# Efficiency of Entropy Testing

Peter Harremoës
Centrum voor Wiskunde en Informatica
P.O. 94079, 1090 GB Amsterdam
The Netherlands
P.Harremoes@cwi.nl

Igor Vajda
Inform. Theory and Automation
Czech Acad. Sciences, Prague
Czech Republic
vajda@utia.caz.cz

***Abstract*— Recently it was shown that Shannon entropy is more Bahadur efficient than any Rényi entropy of order $\alpha > 1$. In this paper we shall show that relative Bahadur efficiency between any two Rényi entropies of orders $\alpha \in \left]0;1\right]$ is 1 when the relative Bahadur efficiency is defined according to [1]. Despite the fact that the relative Bahadur efficiency is 1 it is shown that in a certain sense Shannon entropy is more efficient than Rényi entropy for $\alpha \in \left]0;1\right[$. This indicates that the definition of relative efficiency given in [1] does not fully capture the notion of efficiency.**

## I. Power divergence statistics

Let $M(k)$ denote the set of all discrete probability distributions of the form $P = (p_1, ..., p_k)$ and $M(k|n)$ the subset of possible types. One of the fundamental problems of mathematical statistics can be described as follows. Consider $n$ balls distributed into bins $1, ..., k$ independently according to an unknown probability distribution $P_n = (p_{n1}, \ldots, p_{nk}) \in M(k)$, which may depend on the number of balls $n$. This results in frequency counts $X_{n1}, \ldots, X_{nk}$ such that the vector $\boldsymbol{X}_n = (X_{n1}, \ldots, X_{nk}) \in \{0, 1, \ldots\}^k$ is multinomially distributed with parameters $k$, $n$ and $P_n$. The problem is to decide on the basis of observations $\boldsymbol{X}_n$ whether the unknown law $P_n$ is equal to a given $Q = (q_1, ..., q_k) \in M(k)$ or not.

The observations $\boldsymbol{X}_n$ are represented by the empirical distribution

$$\hat{P}_n = \left(\hat{p}_{n1} \triangleq X_{n1}/n, \ ..., \ \hat{p}_{nk} \triangleq X_{nk}/n\right) \in M(k|n) \quad (1)$$

and a procedure $\mathcal{T}$ on accepting or rejecting a hypothesis based on $\hat{P}_n$ is called a test. The test uses a statistic $T_n(\hat{P}_n, Q)$, which characterizes the goodness-of-fit between the distributions $\hat{P}_n$ and $Q$. The test $\mathcal{T}$ rejects the hypothesis $P_n = Q$ if $T = T_n(\hat{P}_n, Q)$ exceeds a certain level $r_n \in \mathbb{R}$.

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

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

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

$$D_\alpha(P, Q) = \sum_{j=1}^{k} q_j\, \phi_\alpha\left(\frac{p_j}{q_j}\right), \quad \alpha \in \mathbb{R}, \quad (3)$$

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

$$\phi_\alpha(t) = \begin{cases} \frac{t^\alpha - \alpha(t-1) - 1}{\alpha(\alpha-1)}, & \text{for } \alpha \notin \{0, 1\}; \\ -\ln t + t - 1, & \text{for } \alpha = 0; \\ t \ln t - t + 1, & \text{for } \alpha = 1. \end{cases} \quad (4)$$

For details about definition (3) and the properties of power divergences, see [2] and [3]. Important examples of statistics based on power divergences are the Pearson statistic ($\alpha = 2$), the Neyman statistic ($\alpha = -1$), the log-likelihood ratio ($\alpha = 1$), the reversed log-likelihood ratio ($\alpha = 0$) and the Freeman-Tukey statistic ($\alpha = 1/2$). In what follows it is sometimes more convenient to use the statistic $D_\alpha(\hat{P}_n, Q)$ instead of the one-one related $T_\alpha = T_{\alpha,n}$.

In this paper we deal with the question of which of the power divergence statistics $T_\alpha, \alpha \in \mathbb{R}$ is preferable for testing the hypothesis that the true distribution is uniform, i.e. the hypothesis $\mathcal{H} : P_n = U \triangleq (1/k, ..., 1/k) \in M(k|n)$. Then

$$\boldsymbol{X}_n \sim Multinomial_k(n, U) \quad \text{under } \mathcal{H}. \quad (5)$$

The alternative to the hypothesis $\mathcal{H}$ is denoted by $\mathcal{A}_n$, i.e.,

$$\boldsymbol{X}_n \sim Multinomial_k(n, P_n) \quad \text{under } \mathcal{A}_n. \quad (6)$$

*Example 1:* Let $\mu$, $\nu$ be probability measures on the Borel line $(\mathbb{R}, \mathcal{B})$ with continuous distribution functions $F$, $G$ and $Y_1, \ldots, Y_n$ an 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 $\boldsymbol{X}_n = (X_{n1}, \ldots, X_{nk})$ 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 $\boldsymbol{X}_n$ 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 generated by $P_n = (\mu(A_{n1}), \ldots, \mu(A_{nk}))$ 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 $1 \le j \le k$ 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 $1 \le j \le k$. For this partition we get

$$P_n = U = (1/k, ..., 1/k) \in M(k|n) \quad \text{under } \mathcal{H} \quad (7)$$

978-1-4244-2571-6/08/$25.00 ©2008 IEEE

and

$$P_n = (\mu(A_{n1}), \ldots, \mu(A_{nk})) \in M(k) \quad \text{under } \mathcal{A}_n. \tag{8}$$

We see from (7) and (8) that the partitions $\mathcal{P}_n$ generated by quantiles lead exactly to the situation assumed in (5) - (6).

Note that the sequence $P_n$ is contiguous to the sequence of uniform distributions because $\mu$ is absolutely continuous to $\nu$.

*Example 2:* It is often convenient to replace the power divergences $D_\alpha(P, Q)$ of the orders $\alpha > 0$ by the one-one related Rényi divergences [2,4] given by

$$D_\alpha(P\|Q) = \frac{\log\left(\sum p_i^\alpha q_i^{1-\alpha}\right)}{\alpha - 1} \quad \text{when } \alpha \neq 1 \tag{9}$$

and

$$D_1(P\|Q) = \lim_{\alpha\to 1} D_\alpha(P\|Q) = D_1(P, Q). \tag{10}$$

The formulas for divergences $D_\alpha(P\|Q)$ simplify when $Q = U$, e.g.,

$$D_\alpha(P\|U) = \log k - H_\alpha(P) \quad \text{for } P \in M(k) \tag{11}$$

for the Rényi entropy $H_\alpha(P)$ of order $\alpha > 0$, where $H_\alpha(P) = (\alpha - 1)^{-1} \sum p_i^\alpha$ when $\alpha \neq 1$ and $H_1(P) = \lim_{\alpha\to 1} H_\alpha(P)$ coincides with the Shannon entropy $H(P)$.

## II. Bahadur efficiency

In this short report we focus on the typical situation where $k = k_n$ depends on $n$ and increases in the sense $k \to \infty$, but not too fast, so that the average number of observations per bin tends to infinity, i.e.

$$\lim_{n\to\infty} \frac{n}{k} = \infty. \tag{12}$$

This condition implies that $T_\alpha$ is asymptotically Gaussian under the hypothesis $\mathcal{H}$ [5] so that it is easy to calculate for which values of the statistic $T_\alpha$ the hypothesis should be accepted or rejected at a specified significance level $s \in ]0;1[$.

We are interested in the relative asymptotic efficiencies of the power divergence statistics $T_{\alpha_1}$ and $T_{\alpha_2}$ for $0 < \alpha_1 < \alpha_2 < \infty$ . The condition (12) implies that the *Pitman asymptotic relative efficiencies* of all statistics $T_\alpha$, $\alpha \in \mathbb{R}$ coincide [3]. In this situation preferences between these statistics must be based on the Bahadur efficiencies $BE(T_{\alpha_1} \mid T_{\alpha_2})$. Quine and Robinson [1] demonstrated that the log-likelihood ratio statistic $T_1$ is infinitely more Bahadur efficient than the Pearson statistic $T_2$. In [6] we proved that $T_1$ is more Bahadur efficient than any statistic $T_\alpha$ with $\alpha > 1$.

A problem left open in [6] is the relative Bahadur efficiency of the remaining statistics $T_\alpha$, $\alpha \in \mathbb{R}$, in particular the conjecture that the log-likelihood ratio statistic is most Bahadur efficient in the class of all power divergence statistics $T_\alpha$, $\alpha \in \mathbb{R}$. In this paper we solve these problems in the domain $\alpha > 0$. Before defining the Bahadur efficiency, we introduce some important auxiliary concepts.

*Definition 3:* For $\alpha \in \mathbb{R}$ we say that

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

$$\lim_{n\to\infty} D_\alpha(P_n, U) = \Delta_\alpha . \tag{13}$$

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

$$D_\alpha(\hat{P}_n, U) \xrightarrow{p} 0 \quad \text{under } \mathcal{H} \tag{14}$$

while

$$D_\alpha(\hat{P}_n, U) \xrightarrow{p} \Delta_\alpha \quad \text{under } \mathcal{A}_n. \tag{15}$$

Note that the consistency condition (14) is slightly weaker than the one used in [1] or [6], allowing us to get better consistency results without impact on the interpretation and evaluation of the Bahadur efficiency.

The Bahadur condition (13) 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 $\mathcal{H}$. It follows from [7] that the Bahadur condition holds for the model of Example 1.

There is a slight difference in the meaning of (13) when $\alpha \geq 1$ and $\alpha \in ]0;1[$. For $\alpha \geq 1$ the Bahadur condition implies that the sequence of alternatives $P_n$ is asymptotically contiguous with respect to the sequence $U$ in the Le Cam's sense (in the present model this means $\lim_{n\to\infty} p_{nj_n} = 0$ for every sequence $1 \leq j_n \leq k$). This is not the case for $\alpha \in ]0;1[$ where the Bahadur condition is fulfilled even for Dirac's distributions concentrated on single points, and thus asymptotically entirely separated from the sequence of uniform distributions.

The consistency of $D_\alpha(\hat{P}_n, U)$ introduced in (14), (15) 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 a fixed significance level $s \in ]0;1[$ has a power tending to 1. Indeed, under $\mathcal{H}$ we have $D_\alpha(\hat{P}_n, U) \xrightarrow{p} 0$ so that the rejection level of the $D_\alpha(\hat{P}_n, U)$-based test of an asymptotic significance level $s$ tends to 0 for $n \to \infty$ while under $\mathcal{A}_n$ we have $D_\alpha(\hat{P}_n, U) \xrightarrow{p} \Delta_\alpha > 0$.

The above considered 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. In the sequel $\mathsf{P}(B_n)$ denotes the probability of events $B_n$ depending on the random observations $\boldsymbol{X}_n$ (cf. (5) and (6)) and $\mathsf{E}$ the corresponding expectation.

*Definition 4:* If for $\alpha \in \mathbb{R}$ there exists a sequence $c_{\alpha,n} > 0$ and a continuous function $g_\alpha : ]0;\infty[ \to ]0;\infty[$ such that under $\mathcal{H}$

$$\lim_{n\to\infty} -\frac{c_{\alpha,n}}{n} \ln \mathsf{P}(D_\alpha(\hat{P}_n, U) \geq \Delta) = g_\alpha(\Delta) \tag{16}$$

then we say that $g_\alpha$ is the Bahadur function of the statistic $D_\alpha(\hat{P}_n, U)$ generated by $c_{\alpha,n}$.

Next follows the basic definition of the present paper where $\Delta_{\alpha_i}$ are the limits from the Bahadur condition and $g_{\alpha_i}$ and $c_{\alpha_i,n}$ are the functions and sequences from the definition of the Bahadur function.

*Definition 5:* Let for every $\alpha \in \{\alpha_1, \alpha_2\}$ the Bahadur condition hold with a limit $\Delta_\alpha$ considered in (13) and let the statistic $D_\alpha(\hat{P}_n, U)$ be consistent with Bahadur function $g_\alpha$ generated by a sequence $c_{\alpha,n}$. Then the *Bahadur efficiency* $BE(T_{\alpha_1} \mid T_{\alpha_2})$ of the corresponding power divergence statistic $T_{\alpha_1}$ with respect to $T_{\alpha_2}$ is given by the formula

$$BE(T_{\alpha_1} \mid T_{\alpha_2}) = \frac{g_{\alpha_1}(\Delta_{\alpha_1})}{g_{\alpha_2}(\Delta_{\alpha_2})} \lim_{n\to\infty} \frac{c_{\alpha_2,n}}{c_{\alpha_1,n}} \tag{17}$$

provided the limit exists in the extended halfline $[0, \infty]$.[1]

Let the assumptions of the last definition hold. Then the consistency implies that both the $T_{\alpha_i}$-tests of the uniformity hypothesis $\mathcal{H} : U$ will achieve identical powers

$$\pi = \mathsf{P}(D_{\alpha_i}(\hat{P}_n, U) \geq r_{n,i}) \quad \text{for} \quad \pi \in ]0,1[ \quad \text{and} \quad i = 1, 2$$

under $\mathcal{A}_n$ if and only if $r_{n,i} \to \Delta_{\alpha_i}$ for $i \in \{1, 2\}$ as $n \to \infty$. The convergence $r_{n,i} \to \Delta_{\alpha_i}$ leads to the approximate $T_{\alpha_i}$-test significance levels

$$s_{n,i} \triangleq \mathsf{P}(D_{\alpha_i}(\hat{P}_n, U) \geq \Delta_{\alpha_i}) \approx \mathsf{P}(D_{\alpha_i}(\hat{P}_n, U) \geq r_{n,i})$$

for $i = 1, 2$ under $\mathcal{H}$ where $s_{n,i} \to 0$ as $n \to \infty$ for $i = 1, 2$ under $\mathcal{H}$. By (16), the $T_{\alpha_i}$-tests need different sample sizes

$$n_i = \frac{c_{\alpha_i,n}}{g_{\alpha_i}(\Delta_{\alpha_i})} \ln \frac{1}{s_n}, \quad i \in \{1, 2\} \tag{18}$$

to achieve the same approximate test significance levels $s_n = s_{n,1} = s_{n,2}$ with $n$ playing here the role of a formal parameter increasing to $\infty$.

## III. Consistency

The main theorem in this section presents consistency conditions for all power divergence statistics $D_\alpha(\hat{P}_n, U)$, $\alpha > 0$. It extends and refines Theorem 1 in [6]. Consistency conditions for more general classes of $\phi$-divergence statistics can be found in [8].

Throughout this section we skip the subscript $n$ in the probabilities $\hat{p}_{nj}$ and $p_{nj}$ and we use the following fact about the variational distances $V(\hat{P}_n, P_n)$.

*Lemma 6:* If $\lim_{n\to\infty} n/k = \infty$ then $\mathsf{E}\left[V\left(\hat{P}_n, P_n\right)\right] \to 0$ for $n \to \infty$.

*Proof:* By the Cauchy–Schwarz inequality,

$$V(\hat{P}_n, P_n) = \sum_{j=1}^{k} \left|\frac{\hat{p}_j}{p_j} - 1\right| p_j = \sum_{j=1}^{k} \left|\frac{\hat{p}_j}{p_j} - 1\right| p_j^{1/2} p_j^{1/2} \leq$$

$$\left(\sum_{j=1}^{k} \left|\frac{\hat{p}_j}{p_j} - 1\right|^2 p_j\right)^{1/2} \cdot \left(\sum_{j=1}^{k} p_j\right)^{1/2}$$

$$= \left(D_2\left(\hat{P}_n, P_n\right)\right)^{1/2}.$$

Hence

$$\mathsf{E}\left[V\left(\hat{P}_n, P_n\right)\right] \leq \mathsf{E}\left[D_2\left(\hat{P}_n, P_n\right)\right]^{1/2} = \left(\frac{k-1}{n}\right)^{1/2}.$$

■

*Theorem 7:* Let the Bahadur condition (13) hold for all $\alpha \geq 1$. Then $D_\alpha(\hat{P}_n, U)$ is consistent if

$$\alpha \in ]0; 2] \quad \text{and} \quad \lim_{n\to\infty} \frac{n}{k} = \infty, \tag{19}$$

or

$$\alpha > 2 \quad \text{and} \quad \lim_{n\to\infty} \frac{n}{k \log k} = \infty. \tag{20}$$

*Proof:* First consider $\alpha < 1$. The function $\phi_\alpha$ is uniformly continuous on $[0; \infty[$. Thus for any $\varepsilon > 0$ there exists $\delta > 0$ such that all $t, s > 0$ satisfy the relation $|\phi_\alpha(t) - \phi_\alpha(s)| \leq \delta |t - s| + \varepsilon$. Therefore

$$\left|D_\alpha(\hat{P}_n, U) - D_\alpha(P_n, U)\right|$$

$$\leq \sum_{j=1}^{k} \frac{1}{k} \left|\phi_\alpha\left(\frac{\hat{p}_j}{1/k}\right) - \phi_\alpha\left(\frac{p_j}{1/k}\right)\right|$$

$$\leq \sum_{j=1}^{k} \frac{1}{k} \left(\delta \left|\frac{\hat{p}_j}{1/k} - \frac{p_j}{1/k}\right| + \varepsilon\right) = \delta V(\hat{P}_n, P_n) + \varepsilon.$$

Consequently

$$\mathsf{E}\left|D_\alpha(\hat{P}_n, U) - D_\alpha(P_n, U)\right| \leq \delta \mathsf{E}\left[V(\hat{P}_n, P_n)\right] + \varepsilon$$

and

$$\lim_{n\to\infty} \sup \mathsf{E}\left|D_\alpha(\hat{P}_n, U) - D_\alpha(P_n, U)\right| \leq \varepsilon.$$

Since this is true for all $\varepsilon > 0$, the desired relation

$$\lim_{n\to\infty} \mathsf{E}\left|D_\alpha(\hat{P}_n, U) - D_\alpha(P_n, U)\right| = 0$$

holds.

For $\alpha \in [1; 2]$ the desired result was proved in [6]. For $\alpha > 2$ we use the Taylor expansion

$$\hat{p}_j^\alpha = p_j^\alpha + \alpha\, p_j^{\alpha-1}(\hat{p}_j - p_j) + \frac{\alpha(\alpha-1)}{2}\, \xi_j^{\alpha-2}(\hat{p}_j - p_j)^2 \tag{21}$$

where $\xi_j$ is between $p_j$ and $\hat{p}_j$. We need a highly probable upper bound on $\hat{p}_j$. For this choose some number $b > 1$ and consider the random event

$$E_{nj}(b) = \{\hat{p}_j \geq b \max\{p_j, 1/k\}\}.$$

Obviously,

$$\mathsf{P}\left(\cup_j E_{nj}(b)\right) \leq \sum_j \mathsf{P}\left(\hat{p}_j \geq b \max\{p_j, 1/k\}\right)$$

and a large deviation bound on $\mathsf{P}\left(\hat{p}_j \geq b \max\{p_j, 1/k\}\right)$ gives that $\mathsf{P}\left(\cup_j E_{nj}(b)\right) \to 0$ for $n \to \infty$ [8]. Therefore it suffices to prove (13) under the condition that the random events $\cup_j E_{n,j}(b)$ fail to take place, i.e. that

$$\hat{p}_j > b \max\{p_j, 1/k\} \quad \text{for all} \quad 1 \leq j \leq k. \tag{22}$$

[1]In [6] due to a missprint, $\alpha_1$ and $\alpha_2$ were interchanged behind the limit in the Bahadur efficiency formula similar to (17), but it was used in the present correct form.

Under (22) it holds $\xi_j \le \{bp_j, b/k\}$ and, consequently,

$$\xi_j^{\alpha-2} \le (\max\{bp_j, b/k\})^{\alpha-2} \le b^{\alpha-2} p_j^{\alpha-2} + \frac{b^{\alpha-2}}{k^{\alpha-2}}. \quad (23)$$

However, (21) together with (23) implies

$$\left|\hat{p}_j^\alpha - p_j^\alpha\right| \le \alpha\, p_j^{\alpha-1} \left|\hat{p}_j - p_j\right| + \frac{\alpha(\alpha-1)b^{\alpha-2}}{2} \left(p_j^{\alpha-2} + \frac{1}{k^{\alpha-2}}\right)(\hat{p}_j - p_j)^2.$$

The rest of the proof is the same as the proof of [6, Thm. 1, Eq. 40]. ■

## IV. Bahadur function

For $\alpha \in ]0;1]$ let us first calculate the Bahadur function of the Rényi statistics $D_\alpha(\hat{P}_n \| U)$ (cf. (9), (10)). By the argument given in [6], if

$$\lim_{n\to\infty} \frac{n}{k \ln n} = \infty \quad (24)$$

and

$$\inf_{D_\alpha(P\|U)\ge\Delta} D_1(P\|U) = \mathfrak{g}_\alpha(\Delta) \in ]0;\infty[, \qquad \Delta > 0$$

then $\mathfrak{g}_\alpha$ is the desired Bahadur function generated by the constant sequence $c_{n,\alpha} \equiv 1$. However, for $\alpha \in ]0;1]$ the inequality $D_\alpha(P\|U) \le D_1(P\|U)$ implies

$$\inf_{D_\alpha(P\|U)\ge\Delta} D_1(P\|U) \ge \inf_{D_\alpha(P\|U)\ge\Delta} D_\alpha(P\|U) = \Delta.$$

On the other hand, by [9], the minimum of $D_1(P\|U)$ is under the condition $D_\alpha(P\|U) \ge \Delta$ is achieved at a mixture $P$ of two uniform distributions $U_l$ and $U_{l+1}$ supported by nested sets of sizes $l$ and $l+1$ respectively. The support size $l$ is determined by the condition

$$\log\frac{k}{l+1} \le \Delta \le \log\frac{k}{l}$$

where

$$\log\frac{k}{l} - \Delta \le \log\frac{k}{l} - \log\frac{k}{l+1} \le \frac{1}{l}.$$

If $P = sU_l + (1-s)U_{l+1}$ then

$$D_1(P\|U) \le sD_1(U_l\|U) + (1-s)D_1(U_{l+1}\|U) = s\log\frac{k}{l} + (1-s)\log\frac{k}{l+1} \le \log\frac{k}{l} \le \Delta + \frac{1}{l}\ .$$

For a fixed value of $\Delta$ the support size $l$ increases to infinity when $k \to \infty$. Hence

$$\inf_{D_\alpha(P\|U)\ge\Delta} D_1(P\|U) \to \Delta \quad \text{for } k \to \infty$$

and the desired Bahadur function is

$$\mathfrak{g}_\alpha(\Delta) = \Delta. \quad (25)$$

Replacing Rényi statistics $D_\alpha(\hat{P}_n \| U)$ by the power divergence statistics $D_\alpha(\hat{P}_n, U)$, $\alpha \in ]0;1]$ we get

$$D_1(P,U) = \log\frac{k}{l}$$

if $\alpha = 1$ and otherwise

$$D_\alpha(P,U) = \frac{\sum_{i=1}^{l}\left(\frac{1/l}{1/k}\right)^\alpha \frac{1}{k} - 1}{\alpha(\alpha-1)} = \frac{1-\left(\frac{k}{l}\right)^{\alpha-1}}{\alpha(1-\alpha)}.$$

This implies for $0<\alpha<1$

$$\frac{k}{l} = (1-\alpha(1-\alpha)D_\alpha(P,U))^{\frac{1}{\alpha-1}}$$

and, consequently,

$$\inf_{D_\alpha(P,U)\ge\Delta} D_1(P,U) \approx \log\frac{k}{l} = \log\left((1-\alpha(1-\alpha)\Delta)^{\frac{1}{\alpha-1}}\right) = \frac{1}{\alpha-1}\log(1+\alpha(\alpha-1)\Delta).$$

An argument similar to above leads to the conclusion that under (24) the constant sequence $c_{n,\alpha} \equiv 1$ generates for the power divergence statistics $D_\alpha(\hat{P}_n, U)$, $\alpha \in ]0;1]$ the Bahadur functions

$$g_\alpha(\Delta) = \frac{1}{\alpha-1}\log(1+\alpha(\alpha-1)\Delta) \quad \alpha < 1 \quad (26)$$

and

$$g_1(\Delta) = \lim_{\alpha\uparrow 1} g_\alpha(\Delta) = \Delta. \quad (27)$$

## V. Main result and discussion

For the power divergences $D_\alpha(\hat{P}_n, U)$ of the orders $\alpha \ge 1$ we evaluated the Bahadur functions functions $g_\alpha$ and the corresponding generating sequences $c_{\alpha,n}$ in [6]. This enabled us to evaluate the explicit Bahadur efficiencies $BE(\mathcal{T}_{\alpha_1} \mid \mathcal{T}_{\alpha_2})$ of the power divergence statistics $\mathcal{T}_{\alpha_1}, \mathcal{T}_{\alpha_2}$ in the domain $\alpha_1, \alpha_2 \ge 1$. The results of Section IV enable us to extend the corresponding formulas $BE(\mathcal{T}_{\alpha_1} \mid \mathcal{T}_{\alpha_2})$ on the domain $\alpha_1, \alpha_2 > 0$. These formulas, together with formulas for the Bahadur efficiencies of some Rényi divergence statistics

$$\mathcal{T}_{\alpha_1} = 2nD_{\alpha_1}(\hat{P}_n||U) \quad \text{and} \quad \mathcal{T}_{\alpha_2} = 2nD_{\alpha_2}(\hat{P}_n||U) \quad (28)$$

are given in the following main result of this paper.

*Theorem 8:* Let the Bahadur condition (13) hold for some $0 < \alpha_1 < \alpha_2 < \infty$ with the corresponding limits $\Delta_{\alpha_1}, \Delta_{\alpha_2}$.

**(i)** If $\alpha_2 \le 1$ and the assumption (24) holds then the Rényi divergence statistics (28) satisfy the relation

$$BE(\mathcal{T}_{\alpha_1} \mid \mathcal{T}_{\alpha_2}) = \Delta_{\alpha_1}/\Delta_{\alpha_1}. \quad (29)$$

**(ii)** If $\alpha_2 \le 1$ and the assumption (24) holds then the power divergence statistics $\mathcal{T}_{\alpha_1}$ and $\mathcal{T}_{\alpha_2}$ given by (2) satisfy the relation

$$BE(T_{\alpha_1} \mid T_{\alpha_2}) = \frac{(\alpha_2-1)\log(1+\alpha_1(\alpha_1-1)\Delta_{\alpha_1})}{(\alpha_1-1)\log(1+\alpha_2(\alpha_2-1)\Delta_{\alpha_2})}$$

if $\alpha_2 < 1$ and

$$BE(T_{\alpha_1} \mid T_1) = \frac{\log(1+\alpha_1(\alpha_1-1)\Delta_{\alpha_1})}{(\alpha_1-1)\Delta_{\alpha_2}}$$

otherwise.

**(iii)** If $\alpha_2 > 1$ and the assumption

$$\lim_{n\to\infty} \frac{n}{k^{2-1/\alpha_2} \ln n} \tag{30}$$

holds then the power divergence statistics (2) satisfy the relation

$$BE(\mathcal{T}_{\alpha_1} \mid \mathcal{T}_{\alpha_2}) = \infty. \tag{31}$$

*Proof:* **(i), (ii):** Condition (24) implies that $D_{\alpha_j}(\hat{P}_n, U)$ as well as $D_{\alpha_j}(\hat{P}_n||U)$ are consistent for $j = 1, 2$. The rest follows from the definition of Bahadur efficiency and from the results of Section IV.

**(iii):** According to [6] it is sufficient to prove that (30) is sufficient for the consistency of $D_{\alpha_1}(\hat{P}_n, U)$ and $D_{\alpha_2}(\hat{P}_n, U)$. But

$$\frac{n}{k \log k} = \frac{n}{k^{2-1/\alpha_2} \ln n} \cdot \frac{k^{1-1/\alpha_2}}{\log k} \ln n$$

where under (30) each factor tends to infinity when $n \to \infty$. ■

The above theorem may be interpreted in two ways. One interpretation is that all statistics $D_\alpha\left(\hat{P}_n, U\right)$, $\alpha \in ]0;1]$ are equally efficient and an other interpretation is that the definition of efficiency is not sufficiently refined to distinguish the different statistics. Here we shall not introduce a new general definition of efficiency but will discuss an example where $D_1$ is better than $D_\alpha, \alpha \in ]0;1[$ in distinguishing $U$ from certain alternatives.

*Example 9:* We consider alternatives $\mathcal{A}_n$ defined by (6) where $P_n \in M(k)$ is given by

$$P_n = s_n \delta + (1 - s_n) U.$$

We shall fix the power of the tests to some number $p \in ]0;1[$. Then the acceptance region will be approximately

$$\left\{ \hat{P}_n | D_\alpha\left(\hat{P}, U\right) \le D_1(P_n, U) \right\}.$$

Under the null hypothesis the significance level will satisfy

$$\frac{\log P\left\{\hat{P}_n | D_\alpha\left(\hat{P}, U\right) > D_1(P_n, U)\right\}}{n} \approx \inf_{P \in \left\{\hat{P}_n | D_\alpha\left(\hat{P}, U\right) > D_1(P_n, U)\right\}} D_1(P, U).$$

The infimum is achieved for a mixture of uniform distributions on $l$ and $l+1$ points And for such a distribution we have $D_1(P, U) \approx D_\alpha(P, U)$ implying that

$$\inf_{P \in \left\{\hat{P}_n | D_\alpha\left(\hat{P}, U\right) > D_1(P_n, U)\right\}} D_1(P, U) \approx D_1(P_n, U)$$

Thus the order of the sample size is given by

$$n_\alpha \approx \frac{\log P\left\{\hat{P}_n | D_\alpha\left(\hat{P}, U\right) > D_1(P_n, U)\right\}}{D_\alpha(P_n, U)}$$

and for $\alpha_1, \alpha_2 \in ]0;1]$ we have

$$\lim \frac{n_{\alpha_1}}{n_{\alpha_2}} = \lim \frac{\frac{\log P\left\{\hat{P}_n | D_{\alpha_2}\left(\hat{P}, U\right) > D(P_n, U)\right\}}{D_{\alpha_1}(P_n, U)}}{\frac{\log P\left\{\hat{P}_n | D_{\alpha_2}\left(\hat{P}, U\right) > D(P_n, U)\right\}}{D_{\alpha_2}(P_n, U)}} = \lim \frac{D_{\alpha_2}(P_n, U)}{D_{\alpha_1}(P_n, U)}.$$

For simplicity choose $\alpha_1 = 1$ and $s_n$ such that $D_1(P_n, U)$ is constant. Then

$$\lim \frac{n_{\alpha_1}}{n_{\alpha_2}} = \frac{\lim D_{\alpha_2}(P_n, U)}{D_1(P_n, U)}$$

and this limit is 0 which follows from calculations that were essentially carried out in [6, Eq. 66-75].

The conclusion of the example is that there exists a sequence of alternatives that is not contiguous with the uniform distributions such that Shannon entropy is asymptotically infinitely more efficient in characterizing the uniform distribution than any Rényi entropy of order $\alpha \in ]0;1[$. The standard definiton of Bahadur efficiency requires that the Bahadur condition is fulfilled which essentially rules out the possibility of a sequence of alternatives that is not contiguous. Therefore the apparent efficiency of Renyi entropies of order $\alpha \in ]0;1[$ is a direct consequence of the Bahadur condition that might be too stricht for many applications.

## VI. Acknowledgement

P. Harremoës was supported by European Pascal Network and a grant from the Danish Natural Science Council. I. Vajda was supported by grants GAČR 102/07/1131 and MŠMT 1M0572. The authors would also like to thank Łukasz Dębowski for comments that have improved the quality of this paper.

## References

[1] M. P. Quine and J. Robinson, "Efficiencies of chi-square and likelihood ratio goodness-of-fit tests.," *Ann. Statist.*, vol. 13, pp. 727–742, 1985.

[2] F. Liese and I. Vajda, *Convex Statistical Distances*. Leipzig: Teubner, 1987.

[3] T. R. C. Read and N. Cressie, *Goodness of Fit Statistics for Discrete Multivariate Data.* Berlin: Springer, 1988.

[4] A. Rényi, "On measures of entropy and information," in *Proc. 4th Berkeley Symp. Math. Statist. and Prob.*, vol. 1, (Berkely), pp. 547–561, Univ. Calif. Press, 1961.

[5] L. Györfi and I. Vajda, "Asymptotic distributions for goodness-of-fit statistics in a sequence of multinomial models," *Stat. Probab. Letters*, vol. 56, no. 1, pp. 57–67, 2002.

[6] P. Harremoës and I. Vajda, "On the Bahadur-efficient testing of uniformity by means of the entropy," *IEEE Trans. Inform Theory*, vol. 54, pp. 321–331, Jan. 2008.

[7] I. Vajda, "On convergence of information contained in quantized observations," *IEEE Trans. Inform. Theory*, vol. 48, pp. 2163–2172, Aug. 2002.

[8] P. Harremoës and I. Vajda, "Consistence of various $\phi$-divergence statistics," Research Report ÚTIA AV ČR 2218, Institute of Information Theory and Automation, Praha, March 2008.

[9] P. Harremoës and F. Topsøe, "Inequalities between entropy and index of coincidence derived from information diagrams," *IEEE Trans. Inform. Theory*, vol. 47, pp. 2944–2960, Nov. 2001.