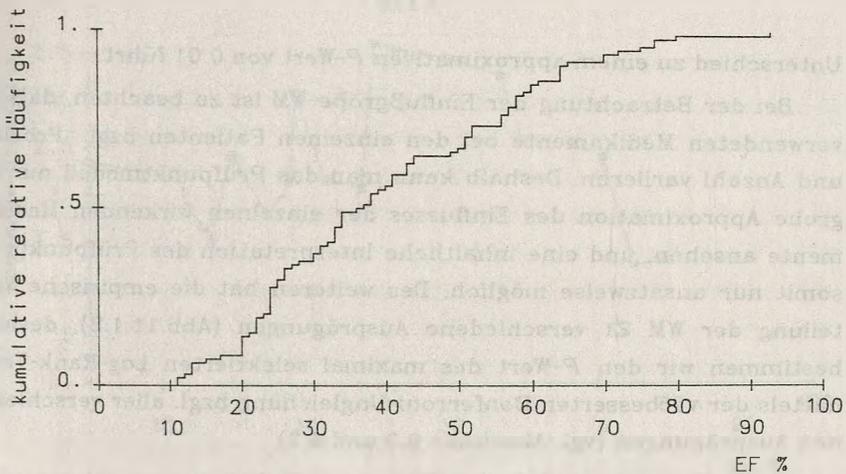


Abb. 11.1.1.: Die empirische Verteilung der linksventrikulären Ejektionsfraktion (EF).

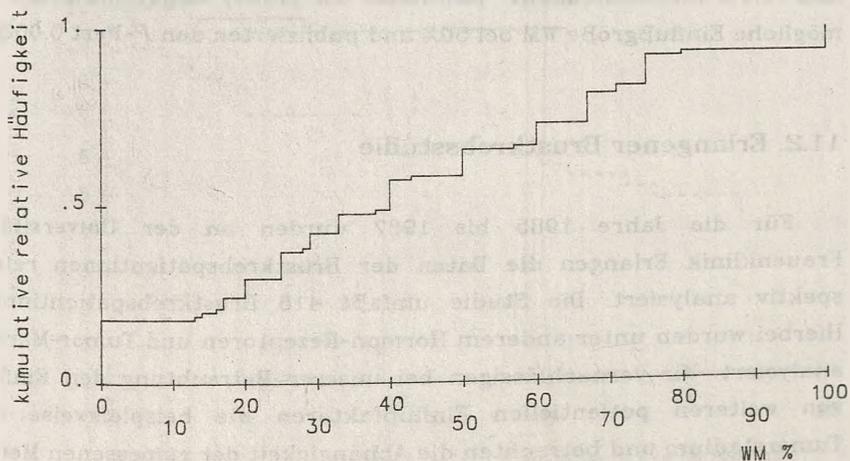


Abb. 11.1.2.: Die empirische Verteilung des Anteils von wirkenden Medikamenten (WM).

Unterschied zu einem approximativen  $P$ -Wert von 0.01 führt.

Bei der Betrachtung der Einflußgröße WM ist zu beachten, daß die verwendeten Medikamente bei den einzelnen Patienten bzgl. Produkt und Anzahl variieren. Deshalb kann man das Prüfpunktmodell nur als grobe Approximation des Einflusses der einzelnen wirkenden Medikamente ansehen, und eine inhaltliche Interpretation des Prüfpunkts ist somit nur ansatzweise möglich. Des weiteren hat die empirische Verteilung der WM 21 verschiedene Ausprägungen (Abb.11.1.2), deshalb bestimmen wir den  $P$ -Wert des maximal selektierten Log-Rank-Tests mittels der verbesserten Bonferroni-Ungleichung bzgl. aller verschiedenen Ausprägungen (vgl. Abschnitt 5.3 und 6.2).

Die Nullhypothese, daß die WM keinen Einfluß hat, ergibt einen  $P$ -Wert von 0.0035. Als maximal selektierten Prüfpunkt erhalten wir 43% WM (Abb.11.1.3), und der Unterschied zwischen den beiden hierdurch gegebenen Gruppen wird von den jeweiligen Kaplan-Meier Schätzern in Abb.11.1.4 veranschaulicht. Hohnloser u.a. (1987) dichotomisierten die mögliche Einflußgröße WM bei 50% und publizierten den  $P$ -Wert 0.0028.

## 11.2. Erlangener Brustkrebsstudie

Für die Jahre 1985 bis 1987 wurden an der Universitäts-Frauenklinik Erlangen die Daten der Brustkrebspatientinnen retrospektiv analysiert. Die Studie umfaßt 418 Brustkrebspatientinnen. Hierbei wurden unter anderem Hormon-Rezeptoren und Tumor-Marker analysiert. Wir vernachlässigen bei unserer Betrachtung den Einfluß von weiteren potentiellen Einflußfaktoren wie beispielsweise das Tumorstadium und betrachten die Abhängigkeit der gemessenen Menge - im folgenden Dichte - der Rezeptoren von Oestrogen und Progesteron vom Alter der Patientinnen als Fragestellung. Die Dichte der Rezeptoren sind interessante Zielgrößen, da sie als mögliche prognostische Kriterien in der Literatur diskutiert werden und in verschiedenen Studien auch als Entscheidungskriterium für die Wahl der Therapie

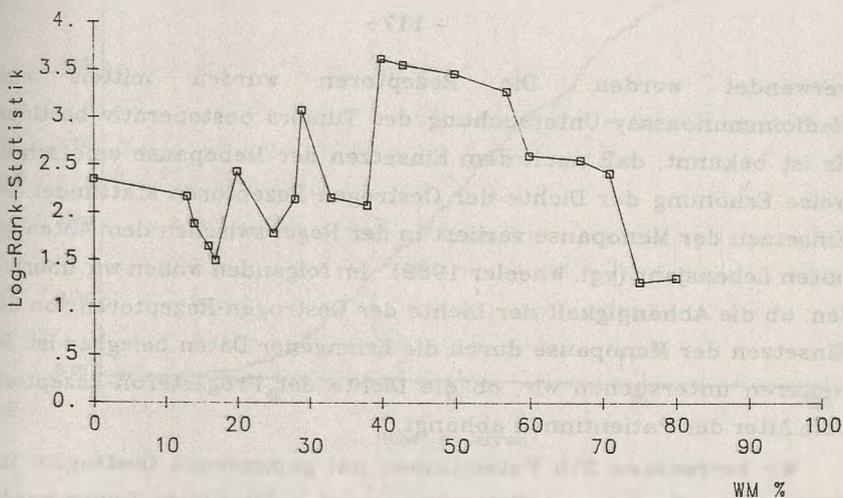


Abb. 11.1.3.: Die standardisierte Log-Rank-Statistik als Prozeß in Abhängigkeit der möglichen Prüfpunkte des Anteils von wirkenden Medikamenten (WM).

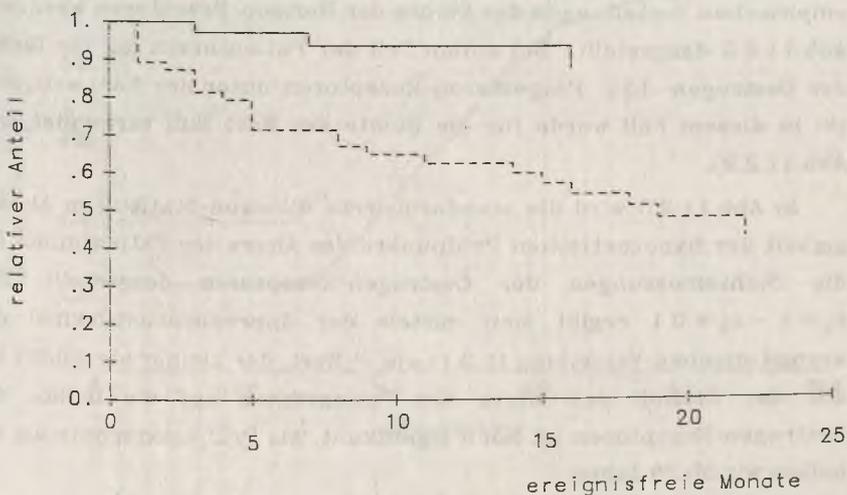


Abb. 11.1.4.: Der prädiktive Wert des maximal selektierten Prüfpunkts 43% WM wird veranschaulicht von den Kaplan-Meier Schätzern für die Patienten mit WM ≤ 43% (gestrichelte Linie) und WM > 43% (durchgezogene Linie).

verwendet werden. Die Rezeptoren wurden mittels einer Radioimmunoassay-Untersuchung des Tumors postoperativ bestimmt. Es ist bekannt, daß nach dem Einsetzen der Menopause eine schrittweise Erhöhung der Dichte der Oestrogen-Rezeptoren stattfindet. Das Einsetzen der Menopause variiert in der Regel zwischen dem 45ten und 55ten Lebensjahr (vgl. Wheeler 1989). Im folgenden wollen wir überprüfen, ob die Abhängigkeit der Dichte der Oestrogen-Rezeptoren von dem Einsetzen der Menopause durch die Erlangener Daten belegbar ist. Des weiteren untersuchen wir, ob die Dichte der Progesteron-Rezeptoren vom Alter der Patientinnen abhängt.

Wir betrachten 378 Patientinnen mit gemessenen Oestrogen- und Progesteron-Rezeptoren. Patientinnen mit beidseitigem Tumor werden nicht berücksichtigt. Die empirische Verteilung des Alters dieser 378 Patientinnen ist in Abb.11.2.1 dargestellt. Die Jahrgänge zwischen dem 45ten und 75ten Lebensjahr sind relativ gleichförmig vertreten. Die empirischen Verteilungen der Dichte der Hormon-Rezeptoren werden in Abb.11.2.2 dargestellt. Bei einem Teil der Patientinnen lag die Dichte der Oestrogen- bzw. Progesteron-Rezeptoren unter der Nachweisgrenze; in diesem Fall wurde für die Dichte der Wert Null verwendet (vgl. Abb.11.2.2).

In Abb.11.2.3 wird die standardisierte Wilcoxon-Statistik in Abhängigkeit der hypothetischen Prüfpunkte des Alters der Patientinnen für die Dichtemessungen der Oestrogen-Rezeptoren dargestellt. Für  $\varepsilon_1 = 1 - \varepsilon_2 = 0.1$  ergibt sich mittels der Approximationsformel der asymptotischen Verteilung (3.3.1) ein  $P$ -Wert, der kleiner als 0.0001 ist; d.h. der Einfluß des Alters der Patientinnen auf die Dichte der Oestrogen-Rezeptoren ist hoch signifikant. Als Prüfpunktschätzung erhalten wir 58.76 Jahre.

Bei der Interpretation des geschätzten Prüfpunkts 58.76 Jahre ist zu beachten, daß die Zunahme der Dichte der Oestrogen-Rezeptoren nach der Menopause nicht abrupt - wie beim Prüfpunktmodell angenommen - sondern eher kontinuierlich abläuft. Deshalb wäre das toxiologische Schwellenwertmodell aus Abschnitt 8.3 eine naheliegende

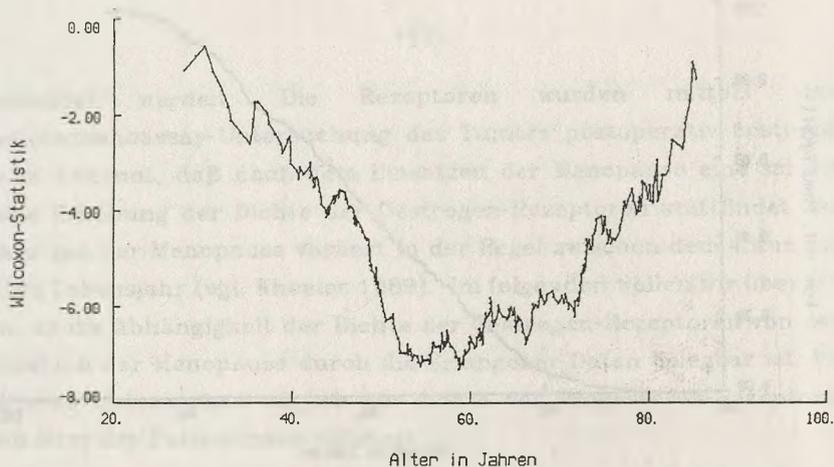


Abb. 11.2.3.: Oestrogen-Rezeptoren: Die standardisierte Wilcoxon-Statistik als Prozeß in Abhängigkeit der möglichen Prüfpunkte bzgl. des Alters der Patientinnen.

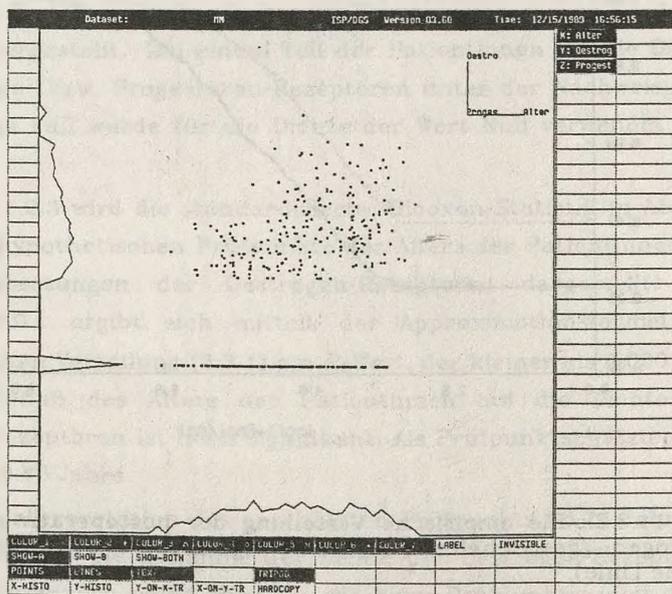


Abb. 11.2.4.: Darstellung mittels PRIM-H; Alter in Jahren gegen Oestrogen  $\log(1+\text{fmol}/\text{mg})$ . Des weiteren wird eine Dichteschätzung der Randverteilungen abgebildet.

Modellierung. Die Monte-Carlo Ergebnisse für die Prüfpunktschätzung aus Abschnitt 9.2 lassen somit vermuten, daß der Zeitpunkt der Menopause durch die maximal selektierte Wilcoxon-Statistik überschätzt wird. Die hohen absoluten Werte der Wilcoxon-Statistik zwischen dem 55ten und 75ten Lebensjahr (Abb.11.2.3) kann man ebenfalls als Hinweis auf eine kontinuierliche Zunahme der Dichte der Oestrogen-Rezeptoren werten. Des weiteren ist zu berücksichtigen, daß das Einsetzen der Menopause individuell variiert (vgl. auch Wheeler 1989). Somit kann man die Prüfpunktschätzung nur als Schätzung eines mittleren Zeitpunkts der Menopause der betrachteten Population von Tumorpatientinnen interpretieren, und die Einflußgröße Lebensalter kann man als eine fehlerhafte Bestimmung eines ideellen 'biologischen Alters' ansehen. Letzterer Aspekt erschwert die Anwendung von Glättungsverfahren zur explorativen Analyse des funktionalen Zusammenhangs zwischen dem (biologischen) Alter der Patientinnen und der Dichte der Oestrogen-Rezeptoren. In der Abb.11.2.4 werden die bivariaten Beobachtungen gegeneinander geplottet, und die empirischen Randverteilungen werden durch die optionale Dichteschätzung der ISP-Prozedur PRIM-H (Huber 1986) veranschaulicht. Zusammenfassend können wir feststellen, daß die Methode der maximal selektierten Wilcoxon-Statistik einen signifikanten Einfluß des Alters der Patientinnen auf die Dichte der Oestrogen-Rezeptoren aufgezeigt hat.

Der mit der Wilcoxon-Statistik gemessene Einfluß des Alters der Patientinnen auf die Dichte der Progesteron-Rezeptoren ist nicht so ausgeprägt, aber signifikant (Abb. 11.2.5,6). Denn für  $\varepsilon_1 = 1 - \varepsilon_2 = 0.1$  erhalten wir den approximativen asymptotischen  $P$ -Wert 0.0072. Der geschätzte Prüfpunkt ist 66.18 Jahre.

Der Pfad der Wilcoxon-Statistik in Abhängigkeit des Alters der Patientinnen hat Extremwerte für hypothetische Prüfpunkte außerhalb des Selektionsintervalls. Auf eine Analyse dieser Beobachtung verzichten wir bei unserer einfachen bivariaten Betrachtung. Ansonsten legt der Verlauf des Pfades der Wilcoxon-Statistik im Selektionsintervall nur die Interpretation des Prüfpunktmodells als erste grobe Approximation

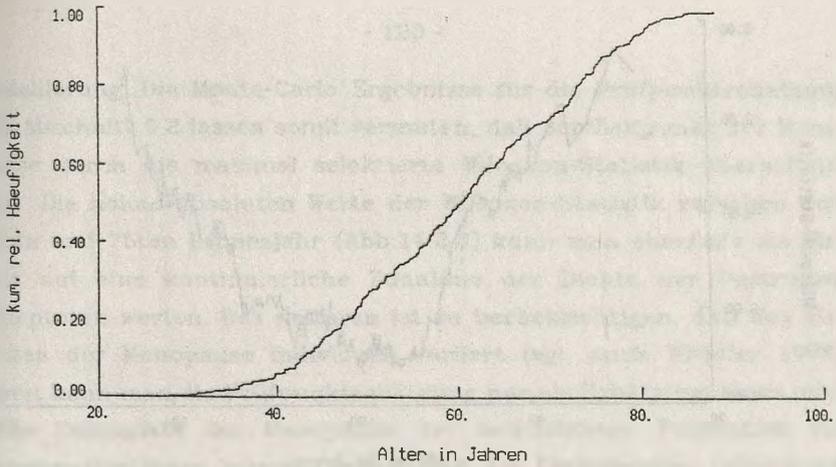


Abb. 11.2.1.: Die empirische Verteilung des Alters der betrachteten Patientinnen.

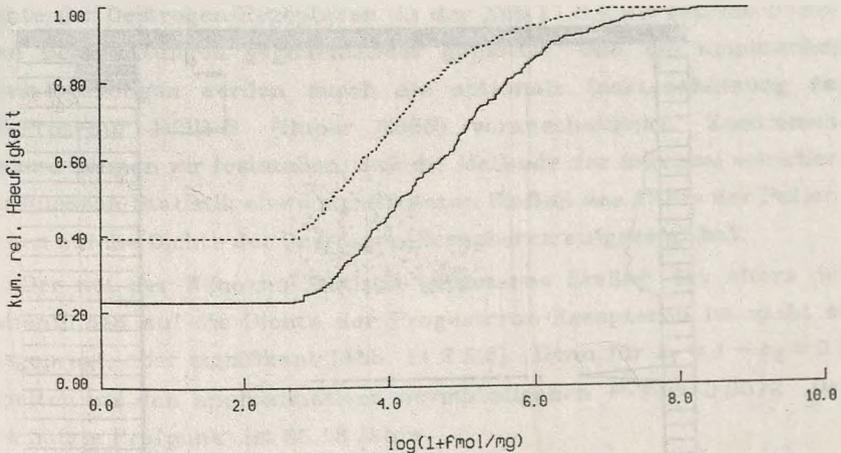


Abb. 11.2.2.: Die empirische Verteilung der postoperativ gemessenen Oestrogen- (gestrichelte Linie) und Progesteron-Rezeptoren (durchgezogene Linie).

eines funktionalen Zusammenhangs nahe. Die Problematik eines individuellen funktionalen Zusammenhangs ergibt sich analog wie bei der Dichte der Oestrogen-Rezeptoren.

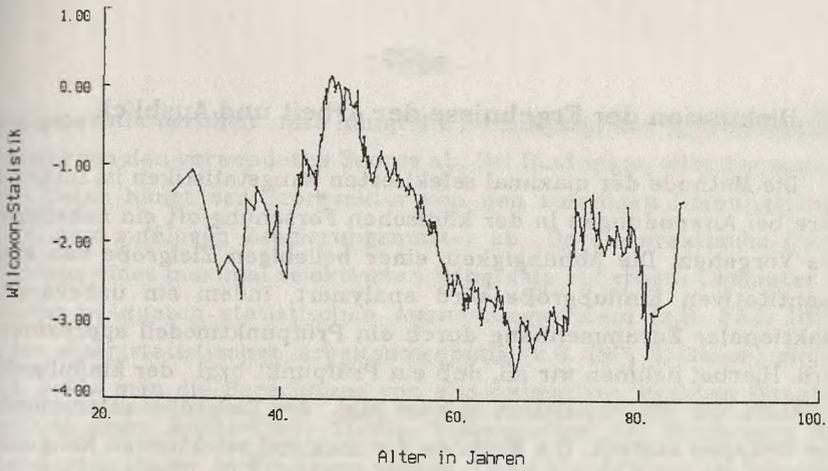


Abb. 11.2.5.: Progesteron-Rezeptoren: Die standardisierte Wilcoxon-Statistik als Prozeß in Abhängigkeit der möglichen Prüfpunkte bzgl. des Alters der Patientinnen.

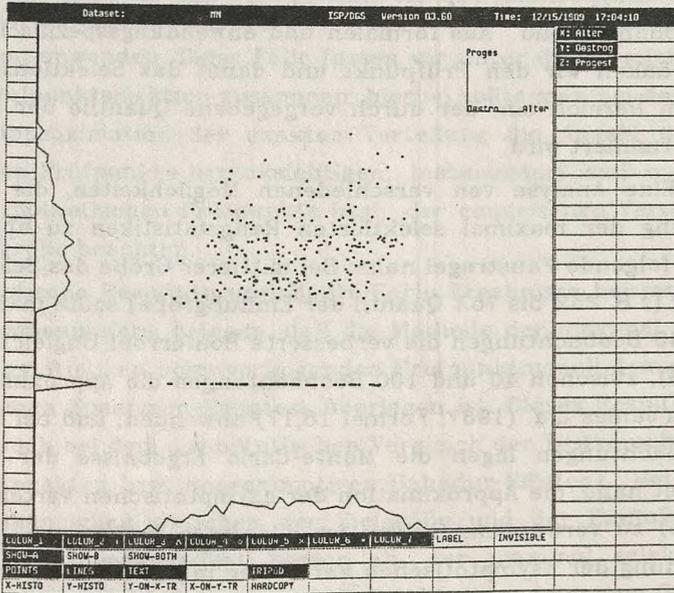


Abb. 11.2.6.: Darstellung mittels PRIM-H; Alter in Jahren gegen Progesteron  $\log(1+\text{fmol}/\text{mg})$ . Des weiteren wird eine Dichteschätzung der Randverteilungen abgebildet.

## 12. Diskussion der Ergebnisse der Arbeit und Ausblick

Die Methode der maximal selektierten Rangstatistiken ist insbesondere bei Anwendungen in der klinischen Forschung oft ein naheliegenderes Vorgehen. Die Abhängigkeit einer beliebigen Zielgröße von einer quantitativen Einflußgröße wird analysiert, indem ein unbekannter funktionaler Zusammenhang durch ein Prüfpunktmodell approximiert wird. Hierbei nehmen wir an, daß ein Prüfpunkt bzgl. der Einflußgröße existiert, der die Population in zwei bzgl. der Zielgröße unterschiedliche Gruppen aufteilt. Die Methode der maximal selektierten Rangstatistiken wird in der vorliegenden Arbeit auch für Zielgrößen mit beobachteten Bindungen oder unter Zensierung eingeführt und diskutiert. Als Nullhypothese formulieren wir, daß die mögliche Einflußgröße keinen Einfluß hat. Diese Nullhypothese ist äquivalent mit der Nullhypothese, daß die Einflußgröße und die Zielgröße stochastisch unabhängig sind. Aus formalen und anwendungsspezifischen Gründen schränken wir den Prüfpunkt und damit das Selektionsintervall auf einen Bereich ein, der durch vorgegebene Quantile der Einflußgröße determiniert wird.

Eine Analyse von verschiedenen Möglichkeiten, die exakte Verteilung der maximal selektierten Rangstatistiken zu approximieren, legt folgende Faustregel nahe: Bei mittlerer Größe des Selektionsintervalls (z.B. 25% bis 75% Quantil der Einflußgröße) sollte man bei weniger als 40 Beobachtungen die verbesserte Bonferroni-Ungleichung (Hunter 1976), zwischen 40 und 100 Beobachtungen die Approximationsformel nach James u.a. (1987; Formel 18,17) anwenden, und bei mehr als 100 Beobachtungen legen die Monte-Carlo Ergebnisse der vorliegenden Arbeit nahe, die Approximation der asymptotischen Verteilung (Jennen 1985) zu verwenden. Hierbei ist allerdings zu betonen, daß die Verwendung der asymptotischen Verteilung in der Regel zu einem konservativen Test führt. Deshalb ist obige Faustregel gemäß der erwünschten Genauigkeit bzw. dem erwünschten Verhalten zu modifizieren. Allerdings ist zu berücksichtigen, daß die vorgeschlagenen Approximationsformeln alle auf der asymptotischen Normalverteilung der jeweiligen

Rangstatistik beruhen. Also hängt die Genauigkeit der Approximationsformel von den verwendeten Scores ab. Bei Bindungen oder bei zensierten Daten hängt der Scorevektor von den zufälligen Bindungslängen bzw. dem zufälligen Zensierungsmuster ab. Da die praktische Durchführung eines maximal selektierten Rangtests auf einem Computer mit einem adäquaten statistischen Auswertungssystem (z.B. SAS, BMDP) oder einer statistischen Arbeitsumgebung (z.B. ISP, S, Gauss) einfach ist, sollte man die Berechnung von approximativen exakten Quantilen mittels einer Monte-Carlo Studie, insbesondere bei Bindungen oder zensierten Daten, in Erwägung ziehen. Diese Empfehlung ist von allgemeiner Bedeutung bei der Anwendung nichtparametrischer Verfahren (z.B. Krauth 1988).

Die Methode der maximal selektierten Rangstatistiken läßt sich auch bei einer apriori vorgegebenen Menge von Prüfpunkten, bei einer diskreten Verteilung der Einflußgröße und bei einer diskret formulierten Selektion anwenden. Diese Fälle fassen wir unter der Bezeichnung endliche Prüfpunktselektion zusammen; hierbei sollte man bei der Auswahl der Approximation der exakten Verteilung die Anzahl der zu betrachtenden Prüfpunkte berücksichtigen. Insbesondere muß man die Lage der hypothetischen Prüfpunkte bzgl. der empirischen Verteilung der Einflußgröße beachten.

Asymptotische Resultate und Monte-Carlo Ergebnisse bei moderatem Stichprobenumfang belegen, daß die Methode der maximal selektierten Rangstatistiken beim vorliegenden Prüfpunktmodell dem nichtparametrischen Spearman-Rangtest überlegen ist. Dieses Resultat erhalten wir auch bei dem asymptotischen Vergleich der Testvorschriften mittels der exakten bzw. approximativen Bahadur-Effizienz. Bei einer linearen Abhängigkeit zwischen der Zielgröße und der Einflußgröße schneidet der Spearman-Test besser ab, der maximal selektierte Wilcoxon-Test ist jedoch nur unwesentlich schlechter in seinen Güteeigenschaften. Für ein toxikologisches Schwellenwertmodell belegen Monte-Carlo Ergebnisse, daß man die Methode der maximal selektierten Rangstatistiken durch eine naheliegende Modifikation verbessern kann.

In diesem Zusammenhang wollen wir noch einmal betonen, daß das Prüfpunktmodell oft nur als eine grobe Approximation eines unbekannt-funktionalen Zusammenhangs zu betrachten ist. Diese bei einem maximal selektierten Rangtest verwendete grobe Approximation des unbekannt-funktionalen Zusammenhangs gestattet in der Regel eine konfirmatorische Interpretation. Demgegenüber kann man den unbekannt-funktionalen Zusammenhang mit der für das jeweilige Anwendungsproblem adäquaten Methode der explorativen Datenanalyse näher untersuchen. Insbesondere die explorative Anpassung und Verifikation von Modellen und die Betrachtung des empirischen Prozesses der verwendeten Rangstatistik sollten bei der Interpretation des geschätzten Prüfpunktes und des Prüfpunktmodells berücksichtigt werden. Des weiteren kann man feststellen, daß die Verwendung von Rangstatistiken und die Dichotomisierung der Einflußgröße mittels des Prüfpunktmodells zu einem gegenüber Ausreißen relativ robusten Verfahren führt. Die Monte-Carlo Ergebnisse bei verschiedenen Verteilungen belegen, daß die Auswahl der Scoreerzeugenden-Funktion eine gewisse Optimierung bzgl. möglicher Verteilungen der Zielgröße erlaubt. In diesem Zusammenhang sollten adaptive Scores (z.B. Behnen und Neuhaus 1989) eine automatische Optimierung ermöglichen.

Bei mehrdimensionalen möglichen Einflußgrößen ist eine naheliegende Anwendungsmöglichkeit für die Methode der maximal selektierten Rangstatistiken durch eine Projektion der Einflußgrößen auf eine eindimensionale Größe gegeben, beispielsweise die erste Hauptkomponente der Einflußgrößen bzw. eine inhaltlich begründete Projektion oder Kombination der Einflußgrößen. Inhaltlich begründete Projektionen bzw. Kombinationen ergeben sich beispielsweise aus der Modellierung des Krankheitsrisikos durch ein logistisches Regressionsmodell (vgl. auch Gail und Green 1976). Falls der funktionale Zusammenhang nur bei einer Einflußgröße unbekannt ist bzw. durch eine Dichotomisierung grob approximiert werden soll, kann man in Analogie zu den parametrischen Regressionsmodellen von Worsley (1983) argumentieren (vgl. auch Kim und Siegmund 1989). In diesem Zusammenhang sollten

Verallgemeinerungen bzgl. des nichtparametrischen Vorgehens von Puri und Sen (1985) möglich sein.

Die Methode der maximal selektierten Rangstatistiken legt eine lokale Verbesserung von Regression-Tree-Algorithmen (z.B. Breiman u.a. 1984) oder Segal (1988) nahe. Wenn der Regression-Tree-Algorithmus auf qualitative und quantitative Einflußgrößen angewandt wird, sind die bei der Variablenselektion benutzten  $P$ -Werte nicht direkt vergleichbar, und der Minimierungseffekt bevorzugt offensichtlich quantitative Einflußgrößen. Eine einfache lokale Verbesserung des Algorithmus kann man über die Berechnung der  $P$ -Werte bei quantitativen Einflußgrößen anhand der verbesserten Bonferroni-Ungleichung (Hunter 1976) bzw. anhand der Approximation der asymptotischen Verteilung (Jennen 1985) erreichen. Diese Verbesserung kann man folgendermaßen konkretisieren: Bei quantitativen Einflußgrößen wird die Selektion zwischen dem 10% und 90% Quantil durchgeführt. Falls bei der Selektion der verwendeten Statistik weniger als 40 Beobachtungen berücksichtigt werden, verwendet man die Approximation anhand der verbesserten Bonferroni-Ungleichung und bei 40 oder mehr Beobachtungen die Approximation der asymptotischen Verteilung.

Zusammenfassend wollen wir noch einmal betonen, daß sich die Anwendungsmöglichkeiten maximal selektierter Rangstatistiken über prognostische, diagnostische, klassifikatorische Fragestellungen bis zu einer lokalen Verbesserung eines Variablenselektionsverfahrens erstrecken.

## Literaturverzeichnis

- Abel, U., Berger, J., und Wiebelt, H. (1984), "CRITLEVEL: An Exploratory Procedure for the Evaluation of Quantitative Prognostic Factors," *Methods of Information in Medicine*, 23, 3, 154-156.
- Bahadur, R.R. (1960), "Stochastic Comparison of Tests," *The Annals of Mathematical Statistics*, 31, 276-295.
- (1967), "Rates of Convergence of Estimates and Test Statistics," *The Annals of Mathematical Statistics*, 38, 303-324.
- (1971), *Some Limit Theorems in Statistics*, SIAM: Philadelphia.
- Bajorski, P. (1987), "Local Bahadur Optimality of Some Rank Tests of Independence," *Statistics and Probability Letters*, 5, 255-262.
- Beasley, J.D., und Springer, S.G. (1985), "The Percentage Points of the Normal Distribution," in Griffiths, P., und Hill, I.D. (Hrsg.), *Applied Statistics Algorithms*, Ellis Horwood: London, 188-191.
- Behnen, K. (1974), *Güteeigenschaften von Rangtests unter Bindungen*, Habilitationsschrift, Mathematische Fakultät, Universität Freiburg.
- Behnen, K., und Neuhaus, G. (1989), *Rank Tests With Estimated Scores and Their Applications*, Teubner: Stuttgart.
- Bhattacharyya, G.K., und Johnson, R.A. (1968), "Nonparametric Tests for Shift at an Unknown Time Point," *The Annals of Mathematical Statistics*, 39, 1731-1743.
- Billingsley, P. (1968), *Convergence of Probability Measures*, John Wiley: New York.
- Bock, H.H. (1985), "On Some Significance Tests in Cluster Analysis," *Journal of Classification*, 2, 77-108.
- Bolviken, E. (1988), "Some Probability Bounds Relevant for Bonferroni Significance Levels Under Gaussian Models," *Scandinavian Journal of Statistics*, 15, 281-297.
- Breiman, L., Friedman, J.H., Olshen, R.A., und Stone, C.J. (1984), *Classification and Regression Trees*, Wadsworth: Belmont, California.
- Carlstein, E. (1988), "Nonparametric Change-Point Estimation," *The Annals of Statistics*, 16, 188-197.
- Chernoff, H., und Teicher, H. (1958), "A Central Limit Theorem for Sequences of Exchangeable Random Variables," *The Annals of Mathematical Statistics*, 29, 118-130.
- Cox, C. (1987), "Threshold Dose-Response Models in Toxicology," *Biometrics*, 43, 511-523.
- Cox, D.R. (1972), "Regression Models and Life Tables" (with discussion), *Journal of the Royal Statistical Society*, B, 74, 187-220.
- Csörgö, M., und Horvath, L. (1987), "Nonparametric Tests for the Changepoint Problem," *Journal of Statistical Planning and Inference*, 17, 1-9.

- (1988), "Invariance Principles for Change-point Problems," *Journal of Multivariate Analysis*, 27, 151-168.
- Cuzick, J. (1985), "A Wilcoxon-Type Test for Trend," *Statistics in Medicine*, 4, 87-90.
- Darkhovskh, B.S. (1976), "A Non-Parametric Method for the A Posteriori Detection of the 'Disorder' Time of a Sequence of Independent Random Variables," *Theory of Probability and its Applications*, 21, 178-183.
- Davies, R.B. (1977), "Hypothesis Testing When a Nuisance Parameter is Present Only Under the Alternative," *Biometrika*, 64, 247-254.
- (1987), "Hypothesis Testing When a Nuisance Parameter is Present Only Under the Alternative," *Biometrika*, 74, 33-43.
- Dümbgen, L. (1989), *The Asymptotic Behaviour of E. Carlstein's Nonparametric Change-point-Estimator*, Preprint Nr. 520, SFB 123, Universität Heidelberg.
- Feder, P.I. (1975), "The Log Likelihood Ratio in Segmented Regression," *The Annals of Statistics*, 3, 87-97.
- Freeman, J.M. (1986), "An Unknown Change Point and Goodness of Fit," *The Statistician*, 35, 335-344.
- Gail, M.H., und Green, S.B. (1976), "A Generalization of the One-Sided Two-Sample Kolmogorov-Smirnov Statistic for Evaluating Diagnostic Tests," *Biometrics*, 32, 561-570.
- Groeneboom, P., Oosterhoff, J., und Ruymgaart, F.H. (1979), "Large Deviation Theorems for Empirical Probability Measures," *The Annals of Probability*, 7, 553-586.
- Hackl, P. (Hrsg., 1989), *Statistical Analysis and Forecasting of Economic Structural Change*, Springer: Berlin.
- Hackl, P., und Katzenbeisser, W. (1989), "Tests Against Nonconstancy in Linear Models Based on Counting Statistics," in Hackl, P. (Hrsg.), *Statistical Analysis and Forecasting of Economic Structural Change*, Springer: Berlin, 53-70.
- Hackl, P., und Westlund, A.H. (1989), "Statistical Analysis of 'Structural Change': An Annotated Bibliography," in Krämer, W. (Hrsg.), *Econometrics of Structural Change*, Physica-Verlag: Heidelberg, 103-128.
- Hajek, J., und Sidak, Z. (1967), *Theory of Rank Tests*, Academic Press: New York.
- Halpern, J. (1982), "Maximally Selected Chi Square Statistics for Small Samples," *Biometrics*, 38, 1017-1023.
- Hawkins, D.M. (1977), "Testing a Sequence of Observations for a Shift in Location," *Journal of the American Statistical Association*, 72, 180-186.
- Hinkley, D.V. (1970), "Inference About the Change-Point in a Sequence of Random Variables," *Biometrika*, 57, 1-16.

- Hohnloser, S.H. (1987), *Der plötzliche Herztod - Diagnostik und Therapien bei Patienten mit maligner ventrikulärer Arrhythmien*, Habilitationsschrift, Medizinische Fakultät, Universität Freiburg.
- Hohnloser, S.H., Raeder, E.A., Podrid, P.J., Graboys, T.B., and Lown, B. (1987), "Predictors of Antiarrhythmic Drug Efficacy in Patients With Malignant Ventricular Tachyarrhythmias," *American Heart Journal*, 114, 1-7.
- Huber, P.J. (1986), *ISP*, Artemis Systems, Inc.: Carlisle.
- Hunter, D. (1976), "An Upper Bound for the Probability of a Union," *Journal of Applied Probability*, 13, 597-603.
- James, B., James, K.L., and Siegmund, D. (1987), "Tests for a Change-Point," *Biometrika*, 74, 71-83.
- James, B., James, K.L., and Siegmund, D. (1988), "Conditional Boundary Crossing Probabilities, With Applications to Change-Point Problems," *The Annals of Probability*, 16, 2, 825-839.
- Jennen, C. (1985), "Second-Order Approximations to the Density, Mean and Variance of Brownian First-Exit Times," *The Annals of Probability*, 13, 126-144.
- Jespersen, N.C.B. (1986), *Dichotomizing a Continuous Covariate in the Cox Regression Model*, Research Report 86/2, Statistical Research Unit, University of Copenhagen.
- Jones, M.P., and Crowley, J. (1989), "A General Class of Nonparametric Tests for Survival Analysis," *Biometrics*, 45, 1, 157-170.
- Kim, H.-J., and Siegmund, D. (1989), "The Likelihood Ratio Test for a Change-Point in Simple Linear Regression," *Biometrika*, 76, 409-423.
- Krämer, W., and Sonnberger, H. (1986), *The Linear Regression Model Under Test*, Physica-Verlag: Heidelberg.
- Krauth, J. (1988), *Distribution-Free Statistics: An Application-Oriented Approach*, Elsevier: Amsterdam.
- Lausen, B., and Schumacher, M. (1988), *Maximally Selected Rank Statistics*, Preprint, Institut für Medizinische Biometrie und Medizinische Informatik, Universität Freiburg.
- Ledwina, T. (1986), "Large Deviations and Bahadur Slopes of Some Rank Tests of Independence," *Sankhya*, 48, A, 188-207.
- (1986b), "On the Limiting Pitman Efficiency of Some Rank Tests of Independence," *Journal of Multivariate Analysis*, 20, 265-271.
- Lombard, F. (1987), "Rank Tests for Changepoint Problems," *Biometrika*, 74, 615-624.
- (1989), "Some Recent Developments in the Analysis of Changepoint Data," *South African Statistical Journal*, 23, 1-21.
- Marcus, M.B., and Shepp, C.A. (1972), "Sample Behaviour of Gaussian Processes," *Sixth Berkeley Symposium on Mathematical Statistics and Probability*, II, 423-439.

- MacNeill, I.B. (1974), "Tests for Change of Parameter at Unknown Times and Distributions of Some Related Functionals on Brownian Motion," *The Annals of Statistics*, 2, 950-962.
- Mehrotra, K.G., und Michalek, J.E. (1982), "A Relationship Between Two Forms of Linear Rank Procedures for Censored Data," *Biometrika*, 69, 674-676.
- Miao, B.Q. und Zhao, L.C. (1988), "Detection of Change Points Using Rank Methods," *Communications in Statistics - Theory and Methods*, 17, 3207-3217.
- Miller, R.G. (1981), *Survival Analysis*, John Wiley: New York.
- Miller, R., und Siegmund, D. (1982), "Maximally Selected Chi Square Statistics," *Biometrics*, 38, 1011-1016.
- Moskilde, L., und Anderson, P. (1973), "The Calcium Infusion Test in Primary Hyperparathyroidism," *Acta Medica Scandinavia*, 193, 331-336.
- Page, E.S. (1954), "Continuous Inspection Schemes," *Biometrika*, 41, 100-115.
- (1955), "A Test for a Change in a Parameter Occuring at an Unknown Point," *Biometrika*, 42, 523-526.
- Pettitt, A.N. (1979), "A Non-Parametric Approach to the Change-Point Problem," *Applied Statistics*, 28, 126-135.
- (1980), "Some Results on Estimating a Change-Point Using Non-Parametric Type Statistics," *Journal of Statistical Computation and Simulation*, 11, 261-272.
- Pollak, M., und Siegmund, D. (1985), "A Diffusion Process and its Application to Detecting a Change in the Drift of Brownian Motion," *Biometrika*, 72, 267-280.
- Praagman, J. (1988), "Bahadur Efficiency of Rank Tests for the Change-Point Problem," *The Annals of Statistics*, 16, 198-217.
- (1989), "Efficiency of Nonparametric Tests for Scale Shift at an Unknown Time Point," *Communications in Statistics - Theory and Methods*, 18, 1325-1338.
- (1989b), "Bahadur Efficiency of Tests for a Shift in Location of Normal Populations," in Hackl, P. (Hrsg.), *Statistical Analysis and Forecasting of Economic Structural Change*, Springer: Berlin, 135-165.
- Puri, M.L., und Sen, P.K. (1985), *Nonparametric Methods in General Linear Models*, John Wiley: New York.
- Quandt, R.E. (1958), "The Estimation of the Parameters of a Linear Regression System Obeying two Separate Regimes," *Journal of the American Statistical Association*, 53, 873-880.
- (1960), "Tests of the Hypothesis That a Linear Regression System Obeys two Separate Regimes," *Journal of the American Statistical Association*, 55, 324-330.

- Randles, R.H., und Wolfe, D.A. (1979), *Introduction to the Theory of Nonparametrics*, John Wiley: New York.
- Rothe, G. (1980), *Die asymptotische relative Effizienz*, Arbeitsbericht Nr.7, Abteilung Statistik, Universität Dortmund.
- (1981), "Some Properties of the Asymptotic Relative Pitman Efficiency," *The Annals of Statistics*, 9, 663-669.
- Schechtman, E. (1982), "A Nonparametric Test for Detecting Changes in Location," *Communications in Statistics - Theory and Methods*, 11, 1475-1482.
- Scariano, S.M., und Watkins, T.A. (1988), "Nonparametric Point Estimators for the Change-Point Problem," *Communications in Statistics - Theory and Methods*, 17,3645-3675.
- Schulze, U. (1987), *Mehrphasenregression*, Akademie-Verlag: Berlin.
- Segal, M. R. (1988), "Regression Trees for Censored Data," *Biometrics*, 44, 35-47.
- Sen, A., und Srivastava, M.S. (1975), "On Tests for Detecting Change in Mean," *The Annals of Statistics*, 3, 98-108.
- Shaban, S.A. (1980), "Change Point Problem and Two-Phase Regression: an Annotated Bibliography," *International Statistical Review*, 48, 83-93.
- Shorack, G.R., und Wellner, J.A. (1986), *Empirical Processes With Applications to Statistics*, John Wiley: New York.
- Siegmund, D. (1985), *Sequential Analysis: Tests and Confidence Intervals*, Springer: New York.
- (1986), "Boundary Crossing Probabilities and Statistical Applications," *The Annals of Statistics*, 14, 361-404.
- (1988), "Confidence Sets in Change-Point Problems," *International Statistical Review*, 56, 31-48.
- Wahrendorf, J., Mahon, G.A.T., und Schumacher, M. (1985), "A Nonparametric Approach in the Statistical Analysis of Mutagenicity Data," *Mutation Research*, 147, 5-13.
- Wheeler, J.M. (1989), "A Multivariate Look at the Menopause," *Journal of Clinical Epidemiology*, 42, 11, 1029-1030.
- Wichmann, B.A., und Hill, I.D. (1985), "An Efficient and Portable Pseudo-Random Number Generator," in Griffiths, P., und Hill, I.D. (Hrsg.), *Applied Statistics Algorithms*, Ellis Horwood: London, 238-242.
- Wieand, H.S. (1976), "A Condition Under Which the Pitman and Bahadur Approach to Efficiency Coincides," *The Annals of Statistics*, 4, 1003-1011.
- Wolfe, D.A., und Schechtman, E. (1984), "Nonparametric Statistical Procedures for the Change-point Problem," *Journal of Statistical Planning and Inference*, 9, 389-396.
- Woodworth, G.G. (1970), "Large Deviation and Bahadur Efficiency of Linear Rank Statistics," *The Annals of Mathematical Statistics*, 41, 251-283.

- Worsley, K.J. (1979), "On the Likelihood Ratio Test for a Shift in Location of Normal Populations," *Journal of the American Statistical Association*, 74, 365-367.
- (1982), "An Improved Bonferroni Inequality and Applications," *Biometrika*, 69, 297-302.
  - (1983), "Testing for a Two-Phase Multiple Regression," *Technometrics*, 25, 35-42.