



**HAL**  
open science

# An adaptive test for the two-sample location problem based on U-statistics

Wolfgang Koessler, Narinder Kumar

► **To cite this version:**

Wolfgang Koessler, Narinder Kumar. An adaptive test for the two-sample location problem based on U-statistics. *Communications in Statistics - Simulation and Computation*, 2008, 37 (07), pp.1329-1346. 10.1080/03610910801983152 . hal-00514325

**HAL Id: hal-00514325**

**<https://hal.science/hal-00514325v1>**

Submitted on 2 Sep 2010

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



**An adaptive test for the two-sample location problem based on U-statistics**

Journal:	<i>Communications in Statistics - Simulation and Computation</i>
Manuscript ID:	LSSP-2007-0093.R2
Manuscript Type:	Original Paper
Date Submitted by the Author:	17-Dec-2007
Complete List of Authors:	Koessler, Wolfgang; Humboldt-Universitaet zu Berlin, Informatik Kumar, Narinder; Panjab University Chandigarh, Statistics
Keywords:	Wilcoxon-Mann-Whitney test, , asymptotic efficacy, asymptotic power, simulation study, tailweight, skewness
Abstract:	For the two-sample location problem we consider a general class of tests, all members of it are based on U-statistics. The asymptotic efficacies are investigated in detail. We construct an adaptive test where all statistics involved are suitably chosen U-statistics. It is shown that the proposed adaptive test has good asymptotic and finite sample power properties.
<p>Note: The following files were submitted by the author for peer review, but cannot be converted to PDF. You must view these files (e.g. movies) online.</p> <p>koessler.tex Figures_KoesslerKumar.zip</p>	



1  
2  
3  
4  
5  
6  
7  
8  
9  
10  
11  
12  
13  
14  
15  
16  
17  
18  
19  
20  
21  
22  
23  
24  
25  
26  
27  
28  
29  
30  
31  
32  
33  
34  
35  
36  
37  
38  
39  
40  
41  
42  
43  
44  
45  
46  
47  
48  
49  
50  
51  
52  
53  
54  
55  
56  
57  
58  
59  
60

For Peer Review Only

# An adaptive test for the two-sample location problem based on U-statistics

W. Kössler, Humboldt-Universität zu Berlin,

koessler@informatik.hu-berlin.de

N. Kumar, Panjab University Chandigarh,

nkumar67@yahoo.com

## Abstract

For the two-sample location problem with continuous data we consider a general class of tests, all members of it are based on U-statistics. The asymptotic efficacies are investigated in detail. We construct an adaptive test where all statistics involved are suitably chosen U-statistics. It is shown that the proposed adaptive test has good asymptotic and finite sample power properties.

**Keywords:** Wilcoxon-Mann-Whitney test, asymptotic efficacy, asymptotic power, simulation study, tailweight, skewness.

**Subject classification:** 62G10, 62G20

## Corresponding author:

Wolfgang Kössler, Humboldt-Universität zu Berlin, Institut für Informatik  
Rudower Chaussee 25, D-12489 Berlin

**Running title:** Adaptive test based on U-statistics

# 1 Introduction

Let  $X_1, \dots, X_{n_1}$  and  $Y_1, \dots, Y_{n_2}$  be independent random samples from populations with absolutely continuous distribution functions  $F(x)$  and  $F(x - \vartheta)$ ,  $\vartheta \in \mathbf{R}$ , respectively. We wish to test

$$H_0 : \quad \vartheta = 0$$

against

$$H_1 : \quad \vartheta > 0.$$

The Wilcoxon-Mann-Whitney test is the most familiar nonparametric test for this problem. This test was generalised to linear rank tests with various other scores, such as the Median test, the normal scores test and the Savage test, see e.g. Hájek, Šidak and Sen (1999).

Another generalisation is to consider a class of tests based on U-statistics. This interesting class of tests has drawn considerable attraction in the literature (cf. e.g. Deshpande and Kochar (1982), Shetty and Govindarajulu (1988), Kumar (1997), Xie and Priebe (2000, 2002), John and Priebe (2007)).

Following Kumar, Singh and Öztürk (2003) a general class  $U_{k,i}$  of U-statistics is defined in Section 2. Local alternatives of the form  $\vartheta = \theta_N = \theta/\sqrt{N}$ ,  $N = n_1 + n_2$ , are considered and the asymptotic efficacies of the tests based on  $U_{k,i}$  are compared in Section 3. It is shown that there are different tests of this type which are efficient for densities with short, medium or long (right or left) tails, respectively. For example, the test based on  $U_{5,1}$  is efficient for densities with short tails, and that based on  $U_{5,3}$  is efficient for densities with long tails. However, the practising statistician has generally no clear idea on the underlying density, thus he/she should apply an adaptive test which takes into account the given data set. In Section 4 two versions of such an adaptive test are proposed, one of them is distribution-free. The adaptive tests first classify the underlying distribution with respect to some measures like (right and left) tailweight and skewness and then select an appropriate test based on U-statistics. Our adaptive test is compared briefly with adaptive tests based on linear rank tests in Section 5 **and with the *t*-test and the Mann-Whitney-Wilcoxon test in Section 6**. In Section 7 a simulation study is performed and the finite sample power is compared with the asymptotic power. It is shown that one of the adaptive tests behaves well also for moderate sample sizes. **A data example is provided in Section 8. We give some conclusions in Section 9.**

## 2 Test statistics

We consider the class of U-statistics, which was proposed by Kumar, Singh and Öztürk (2003). Let  $k$ ,  $1 \leq k \leq \min(n_1, n_2)$ , and  $i$ ,  $i \leq \frac{k+1}{2}$ , be fixed integers. Define

$$\Phi_i(x_1, \dots, x_k, y_1, \dots, y_k) = \begin{cases} 2 & \text{if } x_{(i)k} < y_{(i)k} \text{ and } x_{(k-i+1)k} < y_{(k-i+1)k} \\ 1 & \text{if either } x_{(i)k} < y_{(i)k} \text{ or } x_{(k-i+1)k} < y_{(k-i+1)k} \\ 0 & \text{otherwise,} \end{cases}$$

where  $x_{(i)k}$  is the  $i$ th order statistic in a subsample of size  $k$  from the  $X$ -sample (and likewise for  $y$ 's). Let  $U_{k,i}$  be the U-statistic associated with kernel  $\Phi_i$ , i.e.

$$U_{k,i} = \frac{n_1 n_2}{\binom{n_1}{k} \cdot \binom{n_2}{k}} \sum \Phi_i(X_{r_1}, \dots, X_{r_k}, Y_{s_1}, \dots, Y_{s_k}) - n_1 n_2,$$

where the summation extends over all possible combinations  $(r_1, \dots, r_k)$  of  $k$  integers from  $\{1, \dots, n_1\}$  and all possible combinations  $(s_1, \dots, s_k)$  of  $k$  integers from  $\{1, \dots, n_2\}$ . The null hypothesis  $H_0$  is rejected in favour of  $H_1$  for large values of  $U_{k,i}$ .

**Remark:** The following special cases are of particular interest.

For  $i = 1$  and  $k = 1$  we have the Wilcoxon-Mann-Whitney test.

For  $i = 1$  or  $i = k$  we have the Deshpande-Kocher test (cf. Deshpande and Kocher, 1982).

For  $i = (k + 1)/2$  we have the Kumar-test (cf. Kumar, 1997).

Let

$$\begin{aligned} \varphi_{1,0}^{(i)}(x) &= \mathbf{E}\Phi_i(x, X_2, \dots, X_k, Y_1, \dots, Y_k) \\ \varphi_{0,1}^{(i)}(y) &= \mathbf{E}\Phi_i(X_1, \dots, X_k, y, Y_2, \dots, Y_k) \\ \zeta_{1,0}^{(i)} &= \mathbf{Var}(\varphi_{1,0}^{(i)}(X)) \\ \zeta_{0,1}^{(i)} &= \mathbf{Var}(\varphi_{0,1}^{(i)}(Y)), \end{aligned}$$

where  $\mathbf{E}$  and  $\mathbf{Var}$  denote the expectation and variance respectively. Moreover, let  $F_{(i)k}(\cdot)$  be the cumulative distribution function of the  $i$ th order statistics of a sample of size  $k$ .

**Proposition 2.1** (cf. Kumar, Singh and Öztürk, 2003) *Under assumptions  $N \rightarrow \infty$ ,  $n_1/N \rightarrow \lambda$ ,  $0 < \lambda < 1$  the limiting distribution of  $N^{1/2}(U_{k,i} - \eta_{k,i})/\sigma_{k,i}$  is standard normal, where expectation  $\eta_{k,i} = \mathbf{E}U_{k,i}$  and variance  $\sigma_{k,i}^2 = \mathbf{Var}(U_{k,i})$  have the forms*

$$\eta_{k,i} = n_1 n_2 \left( \int_{-\infty}^{\infty} F_{(i)k}(y) dF_{(i)k}(y - \theta) + \int_{-\infty}^{\infty} F_{(k-i+1)k}(y) dF_{(k-i+1)k}(y - \theta) \right) - n_1 n_2$$

$$\sigma_{k,i}^2 = n_1^2 n_2^2 \left( \frac{k^2 \zeta_{10}^{(i)}}{\lambda} + \frac{k^2 \zeta_{01}^{(i)}}{1 - \lambda} \right).$$

**Remark:** Under  $H_0$  we have  $\eta_{k,i} = 0$  and

$$\sigma_{k,i}^2 = n_1^2 n_2^2 k^2 \frac{\rho_{k,i}}{\lambda(1 - \lambda)},$$

where  $\rho_{k,i}$  depends on  $k$  and  $i$  only. The expression for  $\rho_{k,i}$  is rather long, that is why we do not write it out. It can be found in Kumar, Singh and Öztürk (2003), pp.125-126.

### 3 The asymptotic efficacies

The asymptotic efficacies AE of the statistics  $U_{k,i}$  under the alternative  $\theta_N = N^{-1/2} \cdot \theta$  are given by

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

where  $f(\cdot)$  denotes the probability density function belonging to the c.d.f.  $F(\cdot)$  and

$$C_{k,i}(f) = \frac{\binom{k}{i} i^2}{(k^2 \rho_{k,i})^{1/2}} \cdot \left( \int_{-\infty}^{\infty} (F(x))^{2i-2} (1 - F(x))^{2k-2i} f^2(x) dx + \int_{-\infty}^{\infty} (F(x))^{2k-2i} (1 - F(x))^{2i-2} f^2(x) dx \right)$$

(cf. Kumar, Singh and Öztürk, 2003).

Note that the asymptotic efficacy is defined by the limit of  $\eta_{k,i}^2/\sigma_{k,i}^2$ , cf. Noether (1955).

From Kumar, Singh and Öztürk (2003) we know that for some underlying distributions the test based on  $U_{k,i}$  has highest efficacy if  $k$  is as large as possible as long as it is less than sample size. For example, for the uniform  $i$  should be 1 or  $k$ , and for the double exponential  $k$  should be odd and  $i = (k+1)/2$ . Therefore, for further investigations we have to make some restrictions. We restricted the choice of  $k$  as  $k \leq 5$  ourselves just for illustrative purpose. Of course, computations can be made for values of  $k$  beyond that. We compute the asymptotic Pitman efficacies for all tests  $U_{k,i}$  with  $1 \leq i \leq k$ ,  $k \leq 5$ . Values of the factors  $C_{k,i}^2(f)$  for various densities are presented in Table 1. We also give the corresponding factors for the  $t$ -test (see the first column). The L-DE density was proposed by Policello and Hettmansperger (1976), the U-L by Gastwirth (1965), the RST is named after Ramberg, Schmeiser and Tukey, cf. Ramberg and Schmeiser (1972, 1974),  $CN(\epsilon, \sigma)$  is the scale contaminated normal with contaminating proportion  $\epsilon$ , Mielke denotes the Mielke (1972) density and BT is the density of Box and Tiao (1962).

The bold entries denote, for the given density, the asymptotically best test among the considered tests. (If the  $t$ -test is the asymptotically best, it is denoted in italics. Note that the  $t$ -test is not allowed to be included in the adaptive procedure since the adaptive test requires independence between the two steps, see Proposition 4.1 below.) On the first view we see that the columns for  $U_{3,1}$ ,  $U_{5,1}$  and  $U_{5,3}$  have the most bold entries. This observation gives rise to the idea to base an adaptive test on these few statistics. (The classical test  $U_{1,1}$  is also included in the adaptive test.)

To get a closer idea how to classify densities we apply the method of Hall and Joiner (1982). The content of information in the asymptotic efficacy matrix is analysed by a principal component analysis where the densities are the observations and the efficacies of the  $U_{i,j}$  are the variables. The first principal component explains already 96% of the variability (Figure 1). For better visibility we display the values of the first two (varimax rotated) principal components (Factor 1 and Factor 2). In Figure 1 nearly symmetric densities with short tails are denoted by a green plus, that with short-medium and long-medium tails by a cyan X and a blue star, respectively, and that with long tails with a red dot. Skew densities are denoted by a black plus (short tails) and a yellow dot (medium tails) respectively. The dots denote densities with long tails, the stars and X that with medium tails, and the



plus that with short tails. On the left side we have densities with long tails, in the centre that with medium tails, and on the right that with short tails. For an exact definition what we understand by long, medium and short tails see below. On the first view we see that the  $AE(U_{k,i})$  classify the densities according to their tailweight. The skewness seems to play a marginal role only.

Insert Figure 1 about here

## 4 Adaptive test

There are some proposals for adaptive tests for the two-sample location problem in the literature, see e.g. Hogg (1974, 1982), Hogg, Fisher and Randles (1975), Ruberg (1986) and Büning (1994). We apply the concept of Hogg (1974), that is, to classify at first the type of the underlying density with respect to one measure of skewness  $\hat{s}$  and to three measures of tailweight  $\hat{t}$ ,  $\hat{t}_r$  and  $\hat{t}_l$ , which are defined by

$$\hat{s} = \frac{\hat{Q}(0.95) + \hat{Q}(0.05) - 2 \cdot \hat{Q}(0.5)}{\hat{Q}(0.95) - \hat{Q}(0.05)} \quad (1)$$

$$\hat{t} = \frac{\hat{Q}(0.95) - \hat{Q}(0.05)}{\hat{Q}(0.85) - \hat{Q}(0.15)} \quad (2)$$

$$\hat{t}_l = \frac{\hat{Q}(0.5) - \hat{Q}(0.05)}{\hat{Q}(0.5) - \hat{Q}(0.15)} \quad \hat{t}_r = \frac{\hat{Q}(0.95) - \hat{Q}(0.5)}{\hat{Q}(0.85) - \hat{Q}(0.5)} \quad (3)$$

where  $\hat{Q}(u)$  is the so-called classical quantile estimate of  $F^{-1}(u)$ ,

$$\hat{Q}(u) = \begin{cases} X_{(1)} - (1 - \delta) \cdot (X_{(2)} - X_{(1)}) & \text{if } u < 1/(2 \cdot N) \\ (1 - \delta) \cdot X_{(j)} + \delta \cdot X_{(j+1)} & \text{if } \frac{1}{2 \cdot N} \leq u \leq \frac{2 \cdot N - 1}{2 \cdot N} \\ X_{(N)} + \delta(X_{(N)} - X_{(N-1)}) & \text{if } u > (2 \cdot N - 1)/(2 \cdot N), \end{cases} \quad (4)$$

where  $\delta = N \cdot u + 1/2 - j$  and  $j = \lfloor N \cdot u + 1/2 \rfloor$ . Note that  $\hat{t}_l$  and  $\hat{t}_r$  are measures of left tailweight and right tailweight, respectively. All the measures are estimated from the pooled sample.

In Tables 2 and 3 the values of the corresponding theoretical measures  $s$ ,  $t$ ,  $t_r$  and  $t_l$ , for various selected densities are presented. (For symmetric densities we have  $s = 0$  and  $t = t_r = t_l$ .)

Comparing Table 1 with Tables 2 and 3 roughly we see that  $U_{5,1}$  is the asymptotically best test for symmetric densities with small tailweight,  $U_{3,1}$  for symmetric densities with small to medium tailweight,  $U_{1,1}$  for symmetric densities with medium to larger tailweight, and  $U_{5,3}$  for symmetric densities with large tailweight. The tests  $U_{3,1}$  and  $U_{1,1}$  should be included in an adaptive test since they are the (asymptotically) best for the normal and for the logistic density, respectively (at least among the considered tests).

The measure of skewness gives no clear classification idea. That is why we consider left tailweight  $t_l$  and right tailweight  $t_r$ , and classify densities as densities with partially short tails if  $t_l < 1.55$  or  $t_r < 1.55$ . They are classified to have partially medium tails if  $t_l < 1.8$  or  $t_r < 1.8$  and if they have not partially short tails.

The reasoning of the last two sections gives rise to the following adaptive test.

Define regions  $E_1, \dots, E_7$  of  $\mathbb{R}^4$  which are based on the so called selector statistic  $\hat{S} = (\hat{s}, \hat{t}, \hat{t}_l, \hat{t}_r)$

$$\begin{aligned}
 E_1 &= \{\hat{t} < 1.55, |\hat{s}| \leq 0.2\} && \text{“nearly symmetric, short tails”} \\
 E_2 &= \{1.55 \leq \hat{t} < 1.65, |\hat{s}| \leq 0.2\} && \text{“nearly symmetric, light medium tails”} \\
 E_3 &= \{1.65 \leq \hat{t} \leq 1.8, |\hat{s}| \leq 0.2\} && \text{“nearly symmetric, heavy medium tails”} \\
 E_4 &= \{\hat{t} > 1.8, |\hat{s}| \leq 0.2\} && \text{“nearly symmetric, long tails”} \\
 E_5 &= \{(\hat{t}_l < 1.55 \vee \hat{t}_r < 1.55), |\hat{s}| > 0.2\} && \text{“skew, partially short tails”} \\
 E_6 &= \{(\hat{t}_l > 1.65 \wedge \hat{t}_r > 1.65), |\hat{s}| > 0.2\} && \text{“skew, long tails”} \\
 E_7 &= \{\hat{t}_l \geq 1.55, \hat{t}_r \geq 1.55, |\hat{s}| > 0.2\} \setminus E_6 && \text{“skew, partially medium tails”}
 \end{aligned}$$

where  $\hat{s}$ ,  $\hat{t}$ ,  $\hat{t}_l$  and  $\hat{t}_r$  are given by (1) to (3). Note that there was no density which belongs to class  $E_7$ .

The cutoff values of the regions are determined in such a way that the vast majority of densities is classified correctly, i.e. they fall in the class that has the highest asymptotic power (cf. Tables 2 and 3 with Table 1). For example, the normal (tailweight  $t=1.59$ , cf. Table 2) is classified to  $E_2$ , and the test  $U_{3,1}$ , which is the best among the considered tests (cf. Table 1), is performed. The logistic is mapped to region  $E_3$  and the optimal test  $U_{1,1}$  is performed. Similar observations for the other densities lead to the given cutoff values. In few cases, if the classification is not correct, then the efficacy loss is very small in almost all cases. In Table 1 the chosen test is underlined

if it is not already the (bold) best.

Now, we propose the Adaptive test  $A$  which is based on the four U-statistics  $U_{5,1}$ ,  $U_{3,1}$ ,  $U_{1,1}$ , and  $U_{5,3}$ . We denote the tests by  $(5,1)$ ,  $(3,1)$ ,  $(1,1)$  and  $(5,3)$ , respectively.

$$A = A(\hat{S}) = \begin{cases} (5,1) & \text{if } \hat{S} \in E_1 \cup E_5 \\ (3,1) & \text{if } \hat{S} \in E_2 \cup E_7 \\ (1,1) & \text{if } \hat{S} \in E_3 \\ (5,3) & \text{if } \hat{S} \in E_4 \cup E_6 \end{cases} \quad (5)$$

In Figure 2 the corresponding adaptive scheme is given. As indicated above the skewness plays only a marginal role. It is included in the adaptive scheme implicitly by left and right tailweight.

Insert Figure 2 about here

The two-stage procedure defined above is asymptotically distribution-free since the selector statistic  $\hat{S}$  is based on the order statistic only and the U-statistics are based on the ranks only.

The Adaptive test  $A$  is only *asymptotically* distribution-free because *asymptotic* critical values are used in the adaptive scheme.

**Proposition 4.1** *Let  $\sigma_F$  be the standard deviation of the underlying cdf.  $F$ , if it exists and let  $\{\theta_N\}$  be a sequence of 'near' alternatives with  $\sqrt{N}\theta_N \rightarrow \sigma_F\theta$ . The asymptotic power function of the Adaptive test  $A$  equals*

$$\beta(\theta) = \begin{cases} 1 - \Phi(z_{1-\alpha} - \sqrt{AE(U_{5,1}|f)} \cdot \sigma_F \cdot \theta) & \text{if } f \in E_1 \cup E_5 \\ 1 - \Phi(z_{1-\alpha} - \sqrt{AE(U_{3,1}|f)} \cdot \sigma_F \cdot \theta) & \text{if } f \in E_2 \cup E_7 \\ 1 - \Phi(z_{1-\alpha} - \sqrt{AE(U_{1,1}|f)} \cdot \sigma_F \cdot \theta) & \text{if } f \in E_3 \\ 1 - \Phi(z_{1-\alpha} - \sqrt{AE(U_{5,3}|f)} \cdot \sigma_F \cdot \theta) & \text{if } f \in E_4 \cup E_6 \end{cases}$$

**Proof.** Let be  $h = 1$  if  $(k, i) = (5, 1)$ ,  $h = 2$  if  $(k, i) = (3, 1)$ ,  $h = 3$  if  $(k, i) = (1, 1)$ ,  $h = 4$  if  $(k, i) = (5, 3)$  Let be  $T_1 = U_{5,1}$ ,  $T_2 = U_{3,1}$ ,  $T_3 = U_{1,1}$ ,  $T_4 = U_{5,3}$  and  $D_1 = E_1 \cup E_5$ ,  $D_2 = E_2 \cup E_7$ ,  $D_3 = E_3$ ,  $D_4 = E_4 \cup E_6$ . The proposition follows from the total probability theorem and from the

consistency of the selector statistics, i.e.

$$\begin{aligned} \beta(\theta) &= \sum_{h=1}^4 P_{\sigma_F \theta}(T_h > c_{\alpha h} | T_h \text{ chosen}) \cdot P_{\sigma_F \theta}(T_h \text{ chosen}) \\ &= \sum_{h=1}^4 (1 - \Phi(z_{1-\alpha} - \sqrt{AE(T_h|f)} \cdot \sigma_F \cdot \theta) + o(1)) \cdot \begin{cases} 1 + o(1) & \text{if } f \in D_h \\ o(1) & \text{else} \end{cases} \\ &\sim 1 - \Phi(z_{1-\alpha} - \sqrt{AE(T_h|f)} \cdot \sigma_F \cdot \theta) \quad \text{if } f \in D_h, \end{aligned}$$

where  $c_{\alpha h}$  is the  $(1 - \alpha)$ -quantile of the asymptotic null distribution of  $T_h$ . ■

**Remark:** The factor  $\sigma_F$  is introduced to have similar power values for the various distributions. For the Cauchy the factor  $\sigma_F$  is set to  $\sigma_F = \sigma_{Cau} = F^{-1}(\Phi(1)) = 1.8373$ .

The Adaptive test  $A(\hat{S})$  is based on selector statistics computed from the pooled sample. However, location differences may effect the estimates of tailweight and skewness. That is why we consider also a modification  $A(\hat{S}^*)$  of the adaptive test, where tailweights and skewness are estimated from the single samples. Let  $\hat{s}_i$  and  $\hat{t}_i$ ,  $\hat{t}_{l,i}$  and  $\hat{t}_{r,i}$ ,  $i = 1, 2$  be statistics of the form (1) to (3) for skewness and tailweight, left tailweight and right tailweight, respectively. Applying the  $A(\hat{S}^*)$ -test the selector statistic  $\hat{S}^* = (\hat{s}^*, \hat{t}^*, \hat{t}_l^*, \hat{t}_r^*)$  with

$$\begin{aligned} \hat{s}^* &= \frac{n_1}{N} \hat{s}_1 + \frac{n_2}{N} \hat{s}_2 & \hat{t}^* &= \frac{n_1}{N} \hat{t}_1 + \frac{n_2}{N} \hat{t}_2 \\ \hat{t}_l^* &= \frac{n_1}{N} \hat{t}_{l,1} + \frac{n_2}{N} \hat{t}_{l,2} & \hat{t}_r^* &= \frac{n_1}{N} \hat{t}_{r,1} + \frac{n_2}{N} \hat{t}_{r,2} \end{aligned}$$

is used instead of  $\hat{s}$ ,  $\hat{t}$ ,  $\hat{t}_l$  and  $\hat{t}_r$ . This procedure is also asymptotically distribution-free. However, it is not distribution-free also if the exact critical values are used. This property is due to the fact that the selector statistic is no longer based on the pure order statistic.

For various densities asymptotic power functions (together with finite power functions) are given in Figures 3 and 4.

The blue dotted line is for  $U_{1,1}$ , the violet short-dashed line for  $U_{3,1}$ , the green long-dashed line for  $U_{5,1}$ , the red dashed-dotted line for  $U_{5,3}$  (and the black continuous line for the adaptive test).

**Remark:** Another adaptive test which is based on U-statistics, but with another classes of U-statistics (and in another context) is proposed in Kössler

(2005). It turns out, that the adaptive test presented here is slightly better for the majority of considered densities, especially for short tail densities (uniform, exponential).

**Remark:** Another kind of (nonrestrictive) adaptive tests was constructed by John and Priebe (2007). They take the class of weighted generalised Mann-Whitney-Wilcoxon (WGMWW) statistics proposed by Xie and Priebe (2002), and estimate in a first step the asymptotic efficacies, where the weights are considered as parameters. In a second step these weights are determined in such a way that the estimated asymptotic efficacy is maximised. Then the test based on the resulting WGMWW statistic is performed.

## 5 Comparison to adaptive tests based on linear rank statistics

Restrictive adaptive tests for the two-sample location problem based on linear rank tests are proposed by Hogg (1974, 1982), Hogg, Fisher and Randles (1975), Handl (1986) and Büning (1994). All of them are based on the concept of Hogg (1974), and they use few linear rank statistics, with the following scores: Gastwirth (GA, for short tails), Wilcoxon (WI, for medium tails), Median (Hogg, 1974, and Hogg et al., 1975, Long-tail (LT, Handl, 1986) and Büning, 1994, both for long tails) and Hogg-Fisher-Randles (HFR, for right-skew densities). Since the Median test is known to be bad for most densities (except for the doubleexponential), we restrict to the scores GA, WI, LT and HFR, and call the corresponding adaptive scheme  $B(\hat{S})$  (where the form of  $\hat{S}$  is not of interest here). For some densities we take the respective two asymptotically best tests in the the Adaptive schemes  $A(\hat{S})$  and  $B(\hat{S})$  (cf. Table 4).

For the classical densities considered the U-statistics based test  $A(\hat{S})$  has slightly higher asymptotic power than the test  $B(\hat{S})$ . (For the logistic they are, of course, the same.) For the densities U-L (0.75) and L-D (0.75) that are 'optimal' densities for the tests GA and LT (cf. Büning and Kössler, 1999) it is vice versa.

## 6 Comparison to the $t$ -test and to the Wilcoxon test

To evaluate the adaptive test we compare the AE values for the  $t$ -test, the Wilcoxon test, and for the Adaptive test  $A(S)$ . The test with the highest AE gets rank one, that with the second highest rank two, and that with the lowest AE (among the three tests) rank three, all for the various densities considered. In the case of equal AE values, average ranks are taken. The results are presented in Table 5. The average rank over all 86 densities is 1.23 for the Adaptive test  $A(S)$ , in comparison to 2.10 for the Wilcoxon test and 2.67 for the  $t$ -test. Note that this result is not surprising since the adaptive test  $A(S)$  is constructed in such a way that it has high AE.

## 7 Simulation study

In order to assess whether the asymptotic theory can also be applied for medium to small sample sizes a simulation study (10,000 replications each) is performed. We choose the following six distributions:

- Uniform distribution (density with small tailweight),
- Normal distribution (density with medium tailweight),
- Logistic distribution (density with medium tailweight),
- Doubleexponential distribution (density with large tailweight),
- Cauchy distribution (density with very large tailweight),
- Exponential distribution (very skew density)

We consider the four single tests  $U_{5,1}$ ,  $U_{3,1}$ ,  $U_{1,1}$ ,  $U_{5,3}$  and the two Adaptive tests  $A(\hat{S})$  and  $A(\hat{S}^*)$ . The sample sizes  $n_1 = n_2 = 10, 20, 40$  and the alternatives  $\theta_N = N^{-1/2}\theta\sigma_F$  with various  $\theta$  are considered. Recall that the factor  $\sigma_F$  denotes the standard deviation of the underlying distribution function  $F$  ( $\sigma_{Cauchy} := F^{-1}(\Phi(1)) = 1.8373$ ).

The Adaptive test  $A(\hat{S}^*)$  is slightly anticonservative, also for  $n_1 = n_2 = 40$ , with correspondingly slightly higher power than the Adaptive test  $A(\hat{S})$  (except for the Cauchy). This test is not considered here.

1  
2  
3  
4  
5  
6  
7  
8 For  $n_1 = n_2 = 10$  and  $n_1 = n_2 = 20$  almost all tests (except the test  
9  $U_{1,1}$  which attains the level of significance) are slightly anticonservative, for  
10  $n_1 = n_2 = 10$  the attained level of significance is always less than 0.08, and  
11 for  $n_1 = n_2 = 20$  it is less than 0.06 and therefore  $\epsilon = 0.2$ -robust. (For  
12 the notion of  $\epsilon$ -robustness see e.g. Rasch, Teuscher, and Guiard (2007).) For  
13  $n_1 = n_2 = 40$  the results of the simulation study are summarized in Figures 3  
14 and 4. Again, the blue dotted line is for  $U_{1,1}$ , the violet short-dashed line for  
15  $U_{3,1}$ , the green long-dashed line for  $U_{5,1}$ , the red dashed-dotted line for  $U_{5,3}$   
16 and the continuous line is for the Adaptive test  $A(\hat{S})$ . At first we see that,  
17 for  $n_1 = n_2 = 40$ , the finite power is well approximated by the asymptotic  
18 power (except for the Cauchy). Moreover, it can be seen that, for a given  
19 density, there is always, sometimes together with another test, a single test  
20 which is the best. The test  $U_{5,1}$  is the best for the uniform and for the  
21 exponential (together with the Adaptive test  $A(\hat{S})$ ), the test  $U_{1,1}$  is the best  
22 for the normal and for the logistic density, and the test  $U_{5,3}$  is the best for  
23 the doubleexponential and for the Cauchy. All these facts are not surprising.  
24 Also, not surprisingly, the tests  $U_{5,1}$  and  $U_{5,3}$  may be bad for some densities.  
25 The tests  $U_{1,1}$  and  $A(\hat{S})$  are, over all densities, the best. However, for the  
26 uniform and for the exponential densities the adaptive test is clearly better.  
27

28 For the Cauchy density, somewhat surprisingly, the test  $U_{1,1}$  is clearly  
29 better than the adaptive tests. The Adaptive test  $A(\hat{S}^*)$  is better than  $A(\hat{S})$ .  
30 The reason for these facts is, that for small and moderate sample sizes, the  
31 misclassification rate into the class  $E_1$  is relatively large.  
32

33  
34  
35  
36  
37  
38  
39 Insert Figures 3 and 4 about here  
40  
41

## 42 8 Data example

43 At the 50kms ski-running competition in the Jizerske hory 2002 altogether  
44 2320 runners finished the race. We consider only two age groups, where the  
45 first group contains men with an age less then 40, and the second group men  
46 of an age of 40 to 49. The (perhaps somewhat trivial?) conjecture is that the  
47 running times of the older are longer than that of the younger men. Figure  
48 5 gives a box plot of these data.  
49  
50  
51  
52  
53  
54  
55

1  
2  
3  
4  
5  
6  
7  
8  
9  
10  
11  
12  
13  
14  
15  
16  
17  
18  
19  
20  
21  
22  
23  
24  
25  
26  
27  
28  
29  
30  
31  
32  
33  
34  
35  
36  
37  
38  
39  
40  
41  
42  
43  
44  
45  
46  
47  
48  
49  
50  
51  
52  
53  
54  
55  
56  
57  
58  
59  
60

Insert Figure 5 about here

At first, we estimate skewness and tailweight, cf. (1) and (2), and we obtain  $\hat{s} = 0.116$  and  $\hat{t} = 1.500$ . (The computation of left and right tailweights is not necessary here.) According to our adaptive procedure the test  $U_{5,1}$  is chosen. We obtain  $U_{5,1} = 1.633$  which results in an one-sided p-value of 0.0511. **Although the boxes look very similar we obtain a relatively small p-value. On one hand this fact may be due to the large sample sizes, on the other hand our procedure gives a weak evidence against  $H_0$ .** For comparison we also computed the one-sided p-values for the (here not recommended) Mann-Whitney Wilcoxon test,  $p = 0.181$  and for the  $t$ -test,  $p = 0.188$ .

## 9 Conclusions

What are the results of our study? At first, we see that the finite power of the considered tests based on U-statistics can be well approximated by their asymptotic power. Second, there are modifications of the “classical” Mann-Whitney test MW that may have (considerably) higher power than MW for symmetric as well asymmetric densities. Third, the Adaptive test  $A(\hat{S})$  is a serious alternative for the Wilcoxon-Mann-Whitney test  $U_{1,1}$  for moderate to large sample sizes. Especially, for short-tail densities and for skew densities the adaptive test considered here is better than the adaptive test  $B(\hat{S})$  based on linear rank tests.

Many tests based on U-Statistics have good power properties. They can be applied to a broader class of underlying densities than a linear rank test with fixed scores.

For moderate to large sample sizes (about  $n_1, n_2 \geq 20$ ) the recommendation for the practising statistician is as follows: If the density is known to be nearly normal take the  $t$ -test. If it is only known that it has medium tails then take the Mann-Whitney-Wilcoxon test or the  $t$ -test. (For the robustness of the  $t$ -test see e.g. Büning (1991) or Rasch, Teuscher, and Guiard (2007) and the references therein.) For densities with short tails take the test  $U_{5,1}$ , and for those with long tails the test  $U_{5,3}$ . For a completely unknown density take the adaptive test  $A(\hat{S})$ .



1  
2  
3  
4  
5  
6  
7  
8 For relatively small sample sizes (about  $n_1, n_2 = 10$ ) take the Mann-  
9 Whitney-Wilcoxon test. If it is known that the densities have not too large  
10 tails, then  $t$ -test may be applied. (Of course, the  $t$ -test is allowed only, if  
11 outliers can be excluded.)  
12

13  
14 Note that we considered the location problem for continuous data only.  
15 How the procedures behave for ordinal data may be a topic for further stud-  
16 ies. Comparisons of the Mann-Whitney-Wilcoxon test with the  $t$ -test are  
17 made in the recent paper of Rasch, Teuscher, and Guiard (2007). Further  
18 investigations are also desirable for more general alternatives, such as  $F \neq G$   
19 or for the case of additional scale differences between  $F$  and  $G$ .  
20  
21

22 **Acknowledgements** The authors are grateful to the editor and the referees  
23 who gave valuable hints on an improvement of an earlier version of this paper.  
24  
25

## 26 27 28 29 30 31 32 33 34 35 36 37 38 39 40 41 42 43 44 45 46 47 48 49 50 51 52 53 54 55 56 57 58 59 60

### References

Box, G.E.P., Tiao, G.C. (1962). A further look at robustness via Bayes's  
theorem. *Biometrika*, 49:419-432.

Büning, H. (1991). *Robuste und adaptive Tests*. De Gruyter, Berlin.

Büning, H. (1994). Robust and adaptive tests for the two-sample location  
problem. *OR Spektrum*, 16:33-39.

**Büning, H. and Kössler, W. (1999). The Asymptotic Power of  
Jonckheere-Type Tests for Ordered Alternatives, *Australian &  
New Zealand Journal of Statistics*, vol.41:67-78.**

Deshpande, J.V., Kocher, S.C. (1982). Some competitors of Wilcoxon-Mann-  
Whitney test for location alternatives. *Journal of Indian Statistical Asso-  
ciation*, 19:9-18.

Gastwirth, J.L. (1965). Percentile modifications of two-sample rank tests.  
*Journal of the American Statistical Association*, 60:1127-1141.

Hájek, J., Šidák, Z., Sen, P.K. (1999). *Theory of Rank Tests*, San Diego:  
Academic Press.

Hall, D.L., Joiner, B.L. (1983). Asymptotic Relative Efficiency of  $R$ -estimators  
of Location. *Communications in statistics, Theory and Methods*, 12:739-  
763.

- 1  
2  
3  
4  
5  
6  
7  
8 Handl, A (1986). *Masszahlen zur Klassifizierung von Verteilungen bei der*  
9 *Konstruktion adaptiver verteilungsfreier Tests im unverbundenen Zwei-*  
10 *tichprobenproblem*, unpublished dissertation, Freie Universität Berlin.
- 11  
12 Hogg, R.V. (1974). Adaptive robust procedures: partial review and some  
13 suggestions for future applications and theory. *Journal of the American*  
14 *Statistical Association*, 69:909-923.
- 15  
16 Hogg, R.V. (1982). On adaptive statistical inference. *Communications in*  
17 *Statistics, Theory and Methods*, 11:2531-2542.
- 18  
19 Hogg, R.V., Fisher, D.M., Randles, R.H. (1975). A two-sample adaptive  
20 distribution-free test. *Journal of the American Statistical Association*,  
21 70:656-661.
- 22  
23 John, M., Priebe, C.E. (2007). A data-adaptive methodology for finding an  
24 optimal weighted generalized Mann-Whitney-Wilcoxon statistic. *Compu-*  
25 *tational Statistics & Data Analysis*, 51:4337-4353.
- 26  
27 Kössler, W. (2005). Some  $c$ -Sample Rank Tests of Homogeneity Against  
28 Ordered Alternatives based on U-statistics. *Journal of Nonparametric*  
29 *Statistics*, 17:777-796.
- 30  
31 Kumar, N. (1997). A Class of Two-Sample tests for Location Based on Sub-  
32 Sample Medians. *Communications in Statistics, Theory and Methods*,  
33 26:943-951.
- 34  
35 Kumar, N., Singh, R.S., Öztürk, Ö. (2003). A New Class of Distribution-Free  
36 Tests for Location parameters. *Sequential Analysis*, 22:107-128.
- 37  
38 Mielke, P.W.Jr. (1972). Asymptotic behaviour of two-sample tests based on  
39 powers of ranks for detecting scale and location alternatives. *Journal of*  
40 *the American Statistical Association*, 67:850-854.
- 41  
42 Noether, G.E. (1955). On a Theorem of Pitman. *Annals of Mathematical*  
43 *Statistics*, 26:64-68.
- 44  
45 Policello, G.E., Hettmansperger, T.P. (1976). Adaptive robust procedures  
46 for the one-sample location problem. *Journal of the American Statistical*  
47 *Association*, 71:624-633.
- 48  
49 Ramberg, J.S., Schmeiser, B.W. (1972). An approximate method for gener-  
50 ating symmetric random variables, *Communications of the ACM*, 11:987-  
51 990.
- 52  
53  
54  
55  
56  
57  
58  
59  
60

- 1  
2  
3  
4  
5  
6  
7  
8 Ramberg, J.S., Schmeiser, B.W. (1974). An approximate method for generat-  
9 ing asymmetric random variables, *Communications of the ACM*, 17:78-82.  
10  
11 Rasch, D., Teuscher, F., and Guiard, V. (2007). How robust are tests for  
12 two independent samples? *Journal of Statistical Planning and Inference*,  
13 137:2706-2720.  
14  
15 Ruberg, S.J. (1986). A continuously adaptive nonparametric two-sample  
16 test. *Communications in Statistics, Theory and Methods*, 15:2899-2920.  
17  
18 Shetty, I.D., Govindarajulu, Z. (1988). A Two-Sample Test for Location.  
19 *Communications in Statistics, Theory and Methods*, 27:2389-2401.  
20  
21 Xie, J., Priebe, C.E. (2000). Generalizing the Mann-Whitney-Wilcoxon  
22 Statistic. *Nonparametric Statistics*, 12:661-682.  
23  
24 Xie, J., Priebe, C.E. (2002). A weighted generalization of the Mann-Whitney-  
25 Wilcoxon Statistic. *Journal of Statistical Planning and Inference*, 102:441-  
26 466.  
27  
28  
29  
30  
31  
32  
33  
34  
35  
36  
37  
38  
39  
40  
41  
42  
43  
44  
45  
46  
47  
48  
49  
50  
51  
52  
53  
54  
55  
56  
57  
58  
59  
60

1  
2  
3  
4  
5  
6  
7  
8  
9  
10  
11  
12  
13  
14  
15  
16  
17  
18  
19  
20  
21  
22  
23  
24  
25  
26  
27  
28  
29  
30  
31  
32  
33  
34  
35  
36  
37  
38  
39  
40  
41  
42  
43  
44  
45  
46  
47  
48  
49  
50  
51  
52  
53  
54  
55  
56  
57  
58  
59  
60

Figure 1. The first two principal components

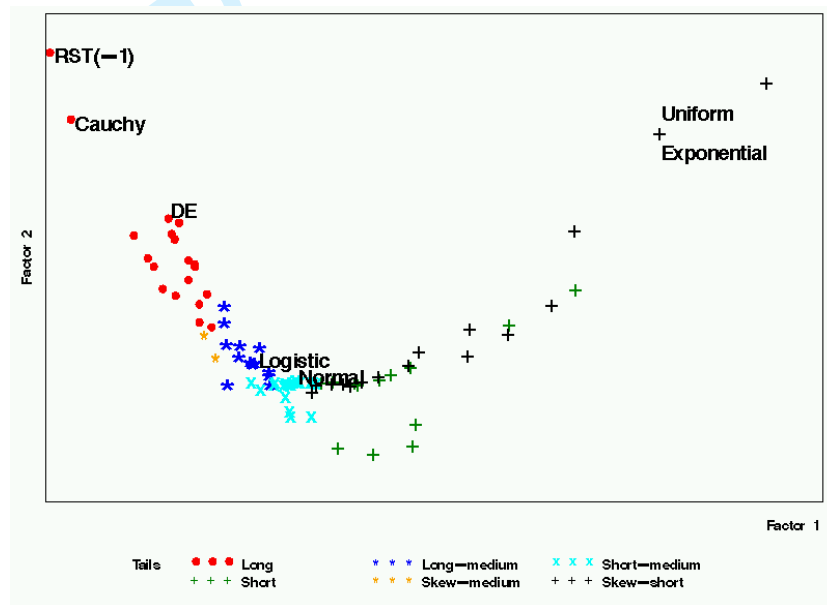
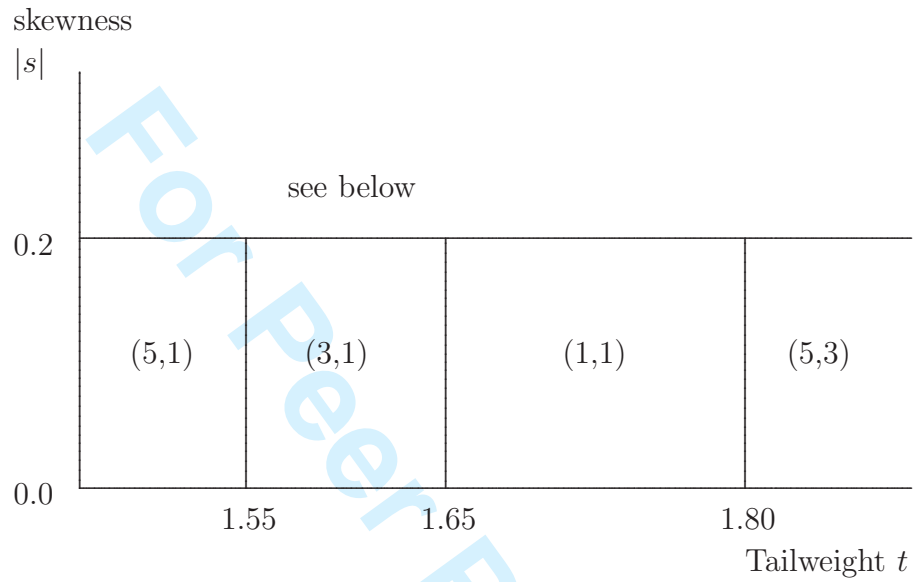
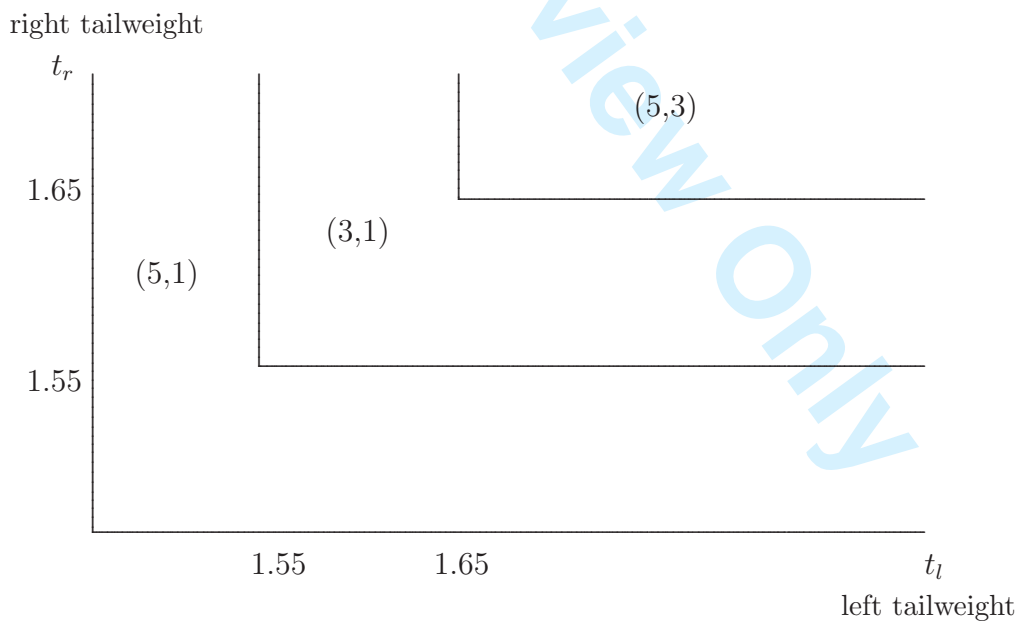


Figure 2. Adaptive scheme.

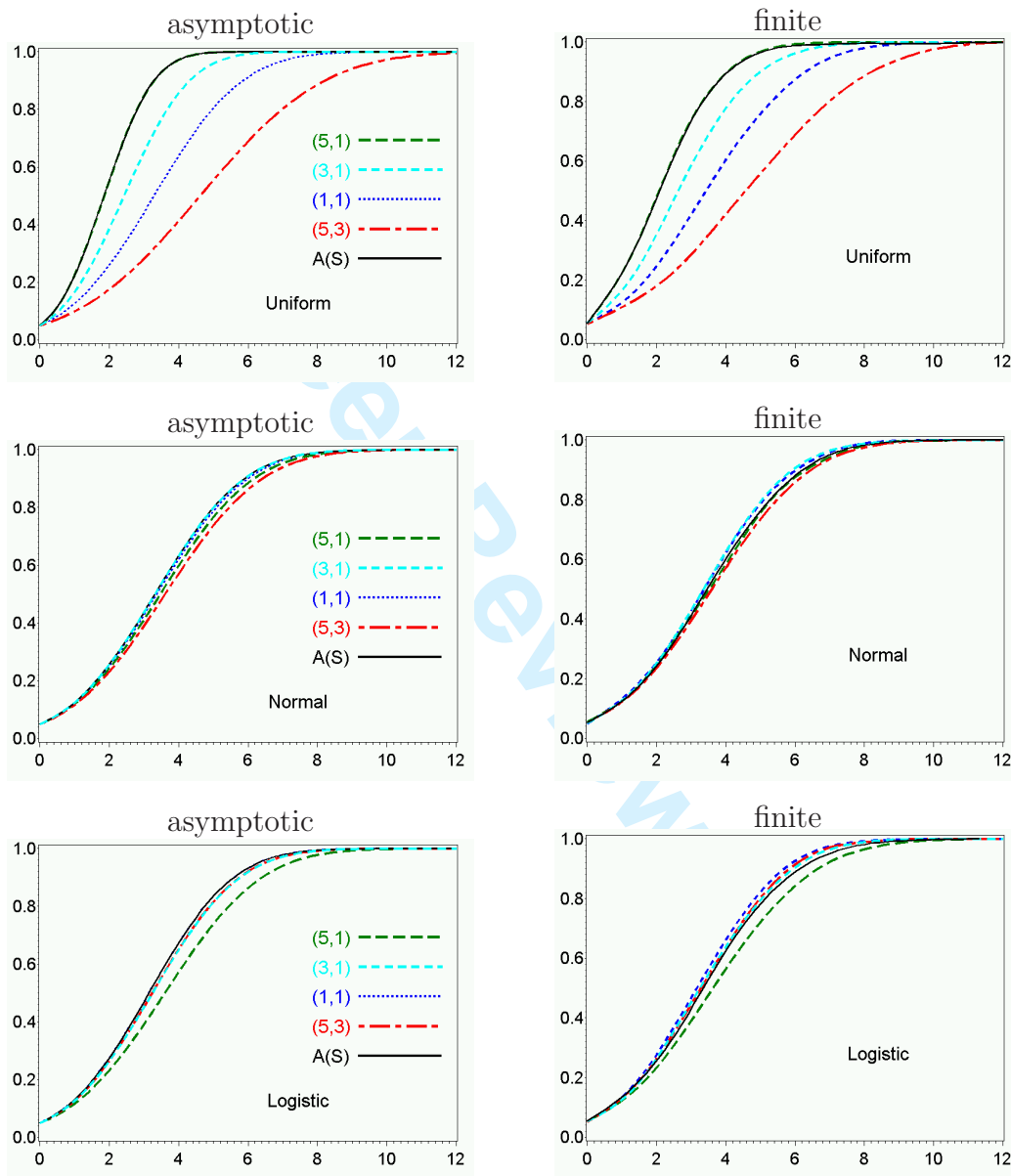
a) symmetric or moderately skew densities



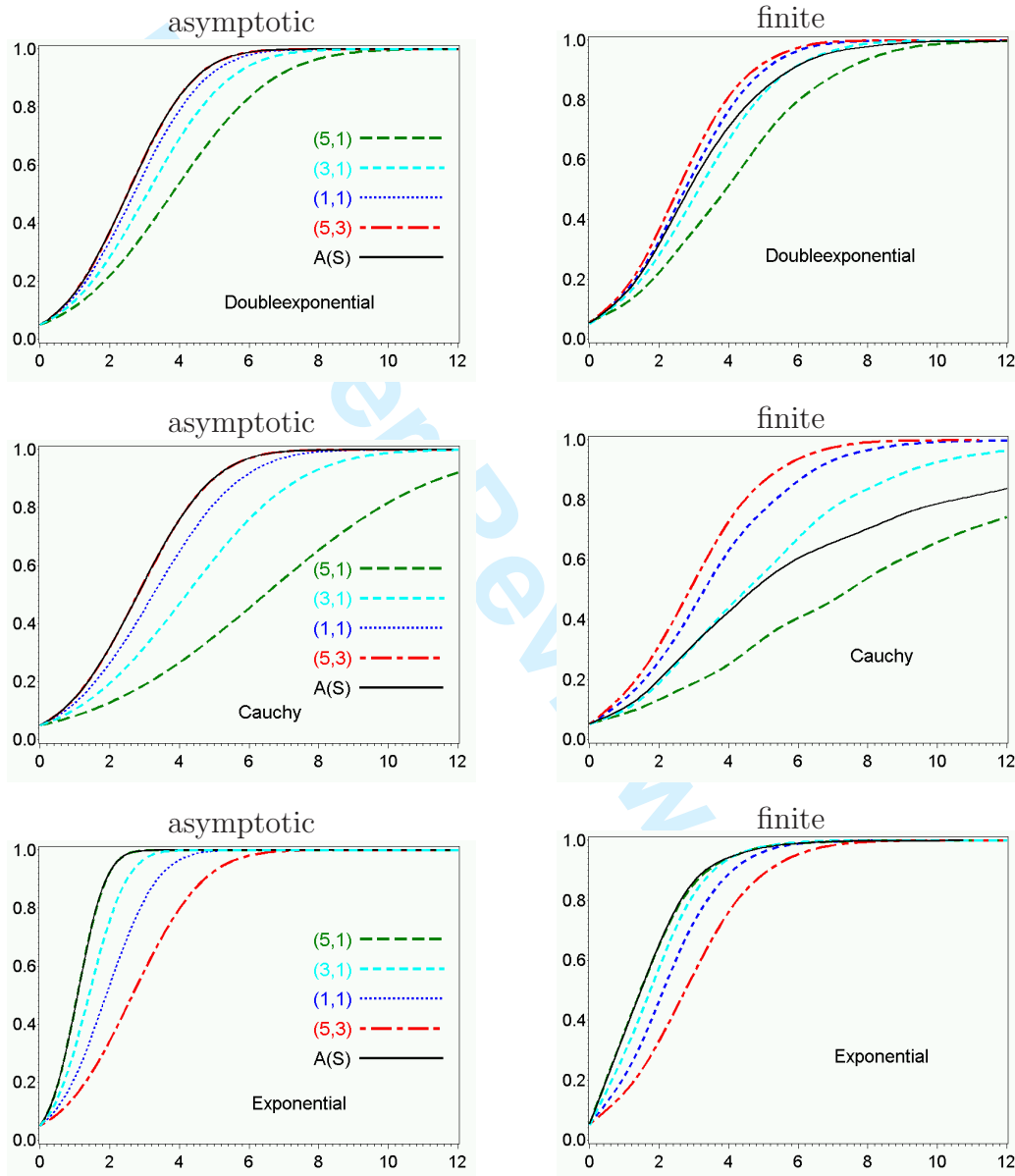
b) very skew densities ( $|\hat{s}| > 0.2$ )



**Figure 3.** The asymptotic and finite ( $n_1 = n_2 = 40$ ) power functions of the tests  $U_{1,1}, U_{3,1}, U_{5,1}, U_{5,3}$  and  $A(\hat{S})$ ; densities: uniform, normal and logistic.



**Figure 4.** The asymptotic and finite ( $n_1 = n_2 = 40$ ) power functions of the tests  $U_{1,1}, U_{3,1}, U_{5,1}, U_{5,3}$  and  $A(\hat{S})$ ; densities: doubleexponential, Cauchy and exponential (Continuation from Figure 3.)



1  
2  
3  
4  
5  
6  
7  
8  
9  
10  
11  
12  
13  
14  
15  
16  
17  
18  
19  
20  
21  
22  
23  
24  
25  
26  
27  
28  
29  
30  
31  
32  
33  
34  
35  
36  
37  
38  
39  
40  
41  
42  
43  
44  
45  
46  
47  
48  
49  
50  
51  
52  
53  
54  
55  
56  
57  
58  
59  
60

Figure 5. Running times for two age groups

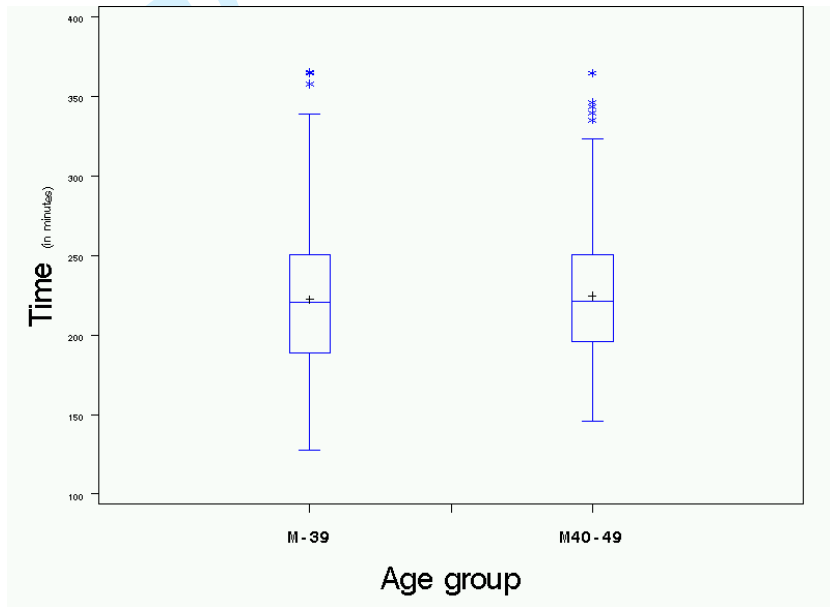




Table 1. Values of the factors  $C_{k,i}^2(f)$ ,  $k \leq 5$ , for various densities  $f$ , together with the corresponding factors for the  $t$ -test

Density	$t$ -test	$U_{1,1}$	$U_{2,1}$	$U_{3,1}$	$U_{3,2}$	$U_{4,1}$	$U_{4,2}$	$U_{5,1}$	$U_{5,2}$	$U_{5,3}$
Uniform	12.00	12.00	14.74	22.09	7.053	30.01	7.535	<b>38.00</b>	9.109	6.095
Normal	<i>1.000</i>	0.954	0.976	<b>0.985</b>	0.873	0.950	0.889	0.900	0.931	0.832
Logistic	0.304	<b>0.333</b>	0.332	0.313	0.324	0.284	0.327	0.254	0.332	0.315
Doubleex	0.500	0.750	0.705	0.576	0.833	0.462	0.816	0.379	0.749	<b>0.866</b>
Cauchy	0.0	0.304	0.268	0.182	0.382	0.118	0.368	0.077	0.318	<b>0.409</b>
$t_2$	0.0	0.520	0.489	0.398	0.572	0.311	0.566	0.244	0.538	<b>0.579</b>
$t_{10}$	0.800	<u>0.843</u>	<b>0.846</b>	0.816	0.802	0.755	0.813	0.688	0.835	0.772
L-DE 0.55	0.499	0.745	0.701	0.574	<b>0.826</b>	0.461	0.809	0.378	0.728	<u>0.822</u>
L-DE 0.61	0.495	0.726	0.686	0.567	0.798	0.458	0.783	0.377	0.728	<b>0.822</b>
L-DE 0.70	0.481	0.672	0.641	0.544	0.720	0.448	0.712	0.373	0.678	<b>0.730</b>
L-DE 0.75	0.468	0.630	0.606	0.524	0.663	0.438	0.659	0.368	0.637	<b>0.665</b>
L-DE 0.80	0.450	0.581	0.563	0.497	<b>0.599</b>	0.423	<b>0.599</b>	0.360	0.587	<u>0.594</u>
L-DE 0.90	0.395	<b>0.464</b>	0.457	0.422	0.460	0.373	0.463	0.328	0.466	0.450
L-DE 0.95	0.355	<b>0.400</b>	0.396	0.372	0.391	0.334	0.394	0.300	0.399	0.380
L-DE 0.97	0.336	<b>0.373</b>	0.371	0.349	0.363	0.315	0.367	0.282	0.372	0.353
L-DE 0.99	0.315	<b>0.347</b>	0.345	0.325	0.337	0.295	0.340	0.264	0.345	0.328
U-L 0.55	0.281	<u>0.298</u>	<b>0.299</b>	0.289	0.282	0.267	0.287	0.242	0.296	0.271
U-L 0.61	0.246	0.250	0.254	<b>0.255</b>	0.228	0.242	0.233	0.224	0.248	0.215
U-L 0.70	0.182	0.173	0.181	0.194	0.147	<b>0.195</b>	0.152	<u>0.188</u>	0.169	0.134
U-L 0.75	0.141	0.130	0.139	0.156	0.105	<b>0.163</b>	0.110	<u>0.161</u>	0.125	0.094
U-L 0.80	0.100	0.090	0.099	0.116	0.069	0.126	0.073	<b>0.130</b>	0.085	0.061
U-L 0.90	0.029	0.026	0.029	0.038	0.018	0.048	0.019	<b>0.054</b>	0.022	0.015
RST (-1)	0.0	0.042	0.037	0.024	0.055	0.015	0.052	0.009	0.044	<b>0.060</b>
RST (-0.5)	0.0	0.464	0.427	0.328	0.533	0.243	0.522	0.182	0.480	<b>0.550</b>
RST (-0.4)	0.193	0.891	0.832	0.665	0.994	0.511	0.980	0.396	0.919	<b>1.016</b>
RST (-0.3)	0.832	1.952	1.850	1.537	2.113	1.230	2.095	0.987	2.003	<b>2.132</b>
RST (-0.2)	3.340	5.426	5.220	4.518	<b>5.678</b>	3.765	5.664	3.135	5.524	<u>5.662</u>
RST (-0.1)	20.92	<u>26.87</u>	26.27	23.73	27.13	20.61	<b>27.22</b>	17.81	27.09	<u>26.71</u>
RST (0.05)	143.7	<u>148.6</u>	<b>149.2</b>	144.1	141.5	133.4	143.3	121.9	147.0	136.7
RST (0.14)	<i>24.18</i>	<u>23.08</u>	23.56	<b>23.73</b>	21.15	22.84	21.55	21.59	22.50	20.19
RST (0.2)	<i>14.04</i>	12.91	13.33	<b>13.81</b>	11.52	13.65	11.78	13.20	12.45	10.91
RST (0.4)	5.815	5.036	5.425	6.196	4.078	6.686	4.222	<b>6.978</b>	4.643	3.762
CN (.01,2)	<i>0.971</i>	0.941	0.960	<b>0.966</b>	0.863	0.928	0.879	0.876	0.918	0.823
CN (.02,2)	0.943	0.927	0.944	<b>0.947</b>	0.853	0.907	0.869	0.853	0.907	0.814
CN (.03,2)	<i>0.917</i>	0.887	<b>0.900</b>	<u>0.893</u>	0.823	0.846	0.837	0.788	0.871	0.787
CN (.05,2)	<i>0.870</i>	0.824	<b>0.831</b>	<u>0.811</u>	0.775	0.757	0.787	0.694	0.815	0.744
CN (.01,3)	0.926	0.934	0.952	<b>0.954</b>	0.859	0.913	0.875	0.859	0.913	0.820
CN (.02,3)	0.862	0.914	<b>0.928</b>	<u>0.924</u>	0.845	0.878	0.860	0.819	0.896	0.807
CN (.03,3)	0.806	0.854	<b>0.861</b>	<u>0.839</u>	0.804	0.780	0.817	0.713	0.846	0.771
CN (.05,3)	0.714	0.761	0.759	<u>0.716</u>	0.738	0.644	0.747	0.569	<b>0.766</b>	0.712
CN (.01,5)	0.806	0.928	<b>0.944</b>	<u>0.943</u>	0.855	0.899	0.871	0.841	0.909	0.817
CN (.02,5)	0.676	0.901	<b>0.913</b>	<u>0.902</u>	0.838	0.850	0.853	0.786	0.887	0.802
CN (.03,5)	0.581	<b>0.825</b>	<b>0.825</b>	<u>0.789</u>	0.788	0.718	0.799	0.642	<b>0.825</b>	0.757
CN (.05,5)	0.455	<u>0.709</u>	0.696	0.631	0.707	0.544	0.713	0.461	<b>0.726</b>	0.686

Continuation from the previous page

Density	<i>t</i> -test	$U_{1,1}$	$U_{2,1}$	$U_{3,1}$	$U_{3,2}$	$U_{4,1}$	$U_{4,2}$	$U_{5,1}$	$U_{5,2}$	$U_{5,3}$
Mielke (0.2)	0.452	0.620	0.591	0.500	0.667	0.413	0.658	0.347	0.620	<b>0.682</b>
Mielke (0.4)	0.408	0.521	0.503	0.439	0.545	0.373	0.541	0.319	0.521	<b>0.548</b>
Mielke (0.6)	0.369	<u>0.444</u>	0.433	0.389	<b>0.452</b>	0.339	<b>0.452</b>	0.295	0.444	0.449
Mielke (0.8)	0.335	<b>0.383</b>	0.377	0.348	0.380	0.309	0.382	0.273	0.382	0.374
Mielke (1.5)	0.241	0.245	<b>0.249</b>	<u>0.246</u>	0.227	0.233	0.231	0.215	0.242	0.215
Mielke (2.0)	0.194	0.188	0.193	<b>0.200</b>	0.167	0.195	0.171	0.185	0.184	0.156
Mielke (5.0)	0.069	0.061	0.067	0.079	0.047	0.087	0.049	<b>0.090</b>	0.057	0.042
Mielke (20)	0.007	0.006	0.007	0.010	0.004	0.013	0.004	<b>0.015</b>	0.005	0.003
BT (0.25)	0.633	<b>0.660</b>	0.658	0.624	0.640	0.572	0.645	0.520	0.650	0.626
BT (0.5)	0.382	<u>0.444</u>	0.433	0.389	<b>0.454</b>	0.340	0.452	0.298	0.441	<b>0.454</b>
BT (0.75)	0.222	0.292	0.279	0.239	0.311	0.200	0.308	0.170	0.291	<b>0.318</b>
BT (1.25)	0.068	0.118	0.116	0.098	0.131	0.081	0.132	0.068	0.123	<b>0.136</b>
BT (1.5)	0.036	0.073	0.077	0.070	0.080	0.061	0.083	0.055	0.080	<b>0.081</b>
Exponential	1.000	3.000	3.684	5.522	1.763	7.502	1.884	<b>9.500</b>	2.277	1.524
Gamma (1.5)	0.667	1.216	1.365	1.705	0.895	1.994	0.936	<b>2.230</b>	1.062	0.808
Gamma (2.0)	0.500	0.750	0.815	0.951	0.595	1.047	0.617	<b>1.112</b>	0.683	0.546
Gamma (2.5)	0.400	0.540	0.578	0.651	0.445	0.694	0.459	<b>0.717</b>	0.502	0.412
Gamma (3.0)	0.333	0.422	0.447	0.493	0.355	0.515	0.365	<b>0.523</b>	0.396	0.330
Gamma (4.0)	0.250	0.293	0.307	0.330	0.252	<b>0.337</b>	0.259	<u>0.336</u>	0.278	0.237
Gamma (5.0)	0.200	0.224	0.234	<u>0.248</u>	0.196	<b>0.250</b>	0.201	0.246	0.214	0.184
Gamma (10)	0.100	0.103	0.106	<u>0.110</u>	0.092	<b>0.108</b>	0.094	0.105	0.100	0.087
Weibull (1.1)	1.296	2.919	3.448	4.793	1.898	6.135	2.008	<b>7.407</b>	2.360	1.673
Weibull (1.5)	2.662	3.391	3.706	4.375	2.662	4.877	2.762	<b>5.249</b>	3.058	2.446
Weibull (2.0)	4.660	4.712	4.980	5.447	3.996	<b>5.665</b>	4.110	<u>5.470</u>	4.428	3.738
Weibull (2.5)	6.937	6.516	6.795	7.203	5.693	<b>7.276</b>	5.835	<u>7.182</u>	6.216	5.364
Weibull (3.0)	<i>9.494</i>	8.732	9.050	<b>9.450</b>	7.736	9.413	7.916	<u>9.174</u>	8.387	7.315
Weibull (4.0)	<i>15.47</i>	14.33	14.78	<b>15.25</b>	12.84	15.01	13.12	<u>14.48</u>	13.84	12.18
Weibull (5.0)	22.61	21.46	22.09	<b>22.70</b>	19.30	22.26	19.71	<u>21.39</u>	20.76	18.32
LogNor (0.05)	<i>398.5</i>	382.5	391.0	<b>395.0</b>	349.3	281.3	356.0	361.2	372.6	332.9
LogNor (0.1)	98.51	95.97	98.21	<b>99.44</b>	87.47	96.18	89.18	<u>91.29</u>	93.41	83.31
LogNor (0.3)	9.705	11.10	11.47	<b>11.89</b>	9.893	11.75	10.12	<u>11.37</u>	10.70	9.363
LogNor (0.5)	2.742	4.328	4.559	4.946	3.690	5.094	3.797	<b>5.104</b>	4.095	3.448
LogNor (0.9)	0.356	1.768	1.984	2.478	1.290	2.874	1.356	<b>3.173</b>	1.561	1.153
LogNor (1.0)	0.214	1.574	1.804	2.350	1.090	2.819	1.154	<b>3.195</b>	1.356	0.959
LogNor (1.1)	0.127	1.445	1.692	2.304	0.944	2.860	1.007	<b>3.328</b>	1.211	0.817
LogNor (1.5)	0.012	1.307	1.690	2.765	0.640	3.924	0.708	<b>5.045</b>	0.947	0.509
RST (.2,.4)	8.622	7.599	8.040	8.812	6.413	9.161	6.608	<b>9.256</b>	7.156	5.976
RST (.2,.49)	7.610	6.638	7.107	8.005	5.449	8.528	5.634	<b>8.800</b>	6.170	5.040
RST (.4,.49)	5.241	4.532	4.933	5.767	3.582	6.353	3.720	<b>6.750</b>	4.127	3.284
RST (-.2,-.4)	0.387	2.061	1.991	1.743	<b>2.135</b>	1.465	2.134	1.227	2.099	<u>2.116</u>
RST (-.2,-.49)	0.021	1.473	1.435	1.283	<b>1.496</b>	1.010	1.502	0.935	1.495	<u>1.471</u>
RST (-.4,-.49)	0.020	0.659	0.613	0.485	0.742	0.369	0.730	0.282	0.681	<b>0.760</b>

**Table 2.** Values of the tailweight  $t$  for various symmetric densities

Density	$t$	Density	$t$	Density	$t$	Density	$t$
Uniform	1.286	Normal	1.587	Logistic	1.697	Doubleexp	1.912
Cauchy	3.217	$t_2$	2.107	$t_{10}$	1.672		
L-DE 0.55	1.911	L-DE 0.61	1.905	L-DE 0.70	1.884	L-DE 0.75	1.864
L-DE 0.80	1.836	L-DE 0.90	1.753	L-DE 0.95	1.712	L-DE 0.97	1.703
L-DE 0.99	1.698						
U-L 0.55	1.668	U-L 0.61	1.623	U-L 0.70	1.534	U-L 0.75	1.474
U-L 0.80	1.409	U-L 0.90	1.300				
RST (-1)	3.451	RST (-0.5)	2.302	RST (-0.4)	2.146	RST (-0.3)	2.010
RST (-0.2)	1.891	RST (-0.1)	1.788	RST (0.05)	1.657	RST (0.14)	1.591
RST (0.2)	1.552	RST (0.4)	1.446				
CN (.01,2)	1.592	CN (.02,2)	1.597	CN (.03,2)	1.601	CN (.05,2)	1.611
CN (.01,3)	1.596	CN (.02,3)	1.605	CN (.03,3)	1.615	CN (.05,3)	1.635
CN (.01,5)	1.600	CN (.02,5)	1.614	CN (.03,5)	1.629	CN (.05,5)	1.665
Mielke 0.2	1.860	Mielke 0.4	1.812	Mielke 0.6	1.770	Mielke 0.8	1.732
Mielke 1.5	1.626	Mielke 2.0	1.571	Mielke 5.0	1.404	Mielke 20.0	1.292
BT (0.25)	1.670	BT(0.5)	1.752	BT(0.75)	1.833	BT(1.25)	1.992
BT (1.5)	2.071						

**Table 3.** Values of skewness  $s$ , tailweight  $t$ , left tailweight  $t_l$  and right tailweight  $t_r$  for various skew densities

Density	Skewness	Tailweight	Left tailweight	Right tailweight
Exponential	0.564	1.697	1.210	1.912
Gamma (1.5)	0.460	1.651	1.284	1.846
Gamma (2.0)	0.401	1.668	1.329	1.810
Gamma (2.5)	0.354	1.621	1.360	1.785
Gamma (3.0)	0.338	1.653	1.382	1.768
Gamma (4.0)	0.298	1.643	1.412	1.743
Gamma (5.0)	0.271	1.636	1.432	1.727
Gamma (10.0)	0.202	1.619	1.480	1.686
Weibull (1.1)	0.508	1.654	1.237	1.859
Weibull (1.5)	0.335	1.571	1.329	1.728
Weibull (2.0)	0.194	1.544	1.411	1.649
Weibull (2.5)	0.103	1.541	1.470	1.605
Weibull (3.0)	0.040	1.546	1.513	1.577
Weibull (4.0)	-0.04	1.559	1.573	1.544
Weibull (5.0)	-0.088	1.571	1.612	1.525
LogNormal (0.05)	0.041	1.588	1.563	1.611
LogNormal (0.1)	0.082	1.591	1.541	1.634
LogNormal (0.3)	0.242	1.626	1.457	1.749
LogNormal (0.5)	0.390	1.695	1.386	1.879
LogNormal (0.9)	0.629	1.940	1.274	2.202
LogNormal (1.0)	0.676	2.024	1.251	2.298
LogNormal (1.1)	0.719	2.117	1.229	2.401
LogNormal (1.5)	0.844	2.588	1.160	2.890
RST (0.2,0.4)	0.029	1.486	1.486	1.486
RST (0.2,0.49)	0.014	1.459	1.469	1.449
RST (0.4,0.49)	-0.011	1.426	1.436	1.417
RST (-0.2,-0.4)	0.375	2.073	1.718	2.289
RST (-0.2,-0.49)	0.490	2.195	1.657	2.470
RST (-0.4,-0.49)	0.153	2.226	2.086	2.342

**Table 4.** The factors  $C_{k,i}^2(f)$ , for the chosen  $U_{k,i}$ -tests, and the corresponding factors for the linear rank tests, for various densities  $f$ , only the entries that are on the first and second place, respectively

density	$A(S)$				$B(S)$			
	$U_{1,1}$	$U_{3,1}$	$U_{5,1}$	$U_{5,3}$	GA	WI	LT	HFR
Uniform		22.0	<b>38.0</b>		<b>24.0</b>	12.0		
Normal	0.954	<b>0.985</b>			0.884	<b>0.954</b>		
Logistic	<b>0.333</b>			0.315		<b>0.333</b>	0.315	
Doublexpon	0.750			<b>0.866</b>		0.750	<b>0.845</b>	
Cauchy	0.304			<b>0.409</b>		0.304	<b>0.407</b>	
Exponential		5.52	<b>9.50</b>		<b>6.00</b>			5.40
U-L (0.75)	0.130		<b>0.161</b>		<b>0.166</b>	0.130		
L-D (0.75)	0.630			<b>0.665</b>		0.630	<b>0.666</b>	

Table 5. Rankings of the  $t$ -test ( $t$ ), the Wilcoxon test ( $W$ ), and the adaptive test ( $A$ )

Density	$t$	$W$	$A$	Density	$t$	$W$	$A$	Density	$t$	$W$	$A$
Uniform	2.5	2.5	1	RST (0.05)	3	1.5	1.5	Exponential	3	2	1
Normal	1	3	2	RST (0.14)	1	3	2	Gamma (1.5)	3	2	1
Logistic	3	1.5	1.5	RST (0.2)	1	3	2	Gamma (2.0)	3	2	1
Doubleex	3	2	1	RST (0.4)	2	3	1	Gamma (2.5)	3	2	1
Cauchy	3	2	1	CN (.01,2)	1	3	2	Gamma (3.0)	3	2	1
$t_2$	3	2	1	CN (.02,2)	2	3	1	Gamma (4.0)	3	2	1
$t_{10}$	3	1.5	1.5	CN (.03,2)	1	3	2	Gamma (5.0)	3	2	1
L-DE 0.55	3	2	1	CN (.05,2)	1	3	2	Gamma (10)	3	2	1
L-DE 0.61	3	2	1	CN (.01,3)	3	2	1	Weibull (1.1)	3	2	1
L-DE 0.70	3	2	1	CN (.02,3)	3	2	1	Weibull (1.5)	3	2	1
L-DE 0.75	3	2	1	CN (.03,3)	3	1	2	Weibull (2.0)	3	2	1
L-DE 0.80	3	2	1	CN (.05,3)	3	1	2	Weibull (2.5)	2	3	1
L-DE 0.90	3	1.5	1.5	CN (.01,5)	3	2	1	Weibull (3.0)	1	3	2
L-DE 0.95	3	1.5	1.5	CN (.02,5)	3	2	1	Weibull (4.0)	1	3	2
L-DE 0.97	3	1.5	1.5	CN (.03,5)	3	1	2	Weibull (5.0)	2	3	1
L-DE 0.99	3	1.5	1.5	CN (.05,5)	3	1.5	1.5	LogNor (0.05)	1	3	2
U-L 0.55	3	1.5	1.5	Mielke (0.2)	3	2	1	LogNor (0.1)	2	3	1
U-L 0.61	3	1.5	1.5	Mielke (0.4)	3	2	1	LogNor (0.3)	3	2	1
U-L 0.70	3	2	1	Mielke (0.6)	3	1.5	1.5	LogNor (0.5)	3	2	1
U-L 0.75	3	2	1	Mielke (0.8)	3	1.5	1.5	LogNor (0.9)	3	2	1
U-L 0.80	3	2	1	Mielke (1.5)	3	2	1	LogNor (1.0)	3	2	1
U-L 0.90	3	2	1	Mielke (2.0)	2	3	1	LogNor (1.1)	3	2	1
RST (-1)	3	2	1	Mielke (5.0)	2	3	1	LogNor (1.5)	3	2	1
RST (-0.5)	3	2	1	Mielke (20)	3	2	1	RST (.2,.4)	2	3	1
RST (-0.4)	3	2	1	BT (0.25)	3	1.5	1.5	RST (.2,.49)	2	3	1
RST (-0.3)	3	2	1	BT (0.5)	3	1.5	1.5	RST (.4,.49)	2	3	1
RST (-0.2)	3	2	1	BT (0.75)	3	2	1	RST (-.2,-.4)	3	2	1
RST (-0.1)	3	1.5	1.5	BT (1.25)	3	2	1	RST (-.2,-.49)	3	2	1
				BT (1.5)	3	2	1	RST (-.4,-.49)	3	2	1
Average	2.67	2.10	1.23								

## Rejoinder to the second report of referee B

Thank you very much for your report on our revised paper. All your comments are considered.

In a new section (Section 6) we compare the adaptive test with the Wilcoxon test and with the  $t$ -test. We think, the improvement of the asymptotic power is convincing.

You are right, the restriction  $k \leq 5$  looks slightly restrictive, and the computer time argument is not satisfactory. However, we have to make some restriction. We think, the arguments presented in the first paragraph of page 5 are sufficient.

We have changed “generalized” and “generalization” to “generalised” and “generalisation”. We are sorry, we cannot see any other American English words. The text is thought to be written in British English.

In Table 1 we added the asymptotic efficacies of the  $t$ -test. We prefer to present the asymptotic efficacies instead of some ARE for the following reasons:

First, the paper is on U-statistics, and if any test is chosen to be the reference test for computing AREs, then it should be the Wilcoxon test. Second, there are some densities for which the AE of the  $t$ -test is zero, all ARE-values in the corresponding row become  $\infty$ . Third, the AE allow easier comparisons. The reader may take his/her favourite test and compare it to the others.

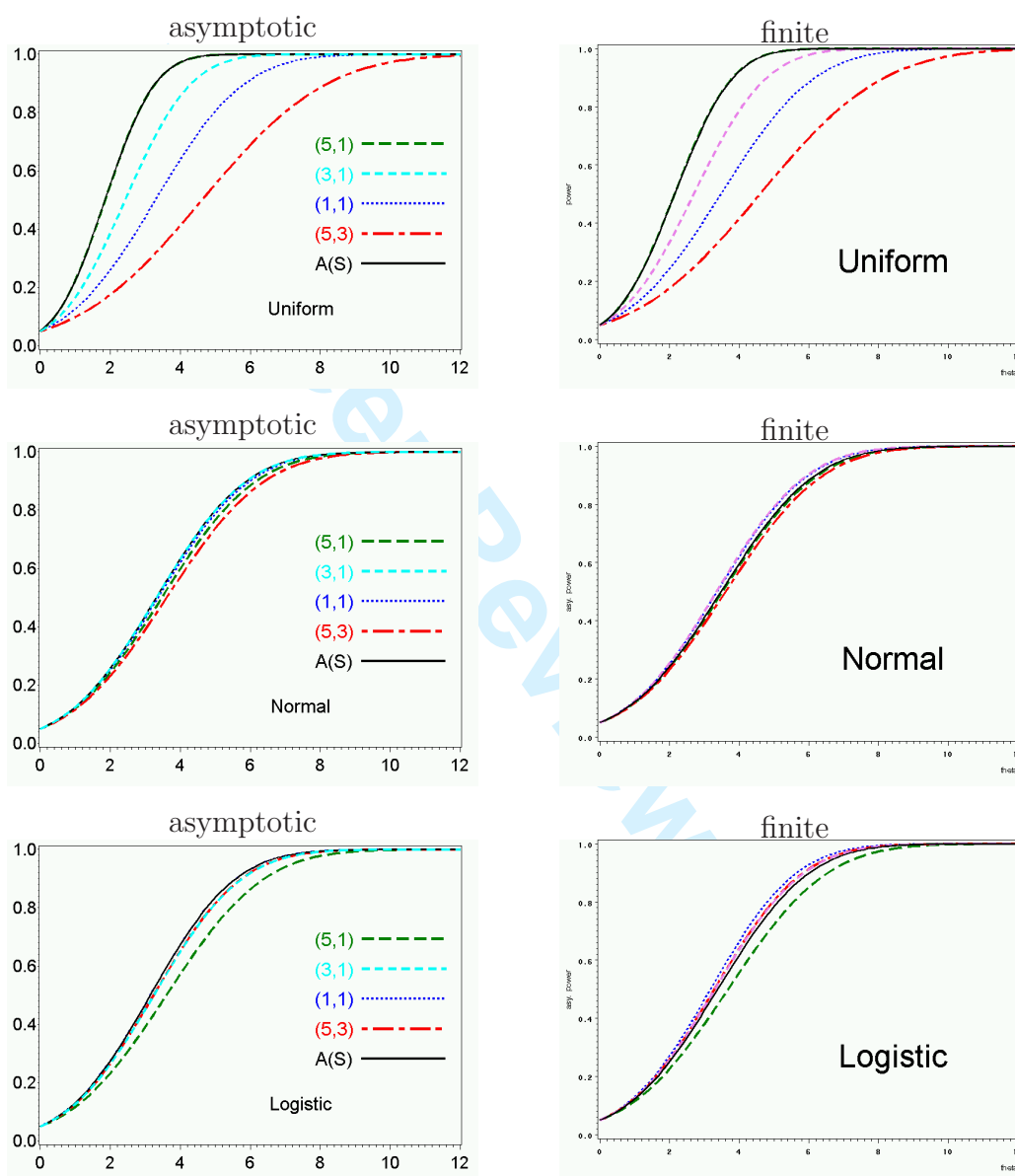
We decided to add no further curves in Figures 3 and 4. One reason is that there are already five lines, and a sixth one will reduce the readability of the curves. Let us illustrate this fact: For the uniform the two lines for the  $t$ -test and  $U_{1,1}$  coincide, for the normal the line for the  $t$ -test lies very slightly above the other lines, but this will not be visible. For the logistic the line for the  $t$ -tests is almost the same as for the tests  $U_{3,1}$  and  $U_{5,3}$ , and for the Cauchy no line exists. An impression of the asymptotic power of the  $t$ -test can be obtained from the corresponding entries in Table 1. Another point is that the  $t$ -test plays only a marginal role in our paper.

We used the SAS package to draw the figures. Concerning the blue dotted and dashed lines this program seems to produce some problems. We are currently working on it. However, we are not able to do it in the narrow time window of two weeks that we have for revision.

In Figure 1 we labelled the dot in the upper left corner. Labels on the axes are removed, since an detailed interpretation of the axes is difficult.

The other mistakes are corrected.

**Figure 3.** The asymptotic and finite ( $n_1 = n_2 = 40$ ) power functions of the tests  $U_{1,1}, U_{3,1}, U_{5,1}, U_{5,3}$  and  $A(\hat{S})$ ; densities: uniform, normal and logistic.





**Figure 4.** The asymptotic and finite ( $n_1 = n_2 = 40$ ) power functions of the tests  $U_{1,1}, U_{3,1}, U_{5,1}, U_{5,3}$  and  $A(\hat{S})$ ; densities: doubleexponential, Cauchy and exponential (Continuation from Figure 3.)

