Integral distribution-free statistics of L_p-type and their asymptotic comparison

Norbert Henze [a], Ya. Yu. Nikitin [b,*], Bruno Ebner [a]

[a] *Institut für Stochastik, University of Karlsruhe (TH) and Karlsruhe Institute of Technology (KIT), Germany*

[b] *Department of Mathematics and Mechanics, St. Petersburg University, Russia.*

Abstract

Generalizing the Cramér-von Mises and the Kolmogorov-Smirnov test, different integral statistics based on L_p-norms are compared with respect to local approximate Bahadur efficiency. Simulation results corroborate the theoretical findings. Several examples illustrate that goodness-of-fit testing based on L_p-norms should receive more attention. It is shown that, given a distribution function F_0 and a specific alternative, one can draw the plot of efficiency as a function of p and determine the value of p giving the maximum efficiency.

Key words: Empirical process, omega-square statistic, Kolmogorov-Smirnov statistic, tail behavior; approximate Bahadur efficiency, shift alternatives, scale alternatives, skew alternatives, test for independence, goodness-of-fit test.

1 Introduction and Summary

Let $X_1, \ldots, X_n$ be independent random variables with common continuous distribution function (cdf) F. A classical problem in statistics is to test the hypothesis $H_0 : F = F_0$, where F_0 is some known continuous cdf, against general alternatives.

* Corresponding author. Department of Mathematics and Mechanics, St. Petersburg University, Universitetski pr. 28, Petrodvorets 198504, Russia. Fax: +7(812)4286944, Phone: +7(812)4284212

Email addresses: N.Henze@math.uni-karlsruhe.de (Norbert Henze), yanikit47@mail.ru (Ya. Yu. Nikitin), ebner@stoch.uni-karlsruhe.de (Bruno Ebner).

Many celebrated distribution-free statistics for this goodness-of-fit testing problem are functionals of the empirical process

$$\xi_n(x) \;=\; \sqrt{n}\,(F_n(x) \;-\; F_0(x))\,, \quad x \in \mathbb{R},$$

where

$$F_n(x) = \frac{1}{n}\sum_{j=1}^{n} \mathbf{1}\{X_j \le x\}$$

is the empirical cdf based on $X_1, \ldots, X_n$.

Prominent examples are the Kolmogorov-Smirnov statistic

$$D_n \;=\; \sup_x |\xi_n(x)|$$

and the Cramér-von Mises statistic

$$\omega_n^2 \;=\; \int_{-\infty}^{\infty} \xi_n^2(x)\; dF_0(x).$$

Many variants of D_n and ω_n^2, together with their limiting properties and efficiencies, have been studied, see, e.g., [32] and [26].

There is, however, a continuing interest in more general integral statistics based on the L_p-norm, i.e., in statistics of the type

$$\begin{aligned}
\omega_{n,p} \;&=\; \left(\int_{-\infty}^{\infty} |\xi_n(x)|^p\; dF_0(x)\right)^{1/p} \\
&= \sqrt{n}\left(\int_{-\infty}^{\infty} |F_n(x) - F_0(x)|^p\; dF_0(x)\right)^{1/p},
\end{aligned}$$

where $1 \le p < \infty$. Notice that the Kolmogorov-Smirnov statistic D_n corresponds to the case $p = \infty$.

For instance, the L_1-variant $\omega_{n,1}$ of the Cramér-von Mises statistic was considered in [30], while its two-sample version was discussed in [31]. The "one-sided" statistic

$$W_{n,k} = \int_{-\infty}^{\infty} \xi_n^k(x)\; dF_0(x),$$

where k is some integer, was introduced in [25] and studied later in [26] and [19]. More general statistics based on a weighted L_p-norm were proposed in [9,10]. These statistics were also investigated in [1] and, from a different point of view, in [7]. Some properties of $\omega_{n,p}$ were studied in [14,15]. Various applications of L_p-norms in statistical classification, estimation and testing have been addressed in [33], [13] and [23].

There is some evidence that the new statistics will have better properties than their classical counterparts. In fact, simulation results (see [31]) show that, at least for certain alternatives, $\omega_{n,1}$ exhibits higher power than $\omega_{n,2}$, while certain merits of $W_{n,3}$ and $W_{n,4}$ are reported in [19]; see also [38] and [39].

Denoting by $X_{(1)} \leq X_{(2)} \leq \ldots \leq X_{(n)}$ the order statistics of $X_1, \ldots, X_n$, and putting $X_{(0)} = -\infty$, $X_{(n+1)} = \infty$, straightforward manipulations of integrals yield the alternative representation

$$\omega_{n,p} = \frac{\sqrt{n}}{(p+1)^{1/p}} \left(\sum_{j=1}^{n+1} \left[\left(F_0(X_{(j)}) - \frac{j-1}{n} \right) \cdot \left| F_0(X_{(j)}) - \frac{j-1}{n} \right|^p \right.\right.$$
$$\left.\left. - \left(F_0(X_{(j-1)}) - \frac{j-1}{n} \right) \cdot \left| F_0(X_{(j-1)}) - \frac{j-1}{n} \right|^p \right] \right)^{1/p},$$

$1 \leq p < \infty$. Hence, a computer routine for implementing the test based on $\omega_{n,p}$ is readily available.

The aim of the present paper is to compare L_p-statistics for different values of p on the basis of asymptotic test efficiency, namely local approximate Bahadur efficiency. This notion of efficiency often coincides with local exact Bahadur efficiency and limiting Pitman efficiency, see [6], [37], and [26]. By using approximate Bahadur efficiency, one also avoids the difficult theory connected with large deviations. If necessary, the pertaining results can be borrowed from [26] and [29].

Theoretical results on a ranking of statistics according to their asymptotic efficiency are corroborated in a Monte Carlo Study. Moreover, the analysis of conditions of local asymptotic optimality (see [26], Ch.6) may suggest the type of alternative against which a given statistic can have maximum efficiency and possibly maximum power.

The paper is organised as follows. We first obtain expressions for the approximate Bahadur efficiency against general alternatives and then calculate these efficiencies for several distributions and shift, scale and skew alternatives. The results are illustrated by graphs that show the dependence of the efficiency as a function of p. Moreover, it is shown that our approach can be extended to cover tests of independence. Finally, we present some results of power simulations which are in accordance with the theoretical findings.

2 Approximate Bahadur efficiency of L_p-statistics

Following [6], let $\{P_\theta : \theta \in \Theta\}$ be a set of probability measures on a measurable space $(S, \mathcal{S})$. Suppose Θ_0 is some subset of Θ, and let H_0 be the hypothesis that $\theta \in \Theta_0$. Assume $\{T_n\}$ is a sequence of real-valued test statistics defined on $(S, \mathcal{S})$, based on a sample of size n. $\{T_n\}$ is called a *standard sequence* if the following conditions hold:

a) There exists a continuous probability cdf G such that, for each $\theta \in \Theta_0$,

$$\lim_{n\to\infty} P_\theta(T_n \le x) = G(x) \qquad \text{for every } \; x.$$

b) There is a constant a, $0 < a < \infty$, such that

$$\log(1 - G(x)) = -\frac{ax^2}{2} \cdot (1 + o(1)) \quad \text{as } \; x \to \infty.$$

c) There exists a real-valued function $b(\theta)$ on $\Theta \setminus \Theta_0$, with $0 < b(\theta) < \infty$, such that, for each $\theta \in \Theta \setminus \Theta_0$,

$$\lim_{n\to\infty} P_\theta \left(\left| \frac{T_n}{\sqrt{n}} - b(\theta) \right| > x \right) = 0 \qquad \text{for every } \; x.$$

For any standard sequence $\{T_n\}$, the so-called *approximate Bahadur slope*

$$c_T^*(\theta) = a \cdot b^2(\theta), \quad \theta \in \Theta \setminus \Theta_0,$$

is a measure of approximate Bahadur efficiency. Usually it is true that $c_T^*(\theta) \sim l(T) \cdot (\theta - \theta_0)^2$ as $\theta \to \theta_0$. In this case $l(T)$ is called the *local (approximate) index* of the sequence $\{T_n\}$.

We will apply this general definition to the family of statistics $\omega_{n,p}$, $1 \le p < \infty$, and the Kolmogorov statistic

$$\omega_{n,\infty} = \sup_{x\in\mathbb{R}} \sqrt{n} \cdot |F_n(x) - F_0(x)|,$$

which arises from (1) in the limit when $p \to \infty$.

The statistics $\{\omega_{n,p}\}$, $1 \le p \le \infty$, are standard statistics in the Bahadur sense: Under H_0, we have convergence of distributions, because

$$\lim_{n\to\infty} P_{F_0}(\omega_{n,p} \le x) = P(\|B\|_p \le x) =: G_p(x), \quad x \in \mathbb{R},$$

where B is the standard Brownian bridge process on $[0, 1]$, and for short,

$$\|B\|_p = \Big\{ \int_0^1 |B(t)|^p dt \Big\}^{1/p} \qquad \text{for } \; 1 \le p < \infty,$$

while $\|B\|_\infty = \sup_t |B(t)|$.

The tail behavior of the distribution of $\|B\|_p$ was found in [35] and rediscovered later by many authors (e.g. [16]). We have

$$\lim_{x\to\infty} \frac{2}{x^2} \cdot \ln(1 - G_p(x)) \;=\; -C(p),$$

where

$$C(p) = \frac{2p\pi}{(1+\frac{p}{2})^{(p-2)/p}} \cdot \left(\frac{\Gamma(1+1/p)}{\Gamma(\frac{1}{2}+1/p)}\right)^2, \quad 1 \le p \ < \infty, \tag{1}$$

and $C(\infty) = 4$. Here and in what follows, $\Gamma(t) = \int_0^\infty x^{t-1}\exp(-x)\,dx$, $t > 0$, denotes the Gamma function.

If the alternative cdf is $F_\theta(x)$, $\theta \neq 0$, and the cdf F_0 has a density f_0, it follows from the Glivenko-Cantelli theorem that under F_θ we have the convergence in probability

$$\frac{\omega_{n,p}}{\sqrt{n}} \to b_p(\theta) := \left\{\int_{-\infty}^{\infty} |F_\theta(x) - F_0(x)|^p \; dF_0(x)\right\}^{1/p}.$$

Suppose now that the alternative cdf is smooth enough to admit the expansion

$$F_\theta(x) - F_0(x) = \frac{d}{d\theta}F_\theta(x)|_{\theta=0}\cdot\theta + o(\theta)$$

as $\theta \to 0$ and the asymptotic equivalence

$$b_p(\theta) \sim \left\{\int_{-\infty}^{\infty} \left|\frac{d}{d\theta}F_\theta(x)\Big|_{\theta=0}\right|^p \; dF_0(x)\right\}^{1/p} \cdot \theta$$

as $\theta \to 0$.

According to Bahadur theory, the approximate slope of the sequence $\{\omega_{n,p}\}$ is $C(p)\cdot b_p^2(\theta)$, and taking into account the previous approximation, the local index of $\{\omega_{n,p}\}$ is

$$l(\omega_p; f_0) = C(p) \cdot \left\{\int_{-\infty}^{\infty} \left|\frac{d}{d\theta}F_\theta(x)\Big|_{\theta=0}\right|^p \; f_0(x)\,dx\right\}^{2/p}. \tag{2}$$

This index reflects the local behavior of the sequence of test statistics at the given alternative F_θ, and one may regard $l(\omega_p; f_0)$ as a measure of local Bahadur efficiency, cf. [26], Ch.2. The value of $l(\omega_p; f_0)$ may depend heavily on f_0 and F_θ. We will study this behavior in the case of three different types of alternatives: shift, scale and skew.

To illustrate the dependence of the test efficiency on p we will plot the local Bahadur efficiencies. With $f_\theta(x)$ denoting the density of F_θ and

$$K(\theta) = \int_{-\infty}^{\infty} \ln\left(\frac{f_\theta(x)}{f_0(x)}\right) f_\theta(x)\,dx$$

denoting the Kullback–Leibler information between F_θ and F_0, these local Bahadur efficiencies arise from dividing the local index by the constant K_0 satisfying

$$K(\theta) \sim \frac{1}{2} \cdot K_0 \cdot \theta^2 \quad \text{as } \theta \to 0, \tag{3}$$

see [26]. With the notable exception of the Laplace distribution figuring in Example 3.2, we have $K_0 = I_0$, where, generically,

$$I_\theta = \int_{-\infty}^{\infty} \left(\frac{\partial}{\partial\theta} f_\theta(x)\right)^2 \Big/ f_\theta(x)\,dx$$

denotes the Fisher information at the point θ. For instance, when we plot the local efficiencies of ω_p-tests in the case of an underlying normal distribution and shift alternatives, the upper bound (the constant K_0 figuring in (3)) is equal to 1. We will give this upper bound in each case when we present a plot.

In contrast to the exact slopes, the approximate Bahadur slopes have no natural bound. Nevertheless, we can compare the latter with the approximate local indices of the locally most powerful parametric tests which, in the cases considered, are equal to K_0. It follows that the local approximate Bahadur efficiency of the statistic $\omega_{n,p}$ is equal to

$$e^B(\omega_p; f_0) = \frac{C(p)}{K_0} \cdot \left\{\int_{-\infty}^{\infty} \left|\frac{d}{d\theta} F_\theta(x)\Big|_{\theta=0}\right|^p f_0(x)\,dx\right\}^{2/p}.$$

3 Shift alternatives

In this section, we consider shift alternatives of the form $F_\theta(x) = F_0(x-\theta)$. If F_0 has a density f_0 that is sufficiently regular, expression (2) can be simplified to give $l(\omega_p; f_0, shift) = C(p) \cdot (\int_{-\infty}^{\infty} f_0^{p+1}(x)\,dx)^{2/p}$ and thus

$$e^B(\omega_p; f_0, shift) = \frac{C(p)}{K_0} \cdot \left\{\int_{-\infty}^{\infty} f_0^{p+1}(x)\,dx\right\}^{2/p}. \tag{4}$$

For any given choice of the density f_0, the integral on the right hand side of (4) can be evaluated. Thus, the local index becomes a function of p showing which values of p yield the maximum efficiency and presumably the maximum power of the test.

Example 3.1: (Normal distribution)
If $f_0(x) = \phi(x) = (2\pi)^{-1/2}\exp(-x^2/2)$ is the standard normal density, we have $K_0 = I_0 = 1$, and straightforward calculations yield

$$e^B(\omega_p; \phi, shift) = l(\omega_p; \phi, shift) = \frac{C(p)}{2\pi(p+1)^{1/p}}$$

which, as a function of p, is monotonically decreasing (see Fig. 3.1) with the limit $\lim_{p\to\infty} e^B(\omega_p; \phi, shift) = 2/\pi \approx 0.637$.

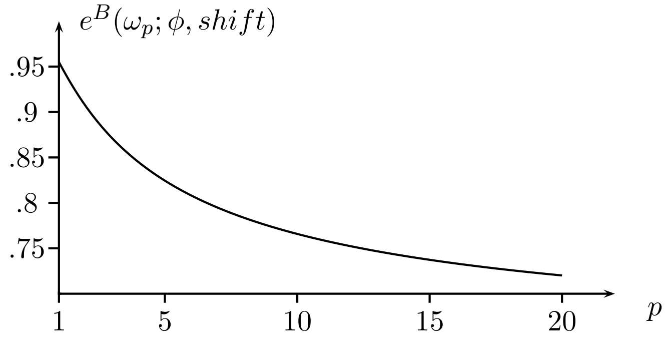


Figure 3.1: Local Bahadur efficiency (Normal distribution: shift alternative)

Example 3.2: (Laplace distribution)
In case of the Laplace distribution with density $f_0(x) = \frac{1}{2}\exp(-|x|)$, $x \in \mathbb{R}$, we again have $K_0 = 1$, and one obtains the slightly similar expression

$$e^B(\omega_p; \text{Laplace}, shift) = l(\omega_p; \text{Laplace}, shift) = \frac{C(p)}{4(p+1)^{2/p}}.$$

In contrast to the case of the normal distribution, the local Bahadur efficiency is an increasing function of p (see Fig. 3.2), and the maximum value 1 is attained in the limit $p \to \infty$, i.e., for the Kolmogorov-Smirnov statistic.

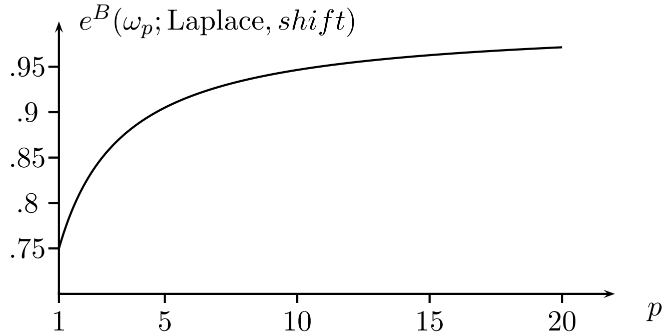


Figure 3.2: Local Bahadur efficiency (Laplace distribution: shift alternative)

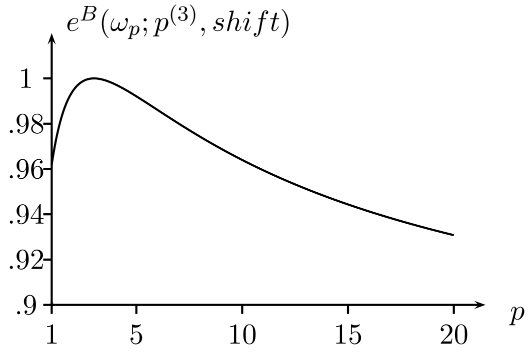
Example 3.3: (Perks' family of distributions)
We now consider the family of densities

$$p^{(s)}(x) = \frac{1}{B(1/2, 1/s)} \cdot \cosh^{-2/s}(x), \quad x \in \mathbb{R},\ s \geq 1,$$

where, generically, $B(a, b) = \Gamma(a)\Gamma(b)/\Gamma(a+b)$ $(a, b > 0)$ denotes the Beta function. This interesting class of densities is a subclass of a family of distributions introduced in [27] (see also [20], p. 143). It includes the logistic and the hyperbolic secant (or Champernowne) distribution, which arise for the special choices $s = 1$ and $s = 2$, respectively. The upper bound for the family of densities $p^{(s)}$ equals $K_0^{(s)} = I_0 = 4/(s(s+2))$.

To obtain the expression (4), one may use formulas 3.512.1 and 8.384.4 of [17]. After some algebraic manipulation, it follows that

$$l(\omega_p; p^{(s)}, shift) = C(p) \cdot \frac{s(s+2)\left(B\left(\frac{p+1}{s}, \frac{p+1}{s}\right)\right)^{2/p}}{(B(1/s, 1/s))^{(2p+2)/p}}.$$



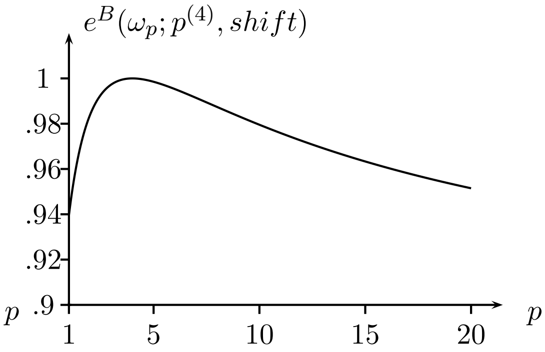


Figure 3.3: Local Bahadur efficiency (Perks' family: shift alternative)

Figure 3.3 shows the values of $l(\omega_p; p^{(s)}, shift)$ as a function of p for the cases $s = 3$ (left) and $s = 4$ (right). The maximum is attained at the points $p = 3$ and $p = 4$, respectively, and from general theory, for any given $s \geq 1$, the maximum value should be attained at $p = s$. This follows from the description of domains of local asymptotic optimality in [26], Ch.6.

Example 3.4: (Cauchy distribution)
For the Cauchy distribution with density $h(x) := (\pi(1+x^2))^{-1},\ x \in \mathbb{R}$, we have $K_0 = 1/2$. So, calculating the expression (4) using formula 3.251.11 of [17], we obtain

$$\left(\int_{-\infty}^{\infty} h^{p+1}(x)dx\right)^{2/p} = \frac{1}{\pi^{\frac{2p+2}{p}}}\left(\frac{\sqrt{\pi}\,\Gamma(p+1/2)}{\Gamma(p+1)}\right)^{2/p}.$$

Hence, it follows that

$$e^B(\omega_p; Cauchy, shift) = \frac{2C(p)}{\pi^{\frac{2p+2}{p}}}\left(\frac{\sqrt{\pi}\Gamma(p+1/2)}{\Gamma(p+1)}\right)^{2/p}.$$

The plot of this efficiency as a function of p is given in Fig. 3.5. Interestingly, $e^B(\omega_p; Cauchy, shift)$ increases for small p and attains its maximum value 0.876 at some point $p_0 \approx 10.2$ that cannot be anticipated from theoretical reasons. For $p > p_0$ the efficiency slowly decreases and, as $p \to \infty$, converges to the value $8/\pi^2 \approx 0.814$, which is the efficiency of the Kolmogorov test (see [26], Table 3, p.80).

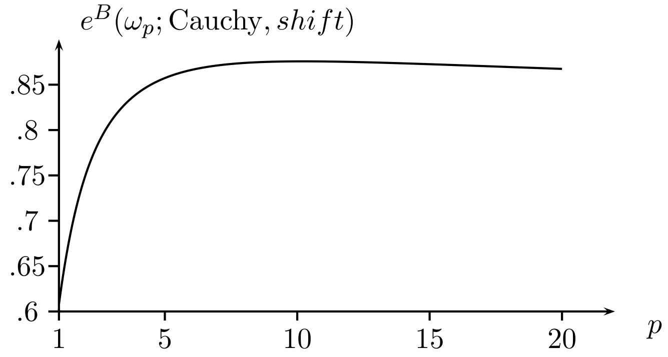


Figure 3.5: Local Bahadur efficiency (Cauchy distribution: shift alternative)

4 Scale and skew alternatives

We now consider scale alternatives of the form $F_\theta(x) = F_0(x\exp(-\theta))$. Expression (2) for the local index then gives

$$l(\omega_p; f_0, scale) = C(p) \cdot \left\{\int_{-\infty}^{\infty} |x|^p f_0(x)^{p+1}dx\right\}^{2/p}.$$

Example 4.1: (Normal distribution, scale alternative)
In the case $f_0(x) = \phi(x) = (2\pi)^{-1/2}\exp(-x^2/2)$ calculations are straightfor-

ward, and we obtain

$$l(\omega_p; \phi, scale) = C(p) \cdot \frac{\Gamma((p+1)/2)^{2/p}}{(\pi(p+1))^{(p+1)/p}}.$$

Since $K_0 = I_0 = 2$, it follows that

$$e^B(\omega_p; \phi, scale) = \frac{C(p)}{2} \cdot \frac{\Gamma((p+1)/2)^{2/p}}{(\pi(p+1))^{(p+1)/p}}.$$

The plot of this efficiency as a function of p is given in Fig. 4.1. Since $\Gamma((p+1)/2)^{2/p} \sim (p+1)/(2e)$ as $p \to \infty$, we have $\lim_{p\to\infty} e^B(\omega_p; \phi, scale) = (\pi e)^{-1} \approx 0.117$.

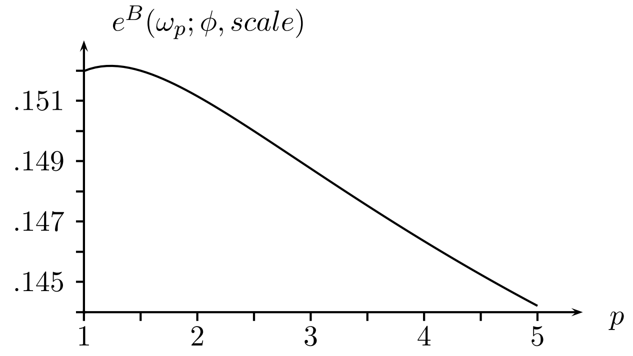


Figure 4.1: Local Bahadur efficiency (Normal distribution: Scale alternative)

Interestingly, $l(\omega_p; f_1, scale)$ is not a monotone function of p. The maximum value is attained at some point $p_0 \approx 1.235$. In fact, the general theory from [26], p. 228-229, forces us to choose scale alternatives of the form $F_0(e^{-\theta} \ln x)$, $x > 0$, if F_0 has some optimal properties as shift alternative. In this way we can multiply the number of examples. The same remark holds for Lehmann alternatives, when we should take the cdf of the form $F_0^{1+\theta}(\ln \ln x)$, $x \geq 1$.

We now pass to skew alternatives in the sense of [4], which are a subject of ongoing interest in the statistical literature (e.g. [18,28,2,3,12,5]).

For any cdf F_0 symmetric around 0 with density f_0, the skewed density $f_\theta(x)$ corresponding to F_0 is defined by

$$f_\theta(x) = 2\, f_0(x)\, F_0(\theta x), \quad x \in \mathbb{R}.$$

In what follows, we assume that f_0 has a finite second moment, denoted by

$$\sigma_0^2 = \int_{-\infty}^{\infty} x^2 f_0(x)\; dx.$$

If F_0 is sufficiently regular (for details see [12]), then

$$F_\theta(x) - F_0(x) \sim 2\theta\, f_0(0) \cdot \int_{-\infty}^{x} u f_0(u)\, du \quad \text{as} \quad \theta \to 0.$$

Invoking formula (2), it follows that

$$l(\omega_p; f_0, skew) = C(p) \cdot 4 f_0^2(0) \cdot \left(\int_{-\infty}^{\infty} \left| \int_{-\infty}^{x} u f_0(u)\, du \right|^p f_0(x)\, dx \right)^{2/p}.$$

Notice that in this case the upper bound K_0 discussed above is equal to $4\, f_0^2(0)\, \sigma_0^2$ (see [12]). Hence the local efficiency is

$$e^B(\omega_p; f_0, skew) = \frac{C(p)}{\sigma_0^2} \cdot \left(\int_{-\infty}^{\infty} \left| \int_{-\infty}^{x} u f_0(u)\, du \right|^p f_0(x)\, dx \right)^{2/p}.$$

Example 4.2: (Uniform distribution, skew alternative)
If $f_0(x) = \mathcal{U}(x) = 2^{-1}\, \mathbf{1}_{[-1,1]}(x)$ is the density of the uniform distribution on $[-1, 1]$, straightforward algebraic manipulations (using again formula 8.384.4 of [17]) give the extremely simple formula

$$l(\omega_p; \mathcal{U}, skew) = C(p) \cdot (B(p+1, p+1))^{2/p}.$$

Since the upper bound K_0 is 1/3, we obtain

$$e^B(\omega_p; \mathcal{U}, skew) = 3\, C(p) \cdot (B(p+1, p+1))^{2/p},$$

which yields the decreasing plot in Fig. 4.2. Since $\lim_{p\to\infty} B(p+1, p+1)^{2/p} = 1/16$, we have $\lim_{p\to\infty} e^B(\omega_p; \mathcal{U}, skew) = 0.75$.

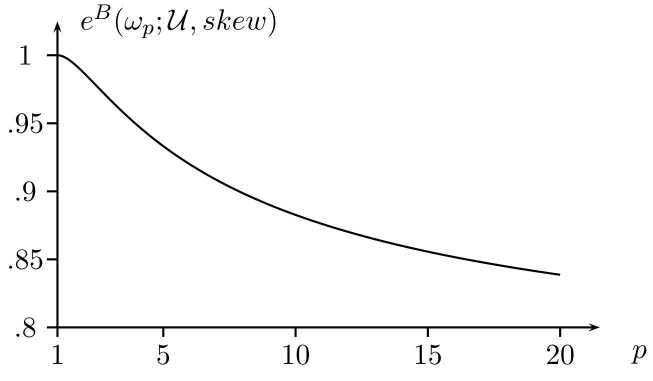


Figure 4.2: Local Bahadur efficiency (Uniform distribution: skew alternative)

Example 4.3: (Symmetrized arcsine-type density, skew alternative)
We now consider the more refined density $f_0(x) = g^{(s)}(x)$, where

$$g^{(s)}(x) = \left(B(1/2, 1/s) \cdot (1-x^2)^{\frac{s-1}{s}}\right)^{-1} \cdot \mathbf{1}_{\{-1,1\}}(x), \quad s \geq 1.$$

Observe that $g^{(s)}$ is the density of $U\sqrt{Y}$, where U and Y are independent, $P(U = 1) = P(U = -1) = 1/2$, and Y has a Beta distribution with parameters $1/2$ and $1/s$. For the special case $s = 2$, Y has the well known arcsine distribution. The calculations are somewhat involved, but the identity

$$\int_{-1}^{x} \frac{u\,du}{(1-u^2)^{(s-1)/s}} = \frac{s}{2} \cdot (1-x^2)^{1/s}, \quad -1 \leq x \leq 1,$$

simplifies matters considerably, and we finally obtain

$$l(\omega_p; g^{(s)}, skew) = C(p) \cdot 4 f_0^2(0) \cdot \frac{s^2}{4\,B^2(\frac{1}{s}, \frac{1}{2})} \cdot \left(B\left(\frac{p+1}{s}, \frac{1}{2}\right) \Big/ B\left(\frac{1}{s}, \frac{1}{2}\right)\right)^{2/p}.$$

Since $\sigma_0^2 = s/(s+2)$, we thus get

$$e^B(\omega_p; g^{(s)}, skew) = \frac{s(s+2)C(p)}{4\,B^2(\frac{1}{s}, \frac{1}{2})} \cdot \left(B\left(\frac{p+1}{s}, \frac{1}{2}\right) \Big/ B\left(\frac{1}{s}, \frac{1}{2}\right)\right)^{2/p}.$$

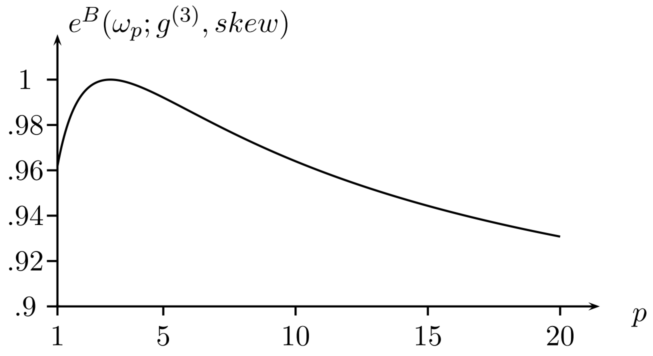


Figure 4.3: Local Bahadur efficiency (Symmetrized arcsine-type density: skew alternative)

It follows from the local optimality theory of integral statistics under skew alternatives developed in [12] that this efficiency should attain its maximum value at $p = s$. The limit value for $p \to \infty$ is $s(s+2)/B^2(1/s, 1/2)$. Figure 4.3 illustrates this situation for the case $s = 3$. We clearly see that the maximum local efficiency is attained at $p = 3$.

5 Testing for Independence

Similar reasoning can be applied to testing for independence. Suppose we observe a sample of i.i.d. two-dimensional observations $(X_1, Y_1), ..., (X_n, Y_n)$

with continuous joint cdf $F(x, y)$ and marginal cdf's $G(x)$ and $H(y)$. Denote by $F_n(x) = n^{-1} \sum_{j=1}^{n} \mathbf{1}\{X_j \le x, Y_j \le y\}$, $G_n(x) = n^{-1} \sum_{j=1}^{n} \mathbf{1}\{X_j \le x\}$ and $H_n(y) = n^{-1} \sum_{j=1}^{n} \mathbf{1}\{Y_j \le y\}$ the corresponding empirical cdf's.

The problem is to test the hypothesis

$$H : F(x, y) = G(x)H(y) \quad \text{for all} \;\; x, \; y$$

of independence. Besides many procedures based on correlation coefficients, there is the classical statistic of omega-square type proposed in [8], namely

$$\Omega_n^2 = n \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} (F_n(x, y) - G_n(x)H_n(y))^2 \, dG_n(x) \, dH_n(y).$$

It was proved in [8] that Ω_n^2 is distribution-free under H and converges weakly to $\int_0^1 \mathbb{B}^2(x, y) \, dxdy$, where $\mathbb{B}(x, y)$ is the Brownian pillow, a centered Gaussian field on the unit square with covariance kernel

$$E \, \mathbb{B}(x, y)\mathbb{B}(s, t) = (\min(x, s) - xs)(\min(y, t) - yt).$$

Generalizations of Ω_n^2, namely the weighted statistics Ω_{n,q_1,q_2}^k, where k is an integer and q_1, q_2 are appropriate weights, were studied in [26].

In this paper we will consider the following statistics of L_p-type for $1 \le p < \infty$:

$$\Omega_{n,p} = \sqrt{n} \left\{ \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} |F_n(x, y) - G_n(x)H_n(y)|^p \, dG_n(x) \, dH_n(y) \right\}^{1/p}.$$

Notice that

$$\Omega_{n,p} = \sqrt{n} \left\{ \frac{1}{n^2} \sum_{i=1}^{n} \sum_{j=1}^{n} \left| F_n(X_{(i)}, Y_{(j)}) - \frac{i}{n} \cdot \frac{j}{n} \right|^p \right\}^{1/p},$$

where $X_{(1)} \le X_{(2)} \le \ldots \le X_{(n)}$ and $Y_{(1)} \le Y_{(2)} \le \ldots \le Y_{(n)}$ are the order statistics of $X_1, \ldots, X_n$ and $Y_1, \ldots, Y_n$, respectively. Moreover, we have

$$F_n(X_{(i)}, Y_{(j)}) = \frac{1}{n} \cdot \sum_{k=1}^{i} \mathbf{1}\{Y_{[k]} \le Y_{(j)}\},$$

where $Y_{[1]}, \ldots, Y_{[n]}$ are the concomitants of $X_{(1)}, \ldots, X_{(n)}$, i.e., $Y_{[k]} = Y_j$ if $X_j = X_{(k)}$. It follows that an algorithm for implementing a test for independence based on $\Omega_{n,p}$ is readily available.

The limiting null distribution of $\Omega_{n,p}$ as $n \to \infty$ is the distribution of

$$\|\mathbb{B}\|_p := \left\{ \int_0^1 \int_0^1 |\mathbb{B}(x, y)|^p \, dx \, dy \right\}^{1/p}.$$

The tail behavior of certain Gaussian fields (including $\|\mathbb{B}\|_p$) was obtained in [24], where it was shown that

$$\lim_{x\to\infty} \frac{2}{x^2} \cdot \ln P(\|\mathbb{B}\|_p > x) \;=\; -C^2(p),$$

with $C(p)$ defined in (1).

As for the alternative, we assume that

$$F(x,y) = G(x)H(y) + \theta \cdot L(F(x), G(y)), \tag{5}$$

where the so-called dependence function $L : [0,1]^2 \to \mathbb{R}$ is zero on the boundary of $[0,1]^2$. Such a model is well known (see [26], Ch. 5).

Arguing as above, we obtain the approximate Bahadur efficiency for $\Omega_{n,p}$ and the corresponding local index, namely

$$l(\Omega_p; L) = C^2(p) \cdot \left\{ \int_0^1 \int_0^1 |L(s,t)|^p \, ds \, dt \right\}^{2/p}.$$

As an example, we consider the special dependency function $L_1(x,y) = \sin \pi x \cdot \sin \pi y$ which, according to the general theory developed in [26], Ch.6, should be locally optimal for the $\Omega_{n,2}$-test. The calculations are straightforward, and one obtains

$$l(\Omega_p; L_1) = C^2(p) \left\{ \int_0^1 |\sin \pi t|^p dt \right\}^{4/p} = C^2(p)(2^p/\pi)^{4/p} B\left(\frac{p+1}{2}, \frac{p+1}{2}\right)^{4/p}.$$

Since the theoretical upper bound here is $\pi^4/4$, the local efficiency takes the form

$$e^B(\Omega_p; L_1) = \frac{64\, C^2(p)}{\pi^{4(1+1/p)}} \cdot B\left(\frac{p+1}{2}, \frac{p+1}{2}\right)^{4/p}.$$

The plot is given in Fig. 5.1, and we see that the maximum efficiency 1 is attained at $p = 2$.

We finish this section by pointing out the possibility of considering *anisotropic* norms in the space L_{p_1,p_2} and the corresponding statistics (possibly weighted)

$$\Omega_{n,p_1,p_2} = \sqrt{n} \left\{ \int_{-\infty}^{\infty} \left(\int_{-\infty}^{\infty} |F_n(x,y) - G_n(x)H_n(y)|^{p_1} \, dG_n(x) \right)^{p_2/p_1} dH_n(y) \right\}^{1/p_2}.$$

The theory of such statistics has not yet been established. It is clear, however,

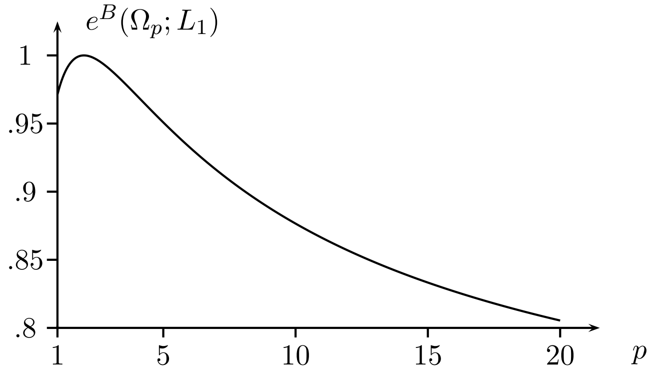


Figure 5.1: Local Bahadur efficiency (testing for independence)

that Ω_{n,p_1,p_2} converges in distribution to

$$\left\{ \int_0^1 \left(\int_0^1 |\mathbb{B}(x,y)|^{p_1}\, dx \right)^{p_2/p_1} dy \right\}^{1/p_2}$$

if the hypothesis H of independence is true. It follows from [22] that the tails of such anisotropic norms decrease at the rate $C(p_1)C(p_2)$. This fact enables us to calculate the approximate Bahadur efficiency. Apparently, the results will be the combination of formulas obtained in the one-dimensional case. For instance, we may predict that the statistic $\Omega_{n,1,2}$ will be locally optimal for the alternatives of type (5) with the dependence function $L(x,y) = x(1-x)\sin \pi y$.

6 Simulations

In this section we present the results of a Monte Carlo study that corroborate the theoretical findings of Sections 3 and 4. The simulations involve random samples of size $n = 100$, drawn from the underlying distribution of the respective example. Critical values of the finite sample null distribution of $\omega_{n,p}$ for each $p \in \{1, 1.5, \ldots, 20\}$ are based on 100000 replications. The nominal level of significance is 1%, and we use 100000 replications throughout our study.

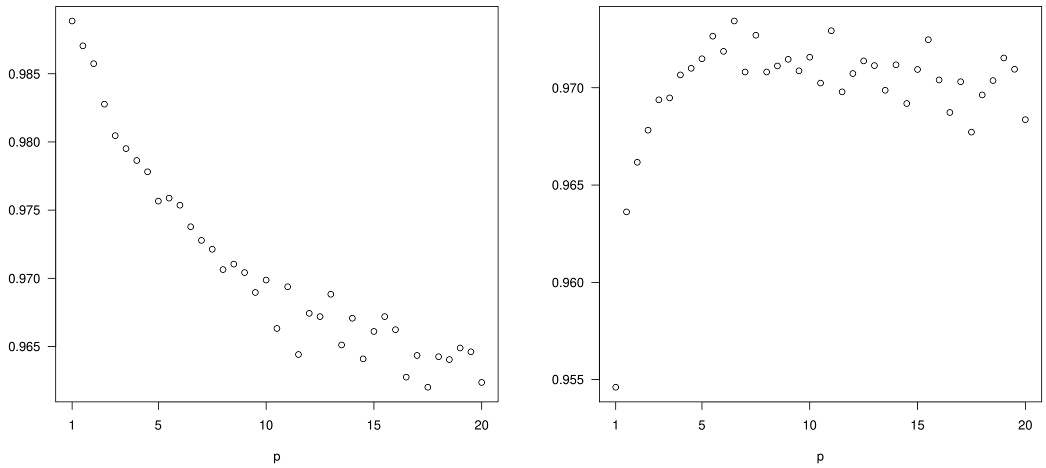


Figure 6.1: Percentages of rejection out of 100000 samples of normal (left) and Laplace (right) shift alternatives

In view of Examples 3.1 and 3.2 we consider a shift alternative with shift parameter $\theta = 0.5$. For the Perks' family alternatives in Example 3.3 we use $\theta = 0.75$. Figure 6.1 shows relative frequencies of rejections of the null hypothesis based on $\omega_{n,p}$ for each $p \in \{1, 1.5, \ldots, 20\}$ against shift alternatives for the normal distribution (left) and the Laplace distribution (right). The plots roughly resemble the curves of local Bahadur efficiencies given in Figures 3.1 and 3.2.

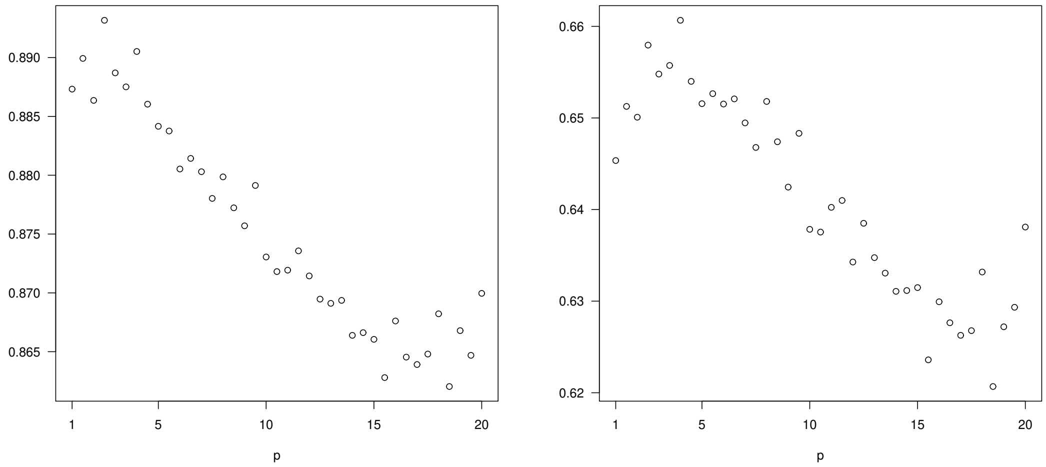


Figure 6.2: Percentages of rejection out of 100000 samples for shift alternatives from Perks' family of distributions, where $s = 3$ (left) and $s = 4$ (right)

Figure 6.2 shows the percentages of rejection against shift alternatives from Perks' family, where the left hand plot corresponds to $s = 3$ and the right hand plot to $s = 4$, cf. Figure 3.3. Once again, the plots roughly resemble the curves of local Bahadur efficiencies. Just like in Figure 6.1 the observed dispersion in the performances is within the width of standard confidence intervals for unknown probabilities. Finally, Figure 6.3 shows the percentages of rejection against shift alternatives from the Cauchy distribution. Also in this case, the results corroborate the theoretical findings.

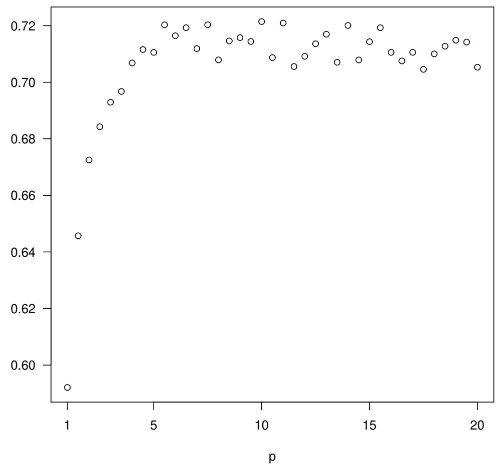


Figure 6.3: Percentages of rejection out of 100000 samples for shift alternatives from the Cauchy distribution

As regards Examples 4.1 and 4.2 we chose a scaling- and skewing factor of both $\theta = 0.75$. The resulting plots of estimated power, as shown in Figure 6.4, also confirm the theoretical results, cf. Figure 4.1 and Figure 4.2.

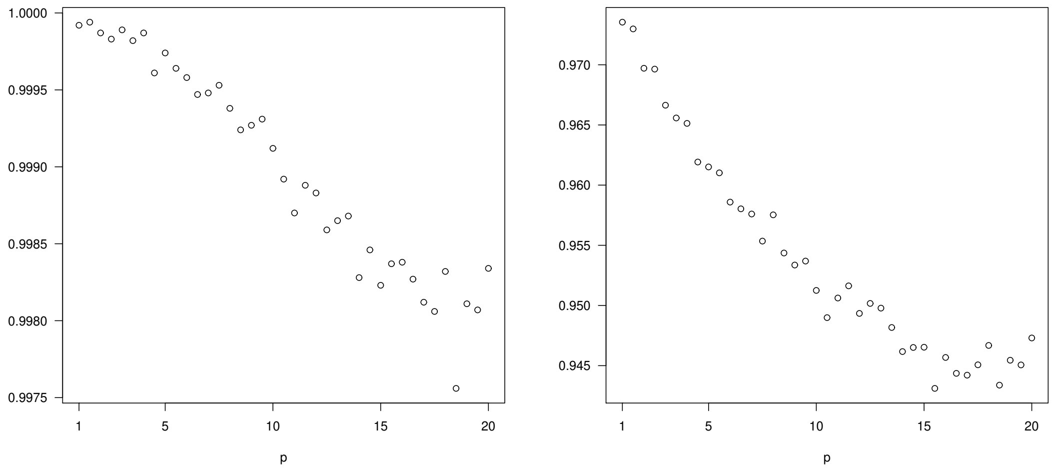


Figure 6.4: Percentages of rejection out of 100000 samples for normal scale alternatives (left) and uniform skew alternatives (right).

Corresponding to Example 4.3, Figure 6.5 shows the plot of estimated power for symmetrized arcsine type alternatives with a skewing parameter of $\theta = 0.95$. Once again, the empirical results are in accordance with the theoretical findings.

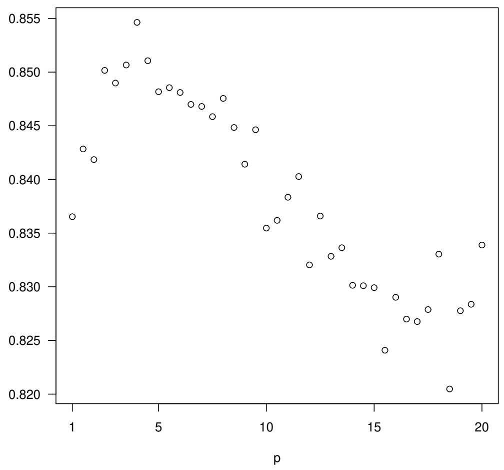


Figure 6.5: Percentages of rejection out of 100000 samples for symmetrized arcsine type skew alternatives.

7 Conclusions

Several examples illustrate that more flexibility with regard to power against specific alternatives may be gained by considering goodness-of-fit tests based on L_p-norms other than $p = 2$ and $p = \infty$. Of course, no universally best value of p is to be expected even under a special setting of alternatives, such as the shift model. This salient feature in nonparametric statistics is due to the fact that the class of alternative distributions is infinite-dimensional. *The results, however, clearly show that, given a distribution function F_0 and a specific alternative, one can draw the plot of efficiency as a function of p and determine the value of p giving maximum efficiency.* More generally, tests based on *weighted* L_p-norms might give additional sensitivity with respect to certain alternatives.

In principle, goodness-of-fit testing based on the L_p-norm is also possible in the context of testing for parametric families of distributions. For example, if $X_1, \ldots, X_n$ is an i.i.d. sequence of random variables with continuous distribution function F, and the problem is to test the hypothesis H_0 that the distribution of X_1 is some unspecified (nondegenerate) normal distribution, standard goodness-of-fit tests are based on the estimated empirical process

$\mathcal{Y}_n(t) = \sqrt{n}(\hat{G}_n(t) - t)$, $0 \le t \le 1$, where

$$\hat{G}_n(t) = \frac{1}{n} \sum_{j=1}^{n} \mathbf{1}\{Y_j \le t\},$$

and

$$Y_j = \Phi\left(\frac{X_j - \bar{X}_n}{S_n}\right), \quad j = 1, \dots, n.$$

Here, Φ denotes the cdf of the standard normal law, and $\bar{X}_n = n^{-1} \sum_{j=1}^n X_j$, $S_n^2 = (n-1)^{-1} \sum_{j=1}^n (X_j - \bar{X}_n)^2$. Generalizing the tests of Kolmogorov-Smirnov and Cramér-von Mises which are based on $\int_0^1 \mathcal{Y}_n^2(t)dt$ and $\sup_{0 \le t \le 1} |\mathcal{Y}_n(t)|$, respectively (see e.g. [34]), one could more generally consider statistics of the type

$$T_{n,p} = \|\mathcal{Y}_n\|^p = \left(\int_0^1 |\mathcal{Y}_n(t)|^p \, dt\right)^{1/p}, \quad 1 \le p < \infty.$$

Both the finite sample and the limit null distribution of $T_{n,p}$ as $n \to \infty$ do not depend on the parameters of the underlying normal law, and the limit null distribution of $T_{n,p}$ is that of $\|\mathcal{Y}\|^p$, where $\mathcal{Y}$ is a centered Gaussian process on $[0, 1]$ having continuous sample paths and covariance kernel

$$K(s,t) = \min(s,t) - st - \frac{1}{2\pi}\left(1 + \frac{J(s)J(t)}{2}\right)\exp\left(-\frac{J^2(s) + J^2(t)}{2}\right),$$

and $J(s) = \Phi^{-1}(s)$ (see e.g. [11]).

The paper [21] contains a study of the relative efficiencies of certain weighted quadratic goodness-of-fit statistics (including the Cramér-von Mises statistic $\omega_{n,2}$) in the context of testing for normality. In a more general setting, [36] contains an investigation of efficiencies of weighted quadratic statistics (including $\omega_{n,2}$) under simple and composite null hypotheses.

Acknowledgements: The authors would like to thank two anonymous referees and Mathew Penrose for valuable comments and suggestions that improved the quality of the paper. The work of Yakov Nikitin was supported by RFBR grant No. 07-01-00159 and grant NSh 638.2008.1.

References

[01] Ahmad, I., 1997. Goodness-of-fit statistics based on weighted L_p-functionals. Stat. & Prob. Lett. 35, 261–268.

[02] Arellano–Valle, R.B., Del Pino, G., and San Martin, E., 2002. Definition and probabilistic properties of skew-distributions. Stat. & Prob. Lett. 58, 111–121.

[03] Arnold, B.C., and Beaver, R., 2002. Skewed multivariate models related to hidden truncation and/or selective reporting. Test, 11, 7–54.

[04] Azzalini, A., 1985. A class of distributions which includes the normal one. Scand. Journ. Stat. 12, 171–178.

[05] Azzalini, A., 2005. The skew-normal distribution and related multivariate families. Scand. Journ. Stat., 32, 159–188.

[06] Bahadur, R.R., 1960. Stochastic comparison of tests. Ann. Math. Stat. 31, 231–260.

[07] Berkes, I., Horvath, L., Shao, Qi-Man, and Steinebach, J., 2000. Strong laws for L_p-norms of empirical and related processes. Period. Math. Hung. 41, 35–69.

[08] Blum, J.R., Kiefer, J., and Rosenblatt, M., 1961. Distribution free tests of independence based on the sample distribution function. Ann. Math. Stat. 32, 485–498.

[09] Csörgő, M., and Horvath, L., 1988. On the distribution of L_p-norms of weighted uniform empirical and quantile processes. Ann. Probab. 16, 142–161.

[10] Csörgő, M., and Horvath, L., 1993. Weighted approximations in Probability and Statistics. Wiley, New York.

[11] Durbin, J., 1973. Weak convergence of the sample distribution function when parameters are estimated. Ann. Stat. 1, 279–290.

[12] Durio, A., and Nikitin, Ya., 2003. Local asymptotic efficiency of some goodness-of-fit tests under skew alternatives. Journ. Stat. Plann. Infer. 115, 171–179.

[13] Durot, C., 2007. On the L_p-error of monotonicity constrained estimators. Ann. Stat. 35, No. 3, 1080–1104.

[14] Fatalov, V.R., 1999. The Laplace method for computing exact asymptotics of distributions of integral statistics. Math. Meth. of Stat. 8, 510–535.

[15] Fatalov, V.R., 1999. Laplace method for probabilities of moderate deviations for ω_n^p-statistics, $p \geq 2$. In: Intern. Seminar on Stability for Stochastic Models, Abstracts, Lublin, 50–51.

[16] Gajek, L., Kaluszka, M., and Lenic, A., 1996. The law of the iterated logarithm for L_p-norms of empirical processes. Stat. & Prob. Lett. 28, 107–110.

[17] Gradshteyn, I.S., and Ryzhik, I.M., 1980. Table of Integrals, Series, and Products. Academic Press, New York.

[18] Henze, N., 1986. A probabilistic representation of the "skew-normal" distribution. Scand. Journ. Stat. 13, 271–275.

[19] Ivanov, V., and Zrelov, P., 1997. Nonparametric Integral Statistics $\omega_n^k = n^{k/2} \int_{-\infty}^{\infty} [S_n(x) - F(x)]^k dF(x)$: Main Properties and Applications. Comp. Math. Appl. 34, 703–726.

[20] Johnson, N.L., Kotz, S., and Balakrishnan, N., 1995. Continuous univariate distributions, Vol. 2. Wiley & Sons, New York.

[21] Koziol, J.A., 1986. Relative efficiencies of goodness of fit procedures for assessing univariate normality. Ann. Inst. Stat. Math. 38, 485–493.

[22] Lifshits, M., Nazarov, A., and Nikitin, Ya., 2003. Tail behavior of anisotropic norms for Gaussian random fields. C.R. Acad. Sci. Paris, Ser. I 336, 85–88.

[23] Mojirsheibani, M., 2007. Some approximations to L_p-statistics of kernel density estimators. Statistics 41, No. 3, 203–220.

[24] Nazarov, A.I., and Nikitin, Ya., 2000. Some extremal problems for Gaussian and empirical random fields (Russian). Proceedings of St. Petersburg Math. Society, 2000, v.8, p.214–230 . English Transl. in Amer. Math. Soc. Transl. Ser.2, ed. by N.N. Uraltseva, v.205, 189–202 (2002).

[25] Nikitin, Ya., 1984. Large deviations and the asymptotic efficiency of statistics of the integral type II. Journ. Sov. Math. 24, 595–603.

[26] Nikitin, Ya., 1995. Asymptotic efficiency of nonparametric tests. Cambridge University Press.

[27] Perks, W.F., 1932. On some experiments in the graduation of mortality statistics. Journ. of the Inst. of Actuaries 58,12–57.

[28] Pewsey, A., 2000. Problems of inference for Azzalini's skew-normal distribution. Journ. of Appl.Stat. 27, 859–870.

[29] Podkorytova, O.A., 1997. Rough large-deviation probability asymptotics of some functionals of the norm kind. Journ. Math. Sci., N.Y. 99, 1161–1172.

[30] Schmid, F., and Trede, M., 1995. A distribution-free test for the two sample problem for general alternatives. Comput. Stat. Data Anal., 20, 409–419.

[31] Schmid, F., and Trede, M., 1996. An L_1-variant of the Cramér-von Mises test. Stat. & Prob. Lett. 26, 91–96.

[32] Shorack, G. R. and Wellner, J. A., 1986. Empirical Processes with Applications to Statistics. Wiley, New York.

[33] Stam, A., 1997. Nontraditional approaches to statistical classification: Some perspectives on L_p-norm methods. Ann. Oper. Res. 74, 1-36.

[34] Stephens, M.A., 1974. EDF-statistics for goodness of fit and some comparisons. Journ. Amer. Stat. Ass. 69, 730–737.

[35] Strassen, V., 1964. An invariance principle for the law of the iterated logarithm. Zeitschr. Wahrscheinlichkeitsth. Verw. Geb. 3, 211–226.

[36] Wells, M.T., 1990. The relative efficiency of goodness-of-fit statistics in the simple and composite hypothesis testing problem. J. Amer. Stat. Assoc. 85, 459–463.

[37] Wieand, H.S., 1976. A condition under which the Pitman and Bahadur approaches to efficiency coincide. Ann. Stat. 4, 1003–1011.

[38] Zrelov, P.V., and Ivanov, V.V., 1991. The relativistic charged particles identification method based on the goodness-of-fit ω_n^3-criterion. Nucl. Instr. and Meth. in Phys. Res. A310, 623–630.

[39] Zrelov, P.V., and Ivanov, V.V., 1992. Test statistic ω_n^3 and its main properties. Trans. of the 11^{th} Prague Conference on Information Theory, Stat. Decision Functions, Random Processes. Prague, Aug. 27–31, 1990, 487–498.