

Adaptive Lokationstests mit U-Statistiken

Wolfgang Kössler
Humboldt-Universität zu Berlin
Institut für Informatik
Rudower Chaussee 25
12489 Berlin
koessler@informatik.hu-berlin.de

Wolf F. Lesener
Humboldt-Universität zu Berlin
Computer- und Medienservice
Rudower Chaussee 26
12489 Berlin
wflesener@cms.hu-berlin.de

Zusammenfassung

Für das Zweistichproben-Lageproblem betrachten wir nichtparametrische Tests und dabei insbesondere solche, die sich mit Hilfe von U-Statistiken darstellen lassen. Die Wahl geeigneter Kerne sichert, dass die U-Statistiken wieder, diesmal aber i.A. nichtlineare, Rangstatistiken sind. Wir konstruieren adaptive Tests nach dem Prinzip von Hogg (1974), bei denen in einer ersten Stufe der Typ der zugrundeliegenden Dichte geschätzt, und in der zweiten Stufe ein geeigneter Rangtest durchgeführt wird. In die erste Stufe geht nur die Ordnungsstatistik der Gesamtstichprobe ein. Da Rang- und Ordnungsstatistiken stochastisch unabhängig sind, hält diese Zweistufen-Prozedur das Signifikanzniveau ein. Die vorgeschlagenen Prozeduren wurden in SAS implementiert. Wir zeigen, dass die (adaptiven) U-Tests gute asymptotische und finite Güteeigenschaften haben.

Schlüsselwörter: Linearer Rangtest, U-Test, Tailstärke, Schiefe, asymptotische absolute Effizienz, asymptotische relative Effizienz, Gütefunktion

1 Einführung

Seien X_1, \dots, X_n und Y_1, \dots, Y_n unabhängige zufällige Stichproben aus Grundgesamtheiten mit absolutstetigen Dichtefunktionen $F(x)$ bzw. $F(x - \mathcal{G})$, $\mathcal{G} \in \mathcal{R}$. Im folgenden nehmen wir an, F ist zweimal stetig differenzierbar auf $(-\infty, \infty)$ mit Ausnahme einer Menge vom Lebesgue Maß Null; f' bezeichnet die Ableitung der Dichte f dort, wo sie existiert, sonst wird sie auf Null gesetzt. Wir testen

$$H_0: \mathcal{G} = 0 \quad \text{gegen} \quad H_1: \mathcal{G} > 0.$$

Der bekannteste nichtparametrische Test ist der Wilcoxon-Mann-Whitney Test (im folgenden abgekürzt durch WMW-Test). Dieser Test wurde verallgemeinert auf lineare Rangtests mit verschiedenen Scores, vgl. z.B. Hájek, u.a. (1999). In SAS implementiert sind neben dem WMW-Test nur der Median-Test, der normal scores Test und der Savage-Test. Die Scores sind jedoch auf spezielle Typen von zugrundeliegenden Dichten zugeschnitten; z.B. ist der normal scores Test für Dichten geeignet, die ähnlich wie die Normal-Dichte aussehen, der Median-Test für die Doppelsexponential-Dichte und der Savage-Test für die Gumbel-Dichte, oder allgemeiner, für linksschiefe Dichten. Deshalb wurden restriktive adaptive Tests konstruiert, bei denen in einem ersten Schritt die

Dichte klassifiziert, und im zweiten Schritt ein geeigneter linearer Rangtest durchgeführt wird, vgl. z.B. Büning (1991). Solche Tests halten das Signifikanzniveau ein und haben hohe Güte.

Die asymptotische Gütefunktion kann, unter der Folge $\{g_N\}$ von lokalen Alternativen, $g_N = g/\sqrt{N}$, $N = n_1 + n_2$, leicht berechnet werden.

Eine zweite Möglichkeit der Verallgemeinerung des WMW-Tests bildet die Klasse der U-Statistiken, die von Hoeffding (1948) eingeführt wurde, und mit der wir uns in dieser Arbeit beschäftigen wollen.

Im zweiten Abschnitt werden zunächst einige bekannte Resultate über lineare Rangtests wiederholt, und im dritten U-Statistiken eingeführt. Adaptive Tests, die auf U-Statistiken beruhen, werden im vierten Abschnitt konstruiert. Im fünften Abschnitt vergleichen wir dann die asymptotischen und finiten Güteeigenschaften des Adaptiven U-Tests und einzelner U-Tests, und zusätzlich eines Adaptiven linearen Rangtests, miteinander. Schlussfolgerungen und Empfehlungen für den Anwender folgen in Abschnitt sechs.

2 Lineare Rangtests

Wir wiederholen einige bekannte Resultate über lineare Rangtests (im folgenden abgekürzt durch R-Tests).

Annahme 1: Wir nehmen an, die Scores $a_N(i)$ erfüllen die Bedingungen

$$\lim_{N \rightarrow \infty} \int_0^1 (a_N(1 + \lfloor uN \rfloor) - \phi(u))^2 du = 0$$

mit quadratisch integrierbaren Scorefunktionen

$$\phi(u, g) := \phi(u) = -\frac{g'(G^{-1}(u))}{g(G^{-1}(u))}. \quad (1)$$

Wir definieren

$$d(f, g) := \int_0^1 \phi'(u, g) \cdot f(F^{-1}(u)) du \quad \text{und} \quad I(g) := \int_0^1 \phi^2(u, g) du,$$

wobei $I(g)$ die Fisher-Information der Dichte g ist, und g durch (1) gegeben ist. ϕ'

bezeichnet die Ableitung von ϕ fast überall. Wir nehmen an, dass $\int_0^1 \phi(u, g) du = 0$ und

$0 < I(g) < \infty$. Weiter verwenden wir die Bezeichnung

$$C(f, g) := d(f, g) \cdot I(g)^{-1/2}.$$

Seien R_{2j} die Ränge von Y_j in der kombinierten X - und Y -Stichprobe, und

$$T_\phi = \sum_{j=1}^{n_2} a_N(R_{2j}) \quad (2)$$

,wobei $N = n_1 + n_2$, lineare Rangstatistiken für das Lageproblem. Große Werte von T_ϕ führen zur Ablehnung von H_0 . Für $\phi(u) = 2u-1$ erhalten wir die Wilcoxon-Version des WMW-Tests.

Satz (Hájek u.a, 1999, Kap.6): Unter H_0 konvergiert T_ϕ/σ gegen eine standardnormal verteilte Zufallsvariable, wobei

$$\sigma^2 = \frac{n_1 n_2}{N} \cdot I(g). \quad (3)$$

Annahme 2: Sei $\theta > 0$ und $\{\theta_N\}$ eine Folge von "benachbarten" Alternativen mit $\theta_N = N^{-1/2}\theta$. Sei $\min(n_1, n_2) \rightarrow \infty$, $n_1/N \rightarrow \lambda$, $0 < \lambda < 1$.

Satz (Hájek u.a, 1999, Kap.7): Unter den Annahmen 1 und 2 ist die lineare Rangstatistik T_ϕ asymptotisch normal verteilt mit Erwartungswert μ und Varianz σ^2 , wobei

$$\mu = \theta \frac{n_1 n_2}{N} d(f, g)$$

und σ^2 durch (3) gegeben ist.

Korollar: Die asymptotische absolute Effizienz (AE) des auf T_ϕ beruhenden R-Tests ist gegeben durch

$$AE(T | f) = \lambda(1 - \lambda)C^2(f, g).$$

Seien T_{ϕ_1} und T_{ϕ_2} zwei R-Statistiken, die auf den scoreerzeugenden Funktionen $\phi_1 = \phi(u, g_1)$ und $\phi_2 = \phi(u, g_2)$ beruhen. Dann ist die asymptotische relative Effizienz (ARE) gegeben durch

$$ARE(T_{\phi_1}, T_{\phi_2} | f) = \frac{AE(T_{\phi_1} | f)}{AE(T_{\phi_2} | f)}.$$

Korollar: Seien $\Phi(x)$ die Verteilungsfunktion und $z_{1-\alpha}$ das $(1-\alpha)$ -Quantil von $N(0,1)$. Die asymptotische Gütefunktion von T_ϕ ist gegeben durch

$$\beta_{T,\phi}(\theta) = 1 - \Phi\left(z_{1-\alpha} - \sqrt{AE(T_\phi | f)} \cdot \theta\right).$$

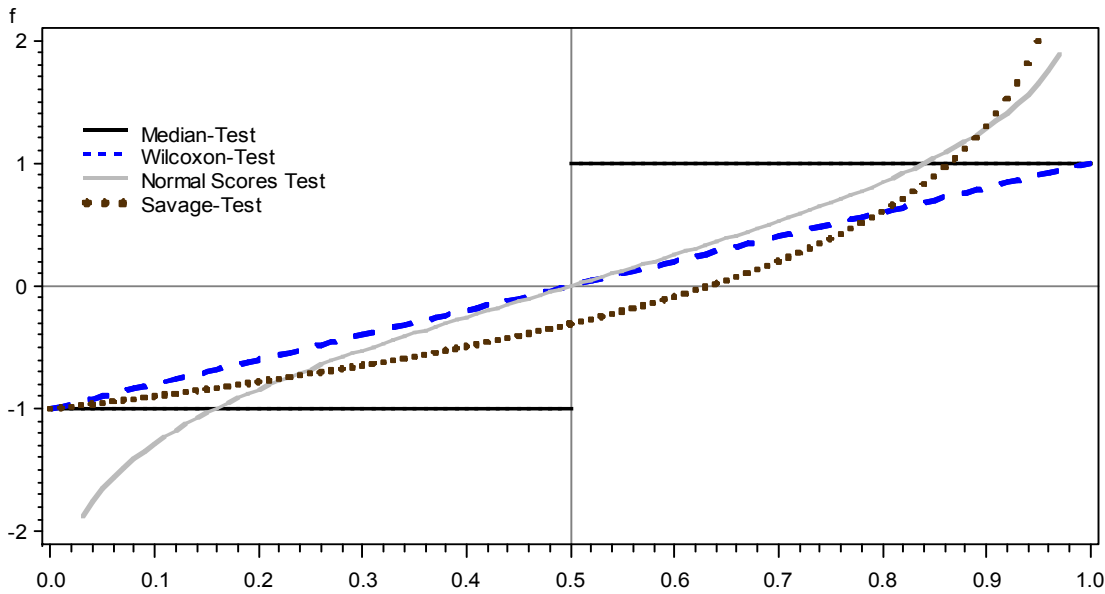


Abbildung 1: In der SAS-Prozedur NPAR1WAY implementierte Scores

In Abbildung 1 sind die Scorefunktionen angegeben, die in SAS implementiert sind. Zusätzlich können benutzerdefinierte Scores mittels der SCORE=Data-Option eingegeben werden.

Weitere Beispiele für Scorefunktionen wurden in einem SAS-Bediensystem (vgl. Deters u.a., 2005) implementiert und im unten genannten Adaptiven R-Test verwendet.

3 U-Statistiken

Sei k eine ganze Zahl mit $k \leq \min(n_1, n_2)$. Wir betrachten die folgende Klasse von U-Statistiken,

$$U = \frac{n_1 n_2}{N} \frac{1}{\binom{n_1}{k} \binom{n_2}{k}} \sum \varphi(X_{\alpha_1}, \dots, X_{\alpha_k}, Y_{\beta_1}, \dots, Y_{\beta_k}) - \frac{n_1 n_2}{2N} \quad (4)$$

wobei die Kernfunktion durch

$$\varphi(X_{\alpha_1}, \dots, X_{\alpha_k}, Y_{\beta_1}, \dots, Y_{\beta_k}) = \begin{cases} 1 & \text{falls } \psi(X_{\alpha_1}, \dots, X_{\alpha_k}) < \psi(Y_{\beta_1}, \dots, Y_{\beta_k}) \\ 0 & \text{sonst} \end{cases}$$

gegeben ist. Die Summe in Gleichung (4) wird über alle möglichen Teilstichproben, jeweils der X -Stichprobe und der Y -Stichprobe, vom Umfang k gebildet. Der Faktor

$\frac{n_1 n_2}{N}$ wurde eingeführt, um für die Varianz dieselbe Größenordnung wie für die linearen

Rangstatistiken zu haben.

Wir verwenden die speziellen ψ -Funktionen

$$\begin{aligned}\psi^{1:1}(X_1) &= X_1 \\ \psi^{3:1}(X_1, X_2, X_3) &= \min(X_1, X_2, X_3) \\ \psi^{3:2}(X_1, X_2, X_3) &= \text{med}(X_1, X_2, X_3) \\ \psi^{3:3}(X_1, X_2, X_3) &= \max(X_1, X_2, X_3) \\ \psi^{5:1}(X_1, \dots, X_5) &= \min(X_1, \dots, X_5) \\ \psi^{5:3}(X_1, \dots, X_5) &= \text{med}(X_1, \dots, X_5) \\ \psi^{5:5}(X_1, \dots, X_5) &= \max(X_1, \dots, X_5)\end{aligned}$$

wobei $\text{med}(\cdot)$ die Medianfunktion bezeichnet. Die entsprechenden ϕ -Funktionen bezeichnen wir mit $\phi^{k:i}$, und die U-Statistiken mit $U_{k:i}$ wobei $(k:i) = (1:1), (3:1), (3:2), (3:3), (5:1), (5:3), (5:5)$. Wir bemerken, dass $U_{1:1}$ die (Mann-Whitney-Version der) WMW-Statistik ist. Seien

$$\begin{aligned}\phi_{1,0}^{k:i}(x) &= E \phi^{k:i}(x, X_2, \dots, X_k, Y_1, \dots, Y_k) \\ \phi_{0,1}^{k:i}(y) &= E \phi^{k:i}(X_1, \dots, X_k, y, Y_2, \dots, Y_k) \\ \zeta_{1,0}^{k:i} &= \text{var} \phi_{1,0}^{k:i}(X) \\ \zeta_{0,1}^{k:i} &= \text{var} \phi_{0,1}^{k:i}(Y).\end{aligned}$$

Dann gilt

Satz (Xie und Priebe, 2000): Unter H_0 sind die U-Statistiken $U_{k:i}$ asymptotisch normal verteilt, mit Erwartungswert Null und Varianz

$$\text{var}(U_{k:i}) = \frac{n_1 n_2}{N} \sigma_{U_{k:i}}^2 \quad (5)$$

wobei

$$\sigma_{U_{k:i}}^2 = k^2 \zeta_{1,0}^{k:i} = k^2 \zeta_{0,1}^{k:i}.$$

Die asymptotischen Varianzen können explizit angegeben werden (siehe Theorem 3 von Xie und Priebe oder Kössler, 2006a, Gleichung (2.26)).

Satz (Xie und Priebe, 2000, S.666, siehe auch Kössler, 2006a, Theorem 2.7.5): Unter Annahme 2 ist die U-Statistik $U_{k:i}$ asymptotisch normal verteilt mit Erwartungswert $\mu_{U_{k:i}}$ und Varianz (5), wobei

$$\mu_{U_{k:i}} = \theta \frac{n_1 n_2}{N} \cdot J_{k:i}(f)$$

und

$$J_{k:i}(f) = \binom{k}{i}^2 \int_{-\infty}^{\infty} F^{2i-2}(x) (1 - F(x))^{2k-2i} f^2(x) dx.$$

Korollar: Die AE des auf $U_{k:i}$ beruhenden Tests ist gegeben durch

$$AE(U_{k:i} | f) = \lambda(1 - \lambda)C_{U_{k:i}}^2(f),$$

wobei
$$C_{U_{k:i}}(f) = \frac{J_{k:i}}{\sigma_{U_{k:i}}}.$$

Einige numerische Werte für $C_{U_{k:i}}(f)$ für verschiedene k, i und Dichten f sind gegeben in Kössler (2010).

Korollar: Die asymptotischen Gütefunktionen der auf den Statistiken $U_{k:i}$ beruhenden Tests sind gegeben durch

$$\beta_{U_{k:i}}(\theta) = 1 - \Phi\left(z_{1-\alpha} - \sqrt{AE(U_{k:i} | f)} \cdot \theta\right).$$

Wir erinnern daran, dass die geeignete Wahl der Kerne uns sichert, dass die verwendeten U-Statistiken auch nur von den Rangzahlen abhängen.

Sei $R_{(s)}$ der Rang von $Y_{(s)}$ in der gemeinsamen Rangzahlenbildung von X - und Y -Stichprobe, wobei $Y_{(s)}$ die s -te Ordnungsstatistik der Y -Stichprobe ist. Dann ist eine Rangrepräsentation der U-Statistiken (vgl. auch Kumar u.a. 2003)

$$\binom{n_1}{k} \binom{n_2}{k} \left(\frac{U_{k:i}}{n_1 n_2} + \frac{1}{2} \right) = \sum_{s=i}^{n_2} \sum_{j=i}^k \binom{s-1}{i-1} \binom{n_2-s}{k-i} \binom{R_{(s)}-s}{j} \binom{n_1+s-R_{(s)}}{k-j}.$$

Bemerkung: SAS-Macros, die diese U-Tests durchführen, können von den Autoren angefordert werden. Sollten Bindungen auftreten, so werden hier die entsprechenden Ränge zufällig gebildet.

4 Adaptive U-Tests

Das Konzept von Hogg (1974) (vgl. Büning, 1991, Handl, 1986) beruht auf der Unabhängigkeit von Rang- und Ordnungsstatistiken, siehe z.B. Randles und Wolfe (1979).

Satz: Die folgende Zwei-Schritt Prozedur






- 1. Klassifikation des Typs der zugrundeliegenden Dichte mit **Ordnungsstatistiken**, und
- 2. abhängig von der Klassifikation führe einen geeigneten **Rangtest** aus.

hält das Signifikanzniveau (asymptotisch) ein.

Um eine geeignete Klassifikation zu finden, untersuchen wir die AE. Dazu bilden wir eine Matrix, bei der die Zeilen die Dichten und die Spalten die asymptotischen absoluten Effizienzen $AE(U_{k:i})$ der U-Statistiken sind, wobei wir alle natürlichen k und i betrachten, für die $1 \leq i \leq k \leq 5$.

Eine Analyse der AE-Matrix liefert uns Erkenntnisse, welche Tests (asymptotisch) gut sind für welche Dichte, vgl. Tabelle 1. In der rechten Hälfte der mit R-Tests bezeichneten Spalte sind die zugehörigen Scorefunktionen visualisiert.

Tabelle 1: R- und U-Tests, die für bestimmte Dichteklassen geeignet sind

Dichteklasse	R-Tests	U-Tests
kurze Tails	GA Gastwirth 1965 	(k,1), (k,k), k ≥ 3
mittlere Tails	WI Wilcoxon 1945 	(1,1) Mann und Whitney 1947
lange Tails	LT Policello u.a. 1976 	$(k, \frac{k+1}{2})$, k ≥ 3 ungerade
rechts-schiefe	HFR Hogg u.a. 1975 "R" kann für rechts stehen 	(k,1), k ≥ 3
links-schiefe	HFL Hogg u.a. 1975 "L" kann für links stehen 	(k,k), k ≥ 3

Wir werden im folgenden die Dichte mit Hilfe von Maßzahlen für Tailstärke und Schiefe klassifizieren (vgl. Groeneveld und Meeden (1984)),

$$t = t_{0.05,0.15}(F) = \frac{F^{-1}(0.95) - F^{-1}(0.05)}{F^{-1}(0.85) - F^{-1}(0.15)} \quad (6)$$

$$s = s_{0.05}(F) = \frac{F^{-1}(0.95) + F^{-1}(0.05) - 2F^{-1}(0.5)}{F^{-1}(0.95) - F^{-1}(0.05)} \quad (7)$$

In Tabelle 2 sind einige numerische Werte für Tailstärke und Schiefe angegeben.

Tabelle 2: Tailstärke und Schiefe einiger Dichten

Symmetrische Dichten		Schiefe Dichten		
Dichte	Tailstärke	Dichte	Tailstärke	Schiefe
Uniform	1.286	Exponential	1.697	0.564
Uni-Logist (GA) ¹	1.474	Gumbel	1.655	0.280
Normal	1.587	neg. Gumbel	1.655	-0.280
Logistic (WI)	1.697	Logist-Exp (HFR) ¹	1.799	0.349
CN1 ²	1.697	CN2 ²	1.592	0.277
Logist-DE (LT) ¹	1.864			
Doppelexp	1.912			

¹ Die fett bezeichneten Dichten Uni-Logist, Logist-DE und Logist-Exp sind die Dichten, für die die GA-, LT- bzw. HFR-scores die optimalen sind.

² CN1 und CN2 bezeichnen kontaminierte Normalverteilungsdichten:
CN1: $0.9N(0,1) + 0.1N(0,9)$ und CN2: $0.5N(1,4) + 0.5N(-1,1)$.

Cauchy | 3.217 || |
Tailstärke und Schiefe werden geschätzt, indem in (6) und (7) Schätzungen für die Quantilfunktion eingesetzt werden,

$$\hat{t} = \frac{\hat{F}^{-1}(0.95) - \hat{F}^{-1}(0.05)}{\hat{F}^{-1}(0.85) - \hat{F}^{-1}(0.15)}$$
$$\hat{s} = \frac{\hat{F}^{-1}(0.95) + \hat{F}^{-1}(0.05) - 2\hat{F}^{-1}(0.5)}{\hat{F}^{-1}(0.95) - \hat{F}^{-1}(0.05)}$$

$$\hat{F}^{-1}(u) = (1 - \varepsilon)X_{(j)} + \varepsilon X_{(j+1)}, \text{ wobei}$$

$$\varepsilon = N \cdot u + 0.5 - j, \quad j = \lfloor N \cdot u + 0.5 \rfloor$$

und $X_{(j)}$ die j -te **Ordnungsstatistik** der zusammengesetzten Stichprobe ist.

Die Abbildung 2 zeigt, dass unsere Tailstärkemaßzahl mit unserer Vorstellung übereinstimmt. Die Tailstärke ist jeweils der Quotient aus den Längen der beiden Abschnitte, wobei im Zähler der längere (dunkle) Abschnitt steht.

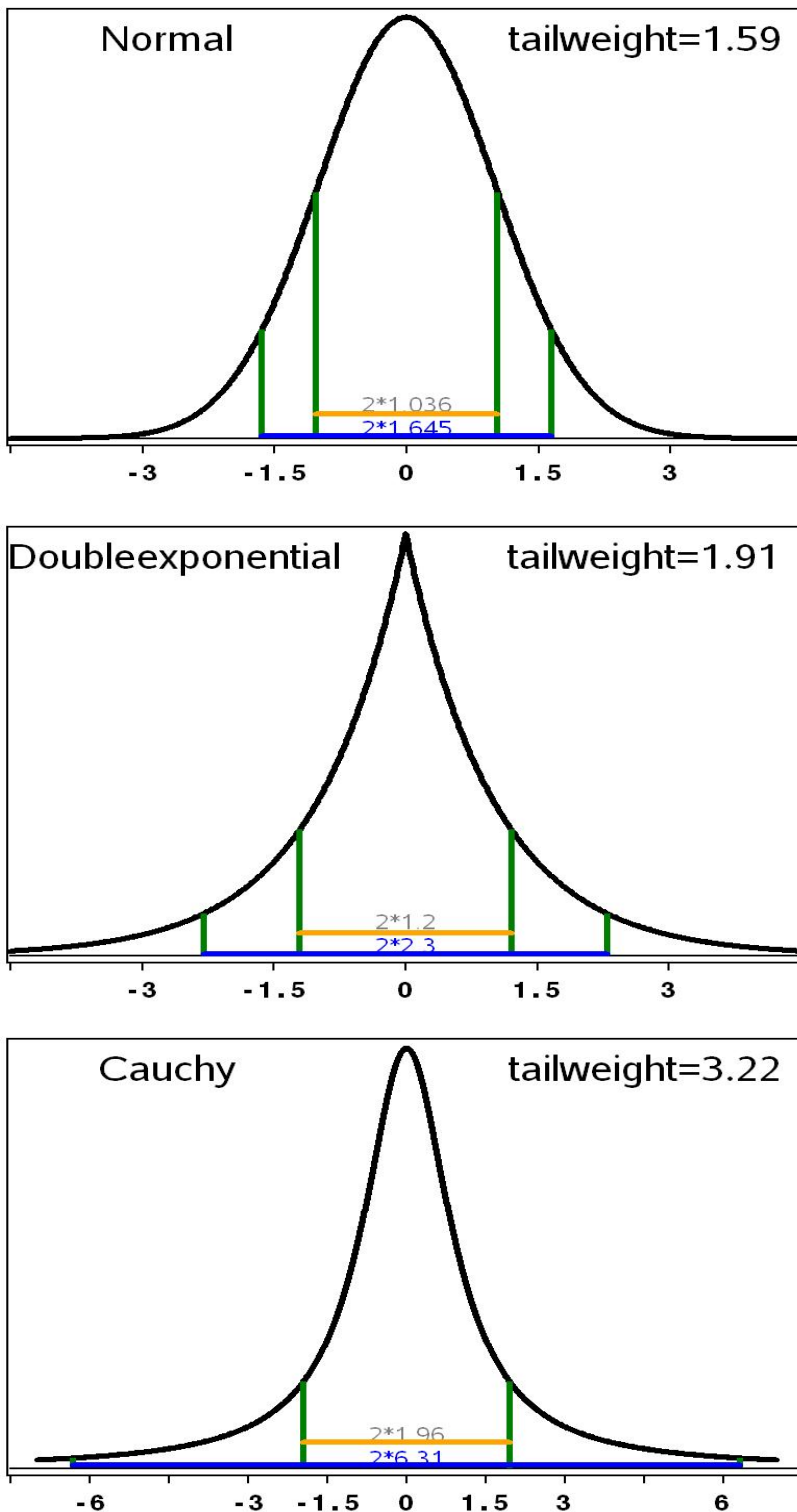


Abbildung 2: Illustration der Tailstärkemaßzahl $t_{0.05, 0.15}$

Zur Konstruktion eines adaptiven Tests definieren wir zunächst grob Regionen bzgl. Tailstärke und Schiefe, hier sind es fünf. Im zweiten Schritt werden den Regionen jeweils geeignete U-Tests zugeordnet. Hier verwenden wir die U-Statistiken $U_{1:1}$, $U_{5:1}$, $U_{5:5}$ und $U_{5:3}$. Zur Definition der Grenzen zwischen den Regionen werden die Tailstärke und Schiefemaße der einzelnen betrachteten Dichten mit den AE verglichen, und dafür gesorgt, dass möglichst jede Dichte in die Region kommt, für die die AE des entsprechenden U-Tests möglichst groß ist. Seien $\hat{S} = (\hat{t}, \hat{s})$ und die Bereiche E_1, \dots, E_5 wie folgt definiert:

$$\begin{array}{ll}
 E_1 = \{\hat{t} < 1.5, |\hat{s}| \leq 0.25\} & \text{“kurze Tails”,} \\
 E_2 = \{1.5 \leq \hat{t} < 1.8, |\hat{s}| \leq 0.25\} & \text{“mittlere Tails”} \\
 E_3 = \{1.8 \leq \hat{t} \leq 2.5, |\hat{s}| \leq 0.25\} & \text{“lange Tails”} \\
 E_4 = \{\hat{s} > 0.25\} & \text{“rechtsschief”} \\
 E_5 = \{\hat{s} < -0.25\} & \text{“linksschief”}
 \end{array}$$

Der resultierende adaptive Test ist gegeben durch

$$A = A(\hat{S}) = \begin{cases} (5:1) & \text{falls } \hat{S} \in E_1 \cup E_4 \\ (1:1) & \text{falls } \hat{S} \in E_2 \\ (5:3) & \text{falls } \hat{S} \in E_3 \\ (5:5) & \text{falls } \hat{S} \in E_5 \end{cases}$$

und in Abbildung 3 schematisiert. Ähnliche adaptive Tests werden in Kössler 2005, 2009, sowie Kössler und Kumar 2008 verwendet. Im Tupel $(k:i)$ bezeichnet k jeweils den Teilstichprobenumfang und i die Nummer der Ordnungsstatistik.

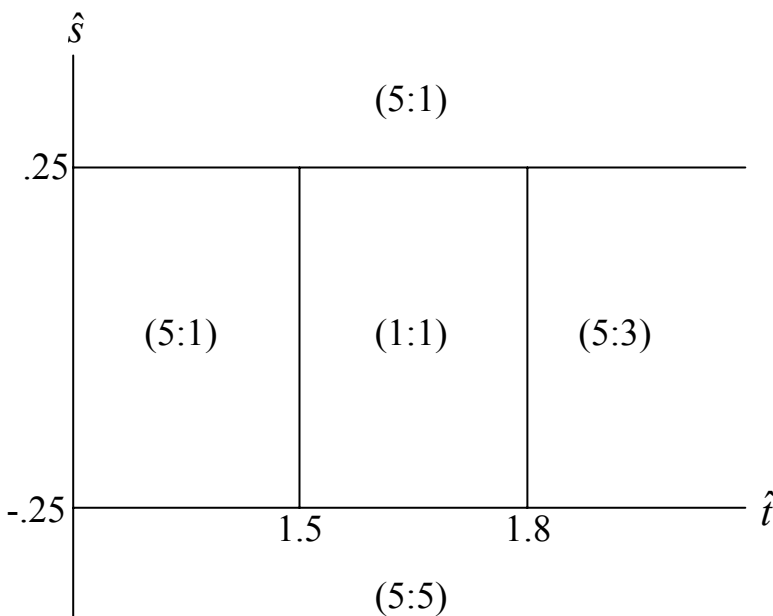


Abbildung 3: Schematische Darstellung des Adaptiven U-Tests

Beispiel: Wir illustrieren das Verfahren anhand eines realen Datensatzes. In Abbildung 4 finden wir ein Boxplot der Laufzeiten des Isergebirgslaufs 2002, eines 50km Skilanglaufs. Wir betrachten nur zwei Altersklassen, die Männer bis 39 Jahre ($n_1=984$), und die Männer im Alter von 40-49 ($n_2=574$). Viele Physiologen meinen, dass sich die Laufzeiten mit wachsendem Alter verlängern. Ist diese Aussage bei Nutzung der verfügbaren Stichprobe nachweisbar? Wir testen also die Nullhypothese, die Laufzeiten unterscheiden sich nicht, gegen die einseitige Alternative, die älteren Männer brauchen länger.

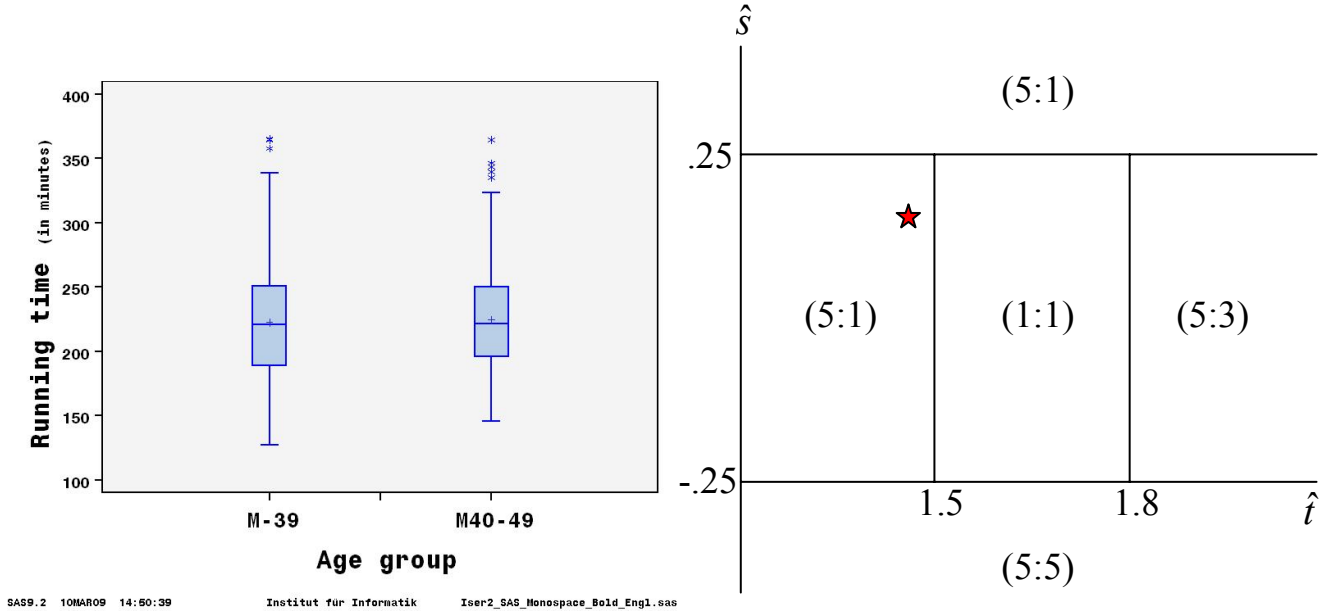


Abbildung 4: Laufzeiten des Isergebirgslaufes 2002 nach zwei Altersklassen

★ = Selektorstatistik (\hat{t}, \hat{s})

Um den adaptiven U-Test durchzuführen, schätzen wir als erstes die Tailstärke und die Schiefe, und erhalten: $\hat{t}=1.499$ und $\hat{s}=0.116$. Demzufolge wird der U-Test $U_{5:1}$ ausgewählt und durchgeführt. Wir erhalten für die standardisierte U-Statistik $U_{5:1}^* = U_{5:1} / \sigma_{5:1} = 3.480$ und damit einen (einseitigen) p -Wert von 0.00025.

Der große Wert für $U_{5:1}^*$ mag etwas überraschend sein. Die Gründe dafür werden uns erst nach einem zweiten Blick auf die Daten klar: 40 der 50 Erstplatzierten (die "Profis") sind in der jüngeren Altersgruppe und in beiden Teilstichproben werden Minima miteinander verglichen.

5 Vergleich der Adaptiven R- und U-Tests

In diesem Abschnitt vergleichen wir den adaptiven U-Test mit einem adaptiven R-Test, vgl. Deters u.a. (2005). In Abbildung 5 sind die beiden adaptiven Tests schematisch dargestellt.

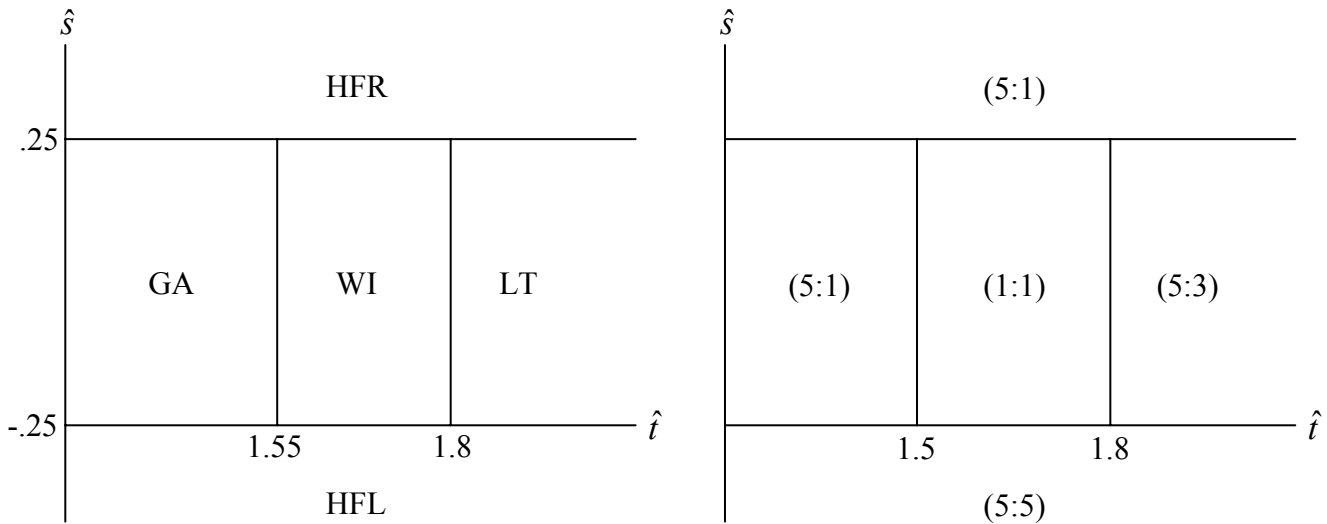


Abbildung 5: Schematische Darstellung der adaptiven R- und U-Tests

Die Abbildungen 6a-d zeigen die asymptotischen Gütefunktionen der beiden adaptiven Tests, zusammen mit denen der drei Einzeltests $U_{1:1}$, $U_{5:1}$ und $U_{5:3}$ für die Dichten:

- Normal und logistisch (Abb. 6a),
- doppel exponential und Cauchy (Abb. 6b),
- Gumbel und kontaminiert normal³ (Abb. 6c),
- uniform und exponential (Abb. 6d).

Die durchgezogene Linie steht jeweils für den adaptiven U-Test, die helle kurz gestrichelte Linie für den Adaptiven R-Test, die gepunktete Linie für den Test $U_{1:1}$, die durch Plus symbolisierte Linie für den Test $U_{5:1}$ und die durch Stern symbolisierte Linie für den Test $U_{5:3}$. Dabei können mehrere Kurven übereinander liegen.

Finite Gütefunktionen ($n_1 = n_2 = 40$) werden in den Abbildungen 7a und 7b dargestellt. Da die Kurven für die normal, logistische, doppel exponential, Gumbel und kontaminierte normal Dichte sehr ähnlich sind, ist hier nur die Güte für die erstgenannte angegeben. Dazu kommen die Gütefunktionen für die extremeren Dichten Cauchy, uniform und exponential.

³ kontaminiert normal: $0.5N(1,4)+0.5N(-1,1)$

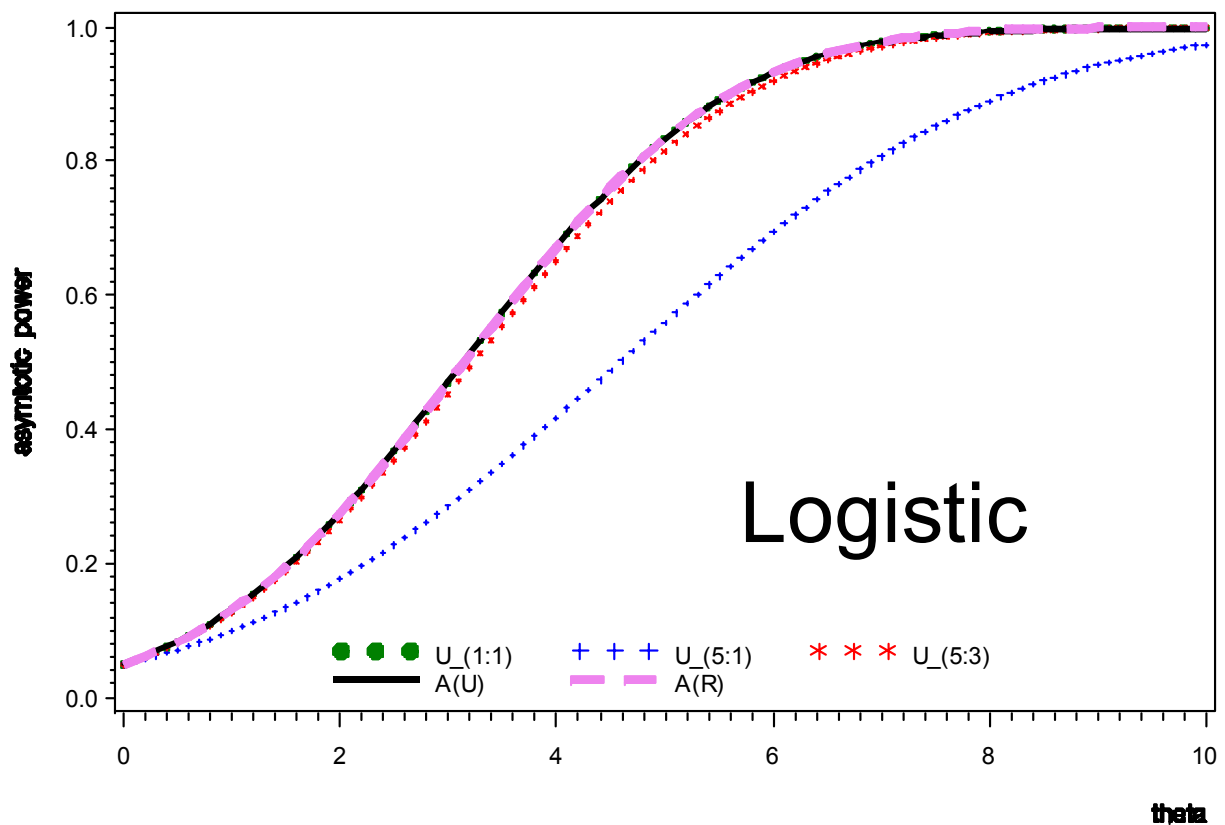
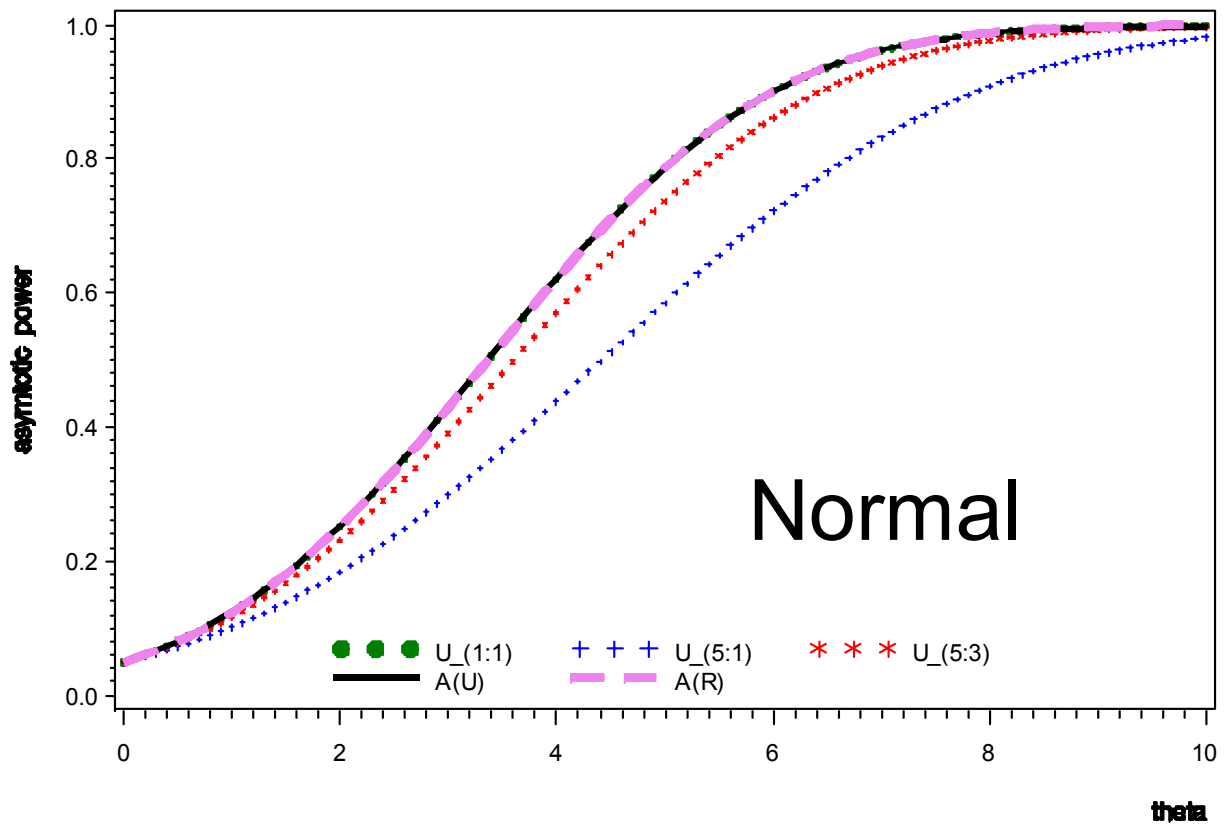


Abbildung 6a: Asymptotische Gütefunktionen der beiden adaptiven Tests R und U, sowie der Tests $U_{1:1}$, $U_{5:1}$ und $U_{5:3}$. Dichten: Normal und Logistisch.

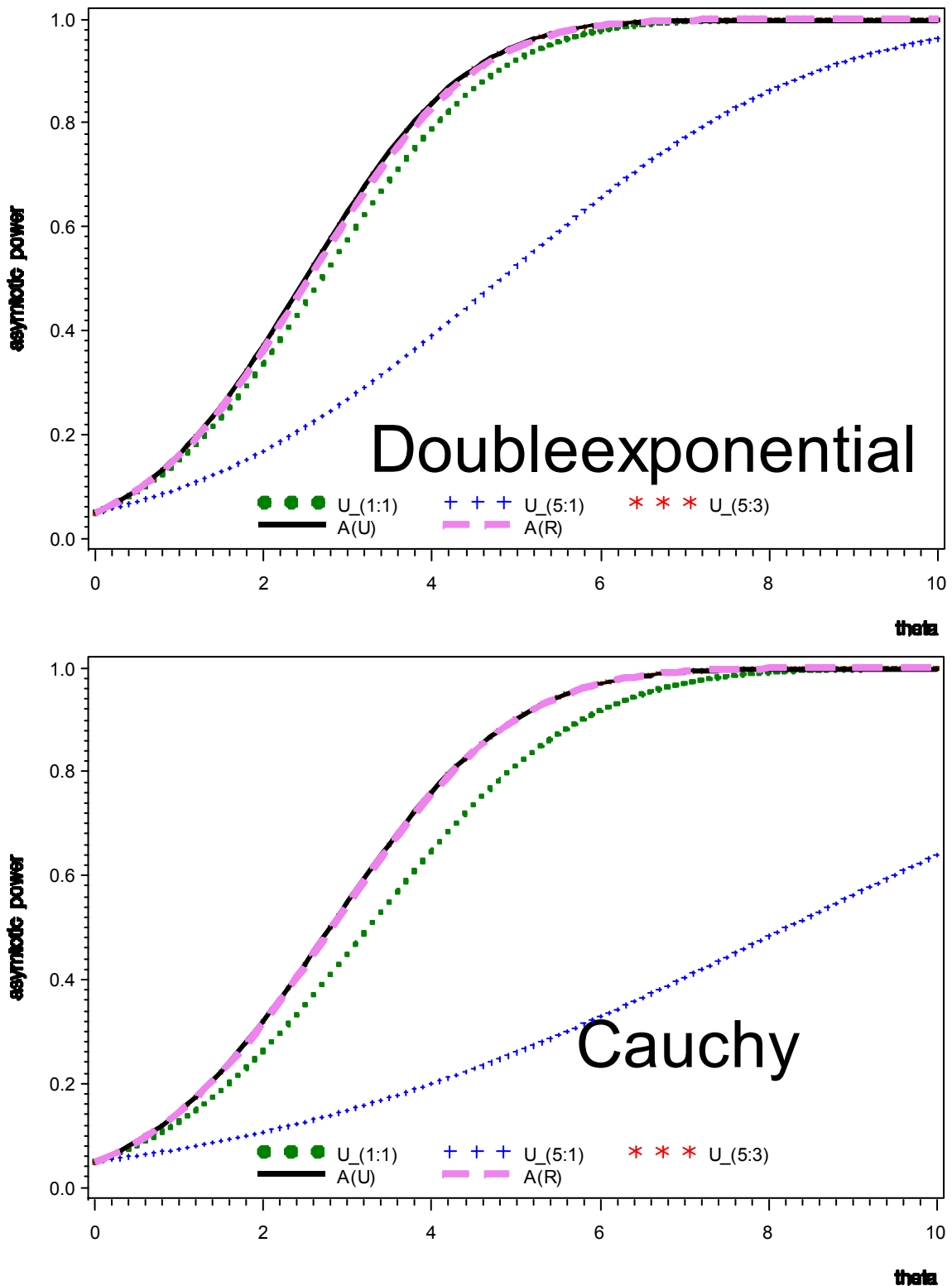


Abbildung 6b: Asymptotische Gütefunktionen der beiden adaptiven Tests R und U, sowie der Tests $U_{1:1}$, $U_{5:1}$ und $U_{5:3}$. Dichten: Doubleexponential und Cauchy.

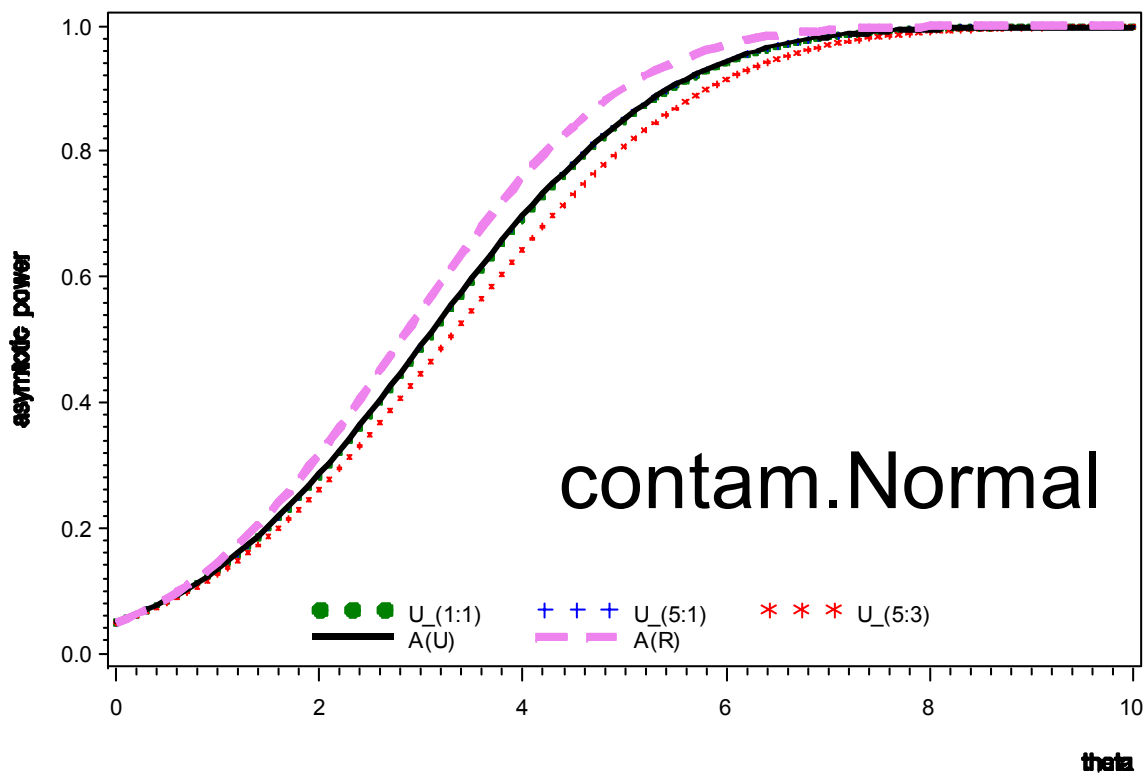
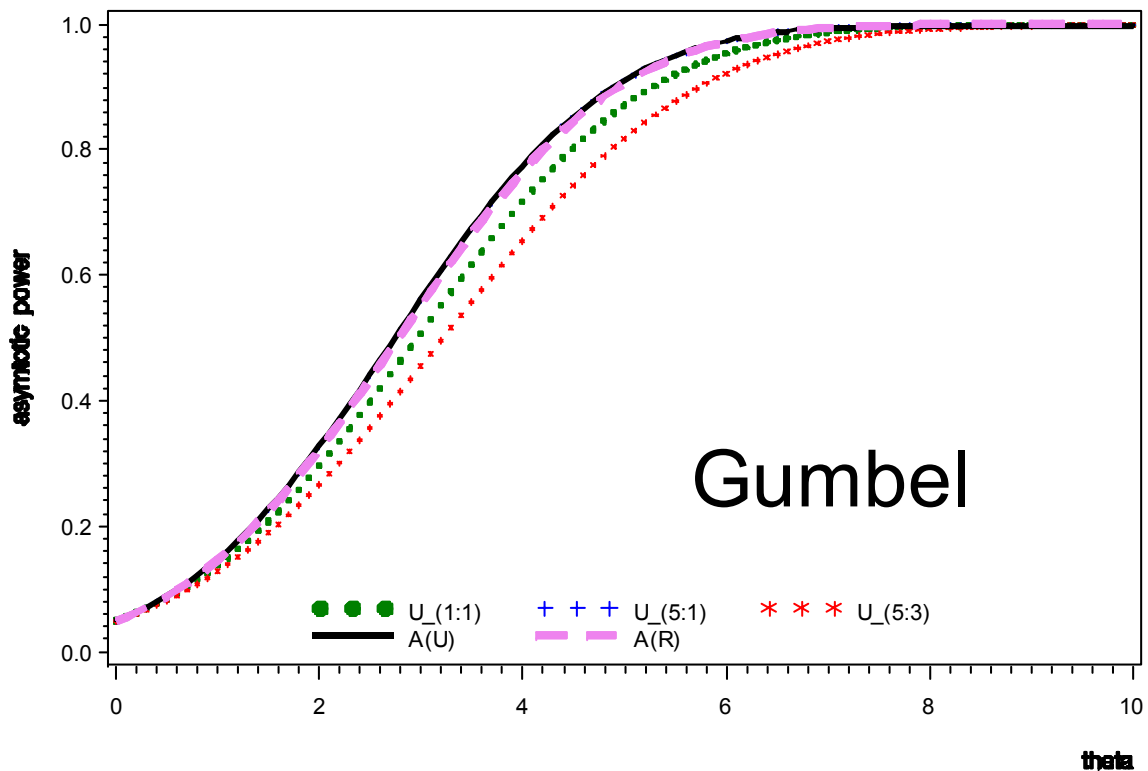


Abbildung 6c: Asymptotische Gütefunktionen der beiden adaptiven Tests R und U, sowie der Tests $U_{1:1}$, $U_{5:1}$ und $U_{5:3}$. Dichten: Gumbel und kontaminiert Normal (CN).

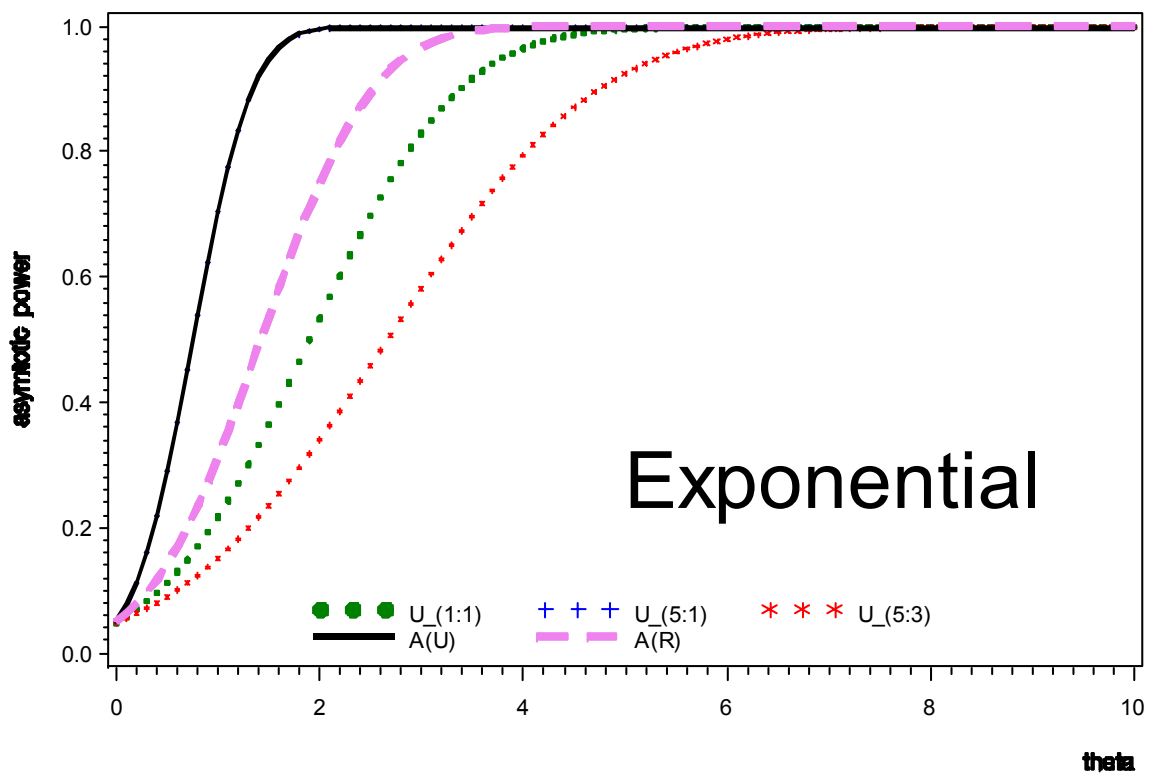
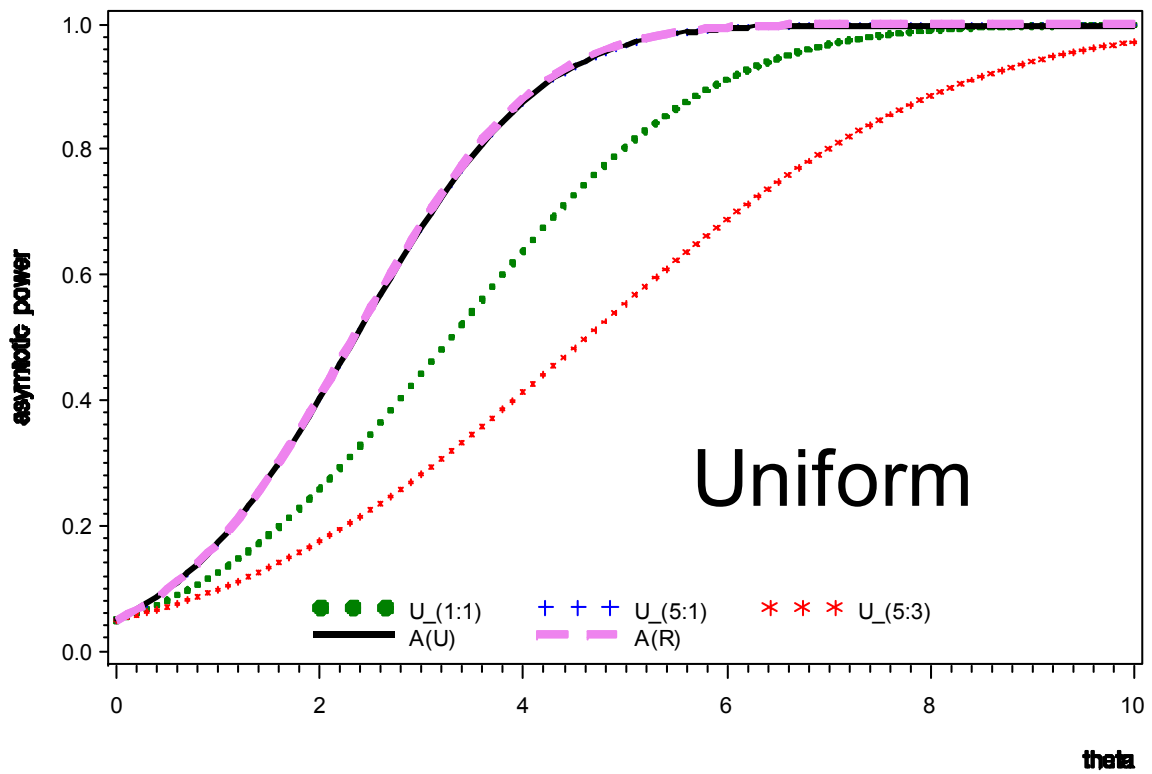


Abbildung 6d: Asymptotische Gütefunktionen der beiden adaptiven Tests R und U, sowie der Tests $U_{1:1}$, $U_{5:1}$ und $U_{5:3}$. Dichten: Uniform und Exponential.

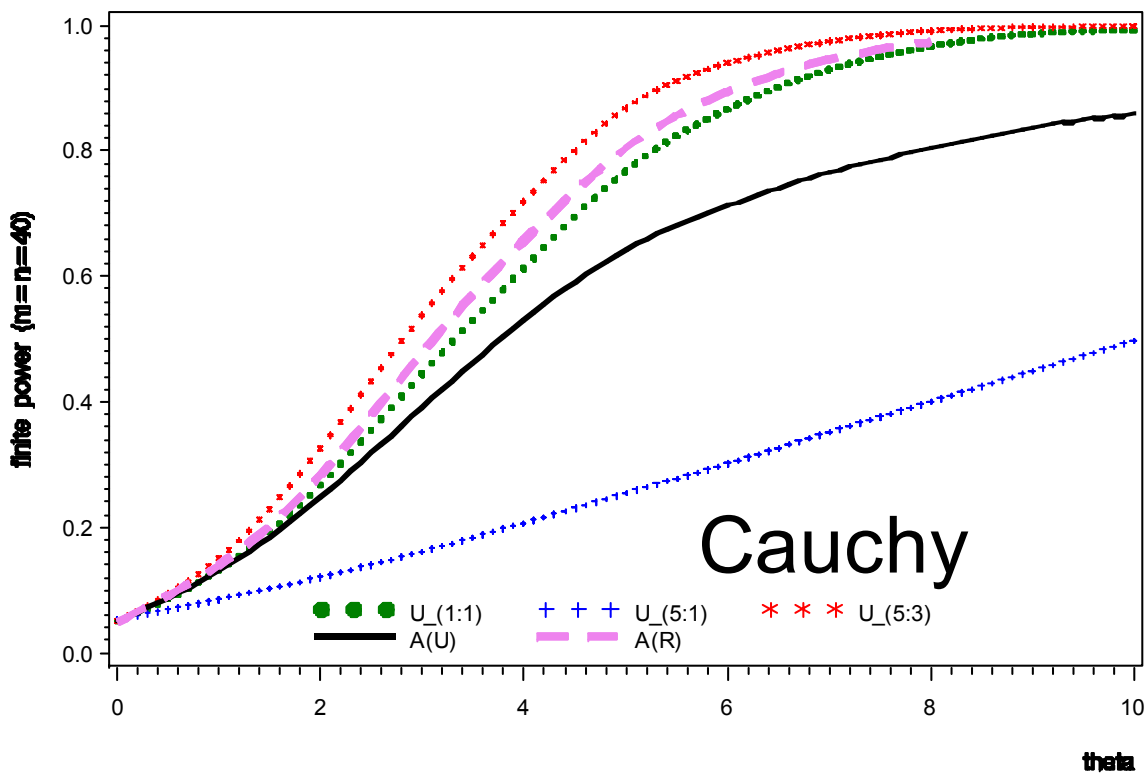
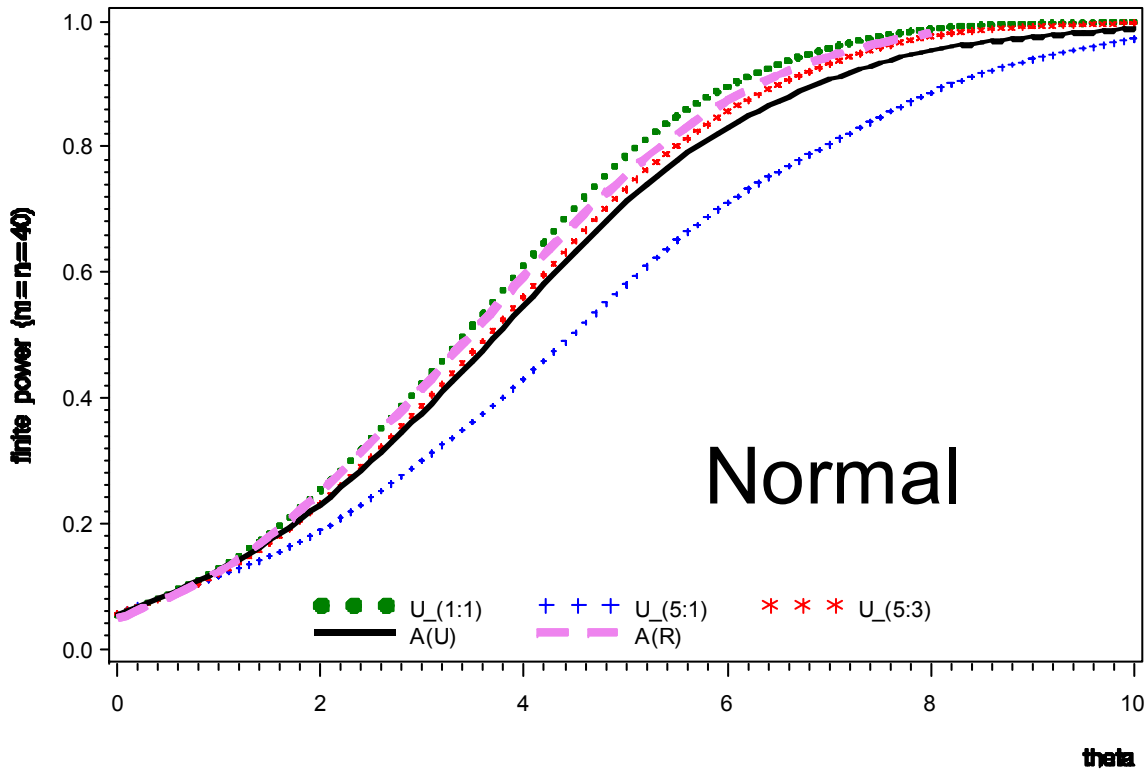


Abbildung 7a: Finite Gütefunktionen ($n_1 = n_2 = 40$) der beiden adaptiven R- und U-Tests, sowie der Tests $U_{1:1}$, $U_{5:1}$ und $U_{5:3}$. Dichten: Normal und Cauchy.

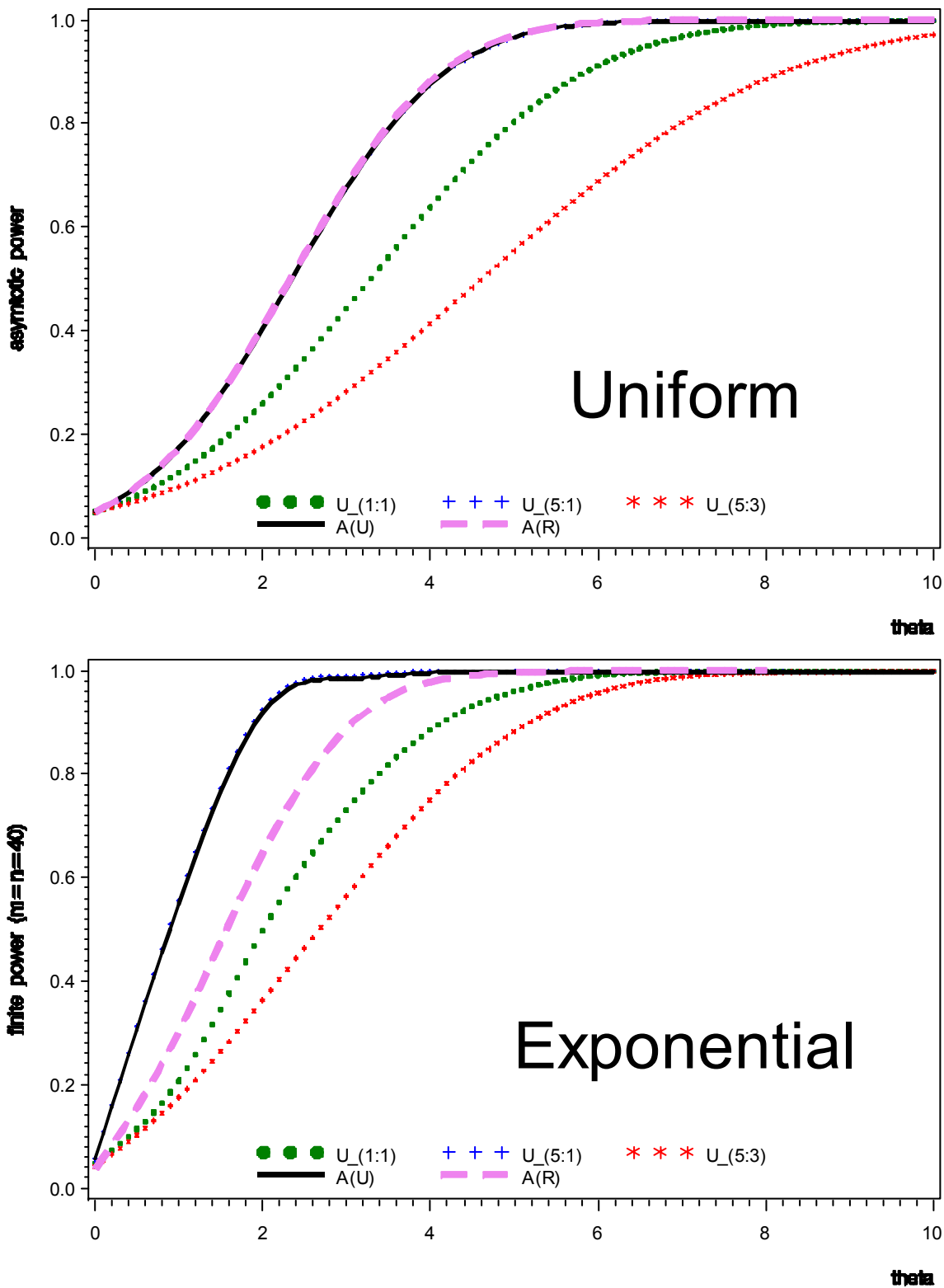


Abbildung 7b: Finite Gütefunktionen ($n_1 = n_2 = 40$) der beiden adaptiven R- und U-Tests, sowie der Tests $U_{1:1}$, $U_{5:1}$ und $U_{5:3}$. Dichten: Uniform und Exponential.

6 Zusammenfassung und Schlussfolgerung

In der vorliegenden Arbeit haben wir das Zweistichproben-Lageproblem betrachtet, und verschiedene U-Statistiken untersucht. Darüber hinaus haben wir einen adaptiven Test konstruiert, der auf U-Statistiken beruht, und bei dem in einer ersten Stufe die zugrunde liegende Dichte klassifiziert, und in der zweiten Stufe ein geeigneter U-Test durchgeführt wird. Da die Klassifikation ausschließlich auf Ordnungsstatistiken beruht, und anschließend ein (evtl. nichtlinearer) Rangtest durchgeführt wird, hält der Adaptive Test das Signifikanzniveau (asymptotisch) ein. Da (asymptotisch) (fast) immer der beste von den vier betrachteten U-Tests ausgewählt wird, ist die asymptotische Güte des Adaptiven Tests gleich der asymptotischen Güte des ausgewählten (fast) besten Rangtests.

Für Stichprobenumfänge von etwa $n_1 = n_2 \geq 40$ und für 'moderate' Dichten sind die Adaptiven R- und U-Tests gleichwertig, wobei Adaptive R-Tests einfacher zu berechnen sind. Sind jedoch 'extreme' Dichten in Betracht zu ziehen, dann erweisen sich Adaptive U-Tests als etwas besser.

Für kleine Stichprobenumfänge sind die adaptiven Tests in der vorliegenden Form nicht geeignet, dann sind exakte Tests oder max-Typ Tests vorzuziehen.

Die vorgestellte Idee der adaptiven Tests kann auf andere statistische Problemstellungen angewendet werden, für Mehrstichprobenprobleme (mit oder ohne Restriktionen) siehe z.B. Büning (1991, 1999), Beier und Büning (1997), Kössler (2005, 2006a, 2010). Max-Typ-Tests in solchen Fragestellungen werden von Neuhäuser u.a. (2004) untersucht, siehe auch Kössler (2010). Skalenprobleme werden in Kössler (1994, 1999) und in Kössler und Kumar (in Arbeit), und kombinierte Lage-Skalenprobleme in Büning und Thadewald (2000) sowie in Kössler (2006b) behandelt.

Literatur

- [1] Beier, F. und Büning, H. (1997). An adaptive test against ordered alternatives. *Comput. Statist. Data Anal.* 25, 441-452.
- [2] Büning, H. (1991). *Robuste und Adaptive Tests*, De Gruyter.
- [3] Büning, H. (1999). Adaptive Jonckheere-type tests for ordered alternatives. *J. Applied Statistics* 26, 541-557.
- [4] Büning, H. und Thadewald, T. (2000). An adaptive two-sample location-scale test of Lepage type for symmetric distributions. *J. Statist. Comput. Simul.* 65, 287-310.
- [5] Deters, D.-J., Kössler, W. und Lesener, W.F. (2005). Ein SAS-Bediensystem für Lokationstests. *Proceedings der 9. KSFE*, 41-61.
- [6] Gastwirth, J.L. (1965). Percentile modifications of two-sample rank tests. *Journal of the American Statistical Association*, 60, 1127-1141.
- [7] Groeneveld, R.A. und Meeden, G. (1984). Measuring Skewness and Kurtosis *Statistician*, 33, 391-399.

- [8] Hájek, J., Šidák, Z. und Sen, P.K. (1999). *Theory of Rank Tests*. Academic Press, San Diego.
- [9] Hall, D.L. und Joiner, B.L. (1982). Representation of the space of distributions useful in robust estimation of location. *The American Statistician* 37, 128-133.
- [10] Hall, D.L. und Joiner, B.L. (1983). Asymptotic relative efficiency of R-estimator of location. *Communications in Statistics - Theory and Methods*, 12, 739-763.
- [11] Handl, A. (1986). *Masszahlen zur Klassifizierung von Verteilungen bei der Konstruktion adaptiver verteilungsfreier Tests im unverbundenen Zweistichprobenproblem*, Dissertation, Freie Universität Berlin.
- [12] Hoeffding, W. (1948). A class of statistics with asymptotically normal distributions. *Annals of Statistics*, 19, 293-325.
- [13] Hogg, R.V. (1974). Adaptive robust procedures: partial review and some suggestions for future applications and theory. *Journal of the American Statistical Association*, 69, 909-923.
- [14] Hogg, R.V., Fisher, D.M. und Randles, R.H. (1975). A two-sample adaptive distribution-free test. *Journal of the American Statistical Association*, 70, 656-661.
- [15] Hušková, M. (1985). A partial review of adaptive procedures. In: *Sequential Methods in Statistics*, 16, Banach Centre Publications, Warschau, 233-243.
- [16] Kössler, W. (1994). Restrictive adaptive tests for the treatment of the two-sample scale problem. *Comput. Statist. Data Anal.* 18, 513-524.
- [17] Kössler, W. (1999). Rank tests in the two-sample scale problem with unequal and unknown locations. *Statistical Papers*, 40, 13-36
- [18] Kössler, W. (2005). Some c-Sample Rank Tests of Homogeneity Against Ordered Alternatives based on U-statistics. *Journal of Nonparametric Statistics*, 17, 777-797.
- [19] Kössler, W. (2006a). *Nonparametric Location Tests Against Restricted Alternatives*, Shaker, Aachen.
- [20] Kössler, W. (2006b). *Asymptotic Power and Efficiency of Lepage-Type Tests for the treatment of Combined Location-Scale alternatives*. Informatik-Bericht Nr. 200, Humboldt-Universität zu Berlin.
- [21] Kössler, W. (2010). Max-type rank tests, U-tests, and adaptive tests for the two-sample location problem - an asymptotic power study. *Computational Statistics and Data Analysis*, 54, 9, 2053-2065.
- [22] Kössler, W. und Kumar, N. (2008). An adaptive test for the two-sample location problem based on U-statistics. *Communications in Statistics - Computation and Simulation*, 37, 1329-1346.
- [23] Kössler, W. und Kumar, N. (2010). An adaptive test for the two-sample scale problem based on U-statistics. (in Vorbereitung)

- [24] Kumar, N., Singh, R.S. und Öztürk, Ö., (2003). A New Class of Distribution-Free Tests for Location parameters, *Sequential Analysis*, 22, 107-128.
- [25] Mann, H.B. und Whitney, D.R. (1947). On a test of whether one or two variables is stochastically larger than the other. *Ann. Math. Statist.*, 18, 50-60.
- [26] Neuhäuser, M., Büning, H. und Hothorn, L. (2004). Maximum Tests versus Adaptive Tests for the Two-Sample Location Problem. *Journal of Applied Statistics* 31, 215-227.
- [27] Policello, G.E. und Hettmansperger, T.P. (1976). Adaptive robust procedures for the one-sample location problem. *Journal of the American Statistical Association*, 71, 624-633.
- [28] Randles, R.H. und Wolfe, D.A. (1991). *Introduction to The Theory of Nonparametric Statistics*. Wiley, New York.
- [29] Wilcoxon, F. (1945). Individual comparisons by ranking methods. *Biometrics*, 1, 80-83.
- [30] Xie, J. und Priebe, C.E. (2000). Generalizing the Mann-Whitney-Wilcoxon Statistic. *Nonparametric Statistics*, 12, 661-682.
- [31] Xie, J. und Priebe, C.E. (2002). A weighted generalization of the Mann-Whitney-Wilcoxon statistic, *Journal of Statistical Planning and Inference*, 102, 441-466.

