LIMIT LAWS FOR f-DISPARITY STATISTICS UNDER LOCAL ALTERNATIVES

I. Vajda

Institute of Information Theory and Automation

Academy of Sciences of the Czech Republic

<vajda@utia.cas.cz>

October 10, 2007

Abstract

This paper simplifies and clarifies the conditions of the well-known 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 there 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 modifying an earlier argument of Györfi and Vajda for the asymptotic equivalence of disparity statistics and Pearson statistic.

1 Introduction. In this paper we study the classical statistical problem of testing the hypothesis $\mathcal{H}$ that a probability measure P governing an independent random sample $Y_1, \ldots, Y_n$ from a measurable observation space $(\Omega, \mathcal{S})$ equals a given measure P_0 on $(\Omega, \mathcal{S})$. If $\mathcal{A}_n = \{A_1, \ldots, A_{k_n}\} \subset \mathcal{S}$ is a sequence of partition of Ω and the statistician observes the quantized values

$$X_j = \#\{1 \le i \le n : Y_i \in A_j\}, \quad 1 \le j \le k_n \tag{1}$$

then the hypothesis $\mathcal{H}$ is represented for every $n = 1, 2, ...$ by the discrete hypothetic distribution

$$p^0 = (p_j^0 = P^0(A_j) : 1 \le j \le k_n)$$

and the alternative $\mathcal{A}$ is represented by the discrete true distribution

$$p = (p_j = P(A_j) : 1 \le j \le k_n).$$

In what follows we admit that the sets A_j, and consequently also the probabilities p_j^0 and p_j, depend on n but we drop the subscript n. Similarly, instead of k_n we write only k. All convergences and asymptotic relations are considered for $n \to \infty$ and we suppose that

$$k = k_n \to \infty \quad \text{and} \quad \frac{k^{\beta+1}}{n} = o(1) \quad \text{for some } \beta \ge 1 \tag{2}$$

as well as

$$k^\beta \, p_{min}^0 \ge \gamma \quad \text{for some } \gamma > 0 \text{ and all } n. \tag{3}$$

The hypothesis $\mathcal{H}$ is usually tested on the basis of observations (1) by means of the f-disparity statistics

$$T_{f,n} = \frac{2n\, D_f(\widehat{p}, p^0)}{f''(1)}, \quad f \in \mathcal{F} \tag{4}$$

where

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

is an f-disparity between the empirical distribution

$$\widehat{p} = (\widehat{p}_1 \equiv X_1/n, \dots \widehat{p}_k \equiv X_k/n) \tag{6}$$

and the hypothetical distribution $p^0 = (p_1^0, \dots, p_k^0)$. For details about the expressions behind the sum in (5) we refer to Menendez et al (1998).

In (4) and elsewhere in the paper, the functions $f \in \mathcal{F}$ are supposed to be continuous on and twice continuously differentiable in the neighborhood of 1 with the second derivative $f''(1) > 0$ and Lipschitz in this neighborhood. Moreover, $f(t) - f'(1)(t-1)$ is assumed to be decreasing on $(0, 1)$ and increasing on $(1, \infty)$ with $f(1) = 0$. Then (5) is an f-disparity of distributions $\widehat{p}$ and p^0. If moreover f is convex on $(0, \infty)$ then (5) is an f-divergence of distributions $\widehat{p}$ and p^0. Well known example is $f(t) = t \ln t$ leading to the information divergence $I(\widehat{p}, p^0)$ and to the *likelihood ratio statistic*

$$I_n = 2n \sum_{j=1}^{k} \widehat{p}_j \, \ln \frac{\widehat{p}_j}{p_j^0} = 2 \sum_{j=1}^{k} X_j \, \ln \frac{X_j}{np_j^0}. \tag{7}$$

Another well known example is $f(t) = (t-1)^2$ leading to the Pearson divergence $\chi^2(\widehat{p}, p^0)$ and to the *Pearson statistic*

$$\chi_n^2 = n \sum_{j=1}^{k} \frac{(\widehat{p}_j - p_j^0)^2}{p_j^0} = \sum_{j=1}^{k} \frac{(X_j - np_j^0)^2}{np_j^0} \,. \tag{8}$$

Györfi and Vajda (2002) investigated the asymptotic normality of the f-disparity statistics (4) under the hypothesis $\mathcal{H}$. They found conditions on k and p^0 such that under $\mathcal{H} : p^0$

$$\frac{T_{f,n} - k}{\sqrt{2k}} \xrightarrow{L} N(0, 1) \tag{9}$$

for all $f \in \mathcal{F}$. In the present paper we investigate the asymptotic normality of the f-disparity statistics (4) under the alternatives $\mathcal{A} : p$ local in the sense

$$\chi^2(p, p^0) = O\left(\frac{\sqrt{k}}{n}\right). \tag{10}$$

This locality means that the χ^2-divergences between $\mathcal{A} : p$ and $\mathcal{H} : p^0$ tend to zero (see the assumption (2)).

Notice that for the classical local alternatives

$$p = \left(1 - \frac{1}{\sqrt{n}}\right) + \frac{1}{\sqrt{n}} \, q \tag{11}$$

obtained by mixing hypothetic distributions p^0 with given fixed alternative distributions $q = (q_1, \dots, q_k)$ we get

$$\chi^2(p, p^0) = \frac{1}{n} \, \chi^2(q, p^0). \tag{12}$$

Hence these classical local alternatives satisfy our locality condition (10) when the χ^2-divergence $\chi^2(q,p^0)$ of both mixed distributions is bounded or increases with the moderate rate $\sqrt{k}$. Since

$$\chi^2(p,p^0) \le \max_{1\le j\le k} |q_j/p_j^0 + 1| \sum_{j=1}^{k} |q_j - p_j^0|$$

$$\le 2\left(\max_{1\le j\le k} q_j/p_j^0 + 1\right), \tag{13}$$

we see from (12) that if all likelihood ratios q_j/p_j^0 increase moderately in the sense

$$\max_{1\le j\le k} q_j/p_j^0 = O(\sqrt{k})$$

then the classical local alternatives (11) are local also in the present sense (10).

2 Main result. Our main result is the a theorem enabling to extend the asymptotic result (9) of Gyorfi and Vajda from hypotheses $\mathcal{H} : p^0$ to the alternatives $\mathcal{A} : p$ local in the sense of (10). More precisely, this theorem enables to extend the asymptotic normality of the Pearson statistic

$$\frac{\chi_n^2 - \mu_n}{\sigma_n} \xrightarrow{L} N(0,1) \tag{14}$$

established for some $\mu_n \in I\!R$ and $\sigma_n > 0$ to the asymptotic normality

$$\frac{T_{f,n} - \mu_n}{\sigma_n} \xrightarrow{L} N(0,1) \tag{15}$$

for all f-disparity statistics $T_{f,n}$ under consideration.

Theorem. Let under the alternatives $\mathcal{A} : p$ satisfying (10) the Pearson statistic χ_n^2 be for some sequences $\mu = \mu_n \in I\!R$ and $\sigma = \sigma_n = O(\sqrt{k})$ asymptotically normal in the sense of (14). Then all f-disparity statistics $T_{f,n}$ are under the same alternatives asymptotically normal in the sense of (15).

Proof. It suffices to prove that the f-disparity statistics $T_{f,n}$ satisfy for all sufficiently small $\varepsilon > 0$ the relation

$$P\left(\frac{|T_{f,n} - \chi_n^2|}{\sqrt{k}} > \varepsilon\right) = o(1)\,. \tag{16}$$

Let us start with the observation that the top four inequalities on p. 65 in Györfi and Vajda (2002) hold for arbitrary distributions $p_n = (p_{n1}, \ldots, p_{nk_n})$ with positive probabilities p_{nj}. This means that for all sufficiently small $\varepsilon > 0$ there exist constants $c(\varepsilon) > 0$ such that for all n

$$P\left(\frac{|T_{f,n} - \chi_n^2|}{\sqrt{k}} > \varepsilon\right) \le c(\varepsilon)\left(\frac{nA_n}{\sqrt{k}} + B_n\right) \tag{17}$$

where

$$A_n = \sum_{j=1}^{k} \frac{\mathsf{E}\,|\widehat{p}_j - p_j^0|^3}{(p_j^0)^2} \tag{18}$$

and

$$B_n = \sum_{j=1}^{k} \frac{\mathsf{E}\,(\widehat{p}_j - p_j^0)^2}{(p_j^0)^2}\,. \tag{19}$$

Applying first (3) and then the Schwarz inequality and (10), (2) we get

$$\begin{aligned}\sum_{j=1}^{k} \frac{p_j}{p_j^0} &\le \sqrt{\frac{k^\beta}{\gamma}} \sum_{j=1}^{k} \frac{p_j - p_j^0}{\sqrt{p_j^0}} + k \\ &= \left(\frac{k^{\beta+1}\,\chi_n^2}{\gamma}\right)^{1/2} + k \\ &= k\left[1 + O\left(\frac{k^{\beta+1/2}}{n}\right)\right]^{1/2} = k(1+o(1)).\end{aligned} \tag{20}$$

Further, from (2) and (3) we obtain the auxiliary inequality

$$n\sum_{j=1}^{k} \frac{|p_j - p_j^0|^3}{(p_j^0)^2} \le \sqrt{\frac{k^\beta}{\gamma}}\,\psi_n$$

where

$$\begin{aligned}\psi_n &\le n\left[\sum_{j=1}^{k} \frac{(p_j - p_j^0)^2}{p_j^0}\right]^{1/2} \sum_{j=1}^{k} \frac{(p_j - p_j^0)^2}{p_j^0} \\ &= \left[n^2\chi^2(p,p^0)^3\right]^{1/2} \le \left(O\left(\frac{k^{3/2}}{n}\right)\right)^{1/2} \qquad \text{(cf. (10))}\end{aligned}$$

so that

$$n\sum_{j=1}^{k} \frac{|p_j - p_j^0|^3}{(p_j^0)^2} = \left(O\left(\frac{k^{\beta+3/2}}{n}\right)\right)^{1/2}. \tag{21}$$

Now, using the Jensen inequality for the convex function $\phi(t) = |t|^3$, we get the elementary inequality

$$|t_1 + t_2|^3 \le 4(|t_1|^3 + |t_2|^3).$$

Hence we can apply in (18) for all $1 \le j \le k$ the inequality

$$\mathsf{E}\,|\widehat{p}_j - p_j^0|^3 \le 4(\mathsf{E}\,|\widehat{p}_j - p_j|^3 + |p_j - p_j^0|^3) \tag{22}$$

and in (19) the inequality

$$\mathsf{E}\,(\widehat{p}_j - p_j^0)^2 \le \mathsf{E}\,(\widehat{p}_j - p_j)^2 + (p_j - p_j^0)^2. \tag{23}$$

Using the fact that the random vector $(X_1, \ldots, X_k) = n(\widehat{p}_1, \ldots \widehat{p}_k)$ is multinomially distributed,

$$(X_1, \ldots, X_k) = n\hat{p} \sim \mathcal{M}_n(p,k), \tag{24}$$

we get

$$\mathsf{E}\,(\widehat{p}_j - p_j)^2 \le \frac{p_j}{n} \tag{25}$$

and using Lemma 2 in Györfi and Vajda (2002) we get also

$$\mathsf{E}\,|\widehat{p}_j - p_j|^3 \le 2\left(\frac{p_j}{n}\right)^{3/2}. \tag{26}$$

Therefore (18), (22) and (26) imply

$$\frac{nA_n}{\sqrt{k}} \le \frac{8}{\sqrt{nk}}\sum_{j=1}^{k}\frac{p_j^{3/2}}{(p_j^0)^2} + \frac{4n}{\sqrt{k}}\sum_{j=1}^{k}\frac{|p_j - p_j^0|^3}{(p_j^0)^2} \tag{27}$$

and (19), (23) and (25) imply

$$B_n \le \frac{1}{n}\sum_{j=1}^{k}\frac{p_j}{(p_j^0)^2} + \sum_{j=1}^{k}\left(\frac{p_j - p_j^0}{p_j^0}\right)^2. \tag{28}$$

By (3) and (**??**),

$$\begin{aligned}\frac{1}{\sqrt{nk}}\sum_{j=1}^{k}\frac{p_j^{3/2}}{(p_j^0)^2} &\le \left(\frac{k^{\beta-1}}{n\gamma}\right)^{1/2}\left(\sum_{j=1}^{k}\frac{p_j}{p_j^0}\right)^{3/2} \\ &= \left(\frac{k^{\beta+1}(1+o(1))}{n\gamma}\right)^{1/2} = o(1) \qquad \text{(cf. (2))}.\end{aligned}$$

By (21) and (2),

$$\frac{n}{\sqrt{k}}\sum_{j=1}^{k}\frac{|p_j - p_j^0|^3}{(p_j^0)^2} = \left(O\left(\frac{k^{\beta+1/2}}{n}\right)\right)^{1/2} = o(1)$$

so that we get from (27)

$$\frac{nA_n}{\sqrt{k}} = o(1). \tag{29}$$

Similarly we get from (28) and (3)

$$\begin{aligned}B_n &\le \frac{k^\beta}{n\gamma}\left[\sum_{j=1}^{k}\frac{p_j}{p_j^0} + \chi^2(p,p^0)\right] \\ &\le \frac{k^\beta}{n\gamma}\left[k(1+o(1)) + O\left(\frac{\sqrt{k}}{n}\right)\right] \qquad \text{(cf. (20) and (10))} \\ &= O\left(\frac{k^{\beta+1}}{n}\right) = o(1) \qquad \text{(cf. (2))}.\end{aligned} \tag{30}$$

By (17), the asymptotic relations (29) and (30) imply the desired relation (16).

3 Discussion.

It is possible to deduce from Theorem 5.1 of Morris (1975) that if the assumptions of our Theorem hold with (10) replaced by the more precise locality condition

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

and $p^0_{\max} = o(1)$ then the asymptotic normality law (14) holds for the Pearson statistic χ^2_n and the sequences

$$\mu = \mu_n = k + \sqrt{k}\Delta \quad \text{and} \quad \sigma = \sigma_n = \sqrt{2k}.$$

Our Theorem under the same conditions guarantees the similar asymptotic normality law

$$\frac{T_{f,n} - k - \sqrt{k}\Delta}{\sqrt{2k}} \xrightarrow{L} N(0,1) \tag{31}$$

for all f-disparity statistics $T_{f,n}$ under consideration. Among others it guarantees the concrete limit law

$$\frac{2I_n - k - \sqrt{k}\Delta}{\sqrt{2k}} \xrightarrow{L} N(0,1) \tag{32}$$

for the likelihod ratio statistic I_n. This law is presented in a more general form inTheorem 5.2 of the cited Morris paper. But the conditions given there theorem are incomparably more complicated than those given here and not easily verifiable in our model. Also the proof of Theorem 5.2 is incomparably more complicated than that given here. On the other hand, our theorem is not only simpler, but it presents the asymptotic normality law for many statistics $T_{f,n}$ different from I_n, e.g. for the *Freeman-Tukey statistic*

$$H^2_n = 4n\sum_{j=1}^{k}\left(\sqrt{\widehat{p}_j} - \sqrt{p^0_j}\right)^2 = 4\sqrt{n}\sum_{j=1}^{k}\left(\sqrt{X_j} - \sqrt{np^0_j}\right)^2$$

obtained for $f(t) = (\sqrt{t}-1)^2$ from the squared Hellinger distance $H^2(\widehat{p},p^0)$.

Asymptotic laws are interesting not only for the Pearson and likelihood ratio statistics χ^2_n and I_n, but also for the remaining statistics $T_{f,n}$ with $f \in \mathcal{F}$. In spite of that the particular statistic I_n is preferable to many others from the asymptotic efficiency point of view (see Quine and Robinson (1985), Györfi et al. (2000), Beirlant et al. (2001) and the more recent results of Harremoës and Vajda (2007,2008)), the non-asymptotic efficiency considerations often prefer $T_{f,n}$ different from I_n, e.g. the Freeman-Tukey's H^2_n.

Acknowledgement. This research was supported by the GAČR grant 103/07/1311 a MŠMT grant 1M0572.

References

- Beirlant, J., Devroye, L., Györfi, L., and Vajda, I. (2001): Large deviations of divergence measures on partitions. *J. Statist. Planning and Infer.*, 93, 1–16.
- 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.
- Györfi, L. and Vajda, I. (2002): Asymptotic distributions for goodness-of-fit statistics in a sequence of multinomial models. *Statist. and Probab. Letters*, 56, 57-67.
- Harremoës, P. and Vajda, I. (2007): Entropy testing is efficient. *Proc. of Internat. Symp. of IEEE on Inform. Theory*, p. 1841-1845, Nice, France, June 24 – 29, 2007.
- Harremoës, P. and Vajda, I. (2008): On the Bahadur-efficient testing of uniformity by means of the entropy. *IEEE Trans. on Inform. Theory,* 48 (in print).
- Menéndez, M. L., Morales, D., Pardo L. and Vajda, I. (1998): Two approaches to grouping of data and related disparity statiustics. *Comm. in Statistics - Theory and Methods*, 27, 609–633.
- Morris, C. (1975): Central limit theorems for multinomial sums. *Ann. of Statistics,* 3, 165–188.
- Quine, M. P. and Robinson, J. (1985): Efficiencies of chi-square and likwlihood ratio tests. *Ann. of Statistics,* 13, 727–742.