On Bahadur Efficiency of Power Divergence Statistics

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

Abstract—It is proved that the information divergence statistic is infinitely more Bahadur efficient than the power divergence statistics of the orders $\alpha > 1$ as long as the sequence of alternatives is contiguous with respect to the sequence of null-hypotheses and the the number of observations per bin increases to infinity is not very slow. This improves the former result in Harremoës and Vajda (2008) where the the sequence of null-hypotheses was assumed to be uniform and the restrictions on on the numbers of observations per bin were sharper. Moreover, this paper evaluates also the Bahadur efficiency of the power divergence statistics of the remaining positive orders $0 < \alpha \leq 1$. The statistics of these orders are mutually Bahadur-comparable and all of them are more Bahadur efficient than the statistics of the orders $\alpha > 1$. A detailed discussion of the technical definitions and conditions is given, some unclear points are resolved, and the results are illustrated by examples.

Index Terms—Bahadur efficiency, consistency, power divergence, Rényi divergence.

I. INTRODUCTION

PROBLEMS of detection, classification and identification are often solved by the method of testing statistical hypotheses. Consider signals $Y_1, Y_2, ..., Y_n$ collected from a random source independently at time instants $i = 1, 2, ..., n$. Signal processing usually requires digitalization based on appropriate quantization. Quantization of the signal space $\mathcal{Y}$ into k disjoint cells (or bins) $\mathcal{Y}_{n1}, \mathcal{Y}_{n2}, ..., \mathcal{Y}_{nk}$ reduces the signals $Y_1, Y_2, ..., Y_n$ into simple k-valued indicators $I_n(Y_1), I_n(Y_2), ..., I_n(Y_n)$ of their cover cells. Various hypotheses about the data source represented by probability measures Q_n on $\mathcal{Y}$ are transformed by the quantization into discrete probability distributions

$$Q_n = (q_{n1} = Q(\mathcal{Y}_{n1}), ..., q_{nk} = Q(\mathcal{Y}_{nk}))$$

on the quantization cells where for no quantization cell $q_{nj} = 0$. These hypothetical distributions need not be the same as the true distributions $P_n = (p_{n1} = P(\mathcal{Y}_{n1}), ..., p_{nk} = P(\mathcal{Y}_{nk}))$. The latter distributions are usually unknown but, by the law of large numbers, they can be approximated by the empirical distributions (vectors of relative cell frequencies)

$$\hat{P}_n = \left(\hat{p}_{n1} = \frac{X_{n1}}{n}, ..., p_{nk} = \frac{X_{nk}}{n}\right) = \frac{\mathbf{X}_n}{n} \quad (1)$$

where X_{nj} is the numbers of the signals $Y_1, Y_2, ..., Y_n$ in $\mathcal{Y}_{nj}$. Formally,

$$X_{nj} = \sum_{i=1}^{n} 1_{\{Y_i \in \mathcal{Y}_{nj}\}} = \sum_{i=1}^{n} 1_{\{I_n(Y_i)=j\}}, \quad 1 \leq j \leq k \quad (2)$$

where 1_A denotes the indicator of the event A. The problem is to decide whether the signals $Y_1, Y_2, ..., Y_n$ are generated by the source $(\mathcal{Y}, Q)$ on the basis of the distributions $\hat{P}_n, Q_n$. A classical method for solving this problem is the method of testing statistical hypotheses in the spirit of Fisher, Neyman and Pearson. In our case the hypothesis is

$$\mathcal{H} : P_n = Q_n \quad (3)$$

and the decision is based either on the *likelihood ratio statistic*

$$\hat{T}_{1,n} = 2 \sum_{j=1}^{k} X_{nj} \ln \frac{X_{nj}}{nq_{nj}} \quad (4)$$

or the *Pearson χ^2-statistic*

$$\hat{T}_{2,n} = \sum_{j=1}^{k} \frac{(X_{nj} - nq_{nj})^2}{nq_{nj}} \quad (5)$$

in the sense that the hypothesis is rejected when the statistic is large, where "large" depends on the required decision error or risk [1].

It is easy to see (c.f. (13), (14) below) that the classical test statistics (4), (5) are of the form

$$\hat{T}_{\alpha,n} = 2n\hat{D}_{\alpha,n} \overset{\text{def}}{=} 2nD_\alpha\left(\hat{P}_n, Q_n\right), \quad \alpha \in \{1, 2\} \quad (6)$$

where $D_\alpha(P, Q)$ for arbitrary $\alpha > 0$ and distributions $P = (p_1, ..., p_k)$, $Q = (q_1, ..., q_k)$ denotes the divergence $D_{\phi_\alpha}(P, Q)$ of Csiszár [2] for the power function

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

and

$$\phi_1(t) = \lim_{\alpha \to 1} \phi_\alpha(t) = t \ln t - t + 1. \quad (8)$$

The power divergences

$$D_\alpha(P, Q) = \frac{1}{\alpha(\alpha - 1)} \left(\sum_{j=1}^{k} p_j^\alpha q_j^{1-\alpha} - 1\right) \quad \alpha \neq 1 \quad (9)$$

or the one-one related Rényi divergences [3]

$$D_\alpha(P\|Q) = \frac{1}{\alpha - 1} \ln \sum_{j=1}^{k} p_j^\alpha q_j^{1-\alpha} \quad \alpha \neq 1 \quad (10)$$

with the common information divergence limit

$$D_1(P, Q) = D_1(P\|Q) = \sum_{j=1}^{k} p_j \ln \frac{p_j}{q_j} \quad (11)$$

are often applied in various areas of information theory. In the present context of detection and identification one can mention

e.g. the work of Kailath [4] who used the Bhattacharryya distance

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

which is one-one related to the Hellinger divergence.

In practical applications it is important to use the statistic $\hat{D}_{\alpha_{\mathrm{opt}},n}$ which is optimal in a sufficiently wide class of divergence statistics $\hat{D}_{\alpha,n}$ containing the standard statistical proposals $\hat{D}_{1,n}$ and $\hat{D}_{2,n}$ appearing in (6). We addressed this problem previously [5]–[7]. Our solution confirmed the classical statistical result of Quine and Robinson [8] who proved that the likelihood ratio statistic $\hat{D}_{1,n}$ is more efficient in the Bahadur sense than the χ^2-statistic $\hat{D}_{2,n}$ and extended the results of Beirlant et al. [9] and Györfi et al. [10] dealing with Bahadur efficiency of several selected power divergence statistics. Namely, we evaluated the Bahadur efficiencies of the statistics $\hat{D}_{n\alpha}$ in the domain $\alpha \geq 1$ for the numbers $k = k_n$ of quantization cells slowly increasing with n when the hypothetical distributions Q_n are uniform and the alternative distributions P_n are *contiguous* in the sense that $\lim_{n\to\infty} D_\alpha\left(P_n, Q_n\right)$ exists and *identifiable* in the sense that this limit is positive. We found that the Bahadur efficiencies decrease with the power parameter in the whole domain $\alpha \geq 1$. In the present paper we sharpen this result by relaxing conditions on the rate of k_n and extend it considerably by admitting non-uniform hypothetical distributions Q_n and by evaluating the Bahadur efficiencies also in the domain $0 < \alpha \leq 1$.

II. BASIC MODEL

Let $M(k)$ denote the set of all probability distributions $P = (p_j : 1 \leq j \leq k)$ and

$$M(k|n) = \left\{P \in M(k) : nP \in \{0,1,\ldots\}^k\right\}$$

its subset called the set of types in information theory. We consider hypothetical distributions $Q_n = (q_{nj} : 1 \leq j \leq k) \in M(k)$ restricted by the condition $q_{nj} > 0$ and arbitrary alternative distributions $P_n = (p_{nj} : 1 \leq j \leq k) \in M(k)$. The $\{0,1,\ldots\}^k$-valued frequency counts $\mathbf{X}_n$ with coordinates introduced in (2) are *multinomially distributed* in the sense

$$\mathbf{X}_n \sim \mathrm{Mult}_k(n, P_n), n = 1,2,\ldots. \qquad (12)$$

Important components of the model are the empirical distributions $\hat{P}_n \in M(k|n)$ defined by (1). Finally, for arbitrary $P \in M(k)$ and arbitrary $Q \in M(k)$ with positive coordinates we consider the power divergences (9)-(8). For their properties we refer to [11]–[13]. In particular, for the empirical and hypothetical distributions $\hat{P}_n, Q_n$ we consider the power divergence statistics $\hat{D}_{\alpha,n} = D_\alpha\left(\hat{P}_n, Q_n\right)$ (c.f. (6)) defined by (9), (11) for all $\alpha > 0$.

Example 1: For $\alpha = 2$, $\alpha = 1$ and $\alpha = 1/2$ we get the special power divergence statistics

$$\hat{D}_{2,n} = \frac{1}{2} \sum_{j=1}^{n} \frac{(\hat{p}_{nj} - q_{nj})^2}{q_{nj}} = \frac{1}{2n}\hat{T}_{2,n}, \qquad (13)$$

$$\hat{D}_{1,n} = \sum_{j=1}^{n} \hat{p}_{nj} \ln \frac{\hat{p}_{nj}}{q_{nj}} = \frac{1}{2n}\hat{T}_{1,n}, \qquad (14)$$

$$\hat{D}_{1/2,n} = 2 \sum_{j=1}^{n} \left(\hat{p}_{nj}^{1/2} - q_{nj}^{1/2}\right)^2 \qquad (15)$$

For testing the hypothesis $\mathcal{H}$ of (3) are usually used the rescaled versions

$$\hat{T}_{\alpha,n} = 2n\hat{D}_{\alpha,n} \qquad (16)$$

distributed under $\mathcal{H}$ asymptotically χ^2 with $k - 1$ degrees of freedom if k is constant and asymptotically normally if $k = k_n$ slowly increases to infinity [14], [15, and references therein] . The statistics (13) and (14) rescaled in this manner were already mentioned in (5) and (4). In (15) is the Hellinger divergence statistics rescaled by $2n$ is known as *Freeman–Tukey statistic*

$$\hat{T}_{1/2,n} = 2n\hat{D}_{1/2,n} = 4 \sum_{j=1}^{k} ((X_{nj})^{1/2} - (nq_{nj})^{1/2})^2. \qquad (17)$$

a) Convention: Unless the hypothesis $\mathcal{H}$ is explicitly assumed, the random variables, convergences and asymptotic relations are considered under the alternative $\mathcal{A}$. Further, unless otherwise explicitly stated, the asymptotic relations are considered for $n \longrightarrow \infty$ and the symbols of the type

$$s_n \longrightarrow s \quad \text{and} \quad s_n(\mathbf{X}_n) \overset{p}{\longrightarrow} s$$

denote the ordinary numerical convergence and the stochastic convergence in probability for $n \longrightarrow \infty$.

In this paper we consider the following assumptions.

A1: The number of cells $k = k_n \leq n$ of the distributions from $M(k)$, $M(k|n)$ depends on the sample size n and increases to infinity. In the rest of the paper the subscript n is suppressed in the symbols containing k.

A2: The hypothetical distributions $Q_n = (q_{nj} > 0 : 1 \leq j \leq k)$ are regular in the sense that $\max_j q_{nj} \to 0$ for $n \to \infty$ and that there exists $\varrho > 0$ such that

$$q_{nj} > \frac{\varrho}{k} \quad \text{for all } 1 \leq j \leq k \text{ and } n = 1,2,\ldots. \qquad (18)$$

A3α: The alternative $\mathcal{A}$: $(P_n : n = 1,2,\ldots)$ is identifiable in the sense that there exits $0 < \Delta_\alpha < \infty$ such that

$$D_{\alpha,n} \overset{def}{=} D_\alpha(P_n, Q_n) \longrightarrow \Delta_\alpha \quad \text{under } \mathcal{A}. \qquad (19)$$

Under **A2**

$$-\ln q_{nj} < \ln \frac{k}{\varrho} \quad \text{and} \quad \ln^2 q_{nj} < \ln^2 \frac{k}{\varrho}. \qquad (20)$$

Further, logical complement to the hypothesis $\mathcal{H}$ is the alternative denoted by $\mathcal{A}$. By (3), under $\mathcal{A}$ the alternative distributions P_n differ from Q_n. Assumption $\mathbf{A3}\alpha$ means that the alternative distributions are neither too close to nor too distant from Q_n in the sense of D_α-divergence for given $\alpha > 0$. Since for all $n = 1, 2, \dots$

$$D_{\alpha,n} = D_\alpha(Q_n, Q_n) \equiv 0 \quad \text{so that } \Delta_\alpha = 0 \quad \text{under } \mathcal{H}$$

it is clear that the hypothesis $\mathcal{A}$ is under $\mathbf{A1}$, $\mathbf{A2}$, $\mathbf{A3}\alpha$ distinguished from the hypothesis $\mathcal{H}$ by achieving a positive D_α-divergence limit Δ_α. In what follows we use the abbreviated notations

$$\mathbf{A}(\alpha) = \{\mathbf{A1}, \mathbf{A2}, \mathbf{A3}\alpha\}, \tag{21}$$

$$\mathbf{A}(\alpha_1, \alpha_2) = \{\mathbf{A1}, \mathbf{A2}, \mathbf{A3}\alpha_1, \mathbf{A3}\alpha_2\} \tag{22}$$

for the combinations of assumptions.

Definition 2: Under $\mathbf{A}(\alpha)$ we say that the statistic $\widehat{D}_{\alpha,n}$ is *consistent* with parameter Δ_α appearing in (19) if

$$\widehat{D}_{\alpha,n} \xrightarrow{p} \Delta_\alpha \quad \text{under } \mathcal{A} \tag{23}$$

and

$$\widehat{D}_{\alpha,n} \xrightarrow{p} 0 \quad \text{under } \mathcal{H} \tag{24}$$

i.e. if $\widehat{D}_{\alpha,n} \xrightarrow{p} \Delta_\alpha$ under both $\mathcal{A}$ and $\mathcal{H}$. If (24) is replaced by the stronger condition that the expectation $\mathsf{E}\widehat{D}_{\alpha,n}$ tends to zero under $\mathcal{H}$, in symbols

$$\mathsf{E}\left[\widehat{D}_{\alpha,n}\Big|\mathcal{H}\right] \longrightarrow 0, \tag{25}$$

then $\widehat{D}_{\alpha,n}$ is said *strongly consistent*.

Definition 3: We say that the statistic $\widehat{D}_{\alpha,n}$ is *Bahadur stable* if there is a continuous function with a Bahadur relative function $\varrho_\alpha :]0, \infty[^2 \to]0, \infty[$ such that the probability of error function

$$\mathsf{e}_{\alpha,n}(\Delta) = \mathsf{P}\left(\widehat{D}_{\alpha,n} > \Delta\Big|\mathcal{H}\right), \quad \Delta > 0 \tag{26}$$

corresponding to the test rejecting $\mathcal{H}$ when $\widehat{D}_{\alpha,n} > \Delta$ satisfies for all $\Delta_1, \Delta_2 > 0$ the relation

$$\frac{\ln \mathsf{e}_{\alpha,n}(\Delta_1)}{\ln \mathsf{e}_{\alpha,n}(\Delta_2)} \longrightarrow \varrho_\alpha(\Delta_1, \Delta_2).$$

If this condition holds then ϱ_α is called the *Bahadur relative function*.

Obviously, the Bahadur relative functions are multiplicative in the sense

$$\varrho_\alpha(\Delta_1, \Delta_2)\,\varrho_\alpha(\Delta_2, \Delta_3) = \varrho_\alpha(\Delta_1, \Delta_3).$$

Statistics that are Bahadur stable have the nice property that the asymptotic behavior of the error function $\mathsf{e}_{\alpha,n}(\Delta)$ is determined by its behavior for just a single argument $\Delta^* > 0$. Indeed, if $\widehat{D}_{\alpha,n}$ is Bahadur stable and if we define for a fixed $\Delta^* > 0$ the sequence

$$c_\alpha^*(n) = -\frac{n}{\ln \mathsf{e}_{\alpha,n}(\Delta^*)} \tag{27}$$

then for all $\Delta > 0$

$$-\frac{c_\alpha^*(n)}{n} \ln \mathsf{e}_{\alpha,n}(\Delta) \longrightarrow \varrho_\alpha(\Delta, \Delta^*) \quad \text{for all } \Delta > 0.$$

Moreover, if the expressions $-c_\alpha(n)/n \ln \mathsf{e}_{\alpha,n}(\Delta)$ converge for a sequence $c_\alpha(n)$ then the ratio $c_\alpha(n)/c_\alpha^*(n)$ tends to a constant.

b) Motivation of the next definition: Suppose that condition $\mathbf{A}(\alpha_1, \alpha_2)$ holds and denote for each $\alpha \in \{\alpha_1, \alpha_2\}$ and $n = 1, 2, \dots$ by $\Delta_\alpha + \varepsilon_{\alpha,n}$ the critical value of the statistics $\widehat{D}_{\alpha_i,n}$ leading to the rejection of $\mathcal{H}$ with a fixed power $0 < \mathsf{p} < 1$. In other words, let

$$\mathsf{p} = \mathsf{P}\left(\widehat{D}_{\alpha,n} > \Delta_\alpha + \varepsilon_{\alpha,n}\right) \quad \text{for all } n = 1, 2, \dots$$

where the sequence $\varepsilon_{\alpha,n} = \varepsilon_{\alpha,n}(\mathsf{p})$ depends on the fixed p. Since the assumed consistency of $\widehat{D}_{\alpha,n}$ implies that $\varepsilon_{\alpha,n}$ tends to zero, the corresponding error probabilities $\mathsf{e}_{\alpha,n}(\Delta_\alpha + \varepsilon_{\alpha,n}) = \mathsf{P}\left(\widehat{D}_{\alpha,n} > \Delta_\alpha + \varepsilon_{\alpha,n}\Big|\mathcal{H}\right)$ can be approximated by $\mathsf{e}_{\alpha,n}(\Delta_\alpha) = \mathsf{P}\left(\widehat{D}_{\alpha,n} > \Delta_\alpha\Big|\mathcal{H}\right)$. By (33),

$$-\frac{c_\alpha(n)}{n} \ln \mathsf{e}_{\alpha,n}(\Delta_\alpha) \longrightarrow g_\alpha(\Delta_\alpha).$$

Hence the error $\mathsf{e}_{\alpha_1,n}(\Delta_{\alpha_1})$ of the statistic $\widehat{D}_{\alpha_1,n}$ tends to zero with the same exponential rate as $\mathsf{e}_{\alpha_2,m_n}(\Delta_{\alpha_2})$ achieved by $\widehat{D}_{\alpha_2,m_n}$ for possibly different sample sizes $m_n \neq n$ with the property $m_n \longrightarrow \infty$ if the corresponding error exponents

$$g_{\alpha_1}(\Delta_{\alpha_1})\frac{n}{c_{\alpha_1}(n)} \quad \text{and} \quad g_{\alpha_2}(\Delta_{\alpha_2})\frac{m_n}{c_{\alpha_2}(m_n)} \tag{28}$$

tend to infinity with the same rate in the sense

$$\frac{m_n}{c_{\alpha_2}(m_n)} = \frac{g_{\alpha_1}(\Delta_{\alpha_1})}{g_{\alpha_2}(\Delta_{\alpha_2})} \cdot \frac{n}{c_{\alpha_1}(n)}\left(1 + o(1)\right). \tag{29}$$

The sample sizes m_n and n needed by the statistics $\widehat{D}_{\alpha_2,n}$ and $\widehat{D}_{\alpha_1,n}$ to achieve the same rate of convergence of errors are thus mutually related by the formula

$$\frac{m_n}{n} = \frac{g_{\alpha_1}(\Delta_{\alpha_1})}{g_{\alpha_2}(\Delta_{\alpha_2})} \cdot \frac{c_{\alpha_2}(m_n)}{c_{\alpha_1}(n)}\left(1 + o(1)\right). \tag{30}$$

Obviously, the statistic $\widehat{D}_{\alpha_1,n}$ is asymptotically less or more efficient than $\widehat{D}_{\alpha_2,n}$ if the ratio m_n/n of sample sizes needed to achieve the same rate of convergence of errors to zero tends to a constant larger or smaller than 1. This motivates the following definition which refers to the typical convergent situation

$$\frac{c_{\alpha_2}(m_n)}{c_{\alpha_1}(n)} \longrightarrow c_{\alpha_2/\alpha_1} \quad \text{for some } 0 \leq c_{\alpha_2/\alpha_1} \leq \infty. \tag{31}$$

Definition 4: If there is a continuous function

$$g_\alpha :]0, \infty[\to]0, \infty[$$

and a sequence $c_\alpha(n)$ such that for all $x > 0$ the error function

$$\mathsf{e}_{\alpha,n}(x) = \mathsf{P}\left(D_{\alpha,n} > x|\mathcal{H}\right), \quad x > 0 \tag{32}$$

satisfies for all $x > 0$ the relation

$$-\frac{c_\alpha(n)}{n} \ln \mathsf{e}_{\alpha,n}(x) \longrightarrow g_\alpha(x) \tag{33}$$

then g_α is called *Bahadur function* of the statistic $D_{\alpha,n}$ generated by $c_\alpha(n)$. If (33) is replaced by the condition

$$- \frac{c_\alpha(n)}{n} \ln \mathsf{e}_{\alpha,n}(x + \varepsilon_n) \longrightarrow g_\alpha(x) \quad \text{for arbitrary } \varepsilon_n \longrightarrow 0 \tag{34}$$

then the function g_α is *strongly Bahadur*.

Definition 5: Let us assume that $\mathbf{A}(\alpha_1, \alpha_2)$ holds and that for each $\alpha \in \{\alpha_1, \alpha_2\}$ the statistic $\widehat{D}_{n,\alpha}$ is consistent with parameter Δ_α and has a Bahadur function g_α generated by a sequence $c_\alpha(n)$ such that (31) is satisfied. Then the *Bahadur efficiency* of $\widehat{D}_{\alpha_1,n}$ with respect to $\widehat{D}_{\alpha_2,n}$ is the number from the interval $[0, \infty]$ defined by the formula

$$\mathrm{BE}\left(\widehat{D}_{\alpha_1,n}; \ \widehat{D}_{\alpha_2,n}\right) = \frac{g_{\alpha_1}(\Delta_{\alpha_1})}{g_{\alpha_2}(\Delta_{\alpha_2})} \cdot c_{\alpha_2/\alpha_1}. \tag{35}$$

Hereafter we shall consider also the slightly modified concept of Bahadur efficiency.

Definition 6: Let in addition to the assumptions of Definition 5, the statistics $\widehat{D}_{\alpha_1,n}, \widehat{D}_{\alpha_2,n}$ be strongly consistent and the functions $g_{\alpha_1}, g_{\alpha_2}$ strongly Bahadur. Then the Bahadur efficiency (35) is said to be *Bahadur efficiency in the strong sense*.

c) Motivation of Definition 6: Let the assumptions of this definition hold then for each $\alpha \in \{\alpha_1, \alpha_2\}, and \ u > 0$ the function

$$L_{\alpha,n}(u) = \mathsf{P}\left(\widehat{T}_{\alpha,n} - \mathsf{E}\left[\widehat{T}_{\alpha,n}\middle|\mathcal{H}\right] \geq u\middle|\mathcal{H}\right), \ (\text{cf. } 26)$$

denotes the level of the error of the statistic

$$\widehat{T}_{\alpha,n} - \mathsf{E}\left[\widehat{T}_{\alpha,n}\middle|\mathcal{H}\right] \equiv 2n\left(\widehat{D}_{\alpha,n} - \mathsf{E}\left[\widehat{D}_{\alpha,n}\middle|\mathcal{H}\right]\right)$$

for critical value $u > 0$. By the assumed strong consistency of $\widehat{D}_{\alpha,n}$,

$$\frac{\mathsf{E}\left[\widehat{T}_{\alpha,n}\middle|\mathcal{H}\right]}{2n} \longrightarrow 0 \quad (\text{cf.}(25)).$$

This means that the sequence $c_\alpha(n)$ generating the strongly Bahadur g_α satisfies for all $t > 0$ the relation

$$- \frac{c_\alpha(n)}{n} \ln \mathsf{P}\left(\widehat{T}_{\alpha,n} \geq \mathsf{E}\left[\widehat{T}_{\alpha,n}\middle|\mathcal{H}\right] + 2nt\middle|\mathcal{H}\right) \longrightarrow g_\alpha(t) \ . \tag{cf. (34)}$$

Consequently, by the argument of Quine and Robinson [8, p. 732],

$$\lim_n - \frac{c_\alpha(n)}{n} \ln L_{\alpha,n}(\widehat{T}_{\alpha,n}) \xrightarrow{p} g_\alpha(\Delta_\alpha).$$

Hence [8], the error level $L_{\alpha_1,n}(\widehat{T}_{\alpha_1,n})$ of the statistic $\widehat{T}_{\alpha_1,n} = 2n\widehat{D}_{\alpha_1,n}$ is asymptotically equivalent to the error level $L_{\alpha_2,m_n}(\widehat{T}_{\alpha_2,m_n})$ of the statistic $\widehat{T}_{\alpha_2,m_n} = 2m_n\widehat{D}_{\alpha_2,m_n}$ achieved by a sample size m_n if the comparability (29) takes place or, in other words, if the sample sizes n and m_n are mutually related by (30). In other words, the concept of Bahadur efficiency introduced in this paper coincides under the stronger assumptions of Definition 6 with the Bahadur efficiency of Quine and Robinson [8].

Harremoës and Vajda [5] assumed the same strong consistency as in Definition 6 but introduced the Bahadur efficiency by the slightly different formula

$$\mathrm{BE}\left(\widehat{D}_{\alpha_1,n}; \ \widehat{D}_{\alpha_2,n}\right) = \frac{g_{\alpha_1}(\Delta_{\alpha_1})}{g_{\alpha_2}(\Delta_{\alpha_2})} \cdot \bar{c}_{\alpha_2/\alpha_1} \tag{36}$$

where[1]

$$\bar{c}_{\alpha_2/\alpha_1} = \lim_{n \longrightarrow \infty} \frac{c_{\alpha_2}(n)}{c_{\alpha_1}(n)}. \tag{37}$$

III. CONSISTENCY

In this section we study the consistency of the class of power divergence statistics $D_\alpha(\widehat{P}_n, Q_n), \alpha > 0$. In the domain $\alpha < 0$ this consistency was studied in the particular case of uniform Q by Harremoës and Vajda [6].

Theorem 7: Let distributions $Q_n \in M(k)$ satisfy the assumption $\mathbf{A}(\alpha)$. Assume that f is uniformly continuous. Then the statistic $D_f(\widehat{P}_n, Q_n)$ is strongly consistent provided

$$\frac{n}{k} \longrightarrow \infty. \tag{38}$$

Proof: Under $\mathcal{H}$ we have $D_f(P_n, Q_n) = D_f(Q_n, Q_n) = 0$. Hence it suffices to prove

$$|\Lambda_{\alpha,n}| \xrightarrow{p} 0 \quad \text{under both } \mathcal{H} \text{ and } \mathcal{A} \tag{39}$$

for $\Lambda_{\alpha,n} = D_f(\widehat{P}_n, Q_n) - D_f(P_n, Q_n)$. For simplicity we skip the subscript n in the symbols $\widehat{P}_n, P_n$, and Q_n, i.e. we substitute

$$\widehat{P}_n = \widehat{P} = (\widehat{p}_j : 1 \leq j \leq k), \quad P_n = P = (p_j : 1 \leq j \leq k). \tag{40}$$

This leads to the simplified formula $\Lambda_{\alpha,n} = D_f(\widehat{P}, Q) - D_f(P, Q)$. We can without loss of generality assume that $D_f(P, Q)$ is constant not only under $\mathcal{H}$ (where the constant is automatically 0) but also under $\mathcal{A}$ (where the assumed detectability implies the convergence $D_f(P, Q) \longrightarrow \Delta_\alpha$ for $0 < \Delta_\alpha < \infty$). In this asymptotic sense we use the equalities

$$D_f(P, Q) = \frac{\sum q_j \left(\frac{p_j}{q_j}\right)^\alpha - 1}{\alpha(\alpha - 1)} = \Delta_\alpha \tag{41}$$

and

$$\Lambda_{\alpha,n} = D_f(\widehat{P}, U) - \Delta_\alpha. \tag{42}$$

Choose some $0 < s < 1$ and define

$$f^s(t) = \begin{cases} f(t) & \text{for } t \geq s, \\ f(s) + f'_+(s)(t - s) & \text{for } 0 \leq t < s. \end{cases}$$

Then

$$0 \leq f(t) - f^s(t) \leq f(0) - f^s(0)$$

so that (9) implies

$$0 \leq D_f(P, Q) - D_{f^s}(P, Q) \leq f(0) - f^s(0).$$

---

[1]Due to a missprint, α_1 and α_2 were interchanged behind the limit in [5, Eq. 30], but the formula was used in the correct form (36). In the Appendix we prove that the conclusions made on the basis of the original formula (36) hold unchanged under the present precised formula (35).

The function f^s is Lipschitz with the Lipschitz constant $\lambda = \max\left\{\left|f'_+(s)\right|, \left|f'(\infty)\right|\right\}$ i.e. $|f(t_1) - f(t_2)| \le \lambda|t_1 - t_2|$ for all $t_1, t_2 \ge 0$. Then

$$
\left|D_{f^s}(\widehat{P}_n, Q) - D_{f^s}(P_n, Q)\right|
$$

$$
= \left|\sum_{j=1}^{k} q_j\, f^s\left(\frac{\widehat{p}_j}{q_j}\right) - \sum_{j=1}^{k} q_j f^s\left(\frac{p_j}{q_j}\right)\right|
$$

$$
= \sum_{j=1}^{k} q_j \left|f^s\left(\frac{\widehat{p}_j}{q_j}\right) - f^s\left(\frac{p_j}{q_j}\right)\right| \le \sum_{j=1}^{k} q_j \lambda \left|\frac{\widehat{p}_j}{q_j} - \frac{p_j}{q_j}\right|
$$

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

where in the last step we used the Schwarz inequality. Since

$$
\mathsf{E}\left[(\widehat{p}_j - p_j)^2\right] = p_j(1 - p_j)/n \le p_j/n \qquad (43)
$$

it holds

$$
\mathsf{E}\left|D_{f^s}(\widehat{P}_n, Q) - D_{f^s}(P_n, Q)\right|
$$

$$
\le \lambda \left(\mathsf{E}\left[\sum_{j=1}^{k} \frac{(\widehat{p}_j - p_j)^2}{p_j}\right]\right)^{1/2} \le \lambda \left(\frac{k}{n}\right)^{1/2}.
$$

Consequently,

$$
\mathsf{E}\left|D_f(\widehat{P}_n, Q) - D_f(P_n, Q)\right|
$$

$$
\le 2\left(f(0) - f^s(0)\right) + \lambda\,(k/n)^{1/2}
$$

so that under (49)

$$
\limsup_{n \to \infty}\ \mathsf{E}\left|D_f(\widehat{P}_n, Q_n) - D_f(P_n, Q_n)\right|
$$

$$
\le 2\left(f(0) - f^s(0)\right).
$$

This holds for all $s > 0$. Since $f(0) - f^s(0) \longrightarrow 0$ for $s \downarrow 0$, we see that in this case (38) implies (39). ∎

The interpretation of condition 38 is that the mean number of observations per bin should tend to infinity under $\mathcal{H}$. Note that this condition does not exclude that we will observe empty cells.

Our results are concentrated in Theorem 9 below. Its proof uses the following auxiliary result.

Lemma 8: For $x, y \ge 0$ and $1 \le \alpha \le 2$ it holds

$$
L_\alpha(x, y) \le \phi_\alpha(y) - \phi_\alpha(x) \le U_\alpha(x, y) \qquad (44)
$$

where

$$
L_\alpha(x, y) = (y - x)\phi'_\alpha(x) \qquad (45)
$$

and

$$
U_\alpha(x, y) = L_\alpha(x, y) + \frac{1}{\alpha}x^{\alpha-2}(y - x)^2. \qquad (46)
$$

Proof: First assume $1 < \alpha < 2$. Since $\frac{1}{\alpha}x^{\alpha-2}(y - x)^2$ is nonnegative, it suffices to prove

$$
\phi_\alpha(y) \ge \phi_\alpha(x) + \phi'_\alpha(x)(y - x) \qquad (47)
$$

and

$$
\phi_\alpha(y) \le \phi_\alpha(x) + \phi'_\alpha(x)(y - x) + \frac{1}{\alpha}x^{\alpha-2}(y - x)^2. \quad (48)
$$

But Inequality (47) is evident since the function $y \to \phi_\alpha(y)$ is convex. We shall prove that the function

$$
f(y) =
$$
$$
\phi_\alpha(y) - \left(\phi_\alpha(x) + \phi'_\alpha(x)(y - x) + \frac{1}{\alpha}x^{\alpha-2}(y - x)^2\right)
$$

is non-positive. First we observe that $f(0) = f(x) = 0$. By differentiating $f(y)$ we get

$$
f'(y) = \frac{y^{\alpha-1} - 1}{\alpha - 1} - \left(\phi'_\alpha(x) + \frac{2}{\alpha}x^{\alpha-2}(y - x)\right)
$$

so that $f'(x) = 0$. Differentiating once more we get

$$
f''(y) = y^{\alpha-2} - \frac{2}{\alpha}x^{\alpha-2}.
$$

Thus $f''(y) > 0$ for $y < x_\alpha \overset{def}{=} (\alpha/2)^{\frac{1}{2-\alpha}} x$ and $f''(y) < 0$ for $y > x_\alpha$. Since $x_\alpha < x$ and $f(y)$ is concave on $[x_\alpha, 1]$, it is maximized on this interval at $y = x$ where $f(x) = 0$. Thus $f(y) \le 0$ on this interval and in particular $f(x_\alpha) \le 0$. This together with $f(0) = 0$ and the convexity of $f(y)$ on the interval $[0, x_\alpha]$ implies $f(y) \le 0$ for $y \in [0, x]$. This completes the proof of the non-positivity of $f(y)$, i.e. the proof of (48). The cases $\alpha = 2$ and $\alpha = 1$ follow by continuity. ∎

The main result of this section is the following theorem.

Theorem 9: Let distributions $Q_n \in M(k)$ satisfy the assumption $\mathbf{A}(\alpha)$. Then the statistic $D_\alpha(\widehat{P}_n, Q_n)$ is strongly consistent provided

$$
0 < \alpha \le 2 \qquad \text{and} \qquad \frac{n}{k} \longrightarrow \infty \qquad (49)
$$

and consistent provided

$$
\alpha > 2 \qquad \text{and} \qquad \frac{n}{k \log k} \longrightarrow \infty. \qquad (50)
$$

Proof: We shall use the same notation as in the proof of Theorem 7. In the proof we treat separately the cases

$$
\begin{aligned}
&\text{i}: 0 < \alpha < 1,\\
&\text{ii}: 1 < \alpha \le 2,\\
&\text{iii}: \alpha = 1,\\
&\text{iv}: \alpha > 2.
\end{aligned}
$$

Case i ($0 < \alpha < 1$): This follows from Theorem 7 because $x \to \phi_\alpha(x)$ is uniformly continuous.

Case ii ($1 < \alpha \le 2$): Here we get from (42)

$$
\Lambda_{\alpha,n} = \sum_{j=1}^{k} q_j\left(\phi_\alpha\left(\frac{\widehat{p}_j}{q_j}\right) - \phi_\alpha\left(\frac{p_j}{q_j}\right)\right) \qquad (51)
$$

so that Lemma 8 implies

$$\sum_{j=1}^{k} q_j L_\alpha \left(\frac{p_j}{q_j}, \frac{\widehat{p}_j}{q_j} \right) \leq \Lambda_{\alpha,n} \leq$$

$$\sum_{j=1}^{k} q_j L_\alpha \left(\frac{p_j}{q_j}, \frac{\widehat{p}_j}{q_j} \right) + \sum_{j=1}^{k} q_j \frac{1}{\alpha} \left(\frac{p_j}{q_j} \right)^{\alpha-2} \left(\frac{\widehat{p}_j}{q_j} - \frac{p_j}{q_j} \right)^2$$

and

$$|\Lambda_{\alpha,n}| \leq \left| \sum_{j=1}^{k} (\widehat{p}_j - p_j) \phi'_\alpha \left(\frac{p_j}{q_j} \right) \right| + \sum_{j=1}^{k} \frac{p_j^{\alpha-2}}{q_j^{\alpha-1}} \frac{(\widehat{p}_j - p_j)^2}{\alpha}.$$

We take the mean and get

$$E\,|\Lambda_{\alpha,n}| \leq$$
$$E \left| \sum_{j=1}^{k} (\widehat{p}_j - p_j) \phi'_\alpha \left(\frac{p_j}{q_j} \right) \right| + \sum_{j=1}^{k} \frac{p_j^{\alpha-2}}{\alpha q_j^{\alpha-1}} E \left[(\widehat{p}_j - p_j)^2 \right]$$

The terms on the right hand side are treated separately.

$$\sum_{j=1}^{k} \frac{p_j^{\alpha-2}}{q_j^{\alpha-1}} \frac{E \left[(\widehat{p}_j - p_j)^2 \right]}{\alpha} = \sum_{j=1}^{k} \frac{p_j^{\alpha-2}}{q_j^{\alpha-1}} \frac{E \left[(n\widehat{p}_j - np_j)^2 \right]}{\alpha n^2}$$
$$= \sum_{j=1}^{k} \frac{p_j^{\alpha-2}}{q_j^{\alpha-1}} \frac{np_j(1-p_j)}{\alpha n^2}$$
$$\leq \frac{1}{\alpha n} \sum_{j=1}^{k} \frac{p_j^{\alpha-1}}{\left(\frac{\rho}{k} \right)^{\alpha-1}}$$
$$\leq \frac{k^{\alpha-1}}{\alpha n \rho^{\alpha-1}} \sum_{j=1}^{k} p_j^{\alpha-1}.$$

The function $P \to \sum_{j=1}^{k} p_j^{\alpha-1}$ is concave so it attains its maximum for $P = (1/k, 1/k, \cdots, 1/k)$. Therefore

$$\sum_{j=1}^{k} \frac{p_j^{\alpha-2}}{q_j^{\alpha-1}} \frac{E \left[(\widehat{p}_j - p_j)^2 \right]}{\alpha} \leq \frac{k^{\alpha-1}}{\alpha n \rho^{\alpha-1}} k \left(\frac{1}{k} \right)^{\alpha-1}$$
$$= \frac{1}{\alpha \rho^{\alpha-1}} \frac{k}{n}.$$

Next we bound the first term.

$$E \left| \sum_{j=1}^{k} (\widehat{p}_j - p_j) \phi'_\alpha \left(\frac{p_j}{q_j} \right) \right|$$
$$\leq E \left[\left(\sum_{j=1}^{k} (\widehat{p}_j - p_j) \phi'_\alpha \left(\frac{p_j}{q_j} \right) \right)^2 \right]^{1/2}$$
$$= \left(\sum_{i,j=1}^{k} Cov\,(\widehat{p}_i, \widehat{p}_j) \phi'_\alpha \left(\frac{p_i}{q_i} \right) \phi'_\alpha \left(\frac{p_j}{q_j} \right) \right)^{1/2}$$
$$= \frac{1}{n} \left(\sum_{i,j=1}^{k} Cov\,(n\widehat{p}_i, n\widehat{p}_j) \phi'_\alpha \left(\frac{p_i}{q_i} \right) \phi'_\alpha \left(\frac{p_j}{q_j} \right) \right)^{1/2}$$
$$= \frac{1}{n} \left(\begin{array}{c} \sum_{i=1}^{k} Var\,(n\widehat{p}_i) \left(\phi'_\alpha \left(\frac{p_i}{q_i} \right) \right)^2 \\ + \sum_{i \neq j} Cov\,(n\widehat{p}_i, n\widehat{p}_j) \phi'_\alpha \left(\frac{p_i}{q_i} \right) \phi'_\alpha \left(\frac{p_j}{q_j} \right) \end{array} \right)^{1/2}.$$

This equals

$$\frac{1}{n} \left(\begin{array}{c} \sum_{i=1}^{k} np_i(1-p_i) \left(\phi'_\alpha \left(\frac{p_i}{q_i} \right) \right)^2 \\ + \sum_{i \neq j} np_i p_j \phi'_\alpha \left(\frac{p_i}{q_i} \right) \phi'_\alpha \left(\frac{p_j}{q_j} \right) \end{array} \right)^{1/2}$$
$$\leq \frac{1}{n^{1/2}} \left(\begin{array}{c} \sum_{i=1}^{k} p_i \left(\phi'_\alpha \left(\frac{p_i}{q_i} \right) \right)^2 \\ + \sum_{i,j} p_i p_j \phi'_\alpha \left(\frac{p_i}{q_i} \right) \phi'_\alpha \left(\frac{p_j}{q_j} \right) \end{array} \right)^{1/2}.$$

This can be bounded as

$$\frac{1}{n^{1/2}} \left(\begin{array}{c} \sum_{i=1}^{k} p_i \left(\phi'_\alpha \left(\frac{p_i}{q_i} \right) \right)^2 \\ + \left(\sum_i p_i \phi'_\alpha \left(\frac{p_i}{q_i} \right) \right)^2 \end{array} \right)^{1/2}$$
$$\leq \left(\frac{2}{n} \sum_{i=1}^{k} p_i \left(\phi'_\alpha \left(\frac{p_i}{q_i} \right) \right)^2 \right)^{1/2}.$$

These bounds can be combined into

$$E\,|\Lambda_{\alpha,n}| \leq \frac{1}{\alpha \rho^{\alpha-1}} \frac{k}{n} + \left(\frac{2}{n} \sum_{i=1}^{k} p_i \left(\phi'_\alpha \left(\frac{p_i}{q_i} \right) \right)^2 \right)^{1/2}. \quad (52)$$

Under (49) the first term tends to zero as $n \to \infty$. The last

term does the same, which is seen from the inequalities

$$
\frac{2}{n} \sum_{i=1}^{k} p_i \left(\frac{\left(\frac{p_i}{q_i}\right)^{\alpha-1} - 1}{\alpha - 1} \right)^2
$$

$$
\leq \frac{2}{n} \sum_{i=1}^{k} p_i \frac{\left(\frac{p_i}{q_i}\right)^{2\alpha-2} + 1}{(\alpha-1)^2}
$$

$$
= \frac{2}{n(\alpha-1)^2} \left(\sum_{i=1}^{k} q_i \left(\frac{p_i}{q_i}\right)^{\alpha} \left(\frac{p_i}{q_i}\right)^{\alpha-1} + 1 \right)
$$

$$
\leq \frac{2}{n(\alpha-1)^2} \sum_{i=1}^{k} q_i \left(\frac{p_i}{q_i}\right)^{\alpha} \left(\frac{1}{\rho/k}\right)^{\alpha-1} + \frac{2}{n(\alpha-1)^2}
$$

$$
= \frac{k^{\alpha-1}}{n} \frac{2}{(\alpha-1)^2 \rho^{\alpha-1}} \sum_{i=1}^{k} q_i \left(\frac{p_i}{q_i}\right)^{\alpha} + \frac{2}{n(\alpha-1)^2}
$$

$$
= \frac{k^{\alpha-1}}{n} \frac{2(\alpha(\alpha-1)\Delta + 1)}{(\alpha-1)^2 \rho^{\alpha-1}} + \frac{2}{n(\alpha-1)^2}.
$$

Case iii ($\alpha = 1$): For $\alpha = 1$ in Inequality 52 we get

$$
\mathsf{E}\,|\Lambda_{1,n}| \leq \frac{k}{n} + \left(\frac{2}{n} \sum_{i=1}^{k} p_i \left(\ln \frac{p_i}{q_i} \right)^2 \right)^{1/2}.
$$

Using $\ln p_i \leq 0$ we find that last term on the right satisfies the relations

$$
\frac{1}{n} \sum_{j=1}^{k} p_i \left(\ln^2 p_i - 2 \ln p_i \ln q_i + \ln^2 q_i \right) \tag{53}
$$

$$
= \frac{1}{n} \sum_{j=1}^{k} p_i \ln^2 p_i - \frac{2}{n} \sum_{j=1}^{k} p_i \ln p_i \ln q_i + \frac{1}{n} \sum_{j=1}^{k} p_i \ln^2 q_i
$$

$$
\leq \frac{1}{n} \sum_{j=1}^{k} p_i \ln^2 p_i + \frac{\ln^2 \frac{k}{\varrho}}{n} \sum_{j=1}^{k} p_i \leq \frac{1}{n} \sum_{j=1}^{k} p_i \ln^2 p_i + \frac{\ln^2 \frac{k}{\varrho}}{n}.
$$

The function $x \to x \ln^2 x$ is concave in the interval $\left[0; e^{-1}\right]$ and convex in the interval $\left[e^{-1}; 1\right]$. Therefore we we can apply the method of [16] to verify that $\sum_{i=1}^{k} p_i \ln^2 p_i$ attains its maximum for a mixture of uniform distributions on k points and on subset of $k-1$ of these points. Thus

$$
\frac{1}{n} \sum_{i=1}^{k} p_i \ln^2 p_j \leq \frac{1}{n} \sum_{i=1}^{k} \frac{1}{k-1} \ln^2 \left(\frac{1}{k}\right)
$$

$$
= \frac{k \ln^2 k}{n(k-1)} \leq \frac{2 \ln^2 k}{n} \tag{54}
$$

and we can conclude that under (49) the first term in (53) tends to zero as n tends to infinity. Obviously, under (49) also the second term in (51) tends to zero so that the desired relation (39) holds.

Case iv ($\alpha > 2$): By **A2**,

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

$$
\geq \frac{1}{\alpha(\alpha-1)} \left(\left(\frac{k}{\varrho}\right)^{\alpha-1} \sum_{j=1}^{k} p_j^\alpha - 1 \right)
$$

so that

$$
\sum_{j=1}^{k} p_j^\alpha \leq (\alpha(\alpha-1)\Delta + 1) \left(\frac{\varrho}{k}\right)^{\alpha-1} \tag{55}
$$

where we replaced $D_\alpha(P,Q)$ by $\Delta = \Delta_\alpha$ in the sense of (41). Further, by the Taylor formula

$$
\widehat{p}_j^\alpha = p_j^\alpha + \alpha\, p_j^{\alpha-1}(\widehat{p}_j - p_j) + \frac{\alpha(\alpha-1)}{2} \xi_j^{\alpha-2}(\widehat{p}_j - p_j)^2 \tag{56}
$$

where ξ_j is between p_j and $\widehat{p}_j$. We shall look for a highly probable upper bound on $\widehat{p}_j$. Choose any $b > 1$ and consider the random event

$$
E_{nj}(b) = \{\widehat{p}_j \geq b \max\{p_j, q_j\}\}.
$$

We shall prove that under (50) it holds

$$
\boldsymbol{\pi}_n(b) \overset{def}{=} \mathsf{P}\left(\cup_j E_{nj}(b)\right) \longrightarrow 0. \tag{57}
$$

The components $X_j = X_{nj}$ of the observation vector $\boldsymbol{X}_n$ defined in Section 1 are approximately *Poisson distributed*, $\mathrm{Po}(np_j)$, so that

$$
\mathsf{P}\left(\widehat{p}_j \geq b \max\{p_j, q_j\}\right) = \mathsf{P}\left(X_j \geq nb \max\{p_j, q_j\}\right)
$$

$$
\leq \exp\{-D_1\left(\mathrm{Po}\left(b \max\{np_j, nq_j\}\right),\ \mathrm{Po}\left(np_j\right)\right)\}
$$

for the divergence $D_1(P,Q)$ defined by (9)-(8) with P, Q replaced by the corresponding Poisson distributions. But

$$
D_1\left(\mathrm{Po}\left(bnq_j\right),\ \mathrm{Po}\left(np_j\right)\right) = np_j \phi_1(b) \tag{58}
$$

for the logarithmic function $\phi_1 \geq 0$ introduced in (7). Since for all $0 \leq p_j, q_j \leq 1$

$$
\phi_1 \left(\frac{b \max\{p_j, q_j\}_j}{p_j} \right) \geq \phi_1(b) > 1 \quad \text{for } b > 1,
$$

it holds

$$
D_1\left(\mathrm{Po}\left(b \max\{np_j, nq_j\}\right),\ \mathrm{Po}\left(np_j\right)\right)
$$

$$
\geq D_1\left(\mathrm{Po}\left(bnq_j\right),\ \mathrm{Po}\left(nq_j\right)\right).
$$

Consequently,

$$
\boldsymbol{\pi}_n(b) \leq \sum_j \mathsf{P}\left(\widehat{p}_j \geq b \max\{p_j, q_j\}\right)
$$

$$
\leq \sum_j \exp\{-D_1\left(\mathrm{Po}\left(b \max\{np_j, nq_j\}\right),\ \mathrm{Po}\left(np_j\right)\right)\}
$$

$$
\leq \sum_j \exp\{-D_1\left(\mathrm{Po}\left(bnq_j\right),\ \mathrm{Po}\left(nq_j\right)\right)\}
$$

$$
= \sum_j \exp\{-nq_j \phi_1(b)\} \qquad (\text{cf. (58)})
$$

$$
\leq k \exp\left\{-n\frac{\varrho}{k}\phi_1(b)\right\} = k^{1 - \frac{n}{k \log k} \varrho \phi_1(b)}. \tag{59}
$$

Assumption (50) implies that the exponent in (59) tends to $-\infty$ so that (57) holds. Therefore it suffices to prove (39) under the condition that for all sufficiently large n the random events $\cup_j E_{nj}(b)$ fail to take place, i.e. that

$$\widehat{p}_j < b \max\{p_j, q_j\} \quad \text{for all} \ \ 1 \leq j \leq k. \quad (60)$$

Let us start with the fact that under (60) it holds $\xi_j \leq \{bp_j, bq_j\}$ and then

$$\xi_j^{\alpha-2} \leq (\max\{bp_j, bq_j\})^{\alpha-2} \leq b^{\alpha-2} p_j^{\alpha-2} + b^{\alpha-2} \frac{\varrho^{\alpha-2}}{k^{\alpha-2}}. \quad (61)$$

Applying this in the Taylor formula (56) we obtain

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

Hence under (60) we get from (51) and Lemma 8

$$|\Lambda_{\alpha,n}| \leq \frac{k^{\alpha-1}}{\alpha(\alpha-1)} \sum_{j=1}^{n} \alpha\, p_j^{\alpha-1} \left|\widehat{p}_j - p_j\right|$$
$$+ \frac{k^{\alpha-1}}{\alpha(\alpha-1)} \sum_{j=1}^{n} \frac{\alpha(\alpha-1)b^{\alpha-2}}{2}\left(p_j^{\alpha-2} + \frac{\varrho^{\alpha-2}}{k^{\alpha-2}}\right)(\widehat{p}_j - p_j)^2.$$

Applying (55) and using Jensen's inequality and the expectation bound (43), we upper bound $\mathbf{E}\,|\Lambda_{\alpha,n}|$ by

$$\frac{(\alpha(\alpha-1)\Delta+1)^{1/2}}{\alpha(\alpha-1)}\left(\frac{\sum p_j^{\alpha-1}}{k^{1-\alpha}n}\right)^{1/2}$$
$$+ \frac{b^{\alpha-2}k^{\alpha-1}}{2}\sum_{j=1}^{k}\left(p_j^{\alpha-2} + \frac{\varrho^{\alpha-2}}{k^{\alpha-2}}\right)\mathbf{E}\left[(\widehat{p}_j - p_j)^2\right]$$
$$\leq \frac{(\alpha(\alpha-1)\Delta+1)^{1/2}}{\alpha(\alpha-1)}\left(\frac{\sum p_j^{\alpha-1}}{k^{1-\alpha}n}\right)^{1/2}$$
$$+ \frac{b^{\alpha-2}k^{\alpha-1}}{2}\sum_{j=1}^{k}\left(p_j^{\alpha-2} + \frac{\varrho^{\alpha-2}}{k^{\alpha-2}}\right)\frac{p_j}{n}$$
$$= \frac{(\alpha(\alpha-1)\Delta+1)^{1/2}}{\alpha(\alpha-1)}\left(\frac{\sum p_j^{\alpha-1}}{k^{1-\alpha}n}\right)^{1/2} \quad (62)$$
$$+ \frac{b^{\alpha-2}}{2}\frac{k^{\alpha-1}\sum_{j=1}^{k}p_j^{\alpha-1}}{n} + \frac{b^{\alpha-2}\varrho^{\alpha-2}}{2}\frac{k}{n}.$$

Obviously, under (60) the desired relation (39) holds if the assumption (50) implies the convergence

$$\frac{\sum p_j^{\alpha-1}}{k^{1-\alpha}n} \to 0. \quad (63)$$

However, by Schwarz inequality and (55),

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

so that the validity of (39) under (50) is obvious and the proof is complete.

∎

Condition 50 is stronger than Condition 38 and implies that for any fixed number $a > 0$ eventually any bins will contain more than a observations.

IV. BAHADUR EFFICIENCY

In this section we study the Bahadur efficiency in the class of power divergence statistics $\hat{D}_{\alpha,n} = D_{\alpha}(\hat{P}_n, Q_n)$, $\alpha > 0$. As before, we use the simplified notations

$$P_n = P, \ Q_n = Q \quad \text{and} \quad k_n = k.$$

The results are concentrated in Theorem 13 below. Its proof is based on the following lemmas. The first two of them make use of the Rényi divergences of orders $\alpha > 0$

$$D_{\alpha}(P\|Q) = \frac{1}{\alpha-1}\ln\sum_{j=1}^{k} p_j^{\alpha} q_j^{1-\alpha},$$
$$D_1(P\|Q) = \lim_{\alpha\to 1} D_{\alpha}(P\|Q) = D(P\|Q)$$

where $D(P\|Q)$ is the classical information divergence denoted above by $D_1(P,Q)$. There is a monotone relationship between the Rényi and power divergences given by the formula

$$D_{\alpha}(P\|Q) = \frac{1}{\alpha-1}\ln(1 + \alpha(\alpha-1)D_{\alpha}(P,Q)), \quad (64)$$
$$D_1(P\|Q) = D_1(P,Q). \quad (65)$$

Lemma 10: Let P and Q be probability vectors on the set $\mathcal{X}$. If $\alpha < \beta$ then

$$D_{\alpha}(P\|Q) \leq D_{\beta}(P\|Q).$$

with equality if and only there exists a subset $A \subseteq \mathcal{X}$ such that $P = Q(\cdot \mid A)$.

since, by Lemma 11,

$$\frac{g_{\alpha_1}(\Delta_{\alpha_1})}{g_{\alpha_2}(\Delta_{\alpha_2})} =$$

$$\begin{cases} \dfrac{\alpha_2 - 1}{\alpha_1 - 1} \cdot \dfrac{\ln\left(1 + \alpha_1(\alpha_1 - 1)\Delta_{\alpha_1}\right)}{\ln\left(1 + \alpha_2(\alpha_2 - 1)\Delta_{\alpha_2}\right)} & \text{when } \alpha_2 < 1 \\[2em] \dfrac{1}{\alpha_1 - 1} \cdot \dfrac{\ln\left(1 + \alpha_1(\alpha_1 - 1)\Delta_{\alpha_1}\right)}{\Delta_{\alpha_2}} & \text{when } \alpha_2 = 1. \end{cases}$$

(75)

The second assertion was for $\alpha_1 = 1$ deduced in Section 2 from the lemmas presented there. The argument was based on the fact that $c_{\alpha_1}(n) = 1$ for $\alpha_1 = 1$. But $c_\alpha(n) = 1$ for all $0 < \alpha \le 1$ so that extension from $\alpha_1 = 1$ to $0 < \alpha_1 < 1$ is straightforward. ∎

Example 14: Let

$$P_n = \left(p_{nj} \stackrel{def}{=} \frac{1_{\{1 \le j \le k/2\}}}{\lfloor k/2 \rfloor}\right), \quad n = 1, 2, \ldots \quad (76)$$

where 1_A is the indicator function, $\lfloor \cdot \rfloor$ stands for the integer part (floor function) and, as before,

$$U = \left(u_j \stackrel{def}{=} 1/k : 1 \le j \le k\right).$$

Then for $\alpha \ne 0, 1$

$$D_\alpha(P_n, U) = \frac{\sum_1^k u_j\left((p_{nj}/u_j)^\alpha - \alpha(p_{nj}/u_j - 1) - 1\right)}{\alpha(\alpha - 1)}$$

$$= \frac{\sum_1^k p_{nj}^\alpha u_j^{1-\alpha} + \sum_1^k (p_{nj} - u_j) - \sum_1^k u_j}{\alpha(\alpha - 1)}$$

$$= \frac{k^{\alpha-1} \sum_1^{\lfloor k/2 \rfloor} \lfloor k/2 \rfloor^{-\alpha} - 1}{\alpha(\alpha - 1)}$$

$$= \frac{k^{\alpha-1} \lfloor k/2 \rfloor / \lfloor k/2 \rfloor^\alpha - 1}{\alpha(\alpha - 1)}$$

$$= \frac{(k/\lfloor k/2 \rfloor)^{\alpha-1} - 1}{\alpha(\alpha - 1)}.$$

Therefore the identifiably condition (19) takes on the form

$$D_\alpha(P_n, U) \longrightarrow$$

$$\begin{cases} \dfrac{2^{\alpha-1} - 1}{\alpha(\alpha - 1)} \stackrel{def}{=} \Delta_\alpha & \text{,if } \alpha > 0, \ \alpha \ne 1 \\[1.5em] \ln 2 \stackrel{def}{=} \Delta_1 & \text{if } \alpha = 1. \end{cases}$$

If $0 < \alpha \le 1$ then Lemma 12 implies

$$g_\alpha(\Delta) = \ln\left(1 + \alpha(\alpha - 1)\Delta\right)/(\alpha - 1)$$

when $0 < \alpha < 1$ and $g_1(\Delta) = \Delta$ when $\alpha = 1$. If moreover (72) then under the alternative (76)

$$\frac{g_\alpha(\Delta_\alpha)}{g_1(\Delta_1)} = \frac{\ln\left(1 + \alpha(\alpha - 1)\frac{2^{\alpha-1}-1}{\alpha(\alpha-1)}\right)}{(\alpha - 1)\ln 2}$$

$$= \frac{\ln\left(1 + 2^{\alpha-1} - 1\right)}{(\alpha - 1)\ln 2} = 1.$$

Hence, by Definition 4, the likelihood ratio statistic $\widehat{D}_{1,n}$ is as Bahadur efficient as any $\widehat{D}_{\alpha,n}$ with $0 < \alpha < 1$. If $\alpha > 1$ then Lemma 12 implies

$$\frac{g_\alpha(\Delta_\alpha)}{g_1(\Delta_1)} = \frac{(2^{\alpha-1} - 1)^{1/\alpha}}{\ln 2} > 1.$$

However, contrary to this prevalence of $g_\alpha(\Delta_\alpha)$ over $g_1(\Delta_1)$, Theorem 13 implies that $\widehat{D}_{1,n}$ is infinitely more Bahadur efficient than $\widehat{D}_{\alpha,n}$.

Example 15: Let us now consider the truncated geometric distribution

$$P_n = (p_{n1}, \ldots, p_{nk}) = c_k(p)(1, p, \ldots, p^k)$$

with parameter $p = p_n \in]0, 1[$. Since

$$1 + p + p^2 + \ldots = \frac{1}{1-p} \quad \text{and} \quad p^{k+1} + p^{k+2} + \ldots = \frac{p^{k+1}}{1-p},$$

it holds

$$1 + p + \cdots + p^k = \frac{1}{1-p} - \frac{p^{k+1}}{1-p} = \frac{1 - p^{k+1}}{1-p} = \frac{1}{c_k(p)}.$$

Hence for all $\alpha \ne 0, 1$

$$\alpha(\alpha - 1)D_{\alpha,n} + 1 = \frac{1}{k} \sum_{j=0}^k \left(\frac{p_{nj}}{1/k}\right)^\alpha$$

$$= \frac{1}{k} \sum_{j=1}^k \frac{k^\alpha (1-p)^\alpha p^{\alpha j}}{(1 - p^{k+1})^\alpha}$$

$$= \frac{(k(1-p))^\alpha}{k(1 - p^{k+1})^\alpha} \sum_{j=0}^k (p^\alpha)^j$$

$$= \frac{(k(1-p))^\alpha}{k(1 - p^{k+1})^\alpha} \cdot \frac{1 - p^{\alpha(k+1)}}{1 - p^\alpha}$$

$$= \frac{(k(1-p))^\alpha}{k(1 - p^\alpha)} \cdot \frac{1 - p^{\alpha(k+1)}}{(1 - p^{k+1})^\alpha}.$$

In the particular case $p = 1 - x/k$ for $x \ne 0$ fixed we get $k(1 - p) = x$ and

$$k(1 - p^\alpha) = k\left(1 - \left(1 - \frac{\alpha x}{k} + o\left(\frac{x}{k}\right)\right)\right) \longrightarrow \alpha x,$$

$$p^{\alpha(k+1)} = \left(1 - \frac{x}{k}\right)^{\alpha(k+1)} \longrightarrow e^{-x\alpha},$$

$$p^{k+1} = \left(1 - \frac{x}{k}\right)^{k+1} \longrightarrow e^{-x}.$$

Therefore

$$\alpha(\alpha - 1)D_{\alpha,n} + 1 = \frac{x^\alpha}{k\left(\frac{\alpha x}{k} + o\left(\frac{x}{k}\right)\right)} \cdot \frac{1 - e^{-x\alpha}}{(1 - e^{-x})^\alpha}$$

$$= \frac{x^\alpha}{\alpha x + o(x)} \cdot \frac{e^{x\alpha} - 1}{(e^x - 1)^\alpha}.$$

Consequently,

$$\alpha(\alpha - 1)\Delta_\alpha + 1 = \frac{x^{\alpha-1}}{\alpha} \cdot \frac{e^{x\alpha} - 1}{(e^x - 1)^\alpha}$$

i.e.,

$$\Delta_\alpha = \frac{x^{\alpha-1}(e^{x\alpha} - 1) - \alpha(e^x - 1)^\alpha}{\alpha^2(\alpha - 1)(e^x - 1)^\alpha} \quad \text{for } \alpha \ne 0, 1.$$

By the L'Hospital rule,

$$\Delta_1 = \ln \frac{x}{e(e^x - 1)} + \frac{xe^x}{e^x - 1},$$
$$\Delta_0 = \frac{\ln(e^x - 1) - \ln x}{2}.$$

From here one can deduce that if $x \to 0$ then

$$\Delta_\alpha \longrightarrow 0 \quad \text{for all} \quad \alpha \in \mathbb{R}.$$

If $x = 1$ then

$$\Delta_\alpha = \frac{e^\alpha - 1 - (e-1)^\alpha}{\alpha^2(\alpha - 1).(e-1)^\alpha} \quad \text{for } \alpha \neq 0, 1$$

and

$$\Delta_1 = \frac{1 - (e-1)\ln(e-1)}{e-1} = 0.035,$$
$$\Delta_0 = \frac{\ln(e-1)}{2} = 0.271.$$

Using Lemma 2 and Theorem 2 in a similar manner as in the previous example, we find that here $\widehat{D}_{1,n}$ is more Bahadur efficient as any $\widehat{D}_{\alpha,n}$ with $0 < \alpha < \infty$, $\alpha \neq 1$.

V. CONTIGUITY

In this paper we proved that the statistics $\hat{D}_{\alpha,n}$ of orders $\alpha > 1$ are less Bahadur efficient than those of the orders $0 < \alpha \leq 1$ and that the latter are mutually comparable in the Bahadur sense. One may have expected $\hat{D}_{1,n}$ to be much more Bahadur efficient than $\hat{D}_{\alpha,n}$ for $0 < \alpha < 1$. In order to understand why this is not the case we have to examine somewhat closer the assumptions of our theory.

Recall that given a sequence of pairs of probability measures $(P_n, Q_n)_{n\in\mathbb{N}}$, $(P_n)_{n\in\mathbb{N}}$ is said to be contiguous with respect to $(Q_n)_{n\in\mathbb{N}}$ if $Q_n(A_n) \to 0$ for $n \to \infty$ implies $P_n(A_n) \to 0$ for $n \to \infty$ and any sequence of sets $(A_n)_{n\in\mathbb{N}}$. When $(P_n)_{n\in\mathbb{N}}$ is contiguous with respect to $(Q_n)_{n\in\mathbb{N}}$ we write $P_n \lhd Q_n$. Let P and Q be probability measures on the same set $\mathcal{X}$ and let $(\mathcal{F}_n)_{n\in\mathbb{N}}$ be an increasing sequence of finite sub-σ-algebras on $\mathcal{X}$ that generates the full σ-algebra on $\mathcal{X}$. If $P_n = P_{|\mathcal{F}_n}$ and $Q_n = Q_{|\mathcal{F}_n}$ then $P_n \lhd Q_n$ if and only if $P \ll Q$ where $\ll$ denotes absolute continuity. For completeness we give the proof of the following simple proposition.

Proposition 16: Let $(P_n, Q_n)_{n\in\mathbb{N}}$ denote a sequence of pairs of probability measures and assume that the sequence $D_1(P_n, Q_n)$ is bounded. Then $P_n \lhd Q_n$.

Proof: Assume that the proposition is false. Then there exist $\varepsilon > 0$ and a subsequence of sets $(A_{n_k})_{k\in\mathbb{N}}$ such that $Q_{n_k}(A_{n_k}) \to 0$ for $k \to \infty$ and $P_{n_k}(A_{n_k}) \geq \varepsilon$ for all $k \in \mathbb{N}$. ∎

In general, a large power α makes the power divergence $D_a(P, Q)$ sensitive to large values of dP/dQ. Therefore the statistics $\widehat{D}_{\alpha,n}$ with large α should be used when the sequence of alternatives P_n may not be contiguous with respect to the sequence of hypotheses Q_n. Conversely, a small power α makes $D_a(P, Q)$ sensitive to small values of dP/dQ. Therefore $\widehat{D}_{\alpha,n}$ with small α should be used when

the sequence of hypotheses Q_n is not contiguous with respect to the sequence alternatives P_n. Our conditions guarantee $P_n \lhd Q_n$ but not the reversed contiguity $Q_n \lhd P_n$. We see that a substantial modification of the conditions is needed in order to guarantee that $\hat{D}_{1,n}$ dominates the divergence statistcs $\hat{D}_{\alpha,n}$ of the orders $0 < \alpha < 1$ in the Bahadur sense.

VI. APPENDIX: RELATIONS TO PREVIOUS RESULTS

As mentioned at the end of Section II, Harremoës and Vajda [5] assumed the same strong consistency as in Definition 4+ but introduced the Bahadur efficiency by the formula (36). The next four lemmas help to clarify the relation between this and the present precised concept of Bahadur efficiency (35).

Under the assumptions of Definition 4, [5] considered the following conditions.

C1: The limit $\bar{c}_{\alpha_2/\alpha_1}$ considered in (37) exists.
C2: Both statistics $\widehat{D}_{\alpha_i,n}$ are strongly consistent and both functions g_{α_i} are strongly Bahadur.

Lemma 17: Let the assumptions of Definition 5 hold. Under **C1** the Bahadur efficiency (36) coincides with the present Bahadur efficiency (35). If moreover **C2** holds then (36) is the Bahadur efficiency in the strong sense.

Proof: The first assertion is clear from (36) and (35). Under **C2** the assumptions of Definition 3+ hold. Hence the second assertion follows from Definition 6. ∎

Lemma 18: Let the assumptions of Definition 3 hold and let $b(\alpha) : \mathcal{I} \longrightarrow]0, 1[$ be increasing and $d_\alpha : \mathcal{I} \longrightarrow]0, \infty[$ arbitrary function on an interval $\mathcal{I}$ covering $\{\alpha_1, \alpha_2\}$. If for every $\alpha \in \{\alpha_1, \alpha_2\}$ the sequence $c_\alpha(n)$ generating the Bahadur function g_α satisfies the asymptotic condition

$$c_\alpha(n) = n^{b(\alpha)}(d_\alpha + o(1)) \tag{77}$$

then (31) holds for $c_{\alpha_2/\alpha_1} = \infty$ and condition **C1** is satisfied.

Proof: Under (77) it suffices to prove that (31) holds for $c_{\alpha_2/\alpha_1} = \infty$, i.e.

$$\lim_{n \to \infty} \frac{c_{\alpha_2}(m_n)}{c_{\alpha_1}(n)} = \infty \tag{78}$$

for m_n defined by (29). By (77),

$$c_{\alpha_2}(m_n) = m_n^{b(\alpha_2)}(d_{\alpha_2} + o(1))$$

and

$$c_{\alpha_1}(n) = n^{b(\alpha_1)}(d_{\alpha_1} + o(1))$$

so that (29) implies

$$m_n^{1-b(\alpha_2)} = n^{1-b(\alpha_1)}(\gamma\delta + o(1))$$

for the finite positive constants

$$\delta = \frac{d_{\alpha_1}}{d_{\alpha_2}} \quad \text{and} \quad \gamma = \frac{g_{\alpha_1}(\Delta_{\alpha_1})}{g_{\alpha_2}(\Delta_{\alpha_2})}.$$

Hence (30) implies

$$\frac{c_{\alpha_2}(m_n)}{c_{\alpha_1}(n)} = \frac{m_n}{n}(\gamma^{-1} + o(1))$$

$$= \frac{n^{\frac{1-b(\alpha_1)}{1-b(\alpha_2)}}}{n}((\gamma\delta)^{\frac{1}{1-b(\alpha_2)}}\gamma^{-1} + o(1))$$

$$= n^{\frac{b(\alpha_2)-b(\alpha_1)}{1-b(\alpha_2)}}(\gamma^{\frac{b(\alpha_2)}{1-b(\alpha_2)}}\delta^{\frac{1}{1-b(\alpha_2)}} + o(1))$$

so that (78) holds. ∎

Lemma 19: Let the assumptions of Definition 5 hold and let for every $\alpha \in \{\alpha_1, \alpha_2\}$ the sequence $c_\alpha(n)$ generating the Bahadur function g_α satisfy the asymptotic condition

$$c_\alpha(n) = \frac{\alpha n^{b(\alpha)}}{\ln n} \qquad (79)$$

for some increasing function $b(\alpha) : \mathcal{I} \longrightarrow]0,1[$ on an interval $\mathcal{I}$ covering $\{\alpha_1, \alpha_2\}$. Then (31) holds for $c_{\alpha_2/\alpha_1} = \infty$ and condition **C1** is satisfied.

Proof: Similarly as before, it suffices to prove the relation (78) for m_n defined by (29). By (79),

$$c_{\alpha_2}(m_n) = \frac{\alpha_2 m_n^{b(\alpha_2)}}{\ln m_n} \quad \text{and} \quad c_{\alpha_1}(n) = \frac{\alpha_1 n^{b(\alpha_1)}}{\ln n}$$

so that (29) implies

$$\frac{\alpha_2 m_n^{1-b(\alpha_2)}}{\ln m_n} = \frac{\alpha_1 n^{1-b(\alpha_1)}}{\ln n}(\gamma + o(1))$$

for the same γ as in the previous proof. Since $1 - b(\alpha_2) < 1 - b(\alpha_1)$, this implies the asymptotic relation

$$\frac{m_n}{n} \longrightarrow \infty. \qquad (80)$$

Similarly as in the previous proof, we get from (30)

$$\frac{c_{\alpha_2}(m_n)}{c_{\alpha_1}(n)} = \frac{m_n}{n}(\gamma^{-1} + o(1)) =$$

$$\left(\frac{\alpha_1 \ln m_n}{\alpha_2 \ln n}\right)^{\frac{1}{1-b(\alpha_2)}} n^{\frac{b(\alpha_2)-b(\alpha_1)}{1-b(\alpha_2)}} \left(\begin{matrix}\gamma^{\frac{b(\alpha_2)}{1-b(\alpha_2)}} \\ +o(1)\end{matrix}\right)$$

$$> n^{\frac{b(\alpha_2)-b(\alpha_1)}{1-b(\alpha_2)}}(\gamma^{\frac{b(\alpha_2)}{1-b(\alpha_2)}} + o(1)).$$

Therefore the desired relation (78) holds. ∎

Lemma 20: Let the assumptions of Definition 3 hold and let for every $\alpha \in \{\alpha_1, \alpha_2\}$ the sequence $c_\alpha(n)$ generating the Bahadur function g_α satisfy the asymptotic condition

$$c_\alpha(n) = \frac{\alpha k^{b(\alpha)}}{\ln k} \qquad (81)$$

where $k = k_n \longrightarrow \infty$ is the sequence considered above and $b(\alpha) : \mathcal{I} \longrightarrow]0,\infty[$ is increasing on an interval $\mathcal{I}$ covering $\{\alpha_1, \alpha_2\}$. Then (31) holds for $c_{\alpha_2/\alpha_1} = \infty$ and condition **C1** is satisfied.

Proof: It suffices to apply Lemma 17 to the sequences

$$c_{\alpha_1}(k) = \frac{\alpha_1 k^{b(\alpha_1)}}{\ln k} \quad \text{and} \quad c_{\alpha_2}(m_k) = \frac{\alpha_2 m_k^{b(\alpha_2)}}{\ln m_k}$$

for m_k defined by the condition

$$\frac{m_k}{c_{\alpha_2}(m_k)} = \frac{g_{\alpha_1}(\Delta_{\alpha_1})}{g_{\alpha_2}(\Delta_{\alpha_2})} \cdot \frac{k}{c_{\alpha_1}(k)}(1 + o(1)) \quad \text{(cf. (29)).} \quad (82)$$

∎

Example 21: Let assumptions of Definition 5 hold for $\alpha_1 = 1$ and $\alpha_2 = \alpha > 1$, and let

$$\frac{k^{b(\alpha)+1}\ln n}{n} \longrightarrow 0 \quad \text{for} \quad b(\alpha) = (\alpha - 1)/\alpha. \qquad (83)$$

By [5, Eq. 51, 76 and 79] and (83) the sequences

$$c_1(n) = 1 \quad \text{and} \quad c_\alpha(n) = \frac{\alpha k^{b(\alpha)}}{\ln k} \qquad (84)$$

generate the Bahadur functions

$$g_1(\Delta) = \Delta \quad \text{and} \quad g_\alpha(\Delta) = (\alpha(\alpha - 1)\Delta)^{1/\alpha}, \quad \Delta > 0. \qquad (85)$$

Here we cannot apply Lemma 18 since $c_1(n)$ is not special case of $c_\alpha(n)$ for $\alpha = 1$. An alternative direct approach can be based on the observation that (29) cannot hold if $\liminf_n m_n < \infty$. In the opposite case $m_n \to \infty$ obviously implies

$$c_{\alpha/1} \overset{def}{=} \lim_n \frac{c_\alpha(m_n)}{c_1(n)} = \infty$$

so that **C1** holds with $\bar{c}_{\alpha_2/\alpha_1} \equiv c_{\alpha/1} = \infty$. Hence Lemma 1 implies that the Bahadur efficiency BE $\left(\widehat{D}_{1,n}; \widehat{D}_{\alpha,n}\right) = \infty$ obtained previously by Harremoës and Vajda [5, Eq. 81] coincides with the Bahadur efficiency of $\widehat{D}_{1,n}$ with respect to $\widehat{D}_{\alpha,n}$ in the present precised sense of (35). Under stronger condition on k than (83), Harremoës and Vajda established also the strong consistency of the statistics $\widehat{D}_{1,n}$ and $\widehat{D}_{\alpha,n}$. One can verify that (85) are strongly Bahadur functions so that **C2** holds as well. Hence, as argued by Lemma 3, we deal here with the Bahadur efficiency in the strong sense.

Example 22: Let assumptions of Definition 5 hold for $\alpha_1 > 1$ and let the function $b(\alpha)$ be defined by (83) for all $\alpha \geq 1$. Harremoës and Vajda (2008) proved that if the sequence k satisfies the condition (83) with $\alpha = \alpha_2$ then for all $\alpha \in \{\alpha_1, \alpha_2\}$ the function $g_\alpha(\Delta)$ given by the second formula in (85) is Bahadur function of the statistics $\widehat{D}_{\alpha,n}$ generated by the sequences $c_\alpha(n)$ from the second formula in (84). Thus in this case the assumptions of Lemma 18 hold. From Lemmas 20 and 17 we conclude that the Bahadur efficiency

$$\text{BE}\left(\widehat{D}_{\alpha_1,n}; \widehat{D}_{\alpha_2,n}\right) = \infty \quad \text{for all} \quad 0 < \alpha_1 < \alpha_2 < \infty$$

obtained in [5, Eq. 81] coincides with the Bahadur efficiency in the present precise sense. Similarly as in the previous example, we can arrive to the conclusion that this is the Bahadur efficiency in the strong sense.

REFERENCES

[1] E. Lehman and G. Castella, *Testing Statistical Hypotheses*. New York: Springer, 3rd ed. ed., 2005.

[2] I. Csiszár, "Eine informationstheoretische Ungleichung und ihre Anwendung auf den Beweis der ergodizität von Markoffschen Ketten," *Publ. Math. Inst. Hungar. Acad.*, vol. 8, pp. 95–108, 1963.

[3] A. Rényi, "On measures of entropy and information," in *Proceedings of the Fourth Berkeley Symposium on Mathematical Statistics and Probability*, vol. 1, pp. 547–561, 1961.

[4] T. Kailath, "The divergence and Bhattacharrya distance in signal selection.," *IEEE Trans. Comm.*, vol. 15, pp. 52–60, 1967.

[5] 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.

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

[7] P. Harremoës and I. Vajda, "Efficiency of entropy testing," in *International Symposium on Information Theory*, pp. 2639–2643, IEEE, July 2008.

[8] 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.

[9] J. Beirlant, L. Devroye, L. Györfi, and I. Vajda, "Large deviations of divergence measures on partitions," *J. Statist. Planning and Infer.*, vol. 93, pp. 1–16, 2001.

[10] L. Györfi, G. Morvai, and I. Vajda, "Information-theoretic methods in testing the goodness-of-fit," in *Proc. International Symposium on Information Theory, Sorrento, Italy, June25-30*, p. 28, 2000.

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

[12] F. Liese and I. Vajda, "On divergence and informations in statistics and information theory," *IEEE Tranns. Inform. Theory*, vol. 52, pp. 4394 – 4412, Oct. 2006.

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

[14] C. Morris, "Central limit theorems for multinomial sums.," *Ann. Statist.*, vol. 3, pp. 165–188, 1975.

[15] 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.

[16] 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.