On the Bahadur-Efficient Testing of Uniformity by Means of the Entropy

Peter Harremoës, Member, IEEE, and Igor Vajda, Fellow, IEEE

Abstract—This paper compares the power divergence statistics of orders $\alpha > 1$ with the information divergence statistic in the problem of testing the uniformity of a distribution. In this problem, the information divergence statistic is equivalent to the entropy statistic. Extending some previously established results about information diagrams, it is proved that the information divergence statistic in this problem is more efficient in the Bahadur sense than any power divergence statistic of order $\alpha > 1$. This means that the entropy provides in this sense the most efficient way of characterizing the uniformity of a distribution.

Index Terms—Bahadur efficiency, entropy, goodness-of-fit, index of coincidence, information diagram, power divergences.

I. POWER DIVERGENCE STATISTICS

Let $M(k)$ denote the set of all discrete probability distributions of the form $P = (p_1, \ldots, p_k)$ and

$$M(k) = \{P \in M(k) : nP \in \{0, 1, \ldots\}\} \tag{1}$$

the subset of types. One of the fundamental problems of mathematical statistics can be described by $n$ balls distributed into boxes $1, \ldots, k$ independently according to an unknown probability law

$$P_n = (p_{n1}, \ldots, p_{nk}) \in M(k) \tag{2}$$

possibly depending on the number of balls $n$. This results in frequency counts $X_{n1}, \ldots, X_{nk}$ the vector of which $X_n = (X_{n1}, \ldots, X_{nk}) \in \{0, 1, \ldots\}^k$ has a multinomial distribution with parameters $k, n, P_n$.

$$X_n \sim \text{Multinomial}_k(n, P_n). \tag{3}$$

The problem is to decide on the basis of observations $X_n$ whether the unknown law (2) is equal to a given $Q = (q_1, \ldots, q_k) \in M(k)$ or not.

The observations $X_n$ are represented by the (random) empirical probability distribution

$$\hat{P}_n = (\hat{p}_{n1} \triangleq X_{n1}/n, \ldots, \hat{p}_{nk} \triangleq X_{nk}/n) \in M(k|n) \tag{4}$$

and the hypothesis $Q$ about $P_n$ is usually decided by means of a procedure $T$ called a test. This procedure uses a statistic $T_n(\hat{P}_n, Q)$ which characterizes the goodness-of-fit between the distributions $P_n$ and $Q$. The test $T$ rejects the hypothesis $P_n = Q$ if $T = T_n(\hat{P}_n, Q)$ exceeds a certain rejection level $r_n \in \mathbb{R}$.

The goodness-of-fit statistic is usually one of the power divergence statistic

$$T_n = T_{\alpha,n} = 2nD_{\alpha}(\hat{P}_n, Q), \quad \alpha \in \mathbb{R} \tag{5}$$

where $D_{\alpha}(P, Q)$ denotes the so-called $\alpha$-divergence (power divergence of order $\alpha$) of distributions $P, Q \in M(k)$ defined by

$$D_{\alpha}(P, Q) = \sum_{j=1}^{k} q_j \phi_{\alpha} \left( \frac{p_j}{q_j} \right), \quad \alpha \in \mathbb{R} \tag{6}$$

for the power function $\phi_{\alpha}$ of order $\alpha \in \mathbb{R}$ given in the domain $t > 0$ by the formula

$$\phi_{\alpha}(t) = \frac{t^\alpha - \alpha(t - 1) - 1}{\alpha(t - 1)}, \quad \text{when } \alpha(t - 1) \neq 0 \tag{7}$$

and by the corresponding limits

$$\phi_{0}(t) = -\ln t + t - 1 \tag{8}$$

$$\phi_{1}(t) = t \ln t - t + 1. \tag{9}$$

For details about the definition (6) and properties of power divergences, see [1] or [2]. Next we cite the best known members of the family of statistics (5) with a reference to the skew symmetry $D_{\alpha}(P, Q) = D_{1-\alpha}(Q, P)$ of the power divergences (6).

Example 1: The quadratic divergences

$$D_2(P, Q) = D_{-1}(Q, P) = \frac{1}{2} \sum_{j=1}^{k} \left( \frac{p_j - q_j}{q_j} \right)^2$$

lead to the well-known Pearson statistic

$$T_2 = T_{2,n} = \sum_{j=1}^{k} \frac{(X_{nj} - nq_j)^2}{nq_j} \tag{10}$$

and Neyman statistics

$$T_{-1} = T_{-1,n} = \sum_{j=1}^{k} \frac{(X_{nj} - nq_j)^2}{X_{nj}}. \tag{11}$$

The logarithmic divergences

$$D_1(P, Q) = D_0(Q, P) = \sum_{j=1}^{k} \frac{p_j \ln p_j}{q_j}$$

are the so-called relative entropies.

Manuscript received April 20, 2007; revised September 21, 2007. The work of P. Harremoës was supported by grants from the Willum Kann Rasmussen Foundation and Danish Natural Science Research Council. The work of I. Vajda was supported by the GAČR under Grant 102/07/1131 and the MSMT under Grant 1M0572. The material in this paper was presented at the IEEE International Symposium on Information Theory (ISIT’07), Nice, France, June 2007.

P. Harremoës is with the Centrum voor Wiskunde en Informatica, 1090 GB Amsterdam, The Netherlands (e-mail: P.Harremoes@cwi.nl).

I. Vajda is with the Institute of Information Theory and Automation, Academy of Sciences of the Czech Republic, 18208 Prague, Czech Republic (e-mail: vajda@utia.cas.cz).

Communicated by K. Kobayashi, Associate Editor for Shannon Theory.

Digital Object Identifier 10.1109/TIT.2007.911155

0018-9448/$25.00 © 2008 IEEE
lead to the log-likelihood ratio statistic
\[ T_1 = T_{1,n} = 2 \sum_{j=1}^{k} X_{nj} \ln \frac{n_{nj}}{n q_j} \quad (12) \]
and reversed log-likelihood ratio statistic
\[ T_0 = T_{0,n} = 2 n q_j \sum_{j=1}^{k} \ln \frac{n_{nj} q_j}{X_{nj}}. \]
The symmetric Hellinger divergence
\[ D_{1/2}(P,Q) = D_{1/2}(Q,P) = 4 \sum_{j=1}^{k} \left( \sqrt{p_j} - \sqrt{q_j} \right)^2 \]
leads to the Freeman–Tukey statistic
\[ T_{1/2} = T_{1/2,n} = 2 \sum_{j=1}^{k} \left( \sqrt{X_{nj}} - \sqrt{n q_j} \right)^2. \quad (13) \]

In this paper, we would like to find the power divergence statistic \( T_{\alpha}, \alpha \in \mathbb{R} \) that is most suitable for testing the hypothesis that the true distribution \( P_n \) is uniform, i.e., the hypothesis \( \mathcal{H} : P_n = U = (1/k, \ldots, 1/k) \). Hence, in our model
\[ X_n \sim \text{Multinomial}_k(n,U) \quad \text{under } \mathcal{H}. \quad (14) \]
The alternative to the hypothesis \( \mathcal{H} \) is denoted by \( A_n \). Thus, by (3)
\[ X_n \sim \text{Multinomial}_k(n,P_n) \quad \text{under } A_n \quad (15) \]
for \( P_n \) assumed in (2).

Next follows a typical example of the hypotheses testing model introduced in (14)–(15).

**Example 2:** Let \( \mu, \nu \) be two different probability measures on the Borel line \( (\mathbb{R}, \mathcal{B}) \) with absolutely continuous distribution functions \( F, G \) and \( Y_1, \ldots, Y_n \) an independent and identically distributed (i.i.d.) sample from the probability space \( (\mathbb{R}, \mathcal{B}, \mu) \). Consider a statistician who knows neither the probability measure \( \mu \) governing the random sample \( Y_1, \ldots, Y_n \) nor this sample itself. Nevertheless, he observes the frequencies \( X_n = (X_{n1}, \ldots, X_{nk}) \) of the samples \( Y_1, \ldots, Y_n \) in an interval partition \( \mathcal{P}_n = \{A_{n1}, \ldots, A_{nk}\} \) of \( \mathbb{R} \) chosen by him. Using \( X_n \), he has to decide about the hypothesis \( \mathcal{H} \) that the unknown probability measure on \( (\mathbb{R}, \mathcal{B}) \) is the given \( \nu \). Thus, for a partition \( \mathcal{P}_n = \{A_{n1}, \ldots, A_{nk}\} \) under his control he obtains the observations
\[ X_n \sim \text{Multinomial}_k(n,P_n) \quad (16) \]
where
\[ P_n = (\mu(A_{n1}), \ldots, \mu(A_{nk})) \]
and his task is to test the hypothesis \( \mathcal{H} : \mu = \nu \). Knowing \( \nu \), he can use the quantile function \( G^{-1} \) of \( \nu \) or, more precisely, the quantiles \( G^{-1}(j/k) \) of the orders \( j/k \) for \( 1 \leq j \leq k \) cutting \( \mathbb{R} \) into a special system of intervals \( \mathcal{P}_n = \{A_{n1}, \ldots, A_{nk}\} \) with the property \( \nu(A_{nj}) = 1/k \) for \( 1 \leq j \leq k \). Hence, for this special partition we get from (16)
\[ P_n = U = (1/k, \ldots, 1/k) \in M(k|n) \text{ under } \mathcal{H} \quad (17) \]
and
\[ P_n = (\mu(A_{n1}), \ldots, \mu(A_{nk})) \in M(k) \text{ under } A_n. \quad (18) \]
We see from (16)–(18) that the quantile-generated partitions \( \mathcal{P}_n \) lead exactly to the situation assumed in (14)–(15). This idea is illustrated in Fig. 1.

The formulas for divergences \( D_{\alpha}(P,Q) \) simplify when \( Q = U \), e.g.,
\[ D_1(P,U) = \ln k - H(P), \quad \text{for } P \in M(k) \quad (19) \]
where \( H(P) \) denotes the Shannon entropy
\[ H(P) = - \sum_{j=1}^{k} p_j \ln p_j. \]
Similarly, (6) and (7) imply for all \( \alpha > 1 \) and \( P \in M(k) \)
\[ D_\alpha(P,U) = \frac{\sum_{j=1}^{k} p_j^{\alpha}(1/k)^{\alpha - 1}}{\alpha(\alpha - 1)} \]
\[ = \frac{1}{\alpha - 1} \sum_{j=1}^{k} p_j^{\alpha - 1} - 1 \]
\[ = \frac{1}{\alpha - 1} IC_\alpha(P) - 1. \quad (20) \]
Here
\[ IC_\alpha(P) = \sum_{j=1}^{k} p_j^{\alpha} \quad \text{for } P \in M(k) \quad (21) \]
is the index of coincidence of \( P \) of order \( \alpha > 1 \) introduced in [3], taking on values between \( k!^{-\alpha} \) (when \( P \) is the uniform...
distribution \( U \) and 1 (when \( P \) is the Dirac distribution \( P_j \) with \( p_j = 1 \) for some \( 1 \leq j \leq k \)).

From (19), we see that the log-likelihood ratio statistics statistic \( T_{1,n} = 2nD_1(P_n, Q) \) is one–one related to the entropy statistic \( 2nH(P_n) \), and from (20) we see that for each \( \alpha > 1 \) the power divergence statistic \( T_{\alpha,n} = 2nD_\alpha(P_n, Q) \) is one–one related to the corresponding IC-statistic \( 2nIC_\alpha(P_n) \). The entropy \( H(P_n) \) as well as the indices of coincidence IC_\alpha(P_n), \( \alpha > 1 \) characterize the uniformity of the distribution \( P_n \). We are interested in the characterization which is most efficient from the statistical point of view.

The rest of the paper is organized as follows. In Section II, the basic idea behind Bahadur efficiency is explained and previous results related to efficiency of certain tests are mentioned. These previous results have been an important inspiration for developing the results of this paper, but they are not used directly. In Section III, conditions for the plug-in estimator of power divergence to be consistent are given. These conditions play an important role in the formulation and the proof of the main results, but the results should also be of independent interest. Section IV is the most technical where Bahadur functions are introduced and the link to a result from [3] is established.

Section V states the main results of the paper and proves it using results of the previous sections. This result is then discussed in Section VI.

II. BAHADUR EFFICIENCY

In the preceding section, we introduced the family of power divergence statistics \( T_{\alpha,n} \), or the one–one related family of statistics \( D_\alpha(P_n, Q), \alpha \in \mathbb{R} \). In the rest of this paper, we are interested in the relative asymptotic efficiencies of these statistics for \( 1 \leq \alpha_1 < \alpha_2 < \infty \) when applied in testing the uniformity hypothesis (14). To this end, we use the concept of Bahadur asymptotic relative efficiency of \( T_{\alpha_1} \) with respect to \( T_{\alpha_2} \) (briefly Bahadur efficiency, in symbols \( BE(T_{\alpha_1} \mid T_{\alpha_2}) \)). Roughly speaking, this efficiency compares the sample sizes \( n_2 \) needed by the \( T_{\alpha_1}-\)tests of the same powers to achieve the same asymptotic sizes. It differs from the Pitman asymptotic relative efficiency of \( T_{\alpha_1} \) with respect to \( T_{\alpha_2} \), which compares the sample sizes \( n_1 \) needed by the \( T_{\alpha_1}-\)tests of the same sizes to achieve the same asymptotic powers (cf. [4, pp. 332–341] or [5]). We use the general concept of Bahadur efficiency introduced in [6] where it was extended the original concept of [7]. Before its formal definition, we briefly review some useful preliminary testing results.

Let us first suppose that \( k \) remains fixed while \( n \) tends to infinity. In this case, the goodness-of-fit statistic (5)–(13) have been studied systematically in [2]. The authors proved under \( \mathcal{H} \) the limit law

\[
T_{\alpha,n} - \frac{k}{\sqrt{2k}} \chi_{\alpha-1}^2 \xrightarrow{d} \mathcal{N}(0,1) \quad \alpha \in \mathbb{R},
\]

where \( \chi_{\alpha-1}^2 \) stands for the \( \chi^2 \)-distributed random variable with \( k-1 \) degrees of freedom and where \( \xrightarrow{d} \) here denotes convergence in distribution. In [2], the authors also proved a modification of (22) under the local alternatives

\[
A_n : P_n = (1 - 1/\sqrt{n})U + P/\sqrt{n} \quad \text{for} \quad P \in M(k) \text{ fixed}.
\]

An extension of (22) to the case where (14) remains valid but \( k = k_n \) increases slowly to \( \infty \) as \( n \to \infty \) in the sense

\[
\lim_{n \to \infty} \frac{\gamma_n}{\sqrt{n}} = 0, \quad \text{for} \quad \gamma_n = \frac{k}{n}.
\]

has been studied for \( \alpha = 2 \) in [8] and for arbitrary positive integers \( \alpha \) in [9].

The asymptotic normality

\[
\frac{T_{\alpha,n} - k}{\sqrt{2k}} \xrightarrow{d} \mathcal{N}(0,1), \quad \text{as} \quad n \to \infty, \quad \alpha \in \mathbb{R}
\]

has been proved under \( \mathcal{H} \) subsequently in [10], [11], and [12] under stronger alternatives to the slow convergence condition (24), namely

\[
\lim_{n \to \infty} \frac{\gamma_n}{\sqrt{n}} = 0, \quad \text{for} \quad \gamma_n = \frac{k^2 \ln^2 n}{n}, \quad \frac{k^2 \ln k}{n}, \quad \text{and} \quad \frac{k^2}{n}
\]

respectively. Extension of (25) to a local alternative of the type (23) can be found for \( \alpha = 1 \) and \( \alpha = 2 \) in [9], and for arbitrary \( \alpha \in \mathbb{R} \) in [13].

If contrary to (24) or (26), \( k \equiv k_n \) increases fast to \( \infty \) in the sense

\[
\lim_{n \to \infty} \frac{\gamma_n}{\sqrt{n}} = \gamma > 0, \quad \text{for} \quad \gamma_n = \frac{k}{n}
\]

then (25) has to be replaced by more complicated limit laws established in [14], [15], [2], and [16]. However, the practical situations where the model satisfies (27) are rare. In our introductory example with distribution of balls, this assumption means that the number of boxes is comparable with the number of balls. Hence, either the frequencies \( X_{n1}, \ldots, X_{nk} \) of balls in all boxes remain bounded as \( n \to \infty \), or majority of the boxes remains empty.

In what follows, we restrict ourselves to the usual situations where \( k = k_n \) satisfies the conditions of the type (24) or (26). The limit laws mentioned above enable us to specify for any \( \alpha \in \mathbb{R} \) the \( T_{\alpha} \)-based test of the hypothesis \( \mathcal{H} \) of an arbitrary asymptotic size \( s \in [0, 1] \). Under the normal law (25), such a test is defined by the rejection rule

\[
T_{\alpha} > r_n(s), \quad \text{for} \quad r_n(s) = k_n + \sqrt{2k_n \Phi^{-1}(1 - s)}
\]

for the quantile of the order \( 1 - s \) of the standard normal distribution function \( \Phi \). We would like to choose the optimal statistic \( T_{\alpha_{opt}} \) from the family \( T_{\alpha} \), \( \alpha \in \mathbb{R} \). This leads to the comparison of the asymptotic relative efficiencies in this family.

If \( k = k_n \) increases slowly as assumed in (24) or (26), then the Pitman asymptotic relative efficiencies of all statistics \( T_{\alpha,n} \), \( \alpha \in \mathbb{R} \) coincide (cf. e.g., [2]). In this situation, preferences between these statistics must be based on the Bahadur efficiencies \( BE(T_{\alpha_1} \mid T_{\alpha_2}) \). The key result in this direction was established in [6] where it was demonstrated that \( BE(T_1 \mid T_2) = \infty \) so that the log-likelihood ratio statistic \( T_1 \) is more Bahadur efficient than the Pearson statistic \( T_2 \). Using the results from [17], this first achievement was extended in [18] where it was proved that the Bahadur efficiencies of the reversed log-likelihood ratio statistic \( T_0 \) and the Neyman statistic \( T_{-1} \) coincide and both are less Bahadur efficient than the Pearson’s \( T_2 \). A problem left open in the previous literature was
to evaluate the Bahadur efficiencies of the remaining statistics $T_{\alpha}$, $\alpha \in \mathbb{R}$, in particular to confirm or reject the conjecture that the log-likelihood ratio statistic $T_1$ is most Bahadur efficient in the class of all power divergence statistics $T_{\alpha}$, $\alpha \in \mathbb{R}$. In this paper, we present a partial solution to this problem for $\alpha \geq 1$. Our solution is based on the results for indices of coincidence established in [3].

The previously mentioned Bahadur efficiency $BE(T_{\alpha_1} \mid T_{\alpha_2})$ is defined under the condition that for $\alpha = \alpha_1$ and $\alpha = \alpha_2$, the statistic $D_{\alpha}(\hat{P}_n, U)$ is consistent and admits the so-called Bahadur function. These two concepts are given in Definitions 1 and 2 below. In what follows, we often use the statistics $D_{\alpha}(\hat{P}_n, U)$ instead of the one–one related $T_{\alpha} = T_{\alpha n}$. Further, by $P(B_n)$ we shall denote the probabilities of events $B_n$ depending on the random observations $X_n$ (cf. (14) and (15)) and by $\mathbf{E}$ the corresponding expectations.

**Definition 1:** For $\alpha \in \mathbb{R}$ we say that

1) the model satisfies the Bahadur condition if there exists $0 < \Delta_\alpha < \infty$ such that under the alternatives $\mathcal{A}_n$

$$
\lim_{n \to \infty} D_{\alpha}(P_n, U) = \Delta_\alpha; 
$$

(29)

2) the statistic $D_{\alpha}(\hat{P}_n, U)$ is consistent if the Bahadur condition holds and for $n \to \infty$

$$
\mathbf{E}D_{\alpha}(\hat{P}_n, U) \to 0 \text{ under } \mathcal{H} 
$$

(30)

and

$$
D_{\alpha}(\hat{P}_n, U) \xrightarrow{p} \Delta_\alpha \text{ under } \mathcal{A}_n. 
$$

(31)

The inequality $0 < \Delta_\alpha < \infty$ in the Bahadur condition means that in term of the statistic $D_{\alpha}(\hat{P}_n, U)$, the alternatives $\mathcal{A}_n$ are neither too near to nor too far from the hypothesis $\mathcal{H}$. The next example demonstrates that in the model of Example 2 this important condition holds.

**Example 3:** Let us consider the typical situation of Example 2 leading to the present statistical testing model. If the probability measure $\mu$ considered there is dominated by $\nu$ then, by [19, Theorem 2]

$$
\lim_{n \to \infty} D_{\alpha}(P_n, U) = \int_{-\infty}^{\infty} \phi_{\alpha} \left( \frac{d\mu}{d\nu} \right) d\nu, \quad \text{for all } \alpha \in \mathbb{R}. 
$$

The integrals are $\alpha$-differences $D_{\alpha}(\mu, \nu)$ of probability measures $\mu$ and $\nu$, see [1]. Thus, (29) holds for $\Delta_\alpha = D_{\alpha}(\mu, \nu)$ when $\mu$ is dominated by $\nu$ and $\Delta_\alpha > 0$ unless $\mu = \nu$ (i.e., $\mathcal{H} = \mathcal{A}_n$ for all $n = 1, 2, \ldots$). This means that if the model of Example 2 is nontrivial then the Bahadur condition holds for all $\alpha \in \mathbb{R}$ such that $D_{\alpha}(\mu, \nu) < \infty$.

The consistency of $D_{\alpha}(\hat{P}_n, U)$ introduced in Definition 1 means that the $D_{\alpha}(\hat{P}_n, U)$-based test of the hypothesis $\mathcal{H} : U$ against the alternative $\mathcal{A}_n : P_n$ of any fixed size has power tending to 1. Indeed, under $\mathcal{H}$, we have $D_{\alpha}(\hat{P}_n, U) \xrightarrow{p} 0$, so that the rejection level $r_n(s)$ of the $D_{\alpha}(\hat{P}_n, U)$-based test of size $s \in [0, 1]$ tends to 0 for $n \to \infty$ while under $\mathcal{A}_n$ we have $D_{\alpha}(\hat{P}_n, U) \xrightarrow{p} \Delta_\alpha > 0$.

**Definition 2:** For $\alpha \in \mathbb{R}$ we say that $g_{\alpha}$ is the Bahadur function for the statistic $T_{\alpha} = 2nD_{\alpha}(\hat{P}_n, U)$ if $g_{\alpha} : \mathbb{R}_+ \to \mathbb{R}_+$ is continuous and there exists a sequence $c_{\alpha,n} > 0$ such that under $\mathcal{H}$

$$
\lim_{n \to \infty} \frac{c_{\alpha,n}}{n} \ln P(D_{\alpha}(\hat{P}_n, U) \geq \Delta) = g_{\alpha}(\Delta), \quad \Delta > 0.
$$

(32)

**Remark 1:** One should note that the Bahadur function depends on the sequence $c_{\alpha,n}$. For the kind of results that we are interested in, the crucial thing is to determine the asymptotic behavior of sequences $c_{\alpha,n}$ admitting the Bahadur function rather than the exact value of this. Nevertheless, we shall calculate the exact value of the Bahadur function for certain sequences because this will allow us to use standard terminology and because the determination of the Bahadur function may be of independent interest.

Next, we consider the following basic definition of the present paper where $\Delta_{\alpha_1}, \Delta_{\alpha_2}$ are the limits from the Bahadur condition and $g_{\alpha_1}, g_{\alpha_2}$ and $c_{\alpha_1,n}, c_{\alpha_2,n}$ are the functions and sequences from the definition of Bahadur function. In this definition, we apply to the power divergence statistics $T_{\alpha_1}$ and $T_{\alpha_2}$ the concept of the Bahadur efficiency $BE(T_{\alpha_1} \mid T_{\alpha_2})$ introduced for more general statistics $T_1$ and $T_2$ in [6, p. 732].

**Definition 3:** Let the statistics $D_{\alpha_1}(\hat{P}_n, U)$ and $D_{\alpha_2}(\hat{P}_n, U)$ be consistent and let the Bahadur functions $g_{\alpha_1}$ and $g_{\alpha_2}$ of the power divergence statistics $T_{\alpha_1}$ and $T_{\alpha_2}$ exist. Then the Bahadur efficiency $BE(T_{\alpha_1} \mid T_{\alpha_2})$ of $T_{\alpha_1}$ with respect to $T_{\alpha_2}$ is defined by

$$
BE(T_{\alpha_1} \mid T_{\alpha_2}) = \frac{g_{\alpha_1}(\Delta_{\alpha_1})}{g_{\alpha_2}(\Delta_{\alpha_2})} \lim_{n \to \infty} \frac{c_{\alpha_1,n}}{c_{\alpha_2,n}}
$$

(33)

provided the limit exists in $[0, \infty]$. Therefore, this efficiency takes on values in the domain $[0; \infty]$.

Assume that the statistics $D_{\alpha_1}(\hat{P}_n, U)$ are consistent for $i \in \{1, 2\}$ and that there exist Bahadur functions $g_{\alpha_i}$ satisfying (32) for some sequences $c_{\alpha_1,n} > 0$. Then the definition of consistency implies that both the $T_{\alpha_i}$-tests of the uniformity hypothesis $\mathcal{H} : U$ will achieve identical powers

$$
\pi = P(D_{\alpha_i}(\hat{P}_n, U) \geq r_{n,i})
$$

for $\pi \in [0, 1]$ and $i \in \{1, 2\}$ under $\mathcal{A}_n$ if and only if $r_{n,i} \xrightarrow{p} \Delta_{\alpha_i}$ for $i \in \{1, 2\}$ as $n \to \infty$. The convergence $r_{n,i} \xrightarrow{p} \Delta_{\alpha_i}$ leads to the approximate $T_{\alpha_i}$-test sizes

$$
s_{n,i} = \frac{c_{\alpha_i,n}}{g_{\alpha_i}(\Delta_{\alpha_i})} \ln \frac{1}{s_{n}}, \quad i \in \{1, 2\}
$$

(34)

to achieve the same approximate test sizes $s_n = s_{n,1} = s_{n,2}$ when $n$ is here playing the role of a formal parameter that increases to $\infty$. Thus, Definition 3 formalizes the concept of
asymptotic relative efficiency announced at the beginning of this section.

Before presenting the main results based on Definition 3 in Section V, we investigate sufficient conditions for consistency of the statistic $D_\alpha(P_n,U)$ in the domain $\alpha \geq 1$ in Section III. Section IV presents conditions for the existence of the corresponding Bahadur functions $g_\alpha$, $\alpha \geq 1$ and explicitly evaluates these functions.

III. CONSISTENCY

When a statistician uses $D_\alpha(P_n,U)$ as a statistic to distinguish between $P_n$ and $U$ then he does so because he considers the plug-in estimator $D_\alpha(P_n,U)$ as a good estimate of $D_\alpha(P_n,U)$. This idea was made precise in Definition 1 dealing with the important concept of consistency of $D_\alpha(P_n,U)$. Our next theorem presents consistency conditions for all statistic $D_\alpha(P_n,U)$, $\alpha \geq 1$. It is based on the following auxiliary result.

**Lemma 1**: For $x \in [0;1]$ and $y \in [0;1]$ and $\alpha \in [1;2]$ we have

$$|y^\alpha - x^\alpha| \leq \alpha x^{\alpha-1} |y-x| + (\alpha - 1)x^{\alpha-2}(y-x)^2.$$  \hspace{1cm} (35)

**Proof**: First we observe that

$$y^\alpha \geq x^\alpha + \alpha x^{\alpha-1}(y-x)$$  \hspace{1cm} (36)

because the function $y \to y^\alpha$ is convex. Next we prove the inequality

$$y^\alpha \leq x^\alpha + \alpha x^{\alpha-1}(y-x) + (\alpha - 1) x^{\alpha-2}(y-x)^2.$$  \hspace{1cm} (37)

The upper and lower bounds in (35) and (36) are illustrated in Fig. 2.

We have to prove that

$$y^\alpha - (x^\alpha + \alpha x^{\alpha-1}(y-x) + (\alpha - 1) x^{\alpha-2}(y-x)^2)$$

is negative. This is obvious for $y = x$ and for $y = 0$. The derivative is

$$\alpha y^\alpha - x^{\alpha-1} + (\alpha - 1)x^{\alpha-2}2(y-x)$$

$$= \alpha y^\alpha + (\alpha - 2)x^{\alpha-1}2 - 2\alpha x^{\alpha-2}.$$

The derivative is 0 for $y = x$. Differentiate once more and get

$$\alpha (\alpha - 1)y^{\alpha-2} + (2-2\alpha)x^{\alpha-2} = (\alpha - 1)(\alpha y^\alpha - 2x^{\alpha-2})$$

which is positive for $y \leq \left(\frac{\alpha}{\alpha-1}\right)^{\frac{1}{\alpha-2}} x \leq x$. Combining (36) and (35) leads to

$$0 \leq y^\alpha - x^\alpha - \alpha x^{\alpha-1}(y-x) \leq (\alpha - 1)x^{\alpha-2}(y-x)^2.$$  \hspace{1cm} (38)

Now 1 $- x^{\alpha-1}$ is increasing in $x$ and equals 0 for $x = 1$ so the lower bound in (37) side is negative and we have

$$|y^\alpha - x^\alpha - \alpha x^{\alpha-1}(y-x)| \leq (\alpha - 1)x^{\alpha-2}(y-x)^2.$$  \hspace{1cm} (39)

The inequality follows because

$$|y^\alpha - x^\alpha| \leq |\alpha x^{\alpha-1}(y-x)| + |\alpha x^{\alpha-1}(y-x)||.$$  \hspace{1cm} (40)

We shall also use the following upper bound a number of times:

$$E(|\hat{p}_j - p_j|^2) \leq \frac{p_j (1-p_j)}{n} \leq \frac{p_j^2}{n}.$$  \hspace{1cm} (41)

For divergence of order 2 it gives

$$E D_2 \left(\hat{P}_n||P\right) = \sum_{j=1}^{k} E \left(\hat{p}_j - p_j\right)^2 \leq \frac{\sum_{j=1}^{k} \left(\frac{1}{n}\right)}{n} \leq \frac{k}{n}.$$  \hspace{1cm} (42)

**Theorem 1**: For all $\alpha \geq 1$ let the Bahadur condition (29) hold. Then $D_\alpha(P_n,U)$ is consistent if

$$\alpha \in [1;2] \text{ and } \lim_{n \to \infty} k \frac{\alpha}{n} = 0.$$  \hspace{1cm} (43)

or

$$\alpha \geq 2 \text{ and } \lim_{n \to \infty} k \frac{\alpha-1}{n} = 0.$$  \hspace{1cm} (44)

**Proof**: Under $\mathcal{H}$, we have $D_\alpha(P_n,U) = D_\alpha(U,U) = 0$. Hence, it suffices to prove that under both $\mathcal{H}$ and $\mathcal{A}_n$

$$\lim_{n \to \infty} E \left[D_\alpha(P_n,U) - D_\alpha(P_n,U)\right] = 0,$$  \hspace{1cm} (45)

for all $\alpha \geq 1$.

Put for brevity $\Lambda_{\alpha,n} = D_\alpha(P_n,U) - D_\alpha(P_n,U)$ and denote the variance by $\text{Var}$ and the covariance by $\text{Cov}$. The cases i) $\alpha = 1$, ii) $\alpha \in [1;2]$, and iii) $\alpha > 2$ have to be treated separately.

i) For $\alpha = 1$, we have

$$\Lambda_{1,n} = \sum_{j=1}^{k} (\hat{p}_j - p_j)^2 \text{ and } \text{Var}(\Lambda_{1,n}) = \sum_{j=1}^{k} \left(\frac{p_j (1-p_j)}{n}\right) \leq \frac{p_j^2}{n}.$$  \hspace{1cm} (46)

Authorized licensed use limited to: IEEE Xplore. Downloaded on April 21,2010 at 06:03:53 UTC from IEEE Xplore. Restrictions apply.
where we dropped the subscript $n$ everywhere in the sum. From (41) we obtain
\[
\Lambda_{1,n} = \sum_{j=1}^{k} \hat{p}_j \ln \hat{p}_j - \sum_{j=1}^{k} (\hat{p}_j - p_j) \ln \frac{1}{p_j}
\]
and hence
\[
|\Lambda_{1,n}| \leq D_1(\hat{P}_n, P_n) + \sum_{j=1}^{k} (\hat{p}_j - p_j) \ln \frac{1}{p_j}.
\]
If $\hat{n}_i = \hat{p}_i n$ is the number of observations of type $i$ then
\[
D_1(\hat{P}_n, P_n) \leq 2D_2(\hat{P}_n, P_n).
\]
Therefore
\[
E|\Lambda_{1,n}| \leq 2E D_2(\hat{P}_n, P_n) + E \left| \sum_{j=1}^{k} (\hat{p}_j - p_j) \ln \frac{1}{p_j} \right|.
\]  
(42)

The last term on the right-hand side can be bounded using Jensen’s Inequality
\[
E \left| \sum_{j=1}^{k} (\hat{p}_j - p_j) \ln \frac{1}{p_j} \right| \leq \left( E \left( \sum_{j=1}^{k} (\hat{p}_j - p_j) \ln \frac{1}{p_j} \right)^2 \right)^{1/2} 
\]
\[
= \left( \sum_{i,j=1}^{k} p_j \ln p_i \text{COV}(\hat{p}_i, \hat{p}_j) \right)^{1/2} 
\]
\[
= \left( \sum_{i,j=1}^{k} p_j \ln p_i \text{COV}(\hat{n}_i, \hat{n}_j) / n^2 \right)^{1/2}. 
\]
(43)

Further
\[
\sum_{i,j=1}^{k} \ln p_j \ln p_i \frac{\text{COV}(\hat{n}_i, \hat{n}_j)}{n^2} 
\]
\[
= \sum_{i=1}^{k} (\ln p_i)^2 \frac{\text{VAR}(\hat{n}_i)}{n^2} + \sum_{i \neq j} \ln p_j \ln p_i \frac{\text{COV}(\hat{n}_i, \hat{n}_j)}{n^2} 
\]
\[
\leq \sum_{i=1}^{k} (\ln p_i)^2 \frac{p_i}{n} + \sum_{i \neq j} \ln p_j \ln p_i \frac{n p_i p_j}{n^2} 
\]
\[
= \frac{1}{n} \sum_{i=1}^{k} p_i (\ln p_i)^2 + \frac{1}{n} \left( \sum_{i=1}^{k} \ln p_i \ln p_i \right)^2. 
\]  
(44)

The function $x \rightarrow x \ln^2 x$ is concave in the interval $[0; e^{-1}]$ and convex in the interval $[e^{-1}; 1]$. Therefore, we are able to use [3, Theorem 3.1] to see that
\[
\sum_{i=1}^{k} p_i (\ln p_i)^2 \text{ attains its maximum for a mixture of uniform distributions on } k \text{ and } k-1 \text{ points. Thus}
\]
\[
\sum_{i=1}^{k} p_i (\ln p_i)^2 \leq \sum_{i=1}^{k} \frac{1}{k-1} \left( \ln \frac{1}{k} \right)^2 
\]
\[
= \frac{k(\ln k)^2}{k-1} \leq 2(\ln k)^2. 
\]  
(45)

The sum $\sum_{i=1}^{k} p_i \ln p_i$ equals minus the entropy, which has maximum $\ln k$. By combining (38), (42), (43), (44), and (45) we get
\[
E|\Lambda_{1,n}| \leq \frac{2k}{n} + \left( \frac{3(\ln k)^2}{n} \right)^{1/2}
\]
and the right-hand side tends to zero under the condition (39) for $n$ tending to infinity. This proves (39) for $\alpha = 1$.

ii) For every $\alpha \in [1; 2]$ we have
\[
\Lambda_{\alpha,n} = \frac{k^{\alpha-1}}{\alpha (\alpha-1)} \sum_{j=1}^{k} (p_j^\alpha - \hat{p}_j^\alpha). 
\]

Using the abbreviation
\[
D_{\alpha,n} = D_{\alpha}(P_n, U)
\]
we obtain
\[
\frac{1}{\alpha} \left| \Lambda_{\alpha,n} \right| 
\]
\[
\leq \frac{k^{\alpha-1}}{\alpha (\alpha-1)} \sum_{j=1}^{k} \left( \frac{\alpha p_j^{\alpha-1} \hat{p}_j - p_j}{\hat{p}_j - p_j} \right)^{1/2} 
\]
\[
\leq \frac{k^{\alpha-1}}{\alpha (\alpha-1)} \left( \sum_{j=1}^{k} \left( p_j^{\alpha/2} \right)^{1/2} \right)^{1/2} \left( \sum_{j=1}^{k} \left( \hat{p}_j - p_j \right)^{1/2} \right)^{1/2} 
\]
\[
+ \frac{k^{\alpha-1}}{\alpha} \sum_{j=1}^{k} p_j^{\alpha-2} (\hat{p}_j - p_j)^2 
\]
\[
= \frac{k^{\alpha-1}}{\alpha (\alpha-1)} \left( \sum_{j=1}^{k} p_j^{\alpha-2} (\hat{p}_j - p_j)^2 \right)^{1/2} 
\]
\[
+ \frac{k^{\alpha-1}}{\alpha} \sum_{j=1}^{k} p_j^{\alpha-2} (\hat{p}_j - p_j)^2 
\]
\[
\text{because } \sum_{j=1}^{k} p_j^{\alpha-2} = [\alpha (\alpha-1) D_{\alpha,n} + 1]/k^{\alpha-1}. \text{ Thus}
\]
\[
E|\Lambda_{\alpha,n}| 
\]
\[
\leq \frac{k^{\alpha-1}}{\alpha (\alpha-1)} \left( \sum_{j=1}^{k} p_j^{\alpha-2} (\hat{p}_j - p_j)^2 \right)^{1/2} 
\]
\[
+ \frac{k^{\alpha-1}}{\alpha} \sum_{j=1}^{k} p_j^{\alpha-2} p_j 
\]
is upper-bounded and that implies that there exists a constant $c < 1$ such that $p_j \leq c$ for all $n, j$. Thus

$$E[\Lambda_{\alpha,n}] \leq \left( \frac{\alpha (\alpha - 1) D_{\alpha,n} + 1}{\alpha - 1} \right)^{1/2} \left( \frac{k^{\alpha-1} \sum_{j=1}^{k} p_j^{\alpha-1}}{n} \right)^{1/2} + \frac{k^{\alpha-1}}{\alpha n} \sum_{j=1}^{k} p_j^{\alpha-1}.$$  

The sequence $D_{\alpha,n}$ is upper-bounded and that implies that there exists a constant $c < 1$ such that $p_j \leq c$ for all $n, j$. Thus

$$E[\Lambda_{\alpha,n}] \leq \left( \frac{\alpha (\alpha - 1) D_{\alpha,n} + 1}{\alpha - 1} \right)^{1/2} \left( \frac{k^{\alpha-1} \sum_{j=1}^{k} p_j^{\alpha-1}}{n} \right)^{1/2} + \frac{k^{\alpha-1}}{\alpha n} \sum_{j=1}^{k} p_j^{\alpha-1}.$$  

Therefore, (40) is not only the model satisfies, for

$$\frac{1}{\alpha - 1} \left( \frac{k^{\alpha-1} \sum_{j=1}^{k} p_j^{\alpha-1}}{n} \right)^{1/2} + \frac{k^{\alpha-1}}{2n},$$  

But $D_{\alpha,n}$ is zero under the hypothesis of a uniform distribution and, by (29), has a finite limit under the alternative. This completes the proof of (40).

Example 4: Assume that for $\alpha = 3$ the model satisfies the Bahadur condition, in particular that (29) holds with $\alpha = 3$. Then

$$ED_3 \left( \hat{P}_n, U \right) = \frac{k^2 E \left( \sum_{j=1}^{k} p_j^3 \right) - 1}{6}$$

where

$$\hat{p}_j^3 = p_j^3 + 3p_j^2 (\hat{p}_j - p_j) + p_j \left( (\hat{p}_j - p_j)^2 + (\hat{p}_j - p_j)^3 \right).$$

Therefore

$$ED_3 \left( \hat{P}_n, U \right) = \frac{k^2 \sum_{j=1}^{k} \left[ p_j^3 \left( (\hat{p}_j - p_j)^2 + (\hat{p}_j - p_j)^3 \right) - 1 \right]}{6} = \frac{k^2 p_j^3 - 1}{6} + \frac{k^2}{6} \sum_{j=1}^{k} \left( 3p_j E(\hat{p}_j - p_j)^2 + E(\hat{p}_j - p_j)^3 \right).$$

By taking mean values we get

$$ED_3 \left( \hat{P}_n, U \right) = D_3 (P_n, U) + \frac{k^2}{6} \sum_{j=1}^{k} \left( 3p_j \frac{p_j (1 - p_j)}{n} + p_j \frac{(1 - p_j)(1 - 2p_j)}{n} \right) = D_3 (P_n, U) + \frac{k^2}{6n} \sum_{j=1}^{k} (p_j - p_j^3) = D_3 (P_n, U) + \frac{k^2}{6n} \sum_{j=1}^{k} (p_j - p_j^3) = D_3 (P_n, U) + \frac{k^2}{6n} \sum_{j=1}^{k} (p_j - p_j^3).$$

By (29), $D_3 (P_n, U)$ is bounded away from $0$ under $A_n$ uniformly for all sufficiently large $n$. Therefore, (40) is not only sufficient but also necessary for the consistency of the statistic $D_\alpha \left( \hat{P}_n, U \right)$ in the special case $\alpha = 3$.

IV. BAHADUR FUNCTIONS

Throughout this section, we consider the statistical testing model (14)–(15) under the hypothesis $H$. This means that $P(B_n)$ denotes the probability of the random events $B_n$ depending on $X_n$ with a multinomial distribution in the sense of (14). As before, we consider $k = k_n$ depending on the sample size $n$ and we study the Bahadur functions (32) corresponding to the statistic $D_\alpha (\hat{P}_n, U)$ for $\alpha \geq 1$.  

Authorized licensed use limited to: IEEE Xplore. Downloaded on April 21,2010 at 06:03:53 UTC from IEEE Xplore. Restrictions apply.
Example 5: Let \( k = k_n \) increase so slowly that
\[
\lim_{n \to \infty} \frac{k \ln n}{n} = 0, \tag{50}
\]
Then (32) holds for the sequence \( c_{\alpha, n} \equiv 1 \) and function
\[
g_1(\Delta) = \Delta, \quad \text{for all } \Delta > 0 \tag{51}
\]
i.e., (51) is the Bahadur function for the log-likelihood ratio statistic
\[
T_1 = 2nD_1(\hat{P}_n, U).
\]
This result was first obtained independently in [20, Corollary 2.4] and [6, Theorem 2]. Using the simple method based on the inequality (52) below, it was obtained in [17, Theorem 2].

According to the result of [21] sharpened in [22, Problem 1.2.11] and in [23, p. 16], for every subset \( A \subseteq M(k) \), the divergence \( D_1(\hat{P}_n, U) \) defined by (11) satisfies the inequality
\[
\inf_{P \in A} \frac{1}{n} \ln P(\hat{P}_n \in A) \leq \frac{k \ln(n + 1)}{n}, \tag{53}
\]
Hence, the approximation of \( \frac{1}{n} \ln P(\hat{P}_n \in A) \) as in (32) by means of the infimum appearing in (52) is possible under the restriction
\[
\lim_{n \to \infty} c_{\alpha, n} \frac{k \ln n}{n} = 0 \tag{54}
\]
on the sequence \( k = k_n \), in addition to (24).

In the rest of this section, we present an alternative to the formula (32) for the Bahadur functions \( g_\alpha, \alpha \in \mathbb{R} \) which is based on the inequality (52). These formulas are given in terms of the Shannon entropy \( H(P) \) maximized on the sets
\[
\mathcal{A}_\alpha(k) = \{ P \in M(k) : D_1(P, U) \geq \Delta \} \tag{55}
\]
and
\[
\mathcal{A}_\alpha(k|n) = \mathcal{A}_\alpha(k) \cap M(k|n) \tag{56}
\]
or, equivalently, in terms of the information divergence \( D_1(\hat{P}_n, U) \) minimized on these sets.

**Lemma 2:** Assume that for some \( \alpha \in \mathbb{R} \) and \( k = k_n \) there exist \( c_{\alpha, n} > 0 \) satisfying (54) such that the sequence of functions
\[
G_{\alpha, \Delta}(k|n) = c_{\alpha, n} \left( \inf_{P \in \mathcal{A}_\alpha(k|n)} D_1(P, U) \right), \quad \Delta > 0 \tag{57}
\]
converges to a positive limit
\[
g_\alpha(\Delta) = \lim_{n \to \infty} G_{\alpha, \Delta}(k|n), \quad \Delta > 0. \tag{58}
\]
Then the Bahadur function for the power divergence statistic \( T_\alpha \) is equal to \( g_\alpha \).

**Proof:** Using (19) and (52) we get that the functions (57) satisfy the inequality
\[
\frac{c_{\alpha, n}}{n} \ln P(D_1(\hat{P}_n, U) \geq \Delta) + G_{\alpha, \Delta}(k|n) \leq \frac{k c_{\alpha, n} \ln(n + 1)}{n}.
\]
Since (54) holds, (58) follows from here and from (33).

In the following assertion, we consider for arbitrary \( \alpha \in \mathbb{R} \), \( k = k_n \), and \( c_{\alpha, n} > 0 \) the sequence of functions
\[
G_{\alpha, \Delta}(k) = c_{\alpha, n} \left( \inf_{P \in \mathcal{A}_\alpha(k)} D_1(P, U) \right), \quad \Delta > 0. \tag{59}
\]
Obviously, \( G_{\alpha, \Delta}(k) \leq G_{\alpha, \Delta}(k|n) \).

**Lemma 3:** Let for some \( \alpha \in \mathbb{R} \) the Bahadur condition hold and \( c_{\alpha, n} > 0 \) satisfy (54). If the corresponding sequences of functions (57) and (59) asymptotically coincide in the sense
\[
\lim_{n \to \infty} \left[ G_{\alpha, \Delta}(k|n) - G_{\alpha, \Delta}(k) \right] = 0 \tag{60}
\]
and at the same time \( G_{\alpha, \Delta}(k) \) converges to a positive limit
\[
g_\alpha(\Delta) = \lim_{n \to \infty} G_{\alpha, \Delta}(k), \quad \Delta > 0 \tag{61}
\]
then \( g_\alpha \) is the Bahadur function for the power divergence statistic \( T_\alpha \).

**Proof:** Clear from the assumption (60) and Lemma 2. \( \square \)

In [3] it was proved that for every \( x \in [k^{1-\alpha}, 1] \) and for the Dirac distribution \( 1 \equiv (1, 0, \ldots, 0) \in M(k) \), the equation
\[
IC_{\alpha}(s1 + (1 - s)U) = x
\]
has a unique solution \( s \in [0; 1] \) and that this solution satisfies the relation
\[
s_{\alpha, k} = \frac{1 - (1 - 1/k)^{1/(\alpha - 1)}}{1 - (1 - 1/k)^{\alpha/(\alpha - 1)}} \in [0; 1]. \tag{64}
\]
This result is illustrated in Fig. 3. It leads to the following lemma using the constants:
\[
\frac{1}{\alpha(\alpha - 1)k} < \Delta \leq \frac{k^{\alpha - 1} - 1}{\alpha(\alpha - 1)}
\]
the equation
\[
\frac{1}{k} (k s + 1 - s)^{\alpha} + (1 - s)^{\alpha} = 1 + \alpha(\alpha - 1) \Delta \tag{66}
\]
has a unique solution \( s \in [0; 1] \) and this solution satisfies the inequality
\[
s_{\alpha,k} < s \leq 1
\]
and the equality
\[
\inf_{P \in A_{\alpha,\Delta}(k)} D_1(P, U) = \frac{1}{k} (ks + 1 - s) \ln(ks + 1 - s) + (1 - s) \ln(1 - s). \tag{68}
\]

**Proof:** By definition of \( A_{\alpha,\Delta}(k) \) in (55) and (20), \( P \in A_{\alpha,\Delta}(k) \) if and only if \( IC_\alpha(P) \geq x \) for
\[
x = k^{1-\alpha}[1 + \alpha(\alpha - 1)\Delta].
\]
By the definition of \( IC_\alpha(P) \) in (21)
\[
IC_\alpha(s1 + (1 - s)U) = k^{1-\alpha} \left( \frac{1}{k} (ks + 1 - s)^\alpha + (1 - s)^\alpha \right) \tag{69}
\]
so that the (62) is equivalent to (66). Further
\[
D_1(s1 - (1 - s)U, U) = \frac{1}{k} (ks + 1 - s) \ln(ks + 1 - s) + (1 - s) \ln(1 - s). \tag{70}
\]
Therefore, by (63) and (62), the relation (68) will be proved if we prove that (66) has a unique solution \( s \in [0; 1] \). One can verify by differentiation that the continuous function
\[
\psi(s) = \frac{1}{k} (ks + 1 - s)^\alpha + (1 - s)^\alpha, \quad s \in [0; 1], \tag{71}
\]
appearing on the left of (66) is decreasing on the interval \([0; s_{\alpha,k}]\) and increasing on the complement \([s_{\alpha,k}; 1]\). Since
\[
\psi(0) = \frac{1}{k} + 1 \quad \text{and} \quad \psi(1) = k^{\alpha-1}
\]
for each \( \Delta \) satisfying (65) the solution \( s \in [0; 1] \) is unique and strictly greater than \( s_{\alpha,k} \). Thus, not only (68) but also (67) is valid.

In the following lemma and everywhere in the sequel, convergence as well as the symbols \( o(\cdot) \) and \( O(\cdot) \) are considered for \( n \to \infty \). We remind that \( k = k_n \) is assumed to satisfy (24).

**Lemma 5:** For every \( \alpha > 1 \) and \( \Delta > 0 \)
\[
\inf_{P \in A_{\alpha,\Delta}(k)} D_1(P, U) = \left( [\alpha(\alpha - 1)\alpha]^{1/\alpha} + o(1) \right) k^{1/\alpha} \ln k^{1/\alpha} - \frac{1}{k}. \tag{72}
\]

**Proof:** Consider arbitrary \( \alpha > 1 \) and \( \varepsilon > 0 \). Since \( k = k_n \) satisfies (24), Lemma 4 implies for all sufficiently large \( n \) that (66) has in the interval \([s_{\alpha,k}; 1]\) a unique solution \( s = s_k \) satisfying (68). Therefore, it suffices to prove that the sequence
\[
x_k = \frac{1}{k} (ks_k + 1 - s_k) \ln(ks_k + 1 - s_k) + (1 - s_k) \ln(1 - s_k) \tag{72}
\]
and the positive constant
\[
\delta = [\alpha(\alpha - 1)\Delta]^{1/\alpha}
\]
satisfy the asymptotic relation
\[
x_k = (\delta + o(1)) k^{1/\alpha} \ln k^{1/\alpha} \tag{74}.
\]
By (64) and (67), \( s_k \) is a positive sequence and (66) with \( s \) replaced by \( s_k \) obviously contradicts the assumption
\[
\limsup_{k \to \infty} s_k > 0.
\]
Therefore, under (24), \( s_k = o(1) \) and, consequently, (66) with \( s \) replaced by \( s_k \) leads to the asymptotic relation
\[
\frac{1}{k} [ks_k + O(1)]^\alpha + 1 + o(1) = 1 + \delta^\alpha.
\]
This relation implies that
\[
s_k = \frac{\delta k^{1/\alpha}}{k} + o \left( \frac{k^{1/\alpha}}{k} \right) \tag{75}
\]
and the desired relation (74) follows from here and from the definition of \( x_k \) in (72).

In the rest of the paper, we are interested in the sequences
\[
c_{\alpha,n} = \frac{k}{k^{1/\alpha} \ln k^{1/\alpha}} \tag{76}
\]
for \( \alpha > 1 \) and \( k = k_n \) satisfying (24).

**Lemma 6:** If \( c_{\alpha,n} \) is given by (76) and (24) is satisfied then (60) holds for every \( \alpha > 1 \) and \( \Delta > 0 \).

**Proof:** Let \( \alpha, \Delta, \) and \( s_k \) be the same as in the previous proof. Further, denote by \( \ell_k \) the integer part of \( n(1 - s_k)/k \)
\[
\ell_k = \left[ \frac{n(1 - s_k)}{k} \right],
\]
and define
\[
\tilde{s}_k = \frac{n - k\ell_k}{n}, \quad \tilde{P}_k = \frac{n}{n} + (1 - \tilde{s}_k)U, \quad \tilde{P}_k = s_k 1 + (1 - s_k)U
\]
where \( 1 \) and \( U \) are the same elements of \( M(k) \) as in (69) and (70). Then
\[
s_k \leq \tilde{s}_k \leq s_k + \frac{k}{n} \tag{77}
\]
and one obtains from (20), (69) and (71)
\[
D_\alpha(\tilde{P}_k, U) = \frac{1}{\alpha(\alpha - 1)} (k^{\alpha - 1} \psi(\tilde{s}_k) - 1) \tag{73}
\]
and
\[
D_\alpha(\tilde{P}_k, U) = \frac{1}{\alpha(\alpha - 1)} (k^{\alpha - 1} \psi(\tilde{s}_k) - 1) \tag{73}.
\]
The distribution \( P_k \) belongs to \( A_\alpha \Delta(k) \) of (55). Indeed, \( s_k \) satisfies (66) and, consequently, \( D_\alpha(\tilde{P}_k, U) = \Delta \). The distribution \( \tilde{P}_k \) belongs to \( M(k|n) \subset M(k) \) defined by (1). Further, in the proof of Lemma 4 we argued that the function \( \psi(s) \) of (71) is increasing in the domain \([s_k; 1] \subset [s_k; 1] \). Therefore, the left-hand side of (77) implies \( \Delta(k,n) \geq \Delta \), which means that \( \tilde{P}_k \) belongs to \( A_\alpha \Delta(k|n) \) of (86). Consequently

\[
\inf_{P \in A_\alpha \Delta(k|n)} D_1(P, U) \leq \inf_{P \in A_\alpha \Delta(k|n)} D_1(\tilde{P}_k, U) = D_1(\tilde{P}_k, U)
\]

where

\[
D_1(\tilde{P}_k, U) = \ln k - H(\tilde{P}_k) = \ln k - H(\tilde{s}_k 1 + (1 - \tilde{s}_k)U) = \tilde{s}_k - \tilde{s}_k - \ln \tilde{s}_k
\]

for \( \tilde{x}_k \) defined by (72) with \( s_k \) replaced by \( \tilde{s}_k \). Further, in the previous proof we deduced for \( \tilde{x}_k \) of (72) the formula (74) from the asymptotic property (75) of \( s_k \). However, under (24), the sequence \( s_k \) satisfies this asymptotic property as well. Therefore, (74) remains to be valid with \( \tilde{x}_k \) replaced by \( \tilde{x}_k \). This means that under (24) takes place the asymptotic relation

\[
\inf_{P \in A_\alpha \Delta(k|n)} D_1(P, U) \leq \left( \alpha(\alpha - 1) \Delta \right)^{1/\alpha} + o(1)
\]

(79)

for some \( \beta \geq 3 \). The case of the remaining statistic \( T_{\alpha_1} \), \( \alpha > 1 \). Theorem 2: Let \( k = k_n \) increase to infinity slowly in the sense that for some \( \alpha \geq 1 \)

\[
\lim_{n \to \infty} \frac{k_n^{2(\alpha - 1)} \ln n}{n \ln k_n} = 0
\]

(80)

Then (33) holds for the sequence \( c_{\alpha,n} \) given by (76) and for the function \( g_\alpha(\Delta) = \left[ \frac{\alpha(\alpha - 1) \Delta}{\alpha} \right]^{1/\alpha} \), \( \Delta > 0 \) (79)

i.e., (79) is the Bahadur function of the statistic \( T_{\alpha} \).

Proof: Let \( \alpha > 1 \) be arbitrary fixed. If \( c_{\alpha,n} \) is given by (76) then (78) implies (24) as well as (54). Hence, it follows from Lemmas 3 and 6 that (32) holds for \( c_{\alpha,n} \) under consideration and for \( g_\alpha \) given by (79). Employing Lemma 4 we find that (61) reduces to (79) which completes the proof.

The particular case of Theorem 2 for \( \alpha = 2 \) was obtained in [6, Theorem 1] by using more complicated analytic methods involving limit theorems for multinomial and Poisson distributions. This particular case has been obtained also by [17] by using similar simple method as here, based on the inequality (52).

V. MAIN RESULTS

The functions \( g_\alpha \) as well as the normalizing sequences \( c_{\alpha,n} \) have been explicitly evaluated in Theorem 2 and Example 5 for all \( \alpha \geq 1 \). Therefore, (33) provides explicit Bahadur efficiencies \( BE(T_{\alpha_1} | T_{\alpha_2}) \) on the whole domain \( \alpha_1, \alpha_2 \geq 1 \). These efficiencies are given in the following main result of this paper.

**Theorem 3:** Let \( 1 \leq \alpha_1 < \alpha_2 < \infty \).

i) If the statistics \( D_{\alpha_1}(\tilde{P}_k, U) \) and \( D_{\alpha_2}(\tilde{P}_k, U) \) are consistent and \( k = k_n \) increases so slowly that

\[
\lim_{n \to \infty} \frac{k^{2-1/\alpha_2} \ln n}{n} = 0
\]

then the Bahadur efficiency of the statistic \( T_{\alpha_1} \) with respect to \( T_{\alpha_2} \) satisfies the relation

\[
BE(T_{\alpha_1} | T_{\alpha_2}) = \infty
\]

(81)

ii) If \( k = k_n \) increases to infinity slowly in the sense that for some \( \beta > 3 \)

\[
\lim_{n \to \infty} \frac{k^{5/2}}{n^{1/2}} = 0
\]

(82)

then (80) and the consistency required in i) hold for all \( 1 \leq \alpha_1 < \alpha_2 \leq \beta + 1 \). Hence, in this case also the Bahadur efficiency relation (81) holds for all \( 1 \leq \alpha_1 < \alpha_2 \leq \beta + 1 \).

**Proof:**

i) Let the assumptions of i) hold for some \( 1 < \alpha_1 < \alpha_2 < \infty \). Then (80) implies (78) for \( \alpha = \alpha_1 \) and \( \alpha = \alpha_2 \). By Theorem 2, the sequences \( c_{\alpha_1,n} \) and \( c_{\alpha_2,n} \) given by (76) for lead to the corresponding Bahadur functions \( g_{\alpha_1} \) and \( g_{\alpha_2} \) given by (32) and to the limit

\[
\lim_{n \to \infty} \frac{c_{\alpha_2,n}}{c_{\alpha_1,n}} = \lim_{n \to \infty} \frac{k^{1/\alpha_1} \ln k^{1/\alpha_2}}{k^{1/\alpha_2} \ln k^{1/\alpha_1}} = \infty
\]

(83)

Relation (81) thus follows from (33) in Definition 3. If the assumptions of i) hold for \( 1 \leq \alpha_1 < \alpha_2 \leq \beta + 1 \) instead of the above considered Bahadur function \( g_{\alpha_1} \) given by (32) we have \( g_{\alpha_1}(\Delta) = \Delta \) given by (51), and instead of \( c_{\alpha_1,n} = k_n^{1/\alpha_1} \ln k_n^{1/\alpha_1} \) given by (76) we have \( c_{\alpha_1,n} = 1 \) given in Example 5. Therefore, the limit

\[
\lim_{n \to \infty} \frac{c_{\alpha_2,n}}{c_{\alpha_1,n}} = \lim_{n \to \infty} \frac{k_n^{1/\alpha_2} \ln k_n^{1/\alpha_2}}{k_n^{1/\alpha_1} \ln k_n^{1/\alpha_1}} = \infty
\]

remains to be infinite as in (83).

ii) If (82) holds for \( \beta > 3 \) then \((\beta - 2)/\beta > 0\) so that (80) implies

\[
0 = \lim_{n \to \infty} \frac{k}{n^{5/2}} = \lim_{n \to \infty} \frac{k^{2} \ln n}{n^{5/2} \sqrt[2]{\beta - 2}} = \lim_{n \to \infty} \frac{k^{2} \ln n}{n^{5/2}}
\]

Therefore, (80) holds for all \( \alpha_2 \geq 1 \). Further, if

\[
\lim_{n \to \infty} \frac{k^3}{n} = 0
\]

then it is easy to verify that the consistency conditions of Theorem 1 are satisfied for all \( 1 \leq \alpha \leq 3 \), and if (82)

\[1\text{Note added in proof}: \text{The condition } \beta > 1 \text{ can be replaced by the condition } \beta > 1 \text{ and only requires a simple modification of the proof.}\]
holds then these conditions are satisfied for all \( 1 \leq \alpha \leq \beta + 1 \). This completes the proof.

VI. DISCUSSION

The special case of (81) with \( \alpha_1 = 1 \) and \( \alpha_2 = 2 \) with increasing \( k = k_n \) has been obtained in [6]. In the present paper, we extended the fact that the log-likelihood ratio statistic \( T_1 \) is more Bahadur efficient than the classical Pearson statistic \( T_2 \) by proving that \( T_1 \) is more Bahadur efficient than any statistics \( T_\alpha \) with \( \alpha > 1 \). Moreover, we found that the Bahadur efficiency of the power divergence statistic \( T_\alpha \) strictly decreases with \( \alpha \) increasing in the domain \([1; \infty] \). In particular, any statistic \( T_\alpha \), \( 1 \leq \alpha < 2 \), is more Bahadur efficient than the Pearson’s \( T_2 \).

One of the aims of this paper was to verify whether there is a statistic \( T_\alpha \), \( \alpha \in \mathbb{R} \) more efficient in the Bahadur sense than \( T_1 \). In this respect, the result of Theorem 3 is negative. All we can say is that, if such a statistic exists, then it is most likely that it is of the form \( T_\alpha \) with \( \alpha \in [0; 1] \). Let us comment this conclusion in more detail.

In spite of the fact that we do not have a systematic result for \( \alpha < 1 \), some fragments of such a result are available. Namely, [17] found the Bahadur functions \( g_\alpha(D) = g_{\alpha-1}(\Delta) = \Delta \) for all \( \Delta > 0 \), under the identical sequences \( \theta_0 = \theta_1 = k_n \) figuring in (32). There is a small problem with the condition (29), because the event \( \min_j \hat{p}_{nj} = 0 \) takes place with a positive probability and implies \( D_\alpha(P_n; U) = \infty \) for all \( \alpha \leq 0 \). Nevertheless, the probability of this unpleasant event tends exponentially to zero, and one can modify the statistic \( T_{0,n} \) and \( T_{-1,n} \) in such a way that the above evaluated Bahadur functions and sequences remain unaltered and, at the same time, the consistency condition (30) hold, see [18] and [24]. Therefore, in the light of present Theorem 2, the result of [17] means that the reversed log-likelihood ratio statistic \( T_0 \) and the Neyman statistic \( T_{-1} \), are mutually equally Bahadur efficient, and each of them is less Bahadur efficient than any \( T_\alpha \), \( \alpha \geq 1 \). This extends the previous result of [18] who found \( T_0 \) and \( T_{-1} \) to be less Bahadur efficient than \( T_2 \). If the low Bahadur efficiency of \( T_0 \) and \( T_{-1} \) is shared by all statistics \( T_\alpha \) of the nonpositive powers \( \alpha \leq 0 \) then the possibility to find \( T_\alpha \) comparable with \( T_1 \) or better is restricted to \( \alpha \in (0, 1) \), as conjectured above.

ACKNOWLEDGMENT

The authors would like to thank Wouter Koolen-Wijkstra for comments that have improved the quality of this paper.

REFERENCES


