Max-Type Rank Tests, U-Tests, and Adaptive Tests for the Two-Sample Location Problem - An Asymptotic Power Study

Wolfgang Kössler
Institut für Informatik
Humboldt-Universität zu Berlin

Abstract

For the two-sample location problem we first consider two types of tests, linear rank tests with various scores, but also some tests based on U-statistics. For both types we construct adaptive tests as well as max-type tests and investigate their asymptotic and finite power properties. It turns out that both the adaptive tests have larger asymptotic power than the max-type tests. For small sample sizes, however, some of the max-type tests are to prefer. U-statistics are convenient if extreme densities may occur.

Keywords: Adaptive tests, U-statistics, Max-type tests, Linear rank tests, Asymptotic power.
Mathematics Subject classification: G2G10, G2G20

1 Introduction

Let $X_1, \ldots, X_{n_1}$ and $Y_1, \ldots, Y_{n_2}$ be independent random samples from a population with absolutely continuous cumulative distribution functions (cdf.) $F(x)$ and $F(x - \vartheta)$, $\vartheta \in \mathbf{R}$, respectively. In the following we assume that F is twice continuously differentiable on $(-\infty, \infty)$ except for a set of Lebesgue measure zero; f' denotes the derivative of the density f where it exists and it is defined to be zero, otherwise. We wish to test:

$H_0: \quad \vartheta = 0 \quad$ against $\quad H_1: \quad \vartheta > 0.$

The most familiar nonparametric test is the Wilcoxon-Mann-Whitney test. This test was generalized to linear rank tests with various other scores, such as the Median test, the normal scores test or the Savage test, see e.g. Hájek, $\check{S}$idák, and Sen (1999). However, the scores are designed for special types of underlying densities, for example the normal scores test is good for normal-like densities, the Median test for the doubleexponential, and the Savage test for the Gumbel, or more generally, for left-skew densities. That is why restrictive adaptive tests are constructed where in a first step the density is classified, and in the second step an appropriate linear rank test is performed, see e.g. Büning (1991). Such tests are proven to be powerful. The asymptotic power function, under the sequence $\{\vartheta_N\}$ of local alternatives, $\vartheta_N = \vartheta/\sqrt{N}$, $N = n_1 + n_2$, is easy to compute.

Another idea is to put the various scores statistics together, and to take the maximum of them. This way we obtain the so-called max-type tests, see e.g. Neuhäuser, Büning and Hothorn (2004). The computation of the asymptotic power function is more difficult. In the present paper we try to approximate it by utilizing the asymtotic correlation structure of the components.

Other possible generalizations of the Wilcoxon-Mann-Whitney statistic are U-statistics. We consider a special subclass of U-statistics, choose some convenient representants of it, and construct an adaptive test as well as various variants of max-type tests. Again, by utilizing the asymptotic correlation structure we are able to approximate the asymptotic power functions.

We show that the adaptive tests are asymptotically the best in most cases. But also a simple max-type test based on linear rank statistics may be useful, especially for smaller sample sizes. U-statistics are convenient for more extreme densities such as the Cauchy or the exponential.

Linear rank tests are considered in Section 2, tests based on U-statistics in Section 3, and adaptive tests and max-type tests based on them in Sections 4 and 5, respectively. In Section 6 some asymptotic and finite power functions are presented and compared. Conclusions are given in Section 7.

All statistics are constructed in such a way that large values are significant, i.e. the tests reject H_0 in favour of H_1 if the corresponding statistic is as least as large as the upper α-quantile of its (asymptotic) null distribution.

2 Linear rank tests

In this section we recall well-known results for linear rank tests for the two-sample location problem.

Assumption 1 *The scores $a_N(i)$ are assumed to satisfy*

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

with square integrable score functions

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

Define

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

where $I(g)$ is the Fisher-information of the density function g defined by (1). ϕ' represents the derivative of ϕ almost everywhere. It is assumed that $\int_0^1 \phi(u,g)\,du = 0$ and $0 < I(g) < \infty$.

We use the notation

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

Let R_{2j} be the rank of Y_j in the combined X- and Y-sample and

$$T = \sum_{j=1}^{n_2} a_N(R_{2j}) \tag{2}$$

with $N = n_1 + n_2$, be linear rank statistics for the location problem.

Proposition 1 (Hájek, $\check{S}$idák, and Sen, 1999, Ch.6) *Under H_0 the limiting distribution of T/σ is standardnormal, where*

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

Assumption 2 *Let be $\theta > 0$ and $\{\theta_N\}$ a sequence of "near" alternatives with $\theta_N = N^{-1/2} \cdot \theta$. Let be $\min(n_1, n_2) \to \infty$, $n_1/N \to \lambda$, $0 < \lambda > 1$.*

Proposition 2 (Hájek, $\check{S}$idák, and Sen, 1999, Ch.7) *Under assumptions 1 and 2 the linear rank statistic T is asymptotically normally distributed with expectation μ and variance σ^2, where*

$$\mu = \theta \frac{n_1 n_2}{N} d(f,g). \tag{3}$$

Corollary 1 *The asymptotic efficacy (AE) of the two-sample linear rank test based on T is given by*

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

Let T_{g_1} and T_{g_2} be are two linear rank statistics based on the score generating functions g_1 and g_2. Then the asymptotic relative efficiency (ARE) is given by

$$ARE(T_{g_1}, T_{g_2}|f) = \frac{AE(T_{g_1}|f)}{AE(T_{g_2}|f)}.$$

For $f = g_2$ we denote

$$ARE(g_1, f) := ARE(T_{g_1}, T_f|f).$$

Next we give some examples of scores. They are used in the scores-based max-type tests as well in the Adaptive test $A(\hat{S})$ (see Section 4).

$$a_{N,GA}(k) = \begin{cases} \frac{4k}{N+1} - 1 & \text{if} \quad k \le \frac{N+1}{4} \\ 0 & \text{if} \quad \frac{N+1}{4} < k < \frac{3(N+1)}{4} \\ \frac{4k}{N+1} - 3 & \text{if} \quad k \ge \frac{3(N+1)}{4} \end{cases} \qquad \text{(Gastwirth scores, GA)}$$

$$a_{N,WI}(k) = \quad k - \frac{N+1}{2} \qquad \text{(Wilcoxon scores, WI)}$$

$$a_{N,LT}(k) = \begin{cases} -1 & \text{if} \quad k < \lfloor \frac{N}{4} \rfloor + 1 \\ \frac{4k}{N+1} - 2 & \text{if} \quad \lfloor \frac{N}{4} \rfloor + 1 \le k \le \lfloor \frac{3(N+1)}{4} \rfloor \\ 1 & \text{if} \quad k > \lfloor \frac{3(N+1)}{4} \rfloor \end{cases} \qquad \text{(Long tail scores, LT)}$$

$$a_{N,HFR}(k) = \begin{cases} \frac{k}{N+1} - \frac{3}{8} & \text{if} \quad k \le \frac{N+1}{2} \\ \frac{1}{8} & \text{if} \quad k > \frac{N+1}{2} \end{cases} \qquad \text{(Hogg-Fisher-Randles scores, HFR)}$$

$$a_{N,SA}(k) = \quad \sum_{j=N-k+1}^{N} \frac{1}{j} - 1 \qquad \text{(Savage scores, SA)}$$

The HFR scores are originally introduced for right-skew densities. For left-skew densities we may use scores $a_{N,HFL}(k) = -a_{N,HFR}(N - k + 1)$ and we call the corresponding test antisymmetric HFR-test, and it is abbreviated by HFL. (The last L stays for *left*, the R in HFR may stay for *right*). Otherwise, the SA scores are originally introduced for left-skew densities. For right-skew densities we may use the scores $a_{N,SAR}(k) = -a_{N,SA}(N - k + 1)$.

The corresponding test statistics are denoted by GA, WI, LT, HFR, SA, HFL and SAR, respectively.

The scores GA, LT and HFR are introduced by Gastwirth (1965), Policello and Hettmansperger (1976), and Hogg, Fisher and Randles (1975), respectively.

Various values for the factors $C(f, g)$ can be found e.g. in Kössler (2006a).

3 U-statistics

Location tests based on U-statistics have good asymptotic power properties (cf. Kössler, 2006a, or Kössler and Kumar, 2008). This fact gives rise to the idea to include some of them in our study.

Let s be an integer with $s < \min(n_1, n_2)$. We consider the following class of U-statistics:

$$U = \frac{n_1 n_2}{N} \frac{1}{\binom{n_1}{s}\binom{n_2}{s}} \sum \phi(X_{\alpha_1}, \ldots, X_{\alpha_s}, Y_{\beta_1}, \ldots, Y_{\beta_s}) - \frac{n_1 n_2}{2N}, \qquad (4)$$

where the kernel function is given by

$$\phi(X_{\alpha_1}, \ldots, X_{\alpha_s}, Y_{\beta_1}, \ldots, Y_{\beta_s}) = \begin{cases} 1 & \text{if} \quad \psi(X_{\alpha_1}, \ldots, X_{\alpha_s}) < \psi(Y_{\beta_1}, \ldots, Y_{\beta_s}) \\ 0 & \text{else.} \end{cases}$$

The sum in eq. (4) is taken over all possible subsamples of size s of the X-sample and of the Y-sample, respectively. The factor $\frac{n_1 n_2}{N}$ is introduced to have variances of the same order as that of the linear rank statistics. We use the special ψ-functions

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

where med(.) denotes the Median function. The corresponding ϕ-functions are denoted by $\phi^{k:s}$, and the U-statistics by $U_{k:s}$ with $(k:s)$=(1:1), (1:3), (2:3), (3:3), (1:5), (3:5), (5:5), respectively. Note that $U_{1:1}$ is the Mann-Whitney-Wilcoxon statistic. Let be

$$\begin{aligned} \phi_{1,0}^{k:s}(x) &= \mathbf{E}\phi^{k:s}(x, X_2, \ldots, X_s, Y_1, \ldots, Y_s) \\ \phi_{0,1}^{k:s}(y) &= \mathbf{E}\phi^{k:s}(X_1, \ldots, X_s, y, Y_2, \ldots, Y_s) \\ \zeta_{1,0}^{k:s} &= \text{var}\phi_{1,0}^{k:s}(X) \\ \zeta_{0,1}^{k:s} &= \text{var}\phi_{0,1}^{k:s}(Y). \end{aligned}$$

Then we have

Proposition 3 (Xie and Priebe, 2000) *Under H_0 the U-statistics $U_{k:s}$ are asymptotcally normally distributed with expectation zero and variance*

$$\mathit{var}(U_{k:s}) = \frac{n_1 n_2}{N} \sigma^2_{U_{k:s}} \qquad (5)$$

with

$$\sigma^2_{U_{k:s}} = s^2\zeta^{k:s}_{1,0} = s^2\zeta^{k:s}_{0,1}.$$

The asymptotic variances may be obtained explicitely by using the formula of Xie and Priebe (cf. their Theorem 3, see also Kössler, 2006a, eq. (2.26)). Since we later on also need the asymptotic covariances we prefer to compute first the various $\phi^{k:s}_{1,0} = \phi^{k:s}_{0,1}$, $\zeta^{k:s}_{1,0} = \zeta^{k:s}_{0,1}$, and then the asymptotic variances and covariances.

Let

$$F_{k:s}(x) = \sum_{i=k}^{s} \binom{s}{i} F^i(x)(1-F(x))^{s-i}$$

be the cdf. of an kth order statistics of a sample of size s. Define $F_{0:s} = 1$ and $F_{k:s} = 0$ if $k > s$. Then (Xie and Priebe, 2000, p. 679)

$$\phi^{k:s}_{1,0}(X_1) = \int_{-\infty}^{\infty} F_{k-1:s-1}\, dF_{k:s} + \int_{-\infty}^{X_1} (F_{k:s-1} - F_{k-1:s-1})\, dF_{k:s}.$$

The asymptotic covariances of the U-statistics are given by (cf. Puri and Sen, 1971, p. 58)

$$\text{cov}(U_{k:s}, U_{k':s'}) = \frac{n_1 n_2}{N} s s' \tau_{k:s,k':s'},$$

where

$$\tau_{k:s,k':s'} = \mathbf{E}\big((\phi^{k:s}_{1,0}(X) - \mathbf{E}\phi^{k:s}_{1,0}(X))(\phi^{k':s'}_{1,0}(X) - \mathbf{E}\phi^{k':s'}_{1,0}(X))\big).$$

Given k, s, k', s', the asymptotic variances, covariances and correlations (under H_0) can be computed analytically. Some numerical values of the correlations are given in Table 5 (see Subsection 5.3).

Proposition 4 (Xie and Priebe, 2000, p.666, see also Kössler, 2006a, p.37) *Under assumption 2 the U-statistic $U_{k:s}$ is asymptotically normally distributed with expectation $\mu_{U_{k:s}}$ and variance given by (5), where*

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

and

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

Table 1: The factors $C_{U_{k:s}}$ for the various tests $U_{k:s}$ and different densities f.

$(k:s)$	Density								
	Lo	N	nGu	Gu	DE	Cau	CN	Uni	Exp
(1:1)	.577	.977	.866	.866	.866	.551	.573	3.464	1.732
(1:3)	.474	.841	.553	.962	.643	.361	.624	3.980	3.317
(2:3)	.569	.934	.819	.819	.913	.618	.549	2.656	1.328
(3:3)	.474	.841	.962	.553	.643	.361	.363	3.980	0.663
(1:5)	.396	.745	.436	.934	.483	.218	.577	4.843	4.359
(3:5)	.561	.912	.797	.797	.931	.640	.538	2.469	1.234
(5:5)	.396	.745	.934	.436	.483	.218	.303	4.843	0.484

Corollary 2 *The asymptotic efficacy of the test based on $U_{k:s}$ is given by*

$$AE(U_{k:s}|f) = \lambda(1-\lambda)C^2_{U_{k:s}}(f).$$

where

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

Some numerical values for $C_{U_{k:s}}(f)$ are presented in Table 1 for the following densities f: logistic (Lo), normal (N), Gumbel ($f_G(x) = e^{-x}e^{-e^{-x}}$, Gu), "negative" Gumbel ($f_{nG}(x) = f_G(-x)$, nGu), Doubleexponential (DE), Cauchy (Cau), a contaminated normal $(F = \frac{1}{2}\mathcal{N}(-1,1) + \frac{1}{2}\mathcal{N}(1,2))$, CN), uniform (Uni), and exponential (Exp).

4 Two adaptive tests

Since we also intend to compare the max-type tests with adaptive tests we first describe them briefly. One concept of adaptive tests is proposed by Hogg (1974). It is based on the independence of rank and order statistics (cf. Randles and Wolfe, 1979, p.388). The density is classified by order statistics, then a rank test is applied. It is quite common to classify the underlying distribution with respect to measures of tailweight and skewness.

There exist many measures of integral type or of quantile type (cf. e.g. Büning,

Table 2: Measures for tailweight $t_{0.05,0.15}(F)$ and skewness $s_{0.05}(F)$ for some distributions.

Symmetric distributions		Skew distributions		
Density	Tailweight	Density	Tailweight	Skewness
Uniform	1.286	Exponential	1.697	0.564
Normal	1.587	Gumbel[1]	1.655	0.280
Logistic	1.697	negGumbel[2]	1.655	-0.280
DoubleExp	1.912	CN[3]	1.592	0.277
Cauchy	3.217			

1 Gumbel denotes the extreme value density $f_G(x) = e^{-x}e^{-e^{-x}}$

2 negGumbel denotes the reflected Gumbel, $f_{nG}(x) = f_G(-x)$

3 CN denotes the contaminated normal $\frac{1}{2}\mathcal{N}(1,4) + \frac{1}{2}\mathcal{N}(-1,1)$.

1991, Handl 1986, Hogg and Lenth, 1984). We choose the measures

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

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

for tailweight and skewness, respectively. These measures are introduced by Groeneveld and Meeden (1984). Some examples are given in Table 2. Some more examples you may find in Kössler (2006b). The table shows that these measures are in accordance with our idea of tailweight and skewness.

Replacing the quantile function $F^{-1}(.)$ by an estimate $\hat{Q}(.)$ we obtain estimates $\hat{t}_{0.05,0.15}$ and $\hat{s}_{0.05}$ of tailweight and skewness. To estimate the quantiles we use the "classical" estimate

$$\hat{Q}(u) = \begin{cases} X_{(1)} - (1-\epsilon)(X_{(2)} - X_{(1)}) & \text{if} \quad u < 1/(2 \cdot L) \\ (1-\epsilon) \cdot X_{(j)} + \epsilon \cdot X_{(j+1)} & \text{if} \quad 1/(2 \cdot L) \le u \le (2 \cdot L - 1)(2 \cdot L) \\ X_{(L)} + \epsilon(X_{(L)} - X_{(L-1)}) & \text{if} \quad u > (2 \cdot L - 1)/(2 \cdot L), \end{cases}$$

where $\epsilon = L \cdot u + 1/2 - j$, $j = \lfloor L \cdot u + 1/2 \rfloor$, and $X_{(i)}$ is the i-th order statistic of a sample of size L.

4.1 An adaptive test based on various scores statistics

It is quite common to use the Gastwirth-scores, Wilcoxon scores and LT-scores for symmetric densities with short tails, medium tails and long tails, respectively. For skew densities the HFR-scores and HFL-scores are used, cf. e.g. Büning (1991) or Kössler (2006a). Denote $\hat{S} = (\hat{t}_{0.05,0.15}, \hat{s}_{0.05})$. We define the Adaptive test by its test statistic $A(\hat{S})$ with the GA, WI, LT and HFR and HFL scores (cf. Figure 1),

$$A(\hat{S}) = \begin{cases} GA & \text{if} \quad \hat{S} \in D_1 := \{\hat{t}_{0.05,0.15} \leq 1.55, \hat{s}_{0.05} \leq 0.25\} \\ WI & \text{if} \quad \hat{S} \in D_2 := \{1.55 < \hat{t}_{0.05,0.15} \leq 1.8, \hat{s}_{0.05} \leq 0.25\} \\ LT & \text{if} \quad \hat{S} \in D_3 := \{\hat{t}_{0.05,0.15} > 1.8, \hat{s}_{0.05} \leq 0.25\} \\ HFR & \text{if} \quad \hat{S} \in D_4 := \{\hat{s}_{0.05} > 0.25\} \\ HFL & \text{if} \quad \hat{S} \in D_5 := \{\hat{s}_{0.05} < -0.25\}. \end{cases}$$

Remark 1 *Note that we also considered the SAR and SA scores instead of the HFR and HFL scores, respectively, but the difference in asymptotic power is very small. We also tried to include the Median scores,* $a_{L,ME}(k) = sign(k - \frac{L+1}{2})$ *in an adaptive test, however the power becomes worse in almost all cases.*

For the one-sided alternative the asymptotic power of the adaptive test is given by

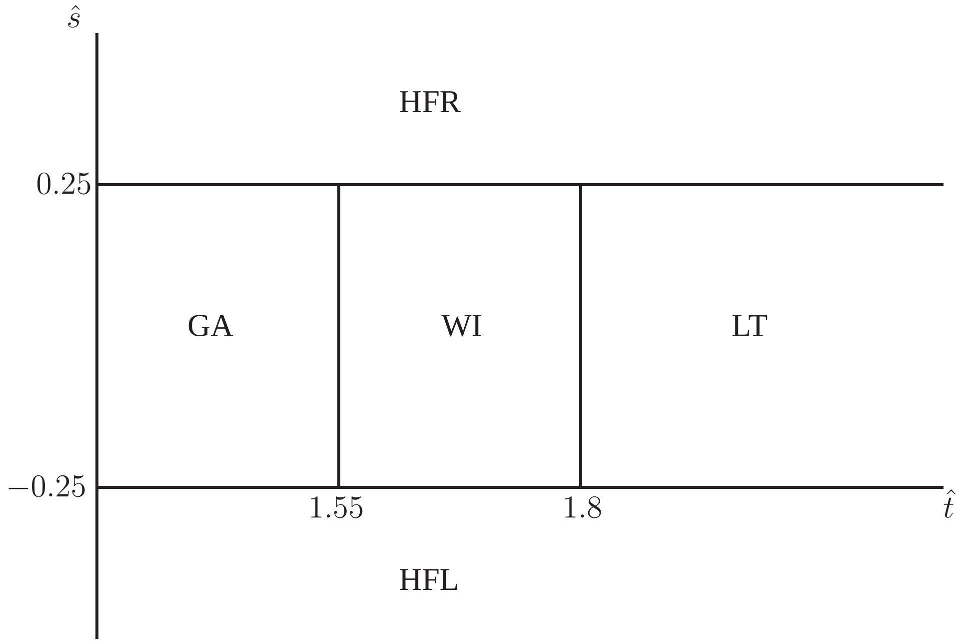
$$\beta_A(t) = 1 - \Phi\big(z_{1-\alpha} - t\lambda(1-\lambda)C(f, g_i)\big) \quad \text{if} \quad f \in D_i$$

where D_i is the region in which the couple $(t_{0.05,0.15}, s_{0.05})$ falls, cf. Kössler, 2006a, p.118).

4.2 An adaptive test based on various U-statistics

Adaptive tests based on U-statistics are proposed by Kössler (2006a) as well as by Kössler and Kumar (2008). Here we propose another variant which is based on some of the statistics $U_{k:s}$, $1 \leq k \leq s \leq 5$, and on the same measures of tailweight

Figure 1: Schematic presentation of the Adaptive test $A(\hat{S})$



and skewness as for the Adaptive test $A(\hat{S})$. Its statistic is given by (cf. Figure 2)

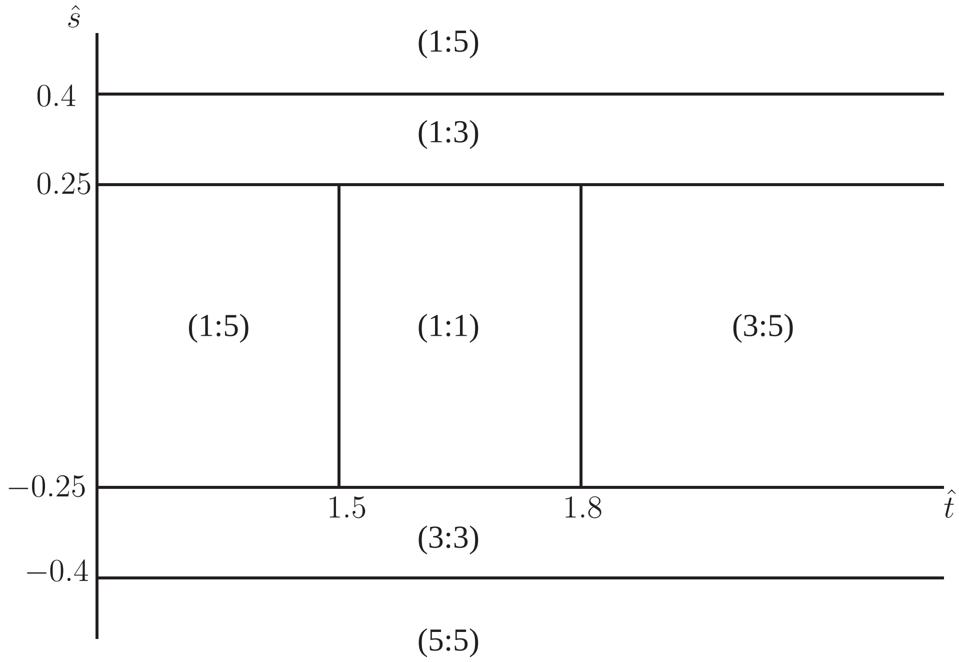
$$B(\hat{S}) = \begin{cases} U_{1:5} & \text{if} \quad \hat{S} \in E_{1:5} := \{\hat{s} > 0.4\} \cup \{\hat{s} \in (-0.25, 0.25), \hat{t} \leq 1.5\} \\ U_{1:3} & \text{if} \quad \hat{S} \in E_{1:3} := \{0.4 \geq \hat{s} \geq 0.25\} \\ U_{1:1} & \text{if} \quad \hat{S} \in E_{1:1} := \{\hat{s} \in (-0.25, 0.25), \hat{t} \in (1.5, 1.8)\} \\ U_{3:5} & \text{if} \quad \hat{S} \in E_{3:5} := \{\hat{s} \in (-0.25, 0.25), \hat{t} \geq 1.8\} \\ U_{3:3} & \text{if} \quad \hat{S} \in E_{3:3} := \{-0.4 \leq \hat{s} \leq -0.25\} \\ U_{5:5} & \text{if} \quad \hat{S} \in E_{5:5} := \{\hat{s} < -0.4\} \end{cases}$$

As for the Adaptive test $A(\hat{S})$ the regions $E_{k:s}$ are motivated by the coefficients $C_{U_{k:s}}(f)$ and by the values for tailweight and skewness for a given density f, i.e. the regions $E_{k:s}$ are defined in such a way that the coefficients $C_{U_{k:s}}(f)$ are large for f in that region.

The asymptotic power of the adaptive test $B(\hat{S})$ is given by

$$\beta_B(t) = 1 - \Phi\big(z_{1-\alpha} - t\lambda(1-\lambda)C_{U_{k:s}}(f)\big) \quad \text{if} \quad f \in E_{k:s},$$

Figure 2: Schematic presentation of the Adaptive tests $B(\hat{S})$



where

$$C_{U_{k:s}}(f) = \frac{1}{\sigma_{U_{k:s}}} \binom{k}{s}^2 s^2 \int_{-\infty}^{\infty} F^{k-2}(x)(1-F(x))^{2s-2k} f^2(x)\,dx$$

Some numerical values for the factors $C_{U_{k:s}}(f)$ are presented in Table 1.

5 Max-type tests

5.1 General max-type tests

Another idea in our paper is to use the maximum of a set of k various scores statistics as a test statistic, i.e

$$T_{max} = \max_{i=1,\ldots,k} T_i^*,$$

where

$$T_i^* = \frac{T_i}{\sqrt{var(T_i)}}$$

are the standardized linear rank statistics.

Denote by $\mathbf{T}^* = (T_1^*, \ldots, T_k^*)$ the vector of k standardized linear rank statistics.

Proposition 5 *Let be* $\mathbf{x} = (x_1, \ldots, x_k)^T$ *and let* $\mathbf{\Sigma}_T^*$ *be the asymptotic covariance matrix of the statistic* $\mathbf{T}^*$ *which is assumed to be regular. Then, for* $\tau > 0$*,*

$$P_0(T_{\max} < \tau) = \underbrace{\int_{-\infty}^{\tau} \cdots \int_{-\infty}^{\tau}}_{k\ times} \frac{det(\mathbf{\Sigma}_T^*)^{-1/2}}{(2\pi)^{(k)/2}} exp(-\frac{1}{2}\mathbf{x}^T \mathbf{\Sigma}_T^{*-1}\mathbf{x})\, d\mathbf{x}. \qquad (6)$$

Corollary 3 *Let* $\tau_{1-\alpha}$ *be the* $(1-\alpha)$*-quantile of the null distribution of* $T_{\max}$ *and* $\boldsymbol{\mu}_T = \boldsymbol{\mu}_T(t)$ *an asymptotic expectation vector of* $\mathbf{T}^*$*. The asymptotic power function of the max-type test* $T_{\max}$ *is given by*

$$\beta_{\max}(t) = 1 - det(\mathbf{\Sigma}_T^*)^{-1/2}(2\pi)^{-(c-1)/2} \cdot \underbrace{\int_{-\infty}^{\tau_{1-\alpha,T}} \cdots \int_{-\infty}^{\tau_{1-\alpha,T}}}_{k\ times} exp(-\frac{1}{2}(\mathbf{x} - \boldsymbol{\mu}_T)^T \mathbf{\Sigma}_T^{*-1}(\mathbf{x} - \boldsymbol{\mu}_T))\, d\mathbf{x}. \qquad (7)$$

The components $\mu_{T,i}(t)$ *of the expectation vector are given by* $\mu_{T,i}(t) = \mu_{T_{g_i}}(t) = \sqrt{\lambda(1-\lambda)}C(f, g_i) \cdot t$*.*

To obtain critical values for the max-type test we have to evaluate a k-dimensional normal integral. In previous attempts with this type of tests a Bonferroni approximation or a simulation based permutation test is used (cf. Neuhäuser, Büning, and Hothorn, 2004).

Our idea is to obtain critical values by using the known asymptotic correlation structure of the components of the max-type test.

5.2 Max-type tests based on various scores statistics

Denote by an asterics the standardized statistics GA, WI, LT, HFR, and HFL.

At first, it seems to be useful to consider at least three different statistics, one defined for symmetric densities, one for right-skewed und one for left-skewed densities. Such a max-type test was proposed, in another context, by Neuhäuser, et al. (2000),

$$MAX_3 = \max\ (WI^*, HFR^*, HFL^*),$$

where the index three stays for the number of components. One may argue that it is useful to have at least two different scores for symmetric densities, one for short tails (GA), and one for long tails (LT). This way we arrive at a max-type test with four components

$$MAX_4 = \max\ (GA^*, LT^*, HFR^*, HFL^*).$$

If we intend to include also the Wilcoxon scores in our max-type test, we would use

$$MAX_5 = \max\ (GA^*, WI^*, LT^*, HFR^*, HFL^*).$$

Note that in MAX_5 we have the same five scores statistics as in the adaptive test $A(\hat{S})$.

Remark 2 *Dropping out the test HFL^* designed for left-skew densities we arrive at*

$$MAX_4' = \max\ (GA^*, WI^*, LT^*, HFR^*).$$

proposed by Neuhäuser, et.al. (2004). Note that the authors used the first Bonferroni inequality to obtain the critical value $t'_{krit,4} = 2.234$ for the test MAX_4'. Therefore their test is rather conservative. However, using our approach, the critical value may be boiled down to $t_{krit} = 1.95$, and the asymptotic power becomes considerably larger. To allow also left-skew densities a slight classification procedure (based on order statistics) might be placed before the application of this test. If the density is classified to be left-skew (for instance by $\hat{s} < -0.25$) then the samples may be reflected at zero, and after that the MAX_4' test can be applied. The resulting procedures is also distribution-free. It can be considered as a combination of an adaptive and a max-type test.

Remark 3 *Note that we considered also all variants where the SAR and SA scores instead of HFR and HFL are used. The differences in asymptotic power are small, the variants with the HFR and HFL scores are slightly better (except, of course, for the Gumbel).*

Remark 4 *On the other hand it may be an idea to simplify the procedure and to apply a max-type test with only two substatistics. We choose the most separate statistics GA and LT ($\rho(GA, LT) = 0.75$) involved in the test MAX_3. In this case it is much easier to compute the asymptotic power functions. They look very similar to that of MAX_3.*

Table 3: The asymptotic correlations for the tests considered in the scores based max-type tests.

	Test						
Test	GA	WI	LT	HFR	HFL	S A	SAR
GA	1.00	.884	.750	.790	.790	.856	.856
WI		1.00	.972	.894	.894	.866	.866
LT			1.00	.869	.869	.796	.796
HFR				1.00	.600		
HFL					1.00		
SA						1.00	.646
SAR							1.00

Denote by $\mathbf{T}^*_{\mathbf{g}} = (T^*_{g_1}, \dots, T^*_{g_k})$ the vector of the standardized linear rank statistics, with pairwise different score functions but applied on the same data.

Proposition 6 *The vector $\mathbf{T}^*_{\mathbf{g}}$ is, under H_0, asymptotically multivariate normal with expectation vector zero and covariance matrix $\mathbf{\Sigma} = (\rho_{ij})$, where the entries are given by*

$$\rho_{ij} = \sqrt{ARE(g_i, g_j)}, \quad i, j = 1, ..., k.$$

Under the alternative, the components of the expectation vector are given by eq. (3) where g is to be replaced by g_i.

Proof: The asymptotic variances and covariances of the linear rank statistics T_{g_i} are (cf. Hájek, Šidák, and Sen, 1999, section 3.3 and chapter 6)

$$\begin{aligned} varT_{g_i} &= \frac{n_1 n_2}{N} I(g_i) \\ cov(T_{g_i}, T_{g_j}) &= \frac{n_1 n_2}{N} d(g_i, g_j). \end{aligned}$$

The rest of the assertion follows from Proposition 2. ■

The asymptotic correlations of the linear rank tests included in the max-type tests are given in Table 3. (For the convenience of the reader the values for the SA and SAR scores are included too.)

Table 4: Asymptotic critical values ($\alpha = 0.05$) of the various max-type tests

MAX_3	MAX_4'	MAX_4	MAX_5	$MAXU_3$	$MAXU_4$	$MAXU_6$
1.92	1.95	2.01	2.02	1.99	2.01	2.11

To obtain asymptotic critical values for these tests we evaluate the multiple integrals (6) by Monte Carlo methods. To do this we simulate the multivariate normal distribution (simulation size M=1,000,000) by using the Cholesky decomposition of the correlation matrix, cf. Tong (1980, ch. 8.1.4), take the maxima, and estimate the critical value by the empirical 0.95-quantile. The simulation is done by using the SAS-package. Values obtained are collected in Table 4, together with that for the max-type tests based on U-statistics (see next subsection).

In a similar way the asymptotic power functions are estimated. Multivariate normally distributed random variables are simulated, then the asymptotic expectations are added (cf. Propositions 1 and 2) and maxima are taken. The power is estimated by the relative number of rejections, i.e. by the relative number of cases with $T_{\max} > t_{krit}$. Simulation size is M=100,000, and we restrict to the case of equal sample sizes. Eight densities are considered, the normal, logistic, doubleexponential, Cauchy, Gumbel, CN (cf. Table 2), uniform and the exponential. They represent symmetric densities with short, medium, long and very long tails as well as skew densities. The factor t in the formula for the asymptotic power function is multiplied by the standard deviation σ_F of the underlying density if it exists, except for the exponential where we set $\sigma_F = 0.5$. (For the Cauchy we set $\sigma_F = \sigma_{Cau} = F^{-1}(\Phi(1)) = 1.8373$.) This way we have similar power values for the various densities. The results are analysed in Section 6.

5.3 Max-type tests based on various U-statistics

Let $U^*_{k:s}$ be the standardized U-statistics. Some ideas of max-type tests based on U-statistics are

$$\begin{aligned} MAXU_6 &= \max(U^*_{1:1}, U^*_{1:3}, U^*_{1:5}, U^*_{3:5}, U^*_{3:3}, U^*_{5:5}) \\ MAXU_4 &= \max(U^*_{1:1}, U^*_{1:3}, U^*_{1:5}, U^*_{3:5}) \\ MAXU_3 &= \max(U^*_{1:3}, U^*_{3:3}, U^*_{3:5}) \end{aligned}$$

where, again, the index stays for the number of components.

Table 5: Asymptotic correlations between some $U_{k:s}$-statistics

(k:s)	(1:1)	(1:3)	(2:3)	(3:3)	(1:5)	(3:5)	(5:5)
(1:1)	1.000	.8207	.9857	.8207	.6863	.9718	.6863
(1:3)		1.000	.7741	.4343	.9638	.7501	.3206
(2:3)			1.000	.7741	.6149	.9973	.6149
(3:3)				1.000	.3206	.7501	.9638
(1:5)					1.000	.5849	.2345
(3:5)						1.000	.5849

All these ideas have certain motivations. In $MAXU_6$ all the considered U-statistics are included, except $U^*_{2:3}$ which has extremely high correlation with $U^*_{3:5}$. The substatistics are the same as that in the Adaptive test $B(\hat{S})$. The other two max-type statistics are included since it may be convenient to have a smaller number of components. In $MAXU_3$ we have two U-statistics for short to medium tails (U^*_{13} and U^*_{33}), and one for long tails (U^*_{35}). Also for right-skew (left-skew) densities the components U^*_{13} (U^*_{33}) may be useful. In $MAXU_4$ we have added the Mann-Whitney statistic $U_{1:1}$. Note that we also considered some other variants but they are found to be asymptotically (slightly) worse.

Proposition 5 and the corollary can also be applied to these types of tests. To obtain asymptotic critical values, again multiple integrals have to be evaluated. This can be done in the same way as in the previous sections. The asymptotic correlations of the U-statistics included in the max-type tests are given in Table 5.

Asymptotic critical values are given in Table 4.

The asymptotic power functions are given by eq. (7), where the factor $C(f, g_i)$ is to replace by the factor $C_{U_{k:s}}(f)$. For a closer investigation we use the same densities as in the previous section. The results are analysed in Section 6.

6 Comparison and discussion of the various ideas

Let us compare the various ideas concerning asymptotic and finite power. In Table 6 we ranked the asymptotic power, at the point $\theta = 4.0$ (for the exponential at $\theta = 2.0$), of the nine considered tests, for all eight considered densities. The worst test gets rank one, the best rank nine. Average ranks are taken in the case of ties, i.e. if the asymptotic powers are equal up to two decimal points.

Table 6: Ranks of the asymptotic power functions of the adaptive and max-type tests for various densities f.

Test	Lo	N	Gu	DE	Cau	CN	Uni	Exp	Sum	Place
MAX_3	5.0	3.0	1.0	2.0	1.5	4.5	1.0	1.0	19.0	
MAX_4	5.0	5.5	2.5	4.0	4.5	4.5	5.0	3.5	27.5	
MAX_4'	5.0	7.0	6.5	6.0	6.0	7.0	6.0	5.0	48.5	3
MAX_5	5.0	5.5	4.5	4.0	4.5	4.5	4.0	2.0	34.0	5
$MAXU_3$	5.0	3.0	2.5	7.0	7.0	2.0	2.0	6.5	35.0	4
$MAXU_4$	2.0	3.0	6.5	4.0	3.0	4.5	3.0	6.5	32.5	6
$MAXU_6$	1.0	1.0	4.5	1.0	1.5	1.0	9.0	8.5	27.5	
$A(\hat{S})$	8.5	8.5	8.0	8.0	8.5	9.0	7.5	3.5	61.5	2
$B(\hat{S})$	8.5	8.5	9.0	9.0	8.5	8.0	7.5	8.5	67.5	1

Clearly, the two adptive tests are, asymptotically, and over all densities, the best, where the adaptive test based on U-statistics is slightly better.

On the third place there is the test MAX_4', which is a combination of an adaptive and max-type test.

Among the max-type tests the test $MAXU_3$ is slightly better than the tests MAX_5 and $MAXU_4$. The other max-type tests are worse. By the way we see that the inclusion of more different components (four or five instead of three for $MAXU_3$) does not necessarily improve the asymptotic power.

In Figure 3 we present the curves of the asymptotic power functions of the asymptotically best tests, the two adaptive tests, together with that of the asymptotically best max-type score test MAX_5 and that of one of the asymtotically best max-type U-statistics test $MAXU_3$.

For moderate densities (not too skew, not too heavy or light tails) the test $MAXU_3$ is, asymptotically, almost as good as the adaptive tests. For more extreme densities (Cauchy, uniform) the adaptive tests are clearly better. For the exponential the tests based on U-statistics are the best.

In order to assess whether the asymptotic theory can also be applied for moderate to small sample sizes a simulation study (10,000 replications each) is performed. We use the same distributions as for the asymptotic case and consider the four asymptotically best tests $A(\hat{S})$, $B(\hat{S})$, MAX_5 and $MAXU_3$ (the same as in Figure 3). Sample sizes of $n_1 = n_2 = 20, 40$ and alternatives $\theta_N = N^{-1/2}\theta\sigma_F$ are considered. For moderate densities (normal, logistic, DE, Gumbel, CN) and for $n_1 = n_2 = 40$ the simulated power values for all the four tests are very similar. In

Figure 3: The asymtotic power functions of the tests $A(\hat{S})$ (red, continuous line), $B(\hat{S})$ (magenta, dash-dot line), MAX_5 (blue, long-dashed line), and $MAXU_3$ (green, dotted line) for various densities.

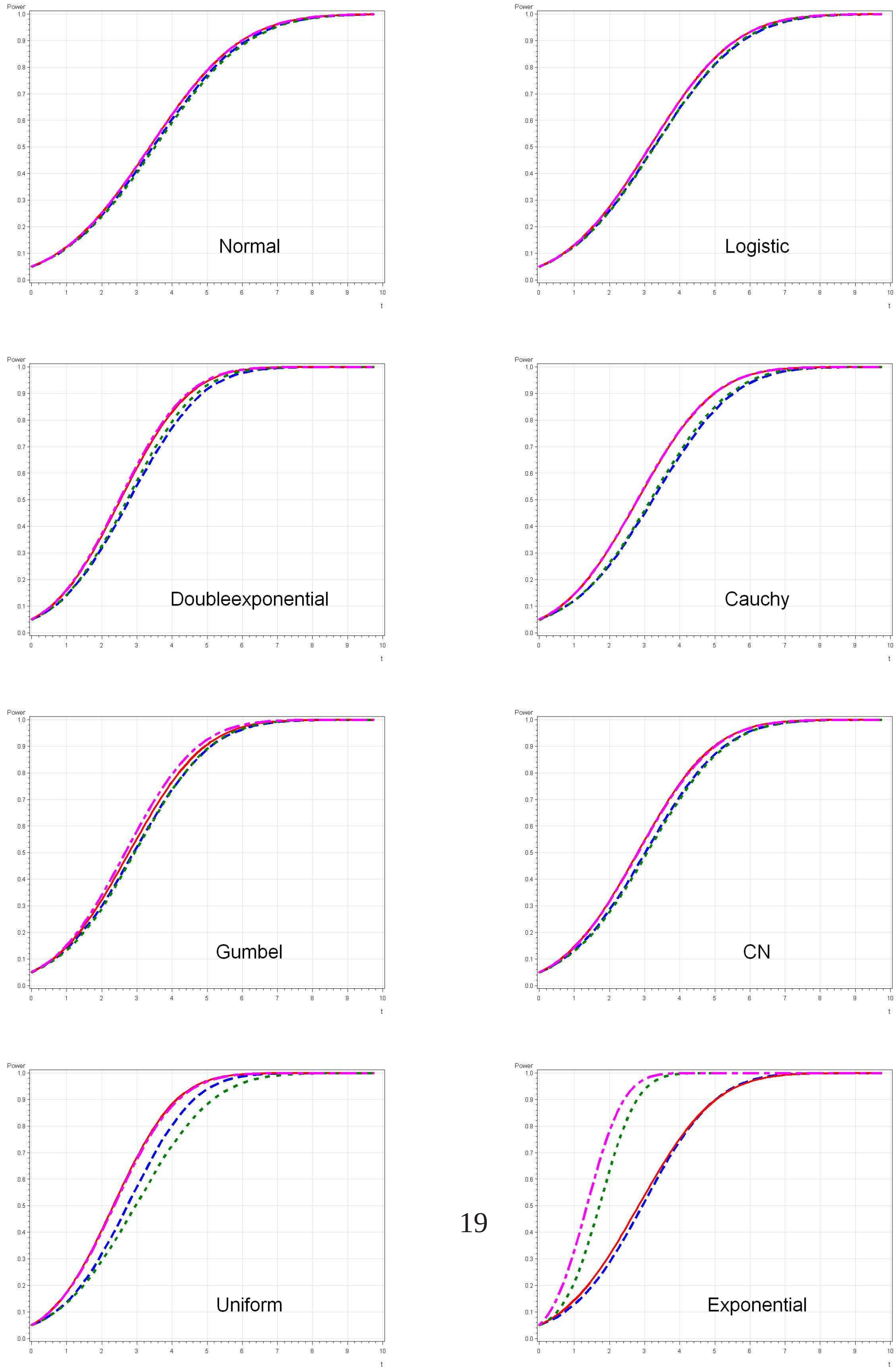


Figure 4: The estimated power functions of the tests $A(\hat{S})$ (red, continuous line), $B(\hat{S})$ (magenta, dash-dot line), MAX_5 (blue, long-dashed line), and $MAXU_3$ (green, dotted line) for various densities.

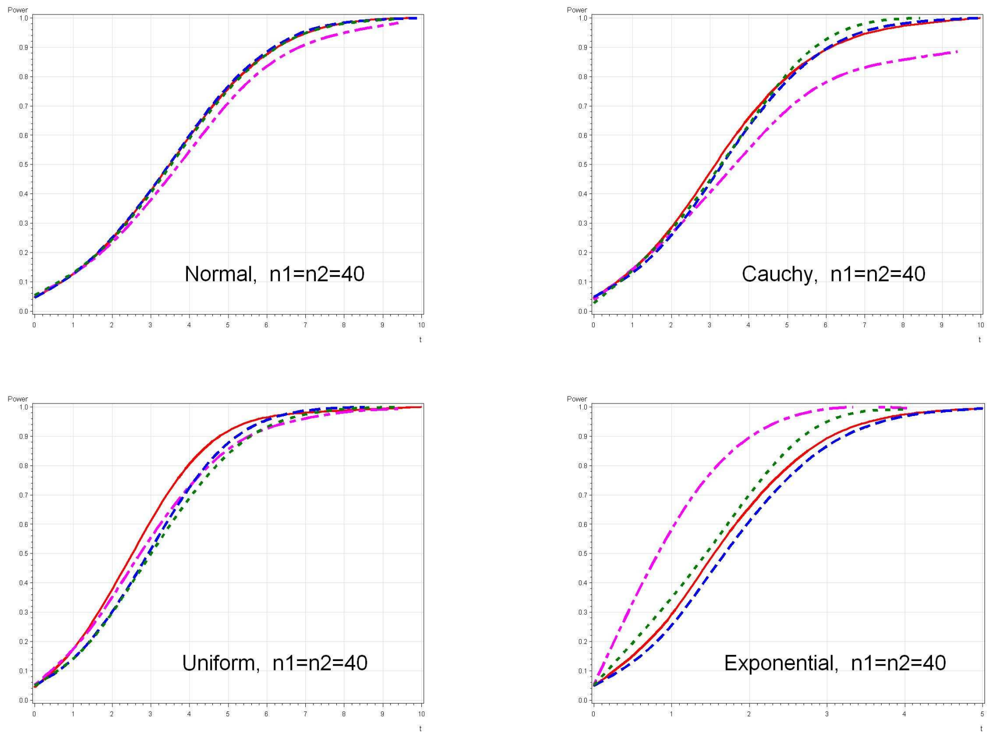


Figure 4 estimated power functions for $n_1 = n_2 = 40$ are presented only for the normal, Cauchy, uniform, and exponential.

For sample sizes $n_1 = n_2 = 20$ all tests are slightly anticonservative (level from 0.05 to 0.055 for the scores type tests, and from 0.05 to 0.063 for the U-statistics based tests), except for the test $MAXU_6$ which satisfies the level.

7 Conclusions

In the present paper we considered the two-sample location problem and investigated various combinations of adaptive tests and max-type tests with linear rank tests and tests based on U-statistics. We established the asymptotic correlations of various linear rank statistics as well as of various U-statistics and obtained asymptotic critical values for the max-type tests. Moreover, we approximated their asymptotic power functions.

When constructing adaptive tests or max-type tests we first have to determine the number of selected substatistics. A large number will result in high asymptotic power for the adaptive test. However, for finite sample sizes, the misclassification probabilities will increase with a rising number of substatistics. On the other hand, increasing the number of components in the max-type test also will increase the critical values and, therefore, may result in reduced (asymptotic) power.

The second point to consider in constructing adaptive tests is to have, for each density, a suitable test statistic (or a suitable representant) in the set of chosen test statistics. We think, that the substatistics based on the scores GA, WI, LT, HFR, and HFL as well as the U-statistics $U_{1:1}, U_{1:3}, U_{1:5}, U_{3:3}, U_{3:5}$, and $U_{5:5}$ are good representants. The scores GA, WI, and LT are good for short, medium, and long tails, respectively. The scores HFR and HFL are good for right-skew and for left-skew densities, respectively. Among the U-statistics, $U_{1:5}$ and $U_{5:5}$ are good for short tails, whereas $U_{1:1}$ and $U_{3:5}$ are good for nearly symmetric densities with medium and long tails, respectively. For very right (left) skew densities the test $U_{1:5}$ ($U_{5:5}$), and for moderately skew densities the test $U_{1:3}$ ($U_{3:3}$) are suitable. This way we have two sets of convenient substatistics of five and six elements, respectively, from which we may choose some to construct our adaptive test or max-type test.

We have shown that the adaptive tests are asymptotically the best in most cases. But also a simple max-type test based on linear rank statistics or on U-statistics may be useful. The tests MAX_5 and $MAXU_3$ are the best among all considered max-type tests. Max-type tests based on linear rank statistics and such based on U-statistics are of same value.

It should be noted that we also considered sum tests but we found that they behave slightly worse (cf. also Neuhäuser, et.al. (2004)).

Summarizing, for the case of an unknown density, all of the four variants, adaptive tests and max-type tests based on linear rank tests or on U-statistics, are justified. The adaptive tests are asymptotically best, and they are proposed for larger sample sizes (about $n_i \geq 40$). For smaller sample sizes (about $n_i \leq 20$) the max-type tests are to prefer. Linear rank statistics are more simple but U-statistics may better smooth the effect of extreme densities.

References

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

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

Groeneveld, R.A. and Meeden, G. (1984). Measuring skewness and kurtosis. *The Statistician*, **33**, 391-399.

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

Handl, A. (1986). Maßzahlen zur Klassifizierung von Verteilungen bei der Konstruktion adaptiver verteilungsfreier Tests im unverbundenen Zweitichprobenproblem, unpublished dissertation, Freie Universität Berlin.

Hogg, R.V., Fisher, D.M. and Randles, R.H. (1975). A two-sample adaptive distribution-free test. *Journal of the American Statistical Association*, **70**, 656-661.

Hogg, R.V. and Lenth, R.V. (1984). A review of some adaptive statistical techniques. *Communications in Statistics, Theory and Methods*, **13**, 1551-1579.

Kössler, W. (2006a). *Nonparametric location tests against restricted alternatives*, Shaker Aachen.

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.

Kössler, W., and Kumar, N. (2008). An adaptive test for the two-sample location problem based on U-statistics, *Communications in Statistics - Simulation and Computation*, **37**, 1329-1346.

Neuhäuser, M., Seidel, D., Hothorn,L., and Urfer (2000). Robust trend tests with application to toxicology. *Environmental and Ecological Statistics*, **7**, 43-56.

Neuhäuser, M., Büning, H., and Hothorn,L. (2004). Maximum Tests versus Adaptive Tests for the Two-Sample Location Problem, *Journal of Applied Statistics*, **31**, 215-227.

Policello, G.E. and Hettmansperger, T.P. (1976). Adaptive robust procedures for the one-sample location problem. *Journal of the American Statistical Association*, **71**, 624-633.

Randles, R.H. and Wolfe, D.A. (1991). *Introduction to The Theory of Nonparametric Statistics*. Wiley, New York.

Tong, Y.L. (1990). *The Multivariate Normal Distribution*. Springer, New York.

Xie, J. and Priebe, C.E. (2000). Generalizing the Mann-Whitney-Wilcoxon Statistic, *Journal of Nonparametric Statistics*, **12**, 661-682.

Xie, J. and Priebe, C.E. (2002). A weighted generalization of the Mann-Whitney-Wilcoxon Statistic, *Journal of Statistical Planning and Inference*, **102**, 441-466.