Tatra Mt. Math. Publ. **26** (2003), 269–280

TATRA MOUNTAINS
Mathematical Publications

ASYMPTOTIC LAWS FOR STOCHASTIC DISPARITY STATISTICS

IGOR VAJDA

ABSTRACT. This paper simplifies and clarifies the conditions of the famous theorem of Morris dealing with the asymptotic distribution of the Pearson goodness-of-fit statistic when the number of partition cells depends on sample size. The local character of alternatives implicitly required by the mentioned theorem is expressed explicitly, in terms of a Pearson distance between the hypothesis and alternative. Moreover, the paper extends the theorem of Morris to a wide class of disparity statistics by proving that appropriately standardized versions of these statistics and the Pearson statistic are asymptotically equivalent. The paper is restricted to the important particular case where the partition cells are equiprobable under the hypothesis.

1. Introduction and basic concepts

Let $\Pi_k \in \mathbb{R}^k$ be the set of all stochastic vectors (discrete probability distributions) $p = (p_1, \dots, p_k)$, and Π_k^0 its relative interior containing stochastic vectors (discrete distributions) $p^0 = \big(p_1^0, \dots, p_k^0\big)$ with nonzero coordinates. We are interested in the disparities

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

of discrete probability distributions $p \in \Pi_k$ and $p^0 \in \Pi_k^0$ where $f(t)$, $t \in (0, \infty)$, is an arbitrary real valued function twice continuously differentiable in a neighborhood of $t = 1$, with $f(1) = 0$, second derivative $f''(1) \neq 0$ and a right-hand limit $f(0) \in [-\infty, \infty]$. Since the function

$$\tilde{f}(t) = f(t) - f'(1)\,(t-1)\,, \qquad t \in (0, \infty)\,,$$

2000 Mathematics Subject Classification: 62G10, 62G20, 62B10.
Keywords: disparity statistics, goodness-of-fit, asymptotic distributions.

This research was supported by the GA CR grant 201/02/1391.

defines the same disparity on $\Pi_k \otimes \Pi_k^0$ as $f(t)$, and since we study the properties of the standardized disparities $D_f(p,p^0)/f''(1)$, we can assume without loss of generality that

$$f(1) = f'(1) = 0 \quad \text{and} \quad f''(1) > 0\,. \tag{2}$$

The second inequality in (2), together with the local continuity of $f''(t)$, means that $f(t)$ is convex (even strictly convex) in a neighborhood of $t = 1$. If $f(t)$ is convex on the whole domain $(0,\infty)$ then (1) is the f-divergence of distributions p and p^0 which is known to be nonnegative, equal to zero if and only if $p = p^0$ (see, e.g., Liese and Vajda (1987)). More generally, if $f(t) - f'(1)(t-1)$ is monotone on the intervals $(0,1]$ and $[1,\infty)$ then $D_f(p,p^0)/f''(1)$ is the f-disparity of distributions p and p^0 in the sense of Lindsay (1995), which is also nonnegative and equal to zero if and only if $p = p^0$.

For example, the functions from the class $\{f_a : a \in \mathbb{R}\}$ defined on $(0,\infty)$ by

$$f_a(t) = \frac{t^{(a+1)/2} - (t-1)(a+1)/2 - 1}{(|a|-1)/2}$$

for $|a| \neq 1$, and by the continuous extensions

$$f_1(t) = t\ln t - t + 1\,, \quad f_{-1}(t) = -\ln t + t - 1$$

for $|a| = 1$, have continuous and positive second derivatives

$$f_a''(t) = t^{(a-3)/2}(|a|+1)/2\,, \qquad a \in \mathbb{R}\,.$$

Thus they define a class of f-divergences under consideration, namely

$$D_{f_a}(p,p^0) = \frac{2}{|a|-1}\left(\sum_{j=1}^{k}\sqrt{p_j^{a+1}\left(p_j^0\right)^{1-a}} - 1\right) \qquad \text{for} \quad |a| \neq 1$$

and

$$D_{f_1}(p,p^0) = \sum_{j=1}^{k} p_j \ln\frac{p_j}{p_j^0}\,, \qquad D_{f_{-1}}(p,p^0) = D_{f_1}(p^0,p)\,.$$

In (7) and (8) below we consider the particular choices $a = 3$ and $a = 1$. On the other hand, the function

$$f(t) = \min\left\{(t-1)^2,\,(1-1/t)^2\right\}$$

satisfies the assumptions considered in (1) too but it is not convex on $(0,\infty)$. However, it is monotone on $(0,1]$ and $[1,\infty)$ so that it defines the f-disparity

$$D_f(p,p^0) = \sum_{i=1}^{k}\frac{p_j^0\left(p_j - p_j^0\right)^2}{\max\left\{\left(p_j^0\right)^2,\, p_j^2\right\}}$$

in the sense of Lindsay (1995). Finally, the function

$$f(t) = t(t-1)^2$$

is neither convex on $(0,\infty)$ nor monotone on $(0,1]$ and $[1,\infty)$. Nevertheless it satisfies the assumptions of (1) so that the corresponding disparity

$$D_f(p,p^0) = \sum_{j=1}^{k} p_j \left(\frac{p_j}{p_j^0} - 1\right)^2$$

belongs to the class considered in the present paper.

We are interested in the statistical tests of simple hypotheses about distributions of i.i.d. random observations based on disparity statistics. Let the observations $X_1,\ldots,X_n$ be independent realizations of a random variable X with an abstract sample space $(\mathcal{X},\mathcal{A},P)$ and let the hypothesis about the observation be represented by a probability measure P^0 on the measurable space $(\mathcal{X},\mathcal{A})$. Disparity goodness-of-fit statistics can be obtained using disjoint partitions

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

of the observation space $\mathcal{X}$, by inserting the hypothetic cell probabilities

$$p_j^0 = P^0(A_j), \qquad 1 \le j \le k,$$

and the empirical cell probabilities (relative cell frequencies)

$$\hat{p}_j = \frac{1}{n}\sum_{i=1}^{n} \mathbb{I}_{A_j}(X_i), \qquad 1 \le j \le k, \tag{3}$$

into the formula (1). Here $\mathbb{I}$ is the indicator of events and we skip the subscript n usually used to denote the dependence of empirical distributions $\hat{p} = (\hat{p}_1,\ldots,\hat{p}_k)$ on the sample size n. The random vector $n\hat{p}$ is multinomially distributed with parameters n and $p = (p_1,\ldots,p_k)$, where

$$p_j = P(A_j), \qquad 1 \le j \le k, \tag{4}$$

are true cell probabilities. Therefore

$$E\,\hat{p}_j = p_j,$$

and one can thus expect smaller values of the disparity statistics

$$D_f(\hat{p},p^0) = \sum_{j=1}^{k} p_j^0\, f\left(\frac{\hat{p}_j}{p_j^0}\right) \tag{5}$$

under the hypothesis $\mathcal{H} \equiv P^0$ than under an alternative $\mathcal{A} \equiv P \ne P^0$. Therefore $\mathcal{H}$ is rejected if $D_f(\hat{p},p^0)$ exceeds a critical value. For a correct setting of the critical values we need asymptotic distributions of the statistics (5). We study these distributions in the simple but most important case where the partition cells are under $\mathcal{H}$ equiprobable, i.e., $p^0 = (p_1^0, \dots, p_k^0)$ satisfies

$$p_j^0 = P^0(A_j) = \frac{1}{k}, \qquad 1 \le j \le k. \tag{6}$$

Unless otherwise explicitly stated, in this paper we consider the statistics $D_f(\hat{p}, p^0)$ when $\hat{p}$ and p^0 are given by (3) and (6). These statistics are defined by (5) for functions $f(t)$ specified at the beginning of this section and satisfying (2). Important particular cases of these statistics, or the corresponding disparity measures $D_f(p, p^0)$, have been studied, e.g., in Liese and Vajda (1987), Read and Cressie (1988), Vajda (1989), Lindsay (1994) and, most recently, in Győrfi and Vajda (2001). We need to mention here at least the Pearson statistic

$$\chi^2(\hat{p}, p^0) = \sum_{j=1}^{k} \frac{(\hat{p}_j - p_j^0)^2}{p_j^0} \tag{7}$$

obtained for $f(t) = (t-1)^2$, the information divergence statistic (log-likelihood statistic)

$$I(\hat{p}, p^0) = \sum_{j=1} \hat{p}_j \ln \frac{\hat{p}_j}{p_j^0} \tag{8}$$

obtained for $f(t) = t \ln t$, and the corresponding disparity measures (divergences) $\chi^2(p, p^0)$ and $I(p, p^0)$.

Asymptotic distributions of the statistics $D_f(\hat{p}, p^0)$ have been studied by Menéndez *et al.* (2001) in the case where the partition size k, and the partitions $\mathcal{P}$ themselves, are fixed, not depending on the sample size n. In this case their Corollary 3.1 implies that if the alternative $p = (p_1, \dots, p_k)$ is local, depending on the sample size n so that there exists a constant $\lambda \ge 0$ with the property

$$\chi^2(p, p^0) = \frac{\lambda}{n} \qquad \text{for all} \quad n \ge 1,$$

then

$$\frac{2n}{f''(1)} D_f(\hat{p}, p^0) \xrightarrow{\mathcal{L}} \chi^2_{k-1}(\lambda) \qquad \text{as } n \to \infty.$$

Here $\chi^2_{k-1}(\lambda)$ stands for a noncentral χ^2-distributed random variable with $k-1$ degrees of freedom and the noncentrality parameter $\lambda \ge 0$.

In the present paper we extend this result. We assume that the partition size $k = k_n$ depends on the sample size n and

$$\lim_{n \to \infty} k = \infty. \tag{9}$$

Then some cells $A_j \in \mathcal{P}$ depend on n too, so that the partition $\mathcal{P}$ itself and, consequently, also the distributions p, p^0 and $\hat{p}$ form triangular schemes depending on n, i.e.,

$$\begin{gathered}\tilde{\mathcal{P}} = \tilde{\mathcal{P}}_n \equiv (A_{n1}, \dots, A_{nk_n}), \\ p = p_n \equiv (p_{n1}, \dots, p_{nk_n}), \quad \hat{p} = \hat{p}_n \equiv (\hat{p}_{n1}, \dots, \hat{p}_{nk_n}).\end{gathered} \tag{10}$$

In the rest of the paper we suppose this dependence on n but we suppress the subscript n in all the objects figuring in (10). Further, unless otherwise explicitly stated, all asymptotic formulas are considered for $n \to \infty$.

By using for large k the standard normal approximation to

$$\frac{\chi^2_{k-1}(\lambda) - \lambda - k}{\sqrt{2k}},$$

one can deduce from the mentioned result of Menéndez *et al.* (2001) that if (9) holds and the above considered sequence $\lambda = n\,\chi^2(p, p^0)$ satisfies for some $\mu \geq 0$ the asymptotic relation

$$\frac{\lambda}{\sqrt{k}} = \mu + o(1),$$

i.e., if

$$\chi^2(p, p^0) = \frac{\sqrt{k}}{n}(\mu + o(1)), \tag{11}$$

then the distribution of $D_f(\hat{p}, p^0)$ should be governed by the asymptotic law

$$\frac{n}{\sqrt{k}}\left(\frac{2D_f(\hat{p}, p^0)}{f''(1)} - \frac{k}{n}\right) \xrightarrow{\mathcal{L}} N(\mu, 2). \tag{12}$$

In the present paper we precise additional assumptions on the partition sizes k and functions f, under which this is valid. The assumptions on f will be satisfied by a very wide class of disparity statistics, including the Pearson and log-likelihood statistics for which the law (12) takes on the form

$$\frac{n}{\sqrt{k}}\left(\chi^2(\hat{p}, p^0) - \frac{\hat{k}}{n}\right) \xrightarrow{\mathcal{L}} N(\mu, 2) \tag{13}$$

and

$$\frac{n}{\sqrt{k}}\left(2I(\hat{p}, p^0) - \frac{k}{n}\right) \xrightarrow{\mathcal{L}} N(\mu, 2). \tag{14}$$

For these two particular statistics, the assumptions concerning the partition sizes k and alternatives p guaranteeing (13) and (14) have been presented by Morris (1975). Our assumptions concerning the particular sizes are somewhat stronger than those of Morris, but they are applicable to a very wide class of disparity statistics. Our assumption (11) concerning the alternatives is considerably simpler and more intuitive than the assumptions of Morris.

2. Main results

Throughout this section we consider the disparity statistics $D_f(\hat{p}, p^0)$ introduced in Section 1 and the distributions $\hat{p}$, p and p_0 defined by (3), (4) and (6) with the parameter k satisfying (9).

Let us first consider the simple particular case where $\chi^2(p, p^0) = 0$ for all $n \geq 1$, i.e., where the hypothesis is true. Then (11) holds with $\mu = 0$. In this case Morris (1975) proved the asymptotic laws (13) and (14) with $\mu = 0$ under the additional assumption

$$\frac{k}{n} = o(1) . \tag{15}$$

Much later Inglot *et al.* (1990) proved the asymptotic law (12) with $\mu = 0$ under the additional assumption that $f(t)$ is three times continuously differentiable in a neighborhood of $t = 1$ and

$$\frac{k^2 \ln^2 n}{n} = o(1) . \tag{16}$$

Pardo and Vajda (2001) proved the same under the additional assumption that $f''(t)$ is locally Lipschitz at $t = 1$ in the sense that for some $\delta \in (0, 1)$ and $L(f) > 0$

$$\left| f''(t) - f''(1) \right| \leq L(f) \, |t - 1| \qquad \text{for} \quad t \in (1 - \delta, 1 + \delta) ,$$

and

$$\frac{k^2 \ln k}{n} = o(1) . \tag{17}$$

Győrfi and Vajda (2001) replaced in this result (17) by the weaker condition

$$\frac{k^2}{n} = o(1) . \tag{18}$$

It is however still stronger than the condition (15) of Morris—this is the price payed for the extension of the two particular laws (13) and (14) to the class of laws (12). On the other hand, (18) is substantially weaker than the condition (16) of Inglot *et al.* It is clear that also the local Lipschitz condition is weaker than the local continuity of the third derivative.

The method of Győrfi and Vajda is very simple. It combines the condition

$$\Delta = o_p(1) \tag{19}$$

imposed on the difference

$$\Delta = \frac{2 D_f(\hat{p}, p^0)}{f''(1)} - \chi^2(\hat{p}, p^0) \tag{20}$$

with what is known from the literature about the asymptotic distribution of the Pearson statistic $\chi^2(\hat{p}, p^0)$. The stochastic convergence of Δ to zero is established using the inequality

$$P\left(\frac{n}{\sqrt{k}}|\Delta| > \varepsilon\right) \le c\, k^{3/2}\left[n\sum_{j=1}^{k} E\left|\hat{p}_j - p_j^0\right|^3 + \sqrt{k}\sum_{j=1}^{k} E\left(\hat{p}_j - p_j^0\right)^2\right] \tag{21}$$

considered for arbitrary $\varepsilon > 0$, where the constant c depends on ε and also on $L(f)$ and δ figuring in the Lipschitz condition above.

Let us now consider the general case where $\chi^2(p, p^0)$ need not be zero. Then the inequality (21) remains valid. We shall use this inequality to obtain conditions for the asymptotic relation $\Delta = o_p(1)$ and combine these conditions with those of Theorem 5.1 of Morris for the validity of (13).

THEOREM. *If $\hat{p}$, p and p^0 defined by* (3), (4) *and* (6) *satisfy* (9), (11) *and* (18) *then the asymptotic law* (12) *holds for all f-disparity statistics* (5) *with the second derivative $f''(t)$ locally Lipschitz at $t = 1$.*

This assertion follows from the following two lemmas.

LEMMA 1. *If the assumptions of Theorem are satisfied then* (19) *holds.*

Proof. It follows from (11) that

$$\max_{1\le j\le k} |p_j - p_j^0| \le \left(\sum_{j=1}^{k}(p_j - p_j^0)^2\right)^{1/2} \le \left(\frac{1}{n\sqrt{k}}(\mu + o(1))\right)^{1/2}. \tag{22}$$

Hence, by (18),

$$\max_{1\le j\le k} p_j \le \frac{1}{k} + \left(\frac{1}{n\sqrt{k}}(\mu + o(1))\right)^{1/2} = \frac{1}{k}(1 + o(1)) \tag{23}$$

and

$$\min_{1\le j\le k} p_j \ge \frac{1}{k} - \left(\frac{1}{n\sqrt{k}}(\mu + o(1))\right)^{1/2} = \frac{1}{k}(1 - o(1)). \tag{24}$$

By using the elementary inequality

$$|t_1 + t_2|^3 \le 4\left(|t_1|^3 + |t_2|^3\right), \qquad t_1, t_2 \in \mathbb{R},$$

obtained from Jensen's inequality for $f(t) = |t|^3$, we get for every $1 \le j \le k$

$$E\left|\hat{p}_j - p_j^0\right|^3 \le 4\left(E\left|\hat{p}_j - p_j\right|^3 + \left|p_j - p_j^0\right|^3\right).$$

Further, by Lemma 2 in Győrfi and Vajda (2002), for all sufficiently large n

$$E|\hat{p}_j - p_j|^3 \le 2\left(\frac{p_j}{n}\right)^{3/2}.$$

Finally, for every $1 \le j \le k$,

$$E\big(\hat{p}_j - p_j^0\big)^2 \le E\big(\hat{p}_j - p_j\big)^2 + \big(p_j - p_j^0\big)^2 \le \frac{p_j}{n} + \big(p_j - p_j^0\big)^2 .$$

Therefore (21) implies for all $\varepsilon > 0$ and sufficiently large n

$$P\left(\frac{n}{\sqrt{k}}|\Delta| > \varepsilon\right) \le c\,k^{3/2}(A+B)\,,$$

where

$$\begin{aligned} A &= 4n\left[2\sum_{j=1}^{k}\Big(\frac{p_j}{n}\Big)^{3/2} + \sum_{k=1}^{k}\big|p_j - p_j^0\big|^3\right] \\ &= 4n\left[\frac{2}{n^{3/2}}\max_{1\le j\le k}\sqrt{p_j} + \frac{1}{k}\,\chi^2(p,p^0)\max_{1\le j\le k}\big|p_j - p_j^0\big|\right] \\ &= 4\left[\frac{2}{\sqrt{n}}\max_{1\le j\le k}\sqrt{p_j} + \frac{1}{\sqrt{k}}\big(\mu + o(1)\big)\max_{1\le j\le k}\big|p_j - p_j^0\big|\right] \quad \text{(cf. (11))} \\ &\le 4\left[\frac{2\big(1+o(1)\big)}{\sqrt{nk}} + \frac{\mu + o(1)}{n^{1/2}\,k^{3/4}}\right] \quad \text{(cf. (22), (23))}\,, \end{aligned}$$

and

$$\begin{aligned} B &= \sqrt{k}\left[\sum_{j=1}^{k}\frac{p_j}{n} + \frac{1}{k}\,\chi^2(p,p^0)\right] \\ &= \frac{\sqrt{k}}{n} + \frac{1}{n}\big(\mu + o(1)\big) \quad \text{(cf. (11))}\,. \end{aligned}$$

From here we see that (18) implies for all $\varepsilon > 0$

$$P\left(\frac{n}{\sqrt{k}}\,|\Delta| > \varepsilon\right) = o(1)$$

which in turn implies (19). □

Lemma 2. *If* $\hat{p}$, p *and* p^0 *defined by* (3), (4) *and* (6) *satisfy* (9), (11) *and* (15) *then the asymptotic law* (13) *holds.*

P r o o f . Theorem 5.1 of M o r r i s (1975), applied to the model defined by (3) and (6) and satisfying (9), reduces to the following assertion. Define

$$m = \frac{k}{n} + \chi^2(p,p^0)\,, \qquad \gamma = \frac{k}{n} + 2\big[1 + \chi^2(p,p^0)\big]\,, \tag{25}$$

$$\theta_j = k\big(p_j - p_j^0\big) - \chi^2(p,p^0)\,, \qquad s_j^2 = n\left(\frac{k}{n} - 2kp_j - \gamma\right)^2 p_j \tag{26}$$

and

$$\sigma_j^2 = 2k\big[1+\chi^2(p,p^0)\big] + s_j^2\,, \qquad \sigma^2 = \sum_{j=1}^k \sigma_j^2\,. \tag{27}$$

Further, let

$$\max_{1\le j\le k} p_j = o(1) \tag{28}$$

and let for some $\alpha > 0$ and all sufficiently large n

$$n \min_{1\le j\le k} p_j \ge \alpha\,. \tag{29}$$

If

$$\frac{n \max\limits_{1\le j\le k} \theta_j^2 p_j}{\sigma^2} = o(1)\,, \tag{30}$$

then

$$\frac{n}{\sigma}\big(\chi^2(\hat p, p^0) - m\big) \xrightarrow{\mathcal{L}} N(0,1)\,. \tag{31}$$

First we prove that under the assumptions of Lemma the conditions (28)–(30) hold. The validity of (28) is clear from (23) and (9). Similarly, (29) follows from (24) and (15). Finally,

$$\begin{aligned}\frac{k}{n} + 2kp_j - \gamma &= 2kp_j - 2\big[1+\chi^2(p,p^0)\big] \qquad \text{(cf. (25))}\\ &= 2\big[kp_j - 1 - \chi^2(p,p^0)\big] = 2\theta_j\,,\end{aligned}$$

so that

$$s_j^2 = 4n\,\theta_j^2\,p_j\,,$$

and (30) holds if the right-hand side of the inequality

$$\frac{n\sum\limits_{j=1}^k \theta_j^2\,p_j}{\sigma^2} \le \frac{n\sum\limits_{j=1}^k \theta_j^2\,p_j}{2k^2\big[1+\chi^2(p,p^0)\big]} \tag{32}$$

is $o(1)$. To prove this, take first into account that by (11) and (15),

$$k^2\big[1+\chi^2(p,p^0)\big] = k^2\big(1+o(1)\big)\,.$$

Further, using a standard algebra and applying (6) and the definition of $\chi^2(p, p^0)$, we obtain

$$\begin{aligned} n\sum_{j=1}^{k} \theta_j^2\, p_j &= n\sum_{j=1}^{k} \big[k\,(p_j - p_j^0) - \chi^2(p,p^0)\big]^2 p_j \\ &= n\big[k^2 \sum_{j=1}^{k} (p_j - p_j^0)^2 p_j - \big(\chi^2(p,p^0)\big)^2\big] \\ &\le n\big[k\,\chi^2(p,p^0) \max_{1\le j\le k} p_j - \big(\chi^2(p,p^0)\big)^2\big] \\ &= n\,\chi^2(p,p^0)\,\big[k \max_{1\le j\le k} p_j - \chi^2(p,p^0)\big] \\ &= \sqrt{k}\big(\mu + o(1)\big)\left[k \max_{1\le j\le k} p_j - \frac{\sqrt{k}}{n}\big(\mu + o(1)\big)\right] \quad \text{(cf. (11))} \\ &\le \sqrt{k}\big(\mu + o(1)\big) \qquad\qquad \text{(cf. (24) and (15))}. \end{aligned}$$

Therefore the right-hand side of (32) is $O(k^{-3/2})$, which is $o(1)$ by (9). Thus we proved that under the assumptions of Lemma, the limit law (31) holds. Since by (25) and (11)

$$m = \frac{k}{n}\big(1 + o(1)\big)$$

and, by what we proved above,

$$\sigma^2 = 2k^2\big(1 + o(1)\big) + 4n\sum_{j=1}^{k} \theta_j^2\, p_j = 2k^2\big(1 + o(1)\big)\,,$$

we see that (13) follows from (31). □

3. Discussion

Lemma 2 is of its own statistical interest. It considerably simplifies the formulation of the conditions of Theorem 5.1 of Morris (1975) and of the assertion of this theorem, and transforms this assertion into a statistically applicable form. Namely, the variables m, γ, θ_j, σ_j and σ introduced in (25)–(27) are not needed at all—they can be replaced by the Pearson divergence $\chi^2(p, p^0)$ and by the model parameters k and n. Further, the rather cumbersome conditions (28)–(30) can be reduced to the simpler conditions (11) and (18). Moreover, the condition (11) clarifies the local behavior of alternatives p implicitly required by the conditions (28)–(30), by presenting this behavior in terms of the Pearson divergence $\chi^2(p, p^0)$, or in terms of the Euclidean distance $\|p - p^0\|$ since (11) is equivalent to

$$\left\|p - p^0\right\|^2 = \frac{1}{n\sqrt{k}} \left(\mu + o(1)\right).$$

Finally, if we compare the statistical outputs (13) of Lemma 2 and (31) of Theorem 5.1 of Morris, we see that (13) involves only the parameters k and n known to the statistician while (31) involves (implicitly, in the variables m and σ) also the unknown true distributions p. In fact, Lemma 2 reduces all what is needed to know about the alternatives to the condition (11) which can be taken as an additional hypothesis in practical applications of this lemma (or of our Theorem).

To say more about the applicability of our Theorem, let us start by pointing out that it automatically provides the limit law (14) for the log-likelihood statistic $I(\hat{p}, p^0)$. This law is not new. It has been first obtained in Theorem 5.2 of Morris dealing with the asymptotic distribution of $I(\hat{p}, p^0)$. However, our Theorem is obtained very easily from Theorem 5.1 of Morris while the proof of Theorem 5.2 consumed more than 4 pages of the Annals. Moreover, verification of the conditions of Theorem 5.2 is not easy and our Theorem simplifies this verification similarly as mentioned above in the case of Theorem 5.1. At the same time, it provides asymptotic laws not only for the statistics $\chi^2(\hat{p}, p^0)$ and $I(\hat{p}, p^0)$, but also for the Neyman statistic $\chi^2(p^0, \hat{p})$, reversed log-likelihood statistic $I(p^0, \hat{p})$ and all other statistics studied previously in the case where the partition size k is fixed (see the references in Section 1).

Our Theorem is also applicable in the studies of asymptotic relative efficiencies of goodness-of-fit tests based on various disparity statistics $D_f(\hat{p}, p^0)$. For example, the Bahadur efficiencies of the tests based on the statistics $\chi^2(\hat{p}, p^0)$, $I(\hat{p}, p^0)$, $\chi^2(p^0, \hat{p})$ and $I(p^0, \hat{p})$ differ considerably (see in this respect Quine and Robinson (1985), Győrfi *et al.* (2000) and Beirlant *et al.* (2001)), and it is still unknown which of the disparity statistics is best in the Bahadur efficiency sense.

REFERENCES

[1] BEIRLANT, J.—DEVROYE, L.—GYŐRFI, L.—VAJDA, I.: *Large deviations of divergence measures on partitions*, J. Statist. Plann. Inference **93** (2001), 1–16.

[2] GYŐRFI, L.—MORVAI, G.—VAJDA, I.: *Information-theoretic methods in testing the goodness-of-fit*, in: Proceedings of the International Symposium of IEEE on Information Theory, Sorrento, Italy, June 25 – 30, 2000, p. 28.

[3] GYŐRFI, L.—VAJDA, I.: *Asymptotic distributions for goodness-of-fit statistics in a sequence of multinomial models*, Statist. Probab. Lett. **56** (2002), 57–67.

[4] INGLOT, T.—JURLEWICZ, T.—LEDWINA, T.: *Asymptotics for multinomial goodness of fit tests for simple hypothesis*, Theory Probab. Appl. **35** (1990), 797–803.

[5] LIESE, F.—VAJDA, I.: *Convex Statistical Distances*, Teubner, Leipzig, 1987.

[6] LINDSAY, G. G.: *Efficiency versus robustness. The case of minimum Hellinger distance and other methods*, Ann. Statist. **22** (1994), 1081–1114.

[7] MENÉNDEZ, M.—MORALES, D.—PARDO, L.—VAJDA, I.: *Approximations to powers of ϕ-disparity goodness-of-fit tests*, Comm. Statist. Theory Methods **30** (2001), 105–134.

[8] MORRIS, C.: *Central limit theorems for multinomial sums*, Ann. Statist. **3** (1975), 165–188.

[9] PARDO, M. C.—VAJDA, I.: *On asymptotic properties of information-theoretic divergences*, IEEE Trans. Inform. Theory **49** (2003), 1860–1868.

[10] QUINE, M. P.—ROBINSON, J.: *Efficiencies of chi-square and likelihood ratio goodness-of-fit tests*, Ann. Statist. **13** (1985), 727–742.

[11] READ, T. R. C.—CRESSIE, N.: *Goodness of Fit Statistics for Discrete Multivariate Data*, Springer-Verlag, Berlin, 1988.

[12] VAJDA, I.: *Theory of Statistical Inference and Information*, Kluwer, Dordrecht, 1989.

Received February 21, 2002

Institute of Information
Theory and Automation
Academy of Sciences of the Czech Republic
Pod vodárenskou věží 4
CZ–182 08 Prague
CZECH REPUBLIC

E-mail: vajda@utia.cas.cz