On efficiencies of decisions about statistical models based on f-divergences of empirical distributions

IGOR VAJDA

ÚTIA AV ČR, Prague

<vajda@utia.cas.cz>

Abstract

Statistical analysis of information sources is usually based on their independent stochastic outputs (signals, images). **Statistical model** of an information source is a hypothetical probability distribution on a given measurable space. True distribution of the outputs is generally unknown. Usually are available digitalized (appropriately quantized or classified) outputs which can be represented by empirical probability distribution. We consider statistical **decisions** about the models consisting in testing hypotheses about the probability distributions.

Decision criteria: General divergence statistics (f-divergences of hypothetical and empirical distributions) including all power divergence statistics such as the classical Pearson statistics or the Neyman , the likelihood ratio and the Freeman-Tukey statistics.

Solved problems: Limit laws (asymptotic distributions) for the divergence statistics leading to the critical values for the asymptotically α-sized tests based on these statistics (i.e. to the tests with guaranteed decision errors of the first kind) and comparison of powers (decision errors of the second kind) of the tests based on various divergence statistics.

The first part of this paper deals with **limit laws** under hypotheses and local alternatives. New extension of the classical Morris (1975) limit law concerning the Pearson statistics to all divergence statistics. Innovative potential of the extension is demonstrated by a comparison with the classical result concerning the important likelihood ratio statistic. The second part of the paper deals with **efficiencies** of the above considered decisions, i.e. with the powers of the tests based on various divergence statistics. It applies both the Pitman and the Bahadur approaches to the relative efficiency of statistics. All divergence statistics are shown to be equally efficient in the Pitman sense. However, some new results are presented demonstrating the maximal Bahadur efficiency of the likelihood ratio statistic in the important class of all power divergence statistics.

1. STATISTICAL MODEL

We consider a statistical model (Ω, $\mathcal{S}$, P) with known measurable observation space (Ω, $\mathcal{S}$) and unknown probability distribution P producing i.i.d. realizations $Y_1, \ldots, Y_n$. Available are only the digitalized (appropriately quantized) data

$$X_j = \sum_{i=1}^{n} 1_{\{A_j\}}(Y_i), \quad 1 \le j \le k$$

for a given partition

$$\mathcal{A} = \{\ A_1, \ldots, A_k\} \subset \mathcal{S}$$

of Ω. We admit that the partition $\mathcal{A}$, the partition sets A_j and the partition sizes k depend on the sample size n, i.e.

$$\mathcal{A} = \mathcal{A}_n,\ A_j = A_{j,n},\ k = k_n. \tag{1}$$

In this paper we study testing the hypothesis $\mathbf{H}$ that the stochastic outputs Y_i of the model are generated by a given distribution P_0 against the alternative $\mathbf{A}$ represented by the true distribution of these outputs. The testing is assumed to be carried out by means of the available data

$$\mathbf{X} = (X_1, \ldots, X_k). \tag{2}$$

This means that, in fact, we study the problem of testing

$$\mathbf{H} \sim p = p^0 \quad \text{against} \quad \mathbf{A} \sim p \quad \text{(true)}$$

where

$$p^0 = (p_j^0 \equiv P^0(A_j) : 1 \leq j \leq k) \tag{3}$$

is a *discrete hypothetical distribution* and

$$p = (p_j \equiv P(A_j) : 1 \leq j \leq k) \tag{4}$$

a *discrete true distribution*, and that the testing is carried out by means of the data (2) uniquely represented by the *discrete empirical distribution*

$$\hat{p} = (\hat{p}_1 \equiv X_1/n, \ldots \hat{p}_k \equiv X_k/n). \tag{5}$$

In view of (1) this means that

$$X_j = X_{j,n}, \quad p_j^0 = p_{j,n}^0, \quad p_j = p_{j,n} \quad \text{and} \quad \hat{p}_j = \hat{p}_{j,n} \tag{6}$$

in (2) - (5).

We study various methods of the testing and preferences between them in the situation where the sample size n increases above all bounds. In this context we respect throughout this paper the following conventions and assumptions.

Conventions. **(i)** The subscripts n considered in (1) and (6) are suppressed and **(ii)** the standard deterministic and stochastic convergences $\longrightarrow$, $\stackrel{p}{\longrightarrow}$ and $\stackrel{d}{\longrightarrow}$ as well as the standard asymptotic expressions of the type $o(1)$ or $O(1)$, are considered for $n \to \infty$.

Assumptions. It holds $k \to \infty$, and for some $\beta \geq 1$ also $k^{\beta+1}/n = o(1)$ and

$$\min_n\ k^\beta\, p_{min}^0 \geq const > 0 \tag{7}$$

where $p_{min}^0 = \min\{p_j^0 : 1 \leq j \leq k\}$.

2. DIVERGENCE STATISTICS

Let us denote by $\mathcal{F}$ the class of all functions $f(t)$ twice differentiable with $f"(t) > 0$ in the domain $t \in (0, \infty)$ which are Lipschitz around $t = 1$ and standardized in the sense $f(1) = 0$. By $f(0) \in (-\infty, \infty]$ we denote the extension for $t \downarrow 0$. This paper studies the following class of statistics.

Definition 1. The *divergence statistics* are defined by the formula

$$\mathcal{D}_{f,n} = \frac{2n\, D_f\ (\hat{p},\ p^0)}{f"(1)}, \qquad f \in \mathcal{F} \tag{8}$$

where

$$D_f(\hat{p}, p^0) = \sum_{j=1}^{k} p_j^0\ f\left(\frac{\hat{p}_j}{p_j^0}\right) \tag{9}$$

is the f-divergence of distributions $\hat{p}$, p^0.

Notice that by (7) it holds $p_j^0 > 0$ in (9). For the properties of the f-divergence (9) see e.g. Liese and Vajda (2006). Next follow some well known examples of the divergence statistics (8).

Example 1 *(classical Pearson statistic).* The quadratic function $f\ (t) = (t-1)^2$ leads to the Pearson divergence $\chi^2(\hat{p},\ p^0)$ and the classical Pearson statistic

$$\chi_n^2 = n\chi^2(\hat{p}, p^0) = n\sum_{j=1}^{k} \frac{(\hat{p}_j\ -\ p_j^0)^2}{p_j^0} = \sum_{j=1}^{k} \frac{(X_j\ -\ np_j^0\ )^2}{np_j^0}.$$

Example 2 *(likelihood ratio statistic).* The logarithmic function $f(t) = t \ln\ t$ leads to the information divergence $I(\hat{p}, p^0)$ and the likelihood ratio statistic

$$\mathcal{I}_n = 2nI(\hat{p},\ p^0) = 2n\sum_{j=1}^{k} \hat{p}_j \ln\ \frac{\hat{p}_j}{p_j^0} = 2\sum_{j=1}^{k} X_j \ln\ \frac{X_j}{np_j^0}. \tag{10}$$

Example 3 *(power divergence statistics).* The class of power functions

$$f_\alpha(t) = \frac{t^\alpha - \alpha(t-1) - 1}{\alpha(\alpha-1)} \quad \text{where } \alpha \in \mathbb{R},\quad \alpha(\alpha-1) \neq 0$$

with the limits

$$f_0(t) = -\ln t + t - 1 \text{ and } f_1(t) = t\ln t - t + 1$$

define power divergences $D_\alpha(\hat{p}, p^0) \equiv D_{f_\alpha}(\hat{p}, p^0)$ for $\alpha \in \mathbb{R}$ and the corresponding power divergence statistics $\mathcal{D}_{\alpha,n} \equiv \mathcal{D}_{f_\alpha,n}$. It is easy to verify that $\chi^2(\hat{p}, p^0) \equiv D_2(\hat{p}, p^0)$ and $I(\hat{p}, p^0) \equiv D_1(\hat{p}, p^0)$ so that also $\chi_n^2 \equiv \mathcal{D}_{2,n}$ and $\mathcal{I}_n \equiv \mathcal{D}_{1,n}$.

3. LIMIT LAWS

Let throughout this section the conditions and assumptions introduced in Section 1 hold. Then for all $f \in \mathcal{F}$

$$\frac{\mathcal{D}_{f,n} - k}{\sqrt{2k}} \xrightarrow{d} N(0,1) \quad \text{under } \mathbf{H} \tag{11}$$

according to Györfi and Vajda (2002).This is extends the classical limit law

$$\frac{\chi_n^2 - k}{\sqrt{2k}} \xrightarrow{d} N(0,1) \quad \text{under } \mathbf{H} \tag{12}$$

of Morris (1975) valid with the present assumption $k^{1+\beta}/n = o(1)$ replaced by the weaker $k/n = o(1)$. It is natural to ask whether a universal asymptotically normal law similar to (11) remains valid also when the hypothetical equality $\mathbf{H} : p = p^0$ is replaced by the alternative $\mathbf{A} \sim p$ local in the sense that p is close to p^0. The answer is yes provided that p tends sufficiently fast to p^0 in terms of their mutual Pearson divergence $\chi^2(\hat{p}, p^0)$. Before going into details note that a partial variant of this answer for the simple but important uniform hypotheses

$$p^0 = (p_j^0 \equiv 1/k : 1 \le j \le k). \tag{13}$$

was obtained previously in Vajda (2003). Here the hypotheses are restricted only by the condition (6).

Definition 2. The alternative $\mathbf{A} \sim p$ is said to be *weakly local* if

$$\chi^2(p, p^0) = O\left(\frac{\sqrt{k}}{n}\right)$$

and *local* if there exists $\Delta \ge 0$ such that

$$\frac{n\chi^2(p,\ p^0)}{\sqrt{k}} \longrightarrow \Delta. \tag{14}$$

Example 4 The classical statistical local alternative is of the form

$$p = \left(1 - \frac{1}{\sqrt{n}}\right) p^0 + \frac{1}{\sqrt{n}} q$$

for some $q = (q_j \equiv Q(A_j) : 1 \le j \le k)$ (see (4)). Since $\chi^2(p, p^0) = \chi^2(q, p^0)/n$, this alternative is weakly local if $\chi^2(q, p^0)/\sqrt{k}$ is bounded and local in the present sense if $\chi^2(q, p^0)/\sqrt{k}$ is convergent.

Under the assumptions considered in this paper (11) can be extended into the following *Universal Asymptotic Normality theorem.*

Theorem 1 (*UAN*). All f-divergence statistics $D_{f,\, n}$ satisfy the limit law

$$\frac{\mathcal{D}_{f,n} - k - \sqrt{k}\Delta}{\sqrt{2k}} \xrightarrow{d} N(0,1) \quad \text{under local } \mathbf{A}. \tag{15}$$

Proof of this theorem is based on the following *Extension lemma*

Lemma 1. If for some $\mu_n \in \mathbb{R}$ and $\sigma_n > 0$

$$\frac{\chi_n^2 - \mu_n}{\sigma_n} \xrightarrow{d} N(0,1) \quad \text{under weakly local } \mathbf{A}$$

then for all divergence statistics $\mathcal{D}_{f,n}$

$$\frac{\mathcal{D}_{f,n} - \mu_n}{\sigma_n} \xrightarrow{d} N(0,1) \quad \text{under weakly local } \mathbf{A}.$$

Proof. Let $\mathbf{A}$ be weakly local. It suffices to prove for all sufficiently small $\varepsilon > 0$

$$\Pr\left(\frac{|\mathcal{D}_{f,\,n} \; - \; \chi_n^2|}{\sqrt{k}} > \varepsilon \,\middle|\, \mathbf{A} \right) = o(1). \tag{16}$$

By inequalities in Györfi and Vajda (2002), for all sufficiently small $\varepsilon > 0$ there exist constants $c(\varepsilon) > 0$ such that for all n

$$\Pr\left(\frac{|\mathcal{D}_{f,\,n} \; - \; \chi_n^2|}{\sqrt{k}} > \varepsilon \,\middle|\, \mathbf{A} \right) \le c(\varepsilon) \left(\frac{nA_n}{\sqrt{k}} \; + \; B_n \right)$$

where

$$A_n \;=\; \sum_{j=1}^{k} \frac{\mathrm{E}\,|\hat{p}_j \; - \; p_j|^3}{(p_j)^2}$$

and

$$B_n \;=\; \sum_{j=1}^{k} \frac{\mathrm{E}\,|\hat{p}_j \; - \; p_j|^3}{(p_j)^2}\,.$$

Using the fact that under $\mathbf{A}$

$$n\hat{p} \equiv \mathbf{X} \sim M_n(p,k) \qquad \text{(see (2))}$$

and employing the properties of the multinomial law $M_n(p,k)$, one obtains from here the desired asymptotic (16).

Proof of Theorem 1. After some effort it is possible to verify that under the assumptions of this paper Theorem 5.1 of Morris (1975) implies

$$\frac{\chi_n^2 - k - \sqrt{k}\Delta}{\sqrt{2k}} \xrightarrow{d} N(0,1) \quad \text{under local } \mathbf{A}.$$

The desired result follows by applying this in the Lemma 1.

Example 5 (*likelihood ratio statistic*). By our UAN theorem

$$\frac{2\mathcal{I}_n - k - \sqrt{k}\Delta}{\sqrt{2k}} \xrightarrow{d} N(0,1).$$

This particular limit law was proved directly in Theorem 5.2 of Morris (1975) under weaker but

(i) less intuitive and much more complicated assumptions

(ii) and the proof was incomparably more complicated.

Our theorem is not only simpler than the mentioned Theorem 5.2, but also universal, e.g. applicable to all statistics $\mathcal{D}_{\alpha,n}$. Among the well known examples different from $\mathcal{D}_{1,n} = \mathcal{I}_n$ and $\mathcal{D}_{2,n} = \chi^2_n$ one can mention the *Freeman-Tukey statistic*

$$\mathcal{D}_{1/2,n} = nH^2(\hat{p}, p^0) = 4n \sum_{j=1}^{k} \left(\sqrt{\hat{p}_j} - \sqrt{p_j^0} \right)^2$$

or the *Neyman statistic* $\mathcal{D}_{-1,n}$ and the *reversed likelihood ratio statistic* $\mathcal{D}_{0,n}$.

4. ASYMPTOTIC RELATIVE EFFICIENCIES

In this section we consider the hypotheses **H** and alternatives **A** introduced in Section 1 and the tests of these hypotheses based on the divergence statistics $\mathcal{D}_{f,n}$ introduced in Section 2 for various functions $f \in \mathcal{F}$. We compare asymptotic efficiencies of the tests rejecting **H** when $\mathcal{D}_{f,n}$ exceeds certain critical value c_n for various $f_1, f_2 \in \mathcal{F}$ under local alternative **A**. The asymptotic efficiencies refer to the powers $\pi_{f,n}(s) = \Pr(\mathcal{D}_{f,n} < c_n | \ \mathbf{A})$ of these tests with critical values c_n satisfying the asymptotic size condition $s = \lim \Pr(\mathcal{D}_{f,n} > c_n | \ \mathbf{H})$ and to the sizes $s_{f,n}(\pi) = \Pr(\mathcal{D}_{f,n} > \tilde{c}_n | \ \mathbf{H})$ of the corresponding tests with critical values $\tilde{c}_n$ satisfying the asymptotic power condition $\pi = \lim \Pr(\mathcal{D}_{f,n} < \tilde{c}_n | \ \mathbf{A})$. Similarly as before, we respect in this section the conditions and assumptions introduced in Section 1. Next follow two classical approaches to the definition of the asymptotic relative efficiency $E(\mathcal{D}_{f_1,n}, \mathcal{D}_{f_2,n})$ depending on parameters $0 < s, \pi < 1$ (see Quine and Robinson (1985).

Definition 3. The *Pitman asymptotic relative efficiency* $PE_s(\mathcal{D}_{f_1,n}, \mathcal{D}_{f_2,n})$ is the limit (if it exists) of the ratio $\pi_{f_1,n}(s)/\pi_{f_2,n}(s)$ of powers of the corresponding divergence tests of equal asymptotic size s. The *Bahadur asymptotic relative efficiency* $BE_\pi(\mathcal{D}_{f_1,n}, \mathcal{D}_{f_2,n})$ is the limit (if it exists) of the ratio $s_{f_1,n}(\pi)/s_{f_2,n}(\pi)$ of the sizes of the corresponding divergence tests of equal asymptotic power π.

Theorem 2. If the alternative **A** is local in the sense of Definition 2 then for all f_1, f_2 and s, π under consideration $PE_s(\mathcal{D}_{f_1,n}, \mathcal{D}_{f_2,n}) = BE_\pi(\mathcal{D}_{f_1,n}, \mathcal{D}_{f_2,n}) = 1$.

Proof. Let Φ be the distribution function of the normal random variable $N(0,1)$, Φ^{-1} the corresponding quantile function and put for every and $\Delta > 0$

$$c_n(s) = k + \sqrt{2k}\Phi^{-1}(1-s), \quad \tilde{c}_n(\pi) = k + \sqrt{k}\Delta + \sqrt{2k}\Phi^{-1}(\pi)$$

By the limit laws (11) and (15), the critical values $c_n = c_n(s)$ and $\tilde{c}_n = \tilde{c}_n(\pi)$ satisfy for all $0 < s, \pi < 1$ and $f \in \mathcal{F}$ the above considered asymptotic size and power conditions and, moreover,

$$s_{f,n}(\pi) = \Pr\left(\frac{-k}{\sqrt{2k}} > \frac{\Delta}{\sqrt{2}} + \Phi^{-1}(\pi) \Big| \ \mathbf{H} \right) \longrightarrow \Phi\left(\frac{\Delta}{\sqrt{2}} + \Phi^{-1}(\pi) \right) \quad \text{for all } f \in \mathcal{F}$$

and

$$\pi_{f,n}(s) = \Pr\left(\frac{\mathcal{D}_{f,n} - k - \sqrt{k}\Delta}{\sqrt{2k}} < \Phi^{-1}(1-s) - \frac{\Delta}{\sqrt{2}}\right| \mathbf{A}\Bigg) \longrightarrow \Phi\left(\Phi^{-1}(\pi) - \frac{\Delta}{\sqrt{2}}\right) \quad \text{for all } f \in \mathcal{F}.$$

The desired result is clear from here.

Obvious reason why Definition 3 fails is the too small (asymptotically vanishing) deviation $\chi^2(p,\, p^0) \longrightarrow 0$ of the local alternative $\mathbf{A} \sim p$ from $\mathbf{H} \sim p^0$ required in Definition 2 and leading to the same asymptotically vanishing deviation $D_f(p,\, p^0) \longrightarrow 0$ in terms of all f-divergences as it is visible from (15). Thus, following Bahadur (1981), in the rest of this section we consider the alternatives $\mathbf{A} \sim p$ satisfying the *large deviation condition*

$$D_f(p, p^0) \longrightarrow \Delta_f > 0 \quad \text{for } f \in \{f_1, f_2\} \subset \mathcal{F}. \tag{17}$$

In accordance with Quine and Robinson (1985) we suppose that the statistics $\mathcal{D}_{f,n}$ are for every $f \in \{f_1, f_2\}$ consistent in the sense

$$\frac{\mathcal{D}_{f,n}}{n} \xrightarrow{\ p\ } \begin{cases} 0 & \text{under } \mathbf{H} \\ \Delta_f & \text{under } \mathbf{A} \end{cases}. \tag{18}$$

This means that the asymptotic power condition $\pi = \lim \Pr\left(\mathcal{D}_{f,n} < \tilde{c}_n \middle|\ \mathbf{A}\right)$ holds for the critical values of the form $\tilde{c}_n = n\Delta_f + o(n)$ so that the test sizes $s_{f,n}(\pi)$ considered in the definition of the Bahadur efficiency above are of the form $s_{f,n}(\pi) = \Pr\left(\mathcal{D}_{f,n} > n\Delta_f + o(n) \middle|\ \mathbf{H}\right)$ $\approx \Pr\left(\mathcal{D}_{f,n} > n\Delta_f \middle|\ \mathbf{H}\right)$. Thus the new concept of relative efficiency in the next definition follows the above stated Bahadur approach, just the small deviation condition $D_f(p,\, p^0) \longrightarrow 0$ on the alternative is replaced by the large deviation condition (17).

Definition 3. Let for every $f \in \{f_1, f_2\}$ the test statistic $\mathcal{D}_{f,n}$ be consistent in the sense of (18) and let there exist a sequence $a_n(f) \longrightarrow \infty$ and a continuous function $g_f : (0, \infty)$ such that for all $\Delta > 0$

$$\Pr\left(\mathcal{D}_{f,n} > n\Delta \middle|\ \mathbf{H}\right) = \exp\left\{-a_n(f)\left[g_f(\Delta) + o(1)\right]\right\} \approx \exp\left\{-a_n(f) g_f(\Delta)\right\}. \tag{19}$$

Then the limit

$$\mathcal{BE}(\mathcal{D}_{f_1,n}, \mathcal{D}_{f_2,n}) = \lim \frac{a_n(f_1)\ g_{f_1}(\Delta_{f_1})}{a_n(f_2)\ g_{f_2}(\Delta_{f_2})} \tag{20}$$

(if it exists) is called the *Bahadur asymptotic relative efficiency* of $\mathcal{D}_{f_1,n}$ with respect to $\mathcal{D}_{f_2,n}$.

Throughout the past decades this concept of efficiency was applied to the tests based on various power divergence statistics $\mathcal{D}_{f_\alpha,n} \equiv \mathcal{D}_{\alpha,n}$, $\alpha \in \mathbb{R}$. The first known result of this kind is $\mathcal{BE}(\mathcal{I}_n, \chi^2_n) \equiv \mathcal{BE}(\mathcal{D}_{1,n}, \mathcal{D}_{2,n}) = \infty$ obtained by Quine and Robinson (1985). Results concerning the Bahadur functions $g_{f_\alpha}(\Delta)$ for some power divergence statistics $\mathcal{D}_{f_\alpha,n}$ can be found in Györfi et al. (2000), Beirlant et al. (2001), and Harremoës and Vajda (2008a). Recently Harremoës and Vajda (2008b) proved the following result.

Theorem 3. If $k^{1+\beta}\ln n/n \longrightarrow 0$ holds instead of $k^{1+\beta}/n \longrightarrow 0$ assumed in Section 1 then the Bahadur efficiency $\mathcal{BE}(\mathcal{D}_{\alpha_1,n},\mathcal{D}_{\alpha_2,n})$ exists for all $0<\alpha_1<\alpha_2$ and is given by the formula

$$\mathcal{BE}(\mathcal{D}_{\alpha_1,n},\mathcal{D}_{\alpha_2,n}) = \begin{cases} g_{\alpha_1}(\Delta_{\alpha_1})/g_{\alpha_2}(\Delta_{\alpha_2}) & \text{if } 0<\alpha_2\leq 1 \\ \infty & \text{if } \quad \alpha_2>1 \end{cases}$$

where

$$g_\alpha(\Delta) = \begin{cases} \ln(1+\alpha(\alpha-1)\Delta)/(\alpha-1) & \text{if } \ 0<\alpha<1 \\ \lim_{\alpha\uparrow 1} g_\alpha(\Delta)=\Delta & \text{if } \quad \alpha=1 \end{cases}$$

and $g_1(\Delta) =$ are the functions corresponding in the sense of (19) to the statistics $\mathcal{D}_{\alpha,n}$.

We see that the Bahadur efficiency is decreasing in the variable $\alpha\in[1,\infty)$, and for small $\Delta_{\alpha_1},\Delta_{\alpha_2}$ it is increasing in $\alpha\in(0,1]$. This rigorously demonstrates the supremacy of the likelihood ratio statistic $\mathcal{I}_n=\mathcal{D}_{1,n}$ over all divergence statistics $\mathcal{D}_{\alpha,n}$ with positive powers α.

Acknowledgement. This paper was prepared with the support of the GA ČR grant 102/07/1131.

REFERENCES

Bahadur, R. R. (1981): *Some Limit Theorems in Statistics*. SIAM, Philadelphia, PA.

Beirlant, J., Devroye, L., Györfi, L., and Vajda, I. (2001): Large deviations of divergence measures on partitions. *J. Statist. Planning and Infer.*,**39**, pp. 1–16.

Györfi, L. and Vajda, I. (2002): Asymptotic distributions for goodness-of-fit statistics in a sequence of multinomial models. *Statist. and Probab. Letters,* **56**, pp. 57-67.

Györfi, L., Morvai, G. and Vajda, I. (2000): Information-theoretic methods in testing the goodness-of-fit. *Proc. of Internat. Symp. of IEEE on Inform. Theory*, p. 28, Sorrento, Italy, June 25 – 30, 2000.

Harremoës, P. and Vajda, I. (2008a): On the Bahadur-efficient testing of uniformity by means of the entropy. *IEEE Trans. on Inform. Theory*, **54**, 321-331.

Harremoës, P. and Vajda, I. (2008b): Efficiency of entropy testing. *Proc. of Internat. Symp. of IEEE on Inform. Theory*, Toronto, Canada, June 2008.

Liese, F. and Vajda, I. (2006): On divergences and informations in statistics and information theory. *IEEE Trans. on Inform. Theory*, **52**, 4394-4412.

Morris, C. (1975): Central limit theorems for multinomial sums. *Ann. of Statistics*, **3**, pp. 165–188.

Quine, M. P. and Robinson, J. (1985): Efficiencies of chi-square and likelihood ratio tests. *Ann. of Statistics*, **13**, pp. 727-742.

Vajda, I. (2003): Asymptotic laws for stochastic disparity statistics. *Tatra Mount. Mathem. Publications*, **26**, pp. 1-12.