Domingo Morales; Leandro Pardo; Igor Vajda

On the optimal number of classes in the Pearson goodness-of-fit tests

Kybernetika, Vol. 41 (2005), No. 6, [677]--698

Persistent URL: http://dml.cz/dmlcz/135686

KYBERNETIKA — VOLUME 41 (2005), NUMBER 6, PAGES 677–698

ON THE OPTIMAL NUMBER OF CLASSES IN THE PEARSON GOODNESS–OF–FIT TESTS

Domingo Morales, Leandro Pardo and Igor Vajda

An asymptotic local power of Pearson chi-squared tests is considered, based on convex mixtures of the null densities with fixed alternative densities when the mixtures tend to the null densities for sample sizes $n \to \infty$. This local power is used to compare the tests with fixed partitions $\mathcal{P}$ of the observation space of small partition sizes $|\mathcal{P}|$ with the tests whose partitions $\mathcal{P} = \mathcal{P}_n$ depend on n and the partition sizes $|\mathcal{P}_n|$ tend to infinity for $n \to \infty$. New conditions are presented under which it is asymptotically optimal to let $|\mathcal{P}|$ tend to infinity with n or to keep it fixed, respectively. Similar conditions are presented under which the tests with fixed $|\mathcal{P}|$ and those with increasing $|\mathcal{P}_n|$ are asymptotically equivalent.

Keywords: Pearson goodness-of-fit test, Pearson-type goodness-of-fit tests, asymptotic local test power, asymptotic equivalence of tests, optimal number of classes

AMS Subject Classification: 62G10, 62G20

1. INTRODUCTION

Let $X_1, \ldots, X_n$ be independent real valued random observations distributed by an unknown probability distribution P. The hypothesis that this distribution is a given Q is frequently tested by the Pearson chi-squared test based on the statistic

$$T_n = n\chi^2(\widehat{P}_n, Q|\,\mathcal{P}) = n \sum_{A \in \mathcal{P}} \frac{(\widehat{P}_n(A) - Q(A))^2}{Q(A)}. \tag{1.1}$$

Here $\widehat{P}_n$ is the empirical distribution defined by means of the relative frequencies

$$\widehat{P}_n(A) = \frac{\#\{X_i \in A\}}{n}, \tag{1.2}$$

$\mathcal{P}$ is a partition of the range of X_i's into $|\mathcal{P}|$ intervals where $1 < |\mathcal{P}| < \infty$, and $\chi^2(\widehat{P}_n, Q|\,\mathcal{P})$ is a special case of the χ^a-divergence defined for arbitrary probability measures μ, ν and $a \geq 1$ by the formula

$$\chi^a(\mu, \nu) = \begin{cases} E_\nu \left|\frac{\mathrm{d}\mu}{\mathrm{d}\nu} - 1\right|^a & \text{if } \mu \ll \nu \\ \infty & \text{otherwise} \end{cases} \tag{1.3}$$

in Vajda [15]. From this formula we obtain the χ^a-divergences

$$\chi^a(\mu,\nu) = \chi^a(P,Q) = \int \frac{|p-q|^a}{q^{a-1}}\,\mathrm{d}x, \qquad a \geq 1, \tag{1.4}$$

when $\mu = P$, $\nu = Q$ are absolutely continuous on the real line with densities p, q and

$$\chi^a(\mu,\nu) = \chi^a(P,Q\,|\,\mathcal{P}) = \sum_{A\in\mathcal{P}} \frac{|P(A)-Q(A)|^a}{Q(A)^{a-1}}, \qquad a \geq 1, \tag{1.5}$$

when μ, ν are restrictions of P, Q on the algebra generated by an interval partition $\mathcal{P}$ of the real line. It is known that for every $a > 1$

$$\chi^a(P,Q\,|\,\mathcal{P}) \leq \chi^a(P,Q) \tag{1.6}$$

and

$$\chi^a(P,Q\,|\,\mathcal{P}) \to \chi^a(P,Q) \quad \text{if } \max_{A\in\mathcal{P}} Q(A) \to 0 \tag{1.7}$$

(cf. Vajda [16] or Berlinet and Vajda [1]).

We see that the Pearson statistic (1.1) measures the fit between the observations and the hypothesis by the χ^2-divergence of the empirical distribution $\widehat{P}_n$ and the hypothetic distribution Q observed on a partition $\mathcal{P}$. In applications of this test, the most sensitive point is the choice of the interval partition $\mathcal{P}$ because the difference between

$$n \inf_{\mathcal{P}} \chi^2(\widehat{P}_n, Q\,|\,\mathcal{P}) \quad \text{and} \quad n \sup_{\mathcal{P}} \chi^2(\widehat{P}_n, Q\,|\,\mathcal{P})$$

is usually infinite. The classical rules mentioned e. g. in the monograph of Greenwood and Nikulin [5] consider the Q-equiprobable partitions $\mathcal{P}$ and recommend $\mathcal{P} = \mathcal{P}_n$ depending on the sample size n where the number of intervals $|\mathcal{P}_n|$ increases with $n \to \infty$ at a rate smaller than $\sqrt{n}$. For example, the rule of Mann and Wald [10] proposes $|\mathcal{P}_n| = n^{2/5}$. However, simulation studies demonstrated that for particular alternatives the powers of the tests with Q-uniform $\mathcal{P}$ and small fixed values of $|\mathcal{P}|$ may be much better than those of the tests with Q-uniform $\mathcal{P}$ and $|\mathcal{P}| = |\mathcal{P}_n|$ tending to infinity as $n \to \infty$.

This paradox was resolved by Kallenberg et al. [8] who assumed that the power of the test with $\mathcal{P} = \mathcal{P}_n$ at a fixed alternative P is reflected by the *asymptotic local power* (ALP) of the same test at the local alternatives

$$P_n = (1-\varepsilon_n)\,Q + \varepsilon_n P \tag{1.8}$$

when $n \to \infty$ and $\varepsilon_n \downarrow 0$.

Notice that the observations $(X_1, \dots, X_n)$ form under the alternatives (1.8) a triangular scheme which can be written as $(X_1^{(n)}, \dots, X_n^{(n)})$. Kallenberg et al. [8] considered the situation where P, Q are absolutely continuous with densities p, q and the sequence $\rho_n = n\varepsilon_n^2$ is bounded away from 0 and ∞, i. e.

$$0 < \lim_{n\to\infty} \inf \rho_n \leq \lim_{n\to\infty} \sup \rho_n < \infty. \tag{1.9}$$

This means that ε_n tends to zero with the rate $1/\sqrt{n}$ as $n \to \infty$. The main theoretic result stated that if the statistical model $\{P, Q\}$ satisfies some regularity conditions then the solution of the problem of asymptotically optimal partitions for the Pearson test depends on the values of the χ^a-divergence $\chi^a(P,Q)$ for a from a neighborhood of $a = 4/3$. Namely, if $\chi^a(P,Q) < \infty$ for some $a > 4/3$ then the above introduced ALP is highest for a fixed $\mathcal{P}$ with small $|\mathcal{P}|$, while if $\chi^a(P,Q) = \infty$ for some $a < 4/3$ then the ALP is highest for $\mathcal{P} = \mathcal{P}_n$ with $|\mathcal{P}_n| \to \infty$ for $n \to \infty$.

This paper completes and extends the results of Kallenberg et al. [8]. It differs mainly in the following three aspects.

(I) It proves a similar characterization of the asymptotic optimality of partitions, but with the χ^a-divergence (1.4) replaced by the moment generating function

$$M_a(P,Q) = \int \frac{p^a}{q^{a-1}}\,\mathrm{d}x = E_Q\left(\exp\{aY\}\right), \quad a \geq 1 \tag{1.10}$$

of the log-likelihood ratio $Y = \ln(p/q)$. The corresponding results are in Theorems 3.1 and 3.3 of Section 3. The advantage of this new characterization is that the moment generating functions $M_a(P,Q)$ are more familiar than the divergences $\chi^a(P,Q)$. Moreover (see e. g. pp. 42–43 in Liese and Vajda [9]), the values of $M_a(P,Q)$ can be explicitly evaluated for all statistical models $\{P,Q\}$ from general exponential families. This cannot be said about the values of $\chi^a(P,Q)$ for $a \neq 2$. For example, the normal distributions P, Q with means $\mu_1 \neq \mu_2$ and variance σ^2 lead to the simple moment generating function

$$M_a(P,Q) = \exp\left\{a(a-1)\frac{(\mu_1-\mu_2)^2}{2\sigma^2}\right\}, \quad a \in \mathbb{R}$$

while the divergence $\chi^a(P,Q)$ has no closed form for $1 \leq a < 2$ or any a different from the even integers $2, 4, 6, \ldots$

(II) The paper presents detailed versions of the proofs sketched in Kallenberg et al. ([8]. In particular, the application of the general theorem of Morris [14] in the problem under consideration represents a nontrivial step.

(III) It extends the results of Kallenberg et al. [8] to the situations where the alternatives P are emphasized less or more than in (1.9) in the sense that

$$\lim_{n\to\infty} \rho_n = 0 \qquad \text{or} \qquad \lim_{n\to\infty} \rho_n = \infty$$

respectively. Application of the theorem of Morris [14] in these extremal situations is more delicate than under (1.9).

Note that instead of (1.9) we in fact suppose

$$\lim_{n\to\infty} \rho_n = \rho \in (0,\infty). \tag{1.11}$$

The analysis is under (1.11) simpler and the results obtained in this case automatically extend to the case (1.9). This means that Theorems 3.1 and 3.3 consider the asymptotic domination of the Pearson statistics $T_n(\mathcal{P})$ over $T_n(\mathcal{P}_n)$ (cf. (1.1)) and vice versa under both (1.11) and (1.9) and, moreover, they extend these relations also to the situations where $\lim_n \rho_n = \rho \in \{0, \infty\}$. But this paper introduces also the asymptotic equivalence of the statistics $T_n(\mathcal{P})$ and $T_n(\mathcal{P}_n)$ and formulates Theorem 3.2 about this equivalence when $\rho = \infty$ and $\rho = 0$

We think that the additional attention paid to the results of Kallenberg et al. [8] is meaningful because these results were not marginalized by the more recent research in this area (see e. g. in Inglot and Janic–Wróblewska [7] and references therein).

Let us add a technical remark that the proofs of all auxiliary results of this paper (called Propositions) are deferred to the Appendix. Let us also note that the end of Section 3 mentions extensions of some of the above discussed results to some non-Pearson statistics $T_n(\mathcal{P})$ and $T_n(\mathcal{P}_n)$ considered e. g. by Drost et al. [2], Menéndez et al. [13] or Mayoral et al. [11]). Section 4 discusses practical statistical consequences of the results considered in the present paper. It also provides arguments in favour of the partitions $\mathcal{P}$ which in the class with fixed $|\mathcal{P}|$ maximize the χ^2-divergence $\chi^2(P, Q|\mathcal{P})$.

2. BASIC CONCEPTS AND AUXILIARY RESULTS

We consider two different probability measures P and Q on the Borel-measurable space $(\mathcal{X}, \mathcal{B})$ where $\mathcal{X} \subseteq \mathbb{R}$ is an interval. We assume that P, Q are absolutely continuous w.r.t. the Lebesgue measure denoted by λ, with the densities

$$p = \frac{\mathrm{d}P}{\mathrm{d}\lambda} \quad \text{and} \quad q = \frac{\mathrm{d}Q}{\mathrm{d}\lambda}$$

and we suppose $q = q(x) > 0$ everywhere on $\mathcal{X}$. Therefore we have $P \ll Q$ (absolute continuity).

By $\chi^2(P, Q)$ we denote the Pearson divergence of P and Q,

$$\chi^2(P, Q) = \int \frac{(p-q)^2}{q} \, \mathrm{d}\lambda \in [0, \infty] \tag{2.1}$$

and by $\mathcal{P}$ we denote finite partitions of $\mathcal{X}$ into $|\mathcal{P}|$ intervals.

Definition 2.1. A partition $\mathcal{P}$ is said (P, Q)-regular if it is Q-uniform and the restrictions $P^{\mathcal{P}}$, $Q^{\mathcal{P}}$ of P, Q on the algebra $\mathcal{A}(\mathcal{P}) \subset \mathcal{B}$ generated by $\mathcal{P}$ do not coincide, i. e. if

$$Q(A) = \frac{1}{|\mathcal{P}|} \quad \text{and} \quad \chi^2(P, Q|\mathcal{P}) \triangleq \sum_{A \in \mathcal{P}} \frac{(P(A) - Q(A))^2}{Q(A)} > 0. \tag{2.2}$$

where $\chi^2(P, Q|\mathcal{P})$ is the Pearson divergence restricted to $\mathcal{P}$.

If $P^{\mathcal{P}}$, $Q^{\mathcal{P}}$ and $\lambda^{\mathcal{P}}$ are restrictions of P, Q and λ on the σalgebra $\mathcal{A}(\mathcal{P}) \subset \mathcal{B}$ generated by $\mathcal{P}$ then $P^{\mathcal{P}}$, $Q^{\mathcal{P}} \ll \lambda^{\mathcal{P}}$ and $\chi^2(P, Q|\mathcal{P})$ is the Pearson divergence of

$P^{\mathcal{P}}$, $Q^{\mathcal{P}}$, i.e.,

$$\chi^2(P^{\mathcal{P}}, Q^{\mathcal{P}}) = \chi^2(P, Q|\,\mathcal{P}).$$

Notice also that the second condition in (2.2) implies $|\mathcal{P}| > 1$.

Definition 2.2. A sequence $\mathcal{P}_n$, $n = 1, 2, \ldots$, of partitions of $\mathcal{X}$ into $|\mathcal{P}_n|$ intervals is (P, Q)-regular if each partition $\mathcal{P}_n$ is (P, Q)-regular and $|\mathcal{P}_n|$ increases to ∞, i.e.

$$\lim_{n\to\infty} |\mathcal{P}_n| = \infty, \tag{2.3}$$

but not too fast in the sense that

$$|\mathcal{P}_n| = O(\sqrt{n}) \quad \text{as} \quad n \to \infty. \tag{2.4}$$

By Theorem 30.B of Halmos [6], the assumed absolute continuity $P \ll Q$ implies that $Q(A_n) \to 0$ only if $P(A_n) \to 0$. Hence (2.3) implies for regular sequences of partitions $\mathcal{P}_n$

$$\max_{A\in\mathcal{P}_n} P(A) \to 0 \tag{2.5}$$

and, consequently,

$$\sum_{A\in\mathcal{P}_n} (P(A) + Q(A))^2 \le 2\left(\max_{A\in\mathcal{P}_n} P(A) + \frac{1}{|\mathcal{P}_n|}\right) \to 0.$$

From here and from (2.2) with $\mathcal{P}$ replaced by $\mathcal{P}_n$ we get

$$\chi^2(P, Q|\mathcal{P}_n) = |\mathcal{P}_n| \sum_{A\in\mathcal{P}_n} (P(A) + Q(A))^2 = o(|\mathcal{P}_n|) \tag{2.6}$$

as $n \to \infty$. Moreover, for regular sequences of partitions $\mathcal{P}_n$

$$\lim_{n\to\infty} \chi^2(P, Q|\mathcal{P}_n) = \chi^2(P, Q). \tag{2.7}$$

This follows from Theorem 2 in Vajda [16] when $\chi^2(P, Q) \in (0, \infty)$, or from Theorem 1 in Berlinet and Vajda [1] when $\chi^2(P, Q) = \infty$, respectively.

The distribution Q represents a hypothesis tested against the local alternatives

$$P_n = (1 - \varepsilon_n)\, Q + \varepsilon_n P \tag{2.8}$$

on the basis of random observations $X_1^{(n)}, \ldots, X_n^{(n)}$ i.i.d. by P_n for a given sequence

$$\varepsilon_n \in (0, 1) \ \text{with} \ \lim_{n\to\infty} \varepsilon_n = 0. \tag{2.9}$$

In what follows we use the sequences

$$\rho_n = n\varepsilon_n^2 \tag{2.10}$$

and

$$\mu_n = \frac{\rho_n \chi^2(P, Q|\mathcal{P}_n)}{\sqrt{|\mathcal{P}_n|}}. \tag{2.11}$$

We restrict ourselves to sequences ε_n and $\mathcal{P}_n$ such that

$$\lim_{n\to\infty} \rho_n = \rho \quad \text{for some} \quad \rho \in [0, \infty] \tag{2.12}$$

and

$$\lim_{n\to\infty} \mu_n = \mu \quad \text{for some} \quad \mu \in [0, \infty]. \tag{2.13}$$

The main attention we pay to the classical Pearson statistic

$$T_n(\mathcal{P}) = n\chi^2(\widehat{P}_n, Q|\mathcal{P}), \; n = 1, 2, \ldots \tag{2.14}$$

where

$$\widehat{P}_n = \frac{1}{n}\sum_{i=1}^{n} \delta_{X_i^{(n)}} \tag{2.15}$$

is the empirical distribution on $(\mathcal{X}, \mathcal{B})$ with the mass uniformly distributed at the observation points and $\mathcal{P}$ is a fixed nontrivial partition not depending on the sample size $n = 1, 2, \ldots$ However, we are also interested in the alternative Pearson test statistics $T_n(\mathcal{P}_n)$ defined by standard sequences of partitions $\mathcal{P}_n$,

$$T_n(\mathcal{P}_n) = n\chi^2(\widehat{P}_n, Q|\mathcal{P}_n), \qquad n = 1, 2, \ldots \tag{2.16}$$

As mentioned above, we shall consider also the extensions of the statistics obtained by replacing the Pearson χ^2-divergence in (2.14), (2.16) by alternative divergence measures, see the end of Section 3.

In accordance with Kallenberg et al. [8], we believe that the preferences between the statistics $T_n(\mathcal{P})$ and $T_n(\mathcal{P}_n)$ when these are used in testing of

$$\mathcal{H} : Q \quad \text{against} \quad \mathcal{A}_n : P_n \tag{2.17}$$

extrapolate in some sense also to the situation where these statistics are used to test

$$\mathcal{H} : Q \quad \text{against} \quad \mathcal{A}_n : P. \tag{2.18}$$

Preferences between the statistics $T_n(\mathcal{P})$ and $T_n(\mathcal{P}_n)$ in the problem (2.17) are rigorously specified in the following definitions.

Definition 2.3. Let $c_n(\alpha)$ or $c_n^*(\alpha)$ be the asymptotically α-size critical values of the statistic $T_n(\mathcal{P})$ or $T_n(\mathcal{P}_n)$, respectively, i. e.

$$\lim_{n\to\infty} \mathsf{P}(T_n(\mathcal{P}) > c_n(\alpha)) = \alpha, \quad \lim_{n\to\infty} \mathsf{P}(T_n(\mathcal{P}_n) > c_n^*(\alpha)) = \alpha, \; \alpha \in (0,1),$$

when $P = Q$. Then the asymptotic local test powers corresponding to these statistics are defined for every $\alpha \in (0, 1)$ as follows

$$\pi(\alpha) = 1 - \limsup_{n\to\infty} \mathsf{P}(T_n(\mathcal{P}) < c_n(\alpha))$$

and

$$\pi^*(\alpha) = 1 - \limsup_{n\to\infty} \mathsf{P}(T_n(\mathcal{P}_n) < c_n^*(\alpha)).$$

Definition 2.4. The statistic $T_n(\mathcal{P})$ is asymptotically better than $T_n(\mathcal{P}_n)$ for the given sequence ε_n (in symbols $T_n(\mathcal{P}) \overset{\varepsilon_n}{\succ} T_n(\mathcal{P}_n)$) if $\pi(\alpha) > \pi^*(\alpha)$ for all $\alpha \in (0,1)$. The reversed inequality defines the preference relation $T_n(\mathcal{P}_n) \overset{\varepsilon_n}{\succ} T_n(\mathcal{P})$. If $\pi(\alpha) = \pi^*(\alpha)$ for all $\alpha \in (0,1)$ then we write $T_n(\mathcal{P}) \overset{\varepsilon_n}{\approx} T_n(\mathcal{P}_n)$ and say that the statistics $T_n(\mathcal{P})$ and $T_n(\mathcal{P}_n)$ are asymptotically equivalent for a given sequence ε_n.

Convention. Unless otherwise explicitly stated, the convergences $\longrightarrow$ and the asymptotic expressions $o(.)$ and $O(.)$ are considered for $n \longrightarrow \infty$.

3. MAIN RESULTS

In this section we evaluate and compare the asymptotic local powers $\pi(\alpha)$ and $\pi^*(\alpha)$ of the Pearson statistics $T_n(\mathcal{P})$ and $T_n(\mathcal{P}_n)$. The powers are evaluated using the limit laws presented in the following two propositions. Remind that there, as in the rest of the paper, the observations are assumed to form triangular schemes distributed by the probability laws (1.8) (see also(2.8)) satisfying (2.12) and (2.13).

In the first proposition $\Phi_k(x)$ stands for the distribution function of the χ^2-distributed random variable with $k \ge 1$ degrees of freedom and $\Phi_k^{-1}(\alpha)$ denotes the corresponding α-quantile. Further, by $\Phi_{k,m}(x)$ we denote the distribution function of the noncentral χ^2 with $k \ge 1$ degrees of freedom and noncentrality $m \ge 0$. It holds $\Phi_{k,0}(x) = \Phi_k(x)$ and, more generally,

$$\Phi_{k,m}(x) = e^{-m} \sum_{i=0}^{\infty} \frac{m^i}{i!}\, \Phi_{k+2i}(x), \quad x \in (0,\infty),\ m \ge 0. \tag{3.1}$$

Remind that the proofs of all propositions are deferred to the Appendix.

Proposition 3.1. Let $T_n(\mathcal{P})$ be the Pearson statistic for a (P,Q)-regular partition $\mathcal{P}$. Then for every $x \in \mathbb{R}$

$$\mathsf{P}\left(T_n(\mathcal{P}) < x\right) \longrightarrow \begin{cases} \Phi_{|\mathcal{P}|-1,\rho\chi^2(P,Q|\mathcal{P})}(x) & \text{if} \quad \rho \in [0,\infty) \\ 0 & \text{if} \quad \rho = \infty \end{cases} \tag{3.2}$$

where ρ is given by (2.12).

By this proposition, $c(\alpha) = \Phi^{-1}_{|\mathcal{P}|-1}(1-\alpha)$ is the critical value of the statistic $T_n(\mathcal{P})$ for the asymptotic test size $\alpha \in (0,1)$. This means that

$$1 - \lim_{n\to\infty} \mathsf{P}\left(T_n(\mathcal{P}) < \Phi^{-1}_{|\mathcal{P}|-1}(1-\alpha)\right)$$

is the asymptotic power $\pi(\alpha)$. Hence (3.2) implies the following.

Corollary 3.1. The asymptotic local power of $T_n(\mathcal{P})$ is given for all $\alpha \in (0,1)$ by the formula

$$\pi(\alpha) = \begin{cases} \alpha & \text{if } \rho = 0 \\ 1 - \Phi_{|\mathcal{P}|-1, \rho\chi^2(P,Q|\mathcal{P})}(\Phi^{-1}_{|\mathcal{P}|-1}(1-\alpha)) & \text{if } \rho \in (0,\infty) \\ 1 & \text{if } \rho = \infty. \end{cases} \tag{3.3}$$

In what follows we denote by $\Phi(x)$ the standard normal distribution function and by $\Phi^{-1}(\alpha)$ the corresponding α-quantile. We also consider the condition

$$\varepsilon_n = O(1/\sqrt{|\mathcal{P}_n|}). \tag{3.4}$$

Notice that this condition imposes no restriction on ε_n when $\rho \in [0,\infty)$ because in this case (3.4) is automatically fulfilled. Indeed, by (2.10) and (2.12), $\rho \in [0,\infty)$ implies

$$\varepsilon_n = O(1/\sqrt{n})$$

and by (2.4) this is stronger than (3.4).The condition (3.4) represents a restriction on ε_n only when $\rho = \infty$.

Proposition 3.2. Let $T_n(\mathcal{P}_n)$ be the Pearson statistic for a (P,Q)-regular sequence of partitions $\mathcal{P}_n$ leading to a limit μ considered in (2.13). If (3.4) holds then for all $x \in \mathbb{R}$

$$\mathsf{P}\left(\frac{T_n(\mathcal{P}_n) - |\mathcal{P}_n|}{\sqrt{2|\mathcal{P}_n|}} < x\right) \to \begin{cases} \Phi\left(x - \mu/\sqrt{2}\right) & \text{if } \mu \in [0,\infty) \\ 0 & \text{if } \mu = \infty \end{cases} \tag{3.5}$$

unless $\rho = \mu = \infty$.

We see that under the assumptions of Proposition 3.2

$$c_n^*(\alpha) = |\mathcal{P}_n| + \sqrt{2|\mathcal{P}_n|}\Phi^{-1}(1-\alpha) \tag{3.6}$$

are critical values of the statistic $T_n(\mathcal{P}_n)$ for the asymptotic test sizes $\alpha \in (0,1)$. Thus

$$1 - \lim_{n\to\infty} \mathsf{P}\left(\frac{T_n(\mathcal{P}_n) - |\mathcal{P}_n|}{\sqrt{2|\mathcal{P}_n|}} < \Phi^{-1}(1-\alpha)\right) = \lim_{n\to\infty} \mathsf{P}\left(\frac{T_n(\mathcal{P}_n) - |\mathcal{P}_n|}{\sqrt{2|\mathcal{P}_n|}} < \Phi^{-1}(\alpha)\right)$$

are the corresponding test powers $\pi(\alpha)$. The limit law (3.5) implies the following assertion.

Corollary 3.2. If the assumptions of Proposition 3.2 hold then the asymptotic local power of $T_n(\mathcal{P}_n)$ is given for all $\alpha \in (0,1)$ by the formula

$$\pi^*(\alpha) = \begin{cases} \alpha & \text{if } \mu = 0 \\ \Phi\left(\frac{\mu}{\sqrt{2}} + \Phi^{-1}(\alpha)\right) & \text{if } \mu \in (0,\infty) \\ 1 & \text{if } \mu = \infty. \end{cases} \tag{3.7}$$

The table given below summarizes the preference relations which can be drawn directly from Corollaries 3.1 and 3.2. The preference in the row for $\rho \in (0, \infty)$ and $\mu = 0$ is drawn with the help of the inequalities $\Phi_{k,\rho}(x) < \Phi_k(x)$ which follow for every $\rho > 0$ and $x \in (0, \infty)$ from (3.1) and from the monotonicity of $\Phi_k(x)$ in the variable $k = 1, 2, \ldots$

Table. Asymptotic preferences between the Pearson statistics for fixed nontrivial partitions $\mathcal{P}$ and (P, Q)-regular sequences of partitions $\mathcal{P}_n$ when the alternatives are specified by sequences ε_n satisfying (2.9), (2.12) and (2.13).

Values of ρ	Values of μ	Preferences
∞	$[0, \infty)$	$T_n(\mathcal{P}) \overset{\varepsilon_n}{\succ} T_n(\mathcal{P}_n)$
$(0, \infty)$	0	$T_n(\mathcal{P}) \overset{\varepsilon_n}{\succ} T_n(\mathcal{P}_n)$
	$(0, \infty)$	$T_n(\mathcal{P}) \overset{\varepsilon_n}{\succ} T_n(\mathcal{P}_n)$ if for $k = \lvert\mathcal{P}\rvert - 1$ and all $\alpha \in (0, 1)$ $\Phi_{k,\rho\chi^2(P,Q\mid\mathcal{P})}\left(\Phi_k^{-1}(1-\alpha)\right) < \Phi\left(\Phi^{-1}(1-\alpha) - \mu/\sqrt{2}\right)$
		$T_n(\mathcal{P}_n) \overset{\varepsilon_n}{\succ} T_n(\mathcal{P})$ if for $k = \lvert\mathcal{P}\rvert - 1$ and all $\alpha \in (0, 1)$ $\Phi_{k,\rho\chi^2(P,Q\mid\mathcal{P})}\left(\Phi_k^{-1}(1-\alpha)\right) > \Phi\left(\Phi^{-1}(1-\alpha) - \mu/\sqrt{2}\right)$
	∞	$T_n(\mathcal{P}_n) \overset{\varepsilon_n}{\succ} T_n(\mathcal{P})$
0	$(0, \infty]$	$T_n(\mathcal{P}_n) \overset{\varepsilon_n}{\succ} T_n(\mathcal{P})$
	0	$T_n(\mathcal{P}_n) \overset{\varepsilon_n}{\approx} T_n(\mathcal{P})$

In the rest of this section we find conditions on models $\{P, Q\}$ and sequences ε_n, $\mathcal{P}_n$ under which the preference relations of the Table take place. A key to the conditions are the following two definitions and the related propositions. They deal with the moment generating function

$$M_a(P, Q) = \int \exp\{a \ln p/q\}\,\mathrm{d}Q = \int \left(\frac{p}{q}\right)^a \mathrm{d}Q, \quad a \in \mathbb{R}, \tag{3.8}$$

of the log-likelihood ratio. The Hölder inequality implies $M_a(P, Q) \le 1$ for $a \in [0, 1]$, but for $a > 1$ the function $M_a(P, Q)$ may be infinite. Note that the standardized moment generators

$$D_a(P, Q) = \frac{M_a(P, Q) - 1}{a(a-1)}, \; a \neq 0, \; a \neq 1, \tag{3.9}$$

are usually called power divergences of P, Q. The remaining power divergences $D_0(P, Q)$ and $D_1(P, Q)$ are obtained by continuous extension of (3.9) from the right and left, respectively (see Proposition 2.14 in Liese and Vajda [9]). It is easy to see that $M_a(Q, P) = M_{1-a}(P, Q)$ and that $D_2(P, Q) = \chi^2(P, Q)/2$.

Definition 3.1. The critical power $a(P, Q) \in [1, \infty]$ of distributions P, Q is defined by

$$a(P, Q) = \sup\{a \in \mathbb{R} : M_a(P, Q) < \infty\}.$$

Thus the critical power $a\,(P,Q)$ separates the powers of the finite power divergences of P,Q from the powers of the infinite power divergences of P,Q. It is the maximal power at which the power divergence of P,Q may still be finite.

Proposition 3.3. Let $a = a\,(P,Q) \geq 1$ be the critical power of P,Q and $\mathcal{P}_n$ a (P,Q)-regular sequence of partitions. Then

$$\frac{\chi^2(P,Q|\,\mathcal{P}_n)}{|\mathcal{P}_n|^b} \to 0 \quad \text{if} \quad b > b_0 \tag{3.10}$$

for $b_0 = (2-a)/a \in (-1,1]$ if $a \in [1,\infty)$ and $b_0 = \lim_{a\to\infty}(2-a)/a = -1$ if $a = \infty$.

In the following definition we consider the function

$$f\,(y) = \frac{p\,(G^{-1}\,(y))}{q\,(G^{-1}\,(y))},\ y \in (0,1)$$

where $G^{-1}\,(y)$ is the quantile function of the distribution Q (the inverse of the distribution function $G(x) = Q((-\infty,x))$ being by assumption increasing on the interval $\mathcal{X}$).Note that the conditions formulated in this definition and an assertion equivalent to the following proposition were first presented in Kallenberg et al. [8].

Definition 3.2. We say that the model $\{P,Q\}$ is regular if $f\,(y)$ can be continuously extended on $[0,1]$ with the extension being bounded except in neighborhoods of finitely many points. If it is not bounded in a right (left) neighborhood of $y \in [0,1]$ then it is assumed that $h(t) = f(y+1/t)$ (or $h(t) = f(y-1/t)$) varies regularly at infinity. This means that for sufficiently large $t > 0$ and some $\beta \in \mathbb{R}$

$$h\,(t) = t^{\beta} g(t) \tag{3.11}$$

where $g(t)$ varies slowly at infinity in the sense

$$\lim_{t\to\infty} \frac{g\,(t\alpha)}{g\,(t)} = 1 \qquad \text{for any } \alpha > 0.$$

Proposition 3.4. Let the model $\{P,Q\}$ be regular with the critical power $a = a\,(P,Q)$, and let $\mathcal{P}_n$ be a (P,Q)-regular sequence of partitions. Then

$$\frac{\chi^2(P,Q|\,\mathcal{P}_n)}{|\mathcal{P}_n|^b} \to \infty \quad \text{if} \quad b < b_0 \tag{3.12}$$

for b_0 defined in Proposition 3.3.

Put $\chi_n^2(P,Q|\,\mathcal{P}_n) = \chi_n^2$ for a (P,Q)-regular sequence $\mathcal{P}_n$ and define $c_0 = (4/3-a)$ / $(2a/3)$ for $a = a\,(P,Q)$. Since $a \in [1,\infty]$, it holds $c_0 \in [-3/2, 1/2]$ where $c_0 = -3/2$ if $a = \infty$. By (2.11) and Propositions 3.3 and 3.4,

$$\frac{\mu_n}{\rho_n\,|\mathcal{P}_n|^c} = \frac{\rho_n \chi_n^2}{\sqrt{|\mathcal{P}_n|}} \frac{1}{\rho_n\,|\mathcal{P}_n|^c} \to \begin{cases} 0 & \text{if } c > c_0 \\ \infty & \text{if } c < c_0 \end{cases}$$

where the convergence to ∞ is justified only for the regular models $\{P,Q\}$. Hence

$$(\rho_n\,|\mathcal{P}_n|^c \to 0 \ \text{ for some } \ c > c_0) \Rightarrow \mu_n \to 0 \tag{3.13}$$

for arbitrary models $\{P,Q\}$ and

$$(\rho_n\,|\mathcal{P}_n|^c \to \infty \ \text{ for some } \ c < c_0) \Rightarrow \mu_n \to \infty \tag{3.14}$$

for the regular models $\{P,Q\}$.

The main results of this paper are the following three theorems. In the proofs we use the property $|\mathcal{P}_n| \to \infty$ of regular sequences of partitions introduced in Definition 2.2.

Theorem 3.1. Let $\mathcal{P}$ be a (P,Q)-regular partition and $\mathcal{P}_n$ a sequence of (P,Q)-regular partitions. If $\rho \in (0,\infty)$ then $T_n(\mathcal{P}) \overset{\varepsilon_n}{\succ} T_n(\mathcal{P}_n)$ provided $a = a\,(P,Q) > 4/3$, i.e. provided $c_0 = (4-3a)/2a$ is negative. This assertion remains valid also when $\rho = \infty$ under the additional conditions (3.4) and

$$\rho_n|\mathcal{P}_n|^c \to 0 \quad \text{for some } \ c > c_0 = (4-3a)/2a. \tag{3.15}$$

Proof. If $a > 4/3$ then $c_0 < 0$. Therefore, if $\rho_n \to \rho \in (0,\infty)$ then $\rho_n\,|\mathcal{P}_n|^c \to 0$ for all $0 > c > c_0$ and, by (3.13), $\mu_n \to \mu = 0$. The desired relation $T_n(\mathcal{P}) \overset{\varepsilon_n}{\succ} T_n(\mathcal{P}_n)$ follows from the row $\rho \in (0,\infty)$, $\mu = 0$ of the Table. If $\rho_n \to \rho = \infty$ then (3.15) and (3.13) imply $\mu_n \to \mu = 0$ and the desired relation $T_n(\mathcal{P}) \overset{\varepsilon_n}{\succ} T_n(\mathcal{P}_n)$ follows from the row $\rho = \infty$, $\mu \in [0,\infty)$ of the Table. □

Theorem 3.2. Let $\mathcal{P}$ be a (P,Q)-regular partition and $\mathcal{P}_n$ a sequence of (P,Q)-regular partitions. If $\rho = 0$ then $T_n(\mathcal{P}) \overset{\varepsilon_n}{\approx} T_n(\mathcal{P}_n)$ as soon as $a = a\,(P,Q) > 4/3$. This statement remains valid also in the models with $a = a\,(P,Q) \le 4/3$, i.e. with $c_0 = (4-3a)/2a$ nonnegative, provided the partitions satisfy the condition (3.15).

Proof. Let $\mathcal{P}_n$, $\{P,Q\}$ be as assumed in the theorem, and let $\rho_n \to 0$. If $a > 4/3$ then, similarly as in the previous proof, $c_0 < 0$ and $\rho_n\,|\mathcal{P}_n|^c \to 0$ holds for all $0 > c > c_0$. Therefore (3.13) implies $\mu_n \to 0$ and the desired equivalence of $T_n(\mathcal{P})$ and $T_n(\mathcal{P}_n)$ follows from the row $\rho = 0$, $\mu = 0$ of the Table. If $a \le 4/3$ then $c_0 \ge 0$ and (3.15) together with (3.13) implies $\mu_n \to 0$ and the rest is the same as before. This completes the proof. □

Theorem 3.3. Let $\mathcal{P}$ be a (P,Q)-regular partition and $\mathcal{P}_n$ a sequence of (P,Q)-regular partitions. If $\rho \in (0,\infty)$ then $T_n(\mathcal{P}_n) \overset{\varepsilon_n}{\succ} T_n(\mathcal{P})$ provided the model $\{P,Q\}$ is regular and $a = a\,(P,Q) < 4/3$. This statement remains valid also when $\rho = 0$ provided the partitions $\mathcal{P}_n$ satisfy the additional condition

$$\rho_n|\mathcal{P}_n|^c \to \infty \quad \text{for some } \ c < c_0 = (4-3a)/2a. \tag{3.16}$$

Proof. If $a < 4/3$ then $c_0 = (4/3 - a) / (2a/3) > 0$. The assumption $\rho_n \to \rho \in (0, \infty)$ implies $\rho_n |\mathcal{P}_n|^c \to \infty$ for all $0 < c < c_0$. Therefore (3.14) implies $\mu_n \to \mu = \infty$ and the desired preference $T_n(\mathcal{P}_n) \overset{\varepsilon_n}{\succ} T_n(\mathcal{P})$ follows from the row $\rho \in (0, \infty)$, $\mu = \infty$ of the Table. If $\rho_n \to \rho = 0$ then (3.16) together with (3.14) implies $\mu_n \to \mu = \infty$ and the desired preference follows from the row $\rho = 0$, $\mu \in (0, \infty]$ of the Table. □

Theorems 3.1 – 3.3 present sufficient conditions for all combinations of values of the limits μ and ρ considered in the Table (and for the preferences corresponding to those combinations) except the case where $\rho \in (0, \infty)$ and $\mu \in (0, \infty)$, i. e. where both these limits are non-zero and finite. In this case we face the problem to compare the functions

$$F_\mu(\alpha) = \Phi\left(\Phi^{-1}(1-\alpha) - \mu/\sqrt{2}\right)$$

and

$$F_{k,\mu_k}(\alpha) = \Phi_{k,\sqrt{k}\mu_k}\left(\Phi_k^{-1}(1-\alpha)\right)$$

where

$$k = |\mathcal{P}_n| - 1 \qquad \text{and} \qquad \mu_k = \frac{\rho\chi^2(P,Q|\mathcal{P})}{\sqrt{|\mathcal{P}|}},$$

i. e. where the parameter k represents the information about the number of intervals in the partition $\mathcal{P}$ and μ_k summarizes the information about the limit ρ and about the Pearson divergence of P, Q achieved on $\mathcal{P}$. The approximations

$$\Phi_{k,m} \approx \Phi\left(\frac{x-k-m}{\sqrt{2k}}\right), \qquad \Phi_k^{-1}(\alpha) \approx k + \sqrt{2k}\Phi^{-1}(\alpha)$$

lead to the approximate formula

$$F_{k,\mu_k}(\alpha) \approx \Phi\left(\Phi^{-1}(1-\alpha) - \mu_k/\sqrt{2}\right).$$

Therefore the preference $T_n(\mathcal{P}) \overset{\varepsilon_n}{\succ} T_n(\mathcal{P}_n)$ can be expected if

$$\mu_k = \frac{\rho\chi^2(P,Q|\mathcal{P})}{\sqrt{|\mathcal{P}|}} > \mu$$

and the reversed preference $T_n(\mathcal{P}_n) \overset{\varepsilon_n}{\succ} T_n(\mathcal{P})$ can be expected if this inequality is reversed. The figure given below presents the curves $F_2(\alpha)$, $F_{k,1/\sqrt{k}}(\alpha)$ and $F_{k,10/\sqrt{k}}(\alpha)$ for $k \in \{1,2,4,8\}$. Since $\mu_k = 1/\sqrt{k}$ is below $\mu = 2$, the curves $F_{k,1/\sqrt{k}}(\alpha)$ are above $F_2(\alpha)$ while $\mu_k = 10/\sqrt{k}$ is above $\mu = 2$ so that the curves $F_{k,10/\sqrt{k}}(\alpha)$ are below $F_2(\alpha)$.

The main purpose of the Figure is to demonstrate numerically that the preference relations in the rows of the Table corresponding to $\rho \in (0, \infty)$ and $\mu \in (0, \infty)$ are possible, i. e. that these rows are not logically empty. The analysis of the functions $F_\mu(\alpha)$ and $F_{k,\mu_k}(\alpha)$ is an interesting problem which however represents a different area of research and requires different techniques than considered in the present paper.

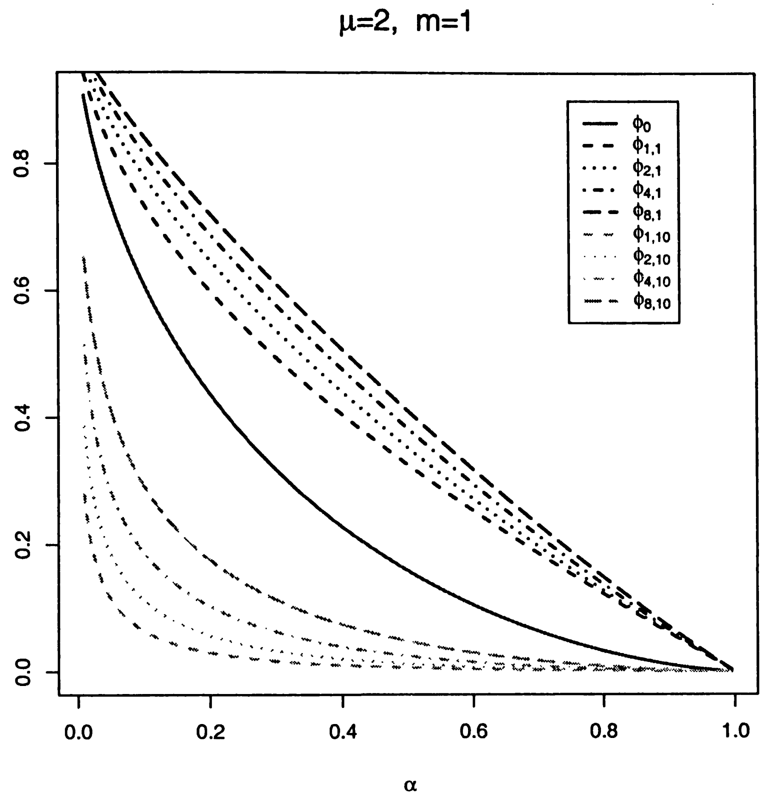

Figure. In the middle is the function $F_2(\alpha)$ for $\alpha \in (0,1)$. Above and below are the functions $F_{k,1/\sqrt{k}}(\alpha)$ and $F_{k,10\sqrt{k}}(\alpha)$ where (bottom to top) $k = 1, 2, 4$ and 8.

Remark 3.1. Let $\phi : (0,\infty) \to \mathbb{R}$ be twice continuously differentiable at 1 with $\phi(1) = 0$, $\phi''(1) > 0$ and $\phi(t) - \phi'(1)(t-1)$ nonincreasing on $(0,1)$ and nondecreasing on $(1,\infty)$. Then

$$D_\phi(P,Q|\mathcal{P}) = \sum_{A\in\mathcal{P}} Q(A)\,\phi\left(\frac{P(A)}{Q(A)}\right) \tag{3.17}$$

is a ϕ-disparity of the distributions P,Q restricted on an interval partition $\mathcal{P}$ of $\mathbb{R}$ in the sense considered e. g. in Menéndez et al. [13]. For the (P,Q)-regular partitions $\mathcal{P}_n$ it is proved in Lemma 2 of Vajda [17] that if (2.13) holds for some $\mu \in [0,\infty)$ then

$$\frac{n}{\sqrt{|\mathcal{P}_n|}}\left(\frac{2D_\phi(\widehat{P}_n,P|\mathcal{P}_n)}{\phi''(1)} - \chi^2(\widehat{P}_n,Q|\mathcal{P}_n)\right) = o_p(1). \tag{3.18}$$

This helps to extend Theorems 3.1 – 3.3 to the Pearson-type tests using the Pearson-type ϕ-disparity statistics $T_n^\phi(\mathcal{P})$, $T_n^\phi(\mathcal{P}_n)$ defined by the formula

$$T_n^\phi(\mathcal{P}) = \frac{n}{\phi''(1)} D_\phi(\widehat{P}_n, P|\,\mathcal{P}). \tag{3.19}$$

Among these is the likelihood ratio statistic studied by Kallenberg et al. [8] and also the other statistics studied by Drost et al. [2], Menéndez et al. [13] and Mayoral et al. [11].

4. STATISTICAL APPLICATIONS

If we want to extrapolate the preferences between the Pearson statistics $T_n(\mathcal{P})$ and $T_n(\mathcal{P}_n)$ under local alternatives P_n to the models with true alternatives P, and the preferences depend on the weights ε_n with which P are represented in P_n, then the priority is naturally given to the preferences valid under large weights ε_n. In our case this means the priority of the preferences observed when $\rho_n = n\varepsilon_n^2 \to \rho = \infty$. Looking on the Table of Section 3 we see that in this case $T_n(\mathcal{P})$ is either strictly better than or at least as good as $T_n(\mathcal{P}_n)$, depending on whether $\mu_n = n\varepsilon_n^2\chi^2(P,Q|\,\mathcal{P}_n)/\sqrt{|\mathcal{P}_n|} \to \mu = \infty$ or μ_n is bounded ($\mu \in [0,\infty)$). Thus the statistic preferred in this case will be the Pearson statistic with a fixed partition $\mathcal{P}$.

This conclusion seems to have a practical meaning when a series of tests is planned for large sample sizes n, $n+1, \dots$ and the problem is to design partitions for the Pearson test: the preferred solution is a fixed (sample-size-independent) partition $\mathcal{P}$. The problem is which partition. If $\rho \in (0,\infty)$ then we see from Corollary 3.1 that the maximal asymptotic power of $T_n(\mathcal{P})$ is achieved by the partition

$$\mathcal{P}^* = \arg\max_{\mathcal{P}} \chi^2(P,Q|\,\mathcal{P}). \tag{4.1}$$

This solution of the problem is intuitively clear – the arg max partition maximizes the Pearson divergence between the hypothetic and true distribution. Therefore it can be heuristically extended also to the models with $\rho = \infty$.

If the partition size $k = |\mathcal{P}|$ is given in advance, then one can write equations for the cutpoints

$$\inf \mathcal{X} < x_1 < \cdots < x_{k-1} < \sup \mathcal{X} \tag{4.2}$$

yielding the arg max partition similar to the equations considered in Mayoral et al. [11] for the partition which maximizes the Fisher information. But if k is not fixed by some extra-theoretical arguments then we face the problem of cardinality of the partition $\mathcal{P}$. By Theorem 6 in Vajda [15] (cf. also Corollary 1.31 in Liese and Vajda [9]),

$$\sup_{\mathcal{P}} \chi^2(P,Q|\,\mathcal{P}) = \chi^2(P,Q)$$

where $|\mathcal{P}|$ tends to infinity with $\chi^2(P,Q|\,\mathcal{P})$ approaching $\chi^2(P,Q)$ except some trivial $\{P,Q\}$. Thus the tendency is to choose $\mathcal{P}$ with large $k = |\mathcal{P}|$. But if both k and n are large and k is so large that the fraction k/n is not negligible then $T_n(\mathcal{P})$ is governed by a limit law different from that of Proposition 3.1 (see Menéndez et al. [12]) so that the conclusions based on this proposition may be misleading.

If $\chi^2(P,Q)$ is finite then a reasonable rule seems to be the arg max partition $\mathcal{P}^*$ with $|\mathcal{P}^*| = k$ large enough to allow a sufficiently close approximation of $\chi^2(P,Q)$ by $\chi^2(P,Q|\,\mathcal{P}^*)$, and small enough to be relatively negligible with respect to the used large sample sizes n.

Namely, for large k and partitions $\mathcal{P}_{(k)}$ with $|\mathcal{P}_{(k)}| = k$ the statistics $T_n(\mathcal{P}_{(k)})$ are characterized by the normal law of Proposition 3.2 rather than by the noncentral χ^2-law of Proposition 3.1. In such cases the asymptotic efficiency is given by $\rho_n\chi^2(P,Q|\,\mathcal{P}_{(k)})/\sqrt{k}$ rather than by $\rho_n\chi^2(P,Q|\,\mathcal{P}_{(k)})$. If

$$k^* = \arg\max_k \frac{\chi^2(P,Q|\,\mathcal{P}_{(k)})}{\sqrt{k}} \tag{4.3}$$

exists and is relatively negligible with respect to the used sample sizes n then k^* can be considered as an optimal partition size of a class of partitions $\{\mathcal{P}_{(k)} : k > 1\}$. This rule can be applied to the class $\{\mathcal{P}_{(k)} : k > 1\}$ of Q–uniform partitions or to the class $\{\mathcal{P}^*_{(k)} : k > 1\}$ of arg max partitions.

Example 4.1. By Theorem 2 in Berlinet and Vajda [1], if the densities $p = \mathrm{d}P/\mathrm{d}\mu$ and $q = \mathrm{d}Q/\mathrm{d}\mu$ satisfy some regularity conditions and $\mathcal{P}$ is the Q-uniform partition with $|\mathcal{P}| = k$ then

$$\chi^2(P,Q|\,\mathcal{P}) = \chi^2(P,Q) - c(P,Q)\frac{1}{k^2} + o\left(\frac{1}{k^2}\right) \quad \text{as } k \to \infty, \tag{4.4}$$

where $c(P,Q) > 0$ is integral of an explicitly given function depending on p and q. From the equation

$$\frac{\mathrm{d}}{\mathrm{d}k}\left(\frac{\chi^2(P,Q) - c(P,Q)k^{-2}}{\sqrt{k}}\right) = 0$$

we get the rule

$$k^*(P,Q) = \sqrt{\frac{5c(P,Q)}{\chi^2(P,Q)}} \tag{4.5}$$

for the choice of the integer representing the optimal partition size. This representation can be accepted if the premises of the method are satisfied.

If $\chi^2(P,Q) = \infty$, and in particular if $\rho \in (0,\infty)$ and

$$\frac{\chi^2(P,Q|\,\mathcal{P}_n)}{\sqrt{|\mathcal{P}_n|}} \to \infty \tag{4.6}$$

for a regular sequence of partitions $\mathcal{P}_n$, then not only the above described method for specification of fixed partition fails, but also the whole concept of optimality of the statistic with fixed partitions collapses. Indeed, then

$$\mu_n = \frac{\rho_n\chi^2(P,Q|\,\mathcal{P}_n)}{\sqrt{|\mathcal{P}_n|}} \to \mu = \infty$$

and, as we see from the row $\rho \in (0, \infty)$, $\mu = \infty$ of the Table, the statistic $T_n(\mathcal{P}_n)$ becomes better than all statistics $T_n(\mathcal{P})$ with fixed partitions $\mathcal{P}$. It has been long ago observed in simulations, and also justified theoretically, that $T_n(\mathcal{P}_n)$ for $\mathcal{P}_n$ with cardinalities $|\mathcal{P}_n|$ slowly increasing with sample sizes are in some models $\{P, Q\}$ more powerful for large n than $T_n(\mathcal{P})$ with fixed partitions (cf. Kallenberg et al. [8] and references therein). The convergence (4.6) was fulfilled in all cases where this took place.

In the practical applications where statistics $T_n(\mathcal{P})$ or $T_n(\mathcal{P}_n)$ are prepared for sample sizes n from some interval, or in the simulations where weights ε_n are selected for n from some interval, one can always embed the selected segments $\mathcal{P}_n$ and ε_n into sequences satisfying $\rho = \infty$, $\mu \in [0, \infty)$ or $\rho \in (0, \infty)$, $\mu = \infty$. Therefore the row $\rho \in (0, \infty)$, $\mu = \infty$ of the Table must be taken seriously in the sense that the statistics $T_n(\mathcal{P}_n)$ are to be preferred to all $T_n(\mathcal{P})$ in the models $\{P, Q\}$ allowing to meet (4.6) by a (P, Q)-regular sequences $\mathcal{P}_n$.

Here no formal rule can be given for the sizes $k_n = |\mathcal{P}_n|$ of the partitions except that for large n the fractions $|\mathcal{P}_n|/n$ should be small (cf. the reference to Menéndez et al. [12] above) and decreasing to zero for $n \to \infty$ as required by (2.4). As soon as the partition sizes $|\mathcal{P}_n|$ are specified, the partitions $\mathcal{P}_n$ themselves can be selected by the arg max rule (4.1).

APPENDIX

Proof of Proposition 3.1. We shall deduce the limit law (3.2) from the asymptotic relation

$$T_n(\mathcal{P}) \geq n\varepsilon_n^2 \left(\chi^2(P, Q|\,\mathcal{P}) + o_p(1)\right) \tag{A.1}$$

when $n\varepsilon_n^2 \to \rho = \infty$ and from the definition

$$T_n(\mathcal{P}) = n\chi^2(\widehat{P}_n, Q|\,\mathcal{P}) \qquad \text{(cf. (2.14))} \tag{A.2}$$

when $n\varepsilon_n^2 \to \rho \in [0, \infty)$. To get (A.1), we use P_n of (2.8), namely we add and subtract $P_n(A) = Q(A) + \varepsilon_n(P(A) - Q(A))$ in the formula

$$\frac{T_n(\mathcal{P})}{n} = \sum_{A \in \mathcal{P}} \frac{\left(\widehat{P}_n(A) - Q(A)\right)^2}{Q(A)}.$$

In this manner we obtain $T_n(\mathcal{P})/n = U_n + V_n + W_n$ for $W_n = \varepsilon_n^2 \chi^2(P, Q|\,\mathcal{P})$ and

$$\begin{aligned} U_n &= \sum_{A \in \mathcal{P}} \frac{\left(\widehat{P}_n(A) - P_n(A)\right)^2}{Q(A)}, \\ V_n &= 2\varepsilon_n \sum_{A \in \mathcal{P}} \frac{\left(\widehat{P}_n(A) - P_n(A)\right)(P(A) - Q(A))}{Q(A)}. \end{aligned}$$

Obviously,

$$\mathsf{E}\, U_n \leq \sum_{A\in\mathcal{P}} \frac{P_n(A)}{nQ(A)} = \frac{|\mathcal{P}|}{n} = o(1) \qquad \text{(cf. (2.2))}$$

and, by the Schwarz and Jensen inequalities,

$$\begin{aligned}\mathsf{E}\,|V_n| &\leq 2\varepsilon_n \left[\chi^2(P,Q|\mathcal{P})\, \mathsf{E}\left(\sum_{A\in\mathcal{P}} \frac{\left(\widehat{P}_n(A)-P_n(A)\right)^2}{Q(A)}\right)\right]^{1/2} \\ &= 2\varepsilon_n \left[\chi^2(P,Q|\mathcal{P})\, \mathsf{E}\, U_n\right]^{1/2} = o(1).\end{aligned}$$

Therefore (A.1) holds and (3.2) follows from (A.1) when $\rho = \infty$. By Corollary 3.1 in Menéndez et al. [13], (3.2) follows from (A.2) when $\rho \in [0,\infty)$. □

Proof of Proposition 3.2. Put for simplicity

$$k_n = |\mathcal{P}_n| \quad \text{and} \quad \chi_n^2 = \chi^2(P,Q|\mathcal{P}_n)$$

and define for μ_n of (2.11) and $A \in \mathcal{P}_n$

$$\begin{aligned} m_n &= k_n + \sqrt{k_n}\,\mu_n, \\ \theta_n(A) &= k_n\varepsilon_n\,(P(A)-Q(A)) - \frac{\sqrt{k_n}}{n}\mu_n, \\ \delta_n^2(A) &= n\theta_n^2(A)P_n(A) \qquad \text{(cf. (2.8))} \\ s_n^2(A) &= 2k_n\frac{P_n^2(A)}{Q(A)} + 4\delta_n^2(A), \\ s_n^2 &= \sum_{A\in\mathcal{P}_n} s_n^2(A) = 2k_n(1+\varepsilon_n^2\chi_n^2) + 4\delta_n^2 \end{aligned}$$

where

$$\delta_n^2 \triangleq \sum_{A\in\mathcal{P}_n} \delta_n^2(A) = \rho_n\chi_n^2 + k_n\rho_n\varepsilon_n \sum_{A\in\mathcal{P}_n} \frac{(P(A)-Q(A))^3}{Q(A)} - \frac{\left(\rho_n\chi_n^2\right)^2}{n} \tag{A.3}$$

(note that the last equality is nontrivial). Applying the inequality

$$\max_{A\in\mathcal{P}_n} |P(A)-Q(A)| \leq \left[\frac{1}{k_n}\sum_{A\in\mathcal{P}_n} \frac{(P(A)-Q(A))^2}{Q(A)}\right]^{1/2}$$

we obtain from (A.3)

$$\delta_n^2 \leq \rho_n\chi_n^2 + k_n\rho_n\varepsilon_n\chi_n^2\sqrt{\chi_n^2/k_n} = \rho_n\chi_n^2 + \sqrt{k_n/n}\left(\rho_n\chi_n^2\right)^{3/2} \tag{A.4}$$

and substituting from (2.11) we get

$$\delta_n^2 \leq \sqrt{k_n}\mu_n + k_n^{3/4}\sqrt{k_n/n}\,\mu_n^{3/2}. \tag{A.5}$$

Now we are ready to apply Theorem 5.1 of Morris [14]. By (2.8),

$$\frac{1-\varepsilon_n}{k_n} \le \min_{A\in\mathcal{P}_n} P_n(A) \le \max_{A\in\mathcal{P}_n} P_n(A) \le \frac{1}{k_n} + \varepsilon_n$$

so that under (2.4)

$$\lim_{n\to\infty} \max_{A\in\mathcal{P}_n} P_n(A) = 0$$

and

$$\lim_{n\to\infty} \min_{A\in\mathcal{P}_n} nP_n(A) = \infty.$$

From here and from (2.3) and (2.4) we see that all assumptions of the mentioned theorem are satisfied. This theorem implies that if the UAN condition

$$\frac{\delta_n^2}{s_n^2} \to 0 \tag{A.6}$$

holds then

$$\mathsf{P}\left(\frac{T_n(\mathcal{P}_n) - m_n}{s_n} < x\right) \to \Phi(x), \; x \in \mathbb{R}.$$

By the present assumptions we have either $\mu_n \to \mu \in (0,\infty)$ or

$$\mu_n = O(\chi^2(P,Q|\,\mathcal{P}_n)/\sqrt{k_n}) = o(\sqrt{k_n}) \qquad \text{(cf. (2.11) and (2.6))},$$

where (2.6) states that $\chi^2(P,Q|\,\mathcal{P}_n) = o(\sqrt{k_n})$. In both cases we see from (A.5) and (2.3), (2.4) that $\delta_n^2 = o(k_n)$. Hence by the definition of s_n^2 and (2.4), (2.6) and (3.3),

$$s_n^2 = 2k_n(1+\varepsilon_n^2 k_n o(1)) + 4\delta_n^2 = 2k_n(1+o(1)) = 2|\mathcal{P}_n|(1+o(1)). \tag{A.7}$$

From here we see that the last limit law is equivalent to

$$\mathsf{P}\left(\frac{T_n(\mathcal{P}_n) - |\mathcal{P}_n|}{\sqrt{2|\mathcal{P}_n|}} < a_n x + b_n\right) \to \Phi(x), \quad x \in \mathbb{R} \tag{A.8}$$

where

$$a_n = \frac{s_n}{\sqrt{2|\mathcal{P}_n|}} = \frac{s_n}{\sqrt{2k_n}} = 1 + o(1) \tag{A.9}$$

and

$$b_n = \frac{(m_n - |\mathcal{P}_n|)}{\sqrt{2|\mathcal{P}_n|}} = \frac{\sqrt{k_n}\mu_n}{\sqrt{2k_n}} = \frac{\mu_n}{\sqrt{2}}. \tag{A.10}$$

Further, by (A.4) and (2.6),

$$\frac{\delta_n^2}{k_n} \le \frac{\rho_n\chi_n^2 + \sqrt{k_n/n}\,(\rho_n\chi_n^2)^{3/2}}{k_n} = \rho_n o(1) + \frac{k_n}{\sqrt{n}}(\rho_n o(1))^{3/2}.$$

By (2.4), the sequence $k_n/\sqrt{n}$ is bounded. Thus we see from (A.7) that if $\rho_n \to \rho < \infty$ then the UAN condition (A.6) holds. Consequently, the desired relation (3.5) follows from (A.8) – (A.10). Let us now suppose that $\rho_n \to \rho = \infty$. Then we get from (A.5)

$$\frac{\delta_n^2}{k_n} \leq \frac{\mu_n}{\sqrt{k_n}} + \frac{k_n}{\sqrt{n}}\left(\frac{\mu_n}{\sqrt{k_n}}\right)$$

and we see from (A.7) that if $\mu_n \to \mu < \infty$ then the UAN condition (A.6) holds too. Thus in this case, similarly as above, the desired relation (3.5) follows from (A.8) – (A.10). □

Proof of Proposition 3.3. Consider the restricted moment generating function

$$M_a(P,Q|\mathcal{P}) = \sum_{A\in\mathcal{P}} P(A)^a Q(A)^{1-a}, \; a \in \mathbb{R}.$$

It is seen from the convexity of $\psi(t) = t^{a_2/a_1}$ for $0 < a_1 < a_2$ that the function $(M_a(P,Q|\mathcal{P}))^{1/a}$ is nondecreasing in the variable $a > 0$. In particular,

$$\chi^2(P,Q|\mathcal{P}) = M_2(P,Q|\mathcal{P}) - 1 \leq (M_a(P,Q|\mathcal{P}))^{2/a} - 1$$

for all $a \geq 2$. But $M_a(P,Q|\mathcal{P}) \leq M_a(P,Q)$ for all $a \geq 1$ and $M_a(P,Q)$ is finite for all $0 < a < a(P,Q)$. Hence it suffices to prove that for every $1 \leq a < 2$ and every $\mathcal{P} = \mathcal{P}_k$ from a regular sequence of partitions

$$\frac{M_2(P,Q|\mathcal{P})}{|\mathcal{P}|^b} \leq (M_a(P,Q))^{2/a} \qquad \text{where } b = (2-a)/a. \tag{A.11}$$

Since $2 = ab + a$, it holds

$$\begin{aligned}\frac{M_2(P,Q|\mathcal{P})}{|\mathcal{P}|^b} &= \frac{1}{|\mathcal{P}|^b}\sum_{A\in\mathcal{P}}\left(\frac{P(A)}{Q(A)}\right)^2 Q(A) \\ &\leq \frac{1}{|\mathcal{P}|^b}\max_{A\in\mathcal{P}}\left(\frac{P(A)}{Q(A)}\right)^{ab}\sum_{A\in\mathcal{P}}\left(\frac{P(A)}{Q(A)}\right)^a Q(A) \\ &\leq \left(\frac{1}{|\mathcal{P}|^{1/a}}\max_{A\in\mathcal{P}}\left(\frac{P(A)}{Q(A)}\right)\right)^{ab} M_a(P,Q).\end{aligned} \tag{A.12}$$

By the Hölder inequality, the relation

$$P(A) = \int_A \frac{p}{q}\,\mathrm{d}Q \leq Q(A)^{1-1/a}\left(\int_A \left(\frac{p}{q}\right)^a \mathrm{d}Q\right)^{1/a}$$

can be extended from $a = 1$ to all $1 \leq a < 2$. Consequently,

$$\begin{aligned}\max_{A\in\mathcal{P}}\frac{P(A)}{Q(A)} &\leq \max_{A\in\mathcal{P}}\left[Q(A)^{-1/a}\right]\left(\int_A \left(\tfrac{p}{q}\right)^a \mathrm{d}Q\right)^{1/a} \\ &\leq |\mathcal{P}|^{1/a}\max_{A\in\mathcal{P}}\left(\int_A \left(\tfrac{p}{q}\right)^a \mathrm{d}Q\right)^{1/a} \quad \text{(cf. (2.2))} \\ &\leq |\mathcal{P}|^{1/a}(M_a(P,Q))^{1/a}.\end{aligned}$$

Inserting this in (A.12) we get (A.11) which completes the proof. □

Proof of Proposition 3.4. The proof of an equivalent proposition was first sketched in Kallenberg et al. [8]. A detailed version of this proof was given in Berlinet and Vajda [1]. A similar detailed version is presented here for the sake of completeness. Since $\chi^2(P,Q|\mathcal{P}_n)$ is positive and tending to $\chi^2(P,Q|\mathcal{P}) > 0$, the assertion is trivial for $a(P,Q) \geq 2$. Hence let $a = a(P,Q)$ be from the semiclosed interval $[1,2)$. Since

$$M_a(P,Q) = \int_0^1 f^a(y)\,\mathrm{d}y$$

for $f(y)$ given on $[0,1]$ in Definition 3.2, $M_2(P,Q) = \infty$ implies that $f(y)$ unbounded on $[0,1]$. Suppose for simplicity that it is unbounded at $y = 0$ and put $h = 1/|\mathcal{P}_n|$. Since

$$\chi^2(P,Q|\mathcal{P}_n) = \frac{1}{h}\sum_{j=1}^{|\mathcal{P}_n|}\left[\left(\int_{(j-1)h}^{jh} f(y)\,\mathrm{d}y\right)^2 - \int_{(j-1)h}^{jh} f(y)\,\mathrm{d}y\right],$$

it suffices to prove that if $0 \leq b < (2-a)/a$ then

$$\frac{h^b}{h}\left(\int_0^h f(y)\,\mathrm{d}y\right)^2 = \left(h^{(b-1)/2}\int_0^h f(y)\,\mathrm{d}y\right)^2 \to \infty \quad \text{for } h \downarrow 0.$$

Since $2/(b+1) > a$, we can choose α between a and $2/(b+1)$ such that

$$\int_0^h f^\alpha(y)\,\mathrm{d}y = \infty \quad \text{while} \quad \int_0^h f(y)\,\mathrm{d}y = P((0,h)) < \infty.$$

By Definition 3.2, $f(1/t) = t^\beta g(t)$ for some $\beta > 0$ and for a function $g(t)$ slowly varying at infinity. Hence the substitution $y = 1/t$ yields the relations

$$\int_{1/h}^\infty t^{\alpha\beta-2} g^a(t)\,\mathrm{d}t = \infty \quad \text{and} \quad \int_{1/h}^\infty t^{\beta-2} g(t)\,\mathrm{d}t < \infty.$$

But $g^\alpha(t)$ is slowly varying at infinity too. Thus the first assertion of the lemma on p. 280 of Feller [3] can be applied to both these relations. The equality implies in this manner $\beta \geq 1/\alpha$ and the inequality implies $\beta \leq 1$. This together with the inequality $\alpha < 2/(b+1)$ implies

$$(b-1)/2 < 1/\alpha - 1 < \beta - 1.$$

Further, by the second assertion of the cited Feller lemma

$$\int_{1/h}^\infty t^{\beta-2} g(t)\,\mathrm{d}t = h^{\beta-1}\Lambda(1/h)$$

where Λ is slowly varying at infinity. But Lemma 2 on p. 277 of Feller implies that if $\varepsilon > 0$ then

$$\Lambda(1/h) = h^{\beta-1}\int_{1/h}^\infty t^{\beta-2} g(t)\,\mathrm{d}t$$

exceeds h^ε for all h sufficiently close to zero. Hence $\Lambda(1/h)\, h^{-\varepsilon} \to \infty$ for $h \downarrow 0$ so that if $\tau < \beta - 1$ then

$$h^\tau \int_{1/h}^{\infty} t^{\beta-2} g(t)\,\mathrm{d}t = h^\tau \int_0^h f(y)\,\mathrm{d}y \to \infty \qquad \text{for } h \downarrow 0.$$

But for $\tau = (b-1)/2$ we obtained above the inequality $\tau < \beta - 1$. Therefore the desired relation

$$h^{(b-1)/2} \int_0^h f(y)\,\mathrm{d}y \to \infty \qquad \text{for } h \downarrow 0$$

is valid. $\square$

ACKNOWLEDGEMENT

The research was supported by the Grant Agency of the Academy of Sciences of the Czech Republic under Grant A1075403, by the Spanish Ministry of Education and Science under Grants BMF2003-00892, BMF2003-04820, and by the Valencian Government under Grant GV04B-670.

(Received March 21, 2005.)

REFERENCES

[1] A. Berlinet and I. Vajda: On asymptotic sufficiency and optimality of quantizations. J. Statist. Plann. Inference *135* (2005), to appear.

[2] F. C. Drost, W. C. M. Kallenberg, D. S. Moore, and J. Oosterhoff: Power approximations to multinomial tests of fit. J. Amer. Statist. Assoc. *89* (1989), 130–141.

[3] W. Feller: An Introduction to Probability and its Applications, Vol. 2. Second edition. Wiley, New York 1966.

[4] T. S. Ferguson: Course in Large Sample Theory. Chapman & Hall, London 1996.

[5] P. E. Greenwood and M. S. Nikulin: A Guide to Chi-Squared Testing. Wiley, New York 1996.

[6] P. Halmos: Measure Theory. Academic Press, New York 1964.

[7] T. Inglot and A. Janic–Wróblewska: Data-driven chi-square test for uniformity with unequal cells. J. Statist. Comput. Simul. *73* (2003), 545–561.

[8] W. C. M. Kallenberg, J. Oosterhoff, and B. F. Schriever: The number of classes in chi-squared goodness-of-fit tests. J. Amer. Statist. Assoc. *80* (1985), 959–968.

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

[10] H. B. Mann and A. Wald: On the choice of the number of intervals in the application of the chi-squared test. Ann. Math. Statist. *13* (1942), 306–317.

[11] A. M. Mayoral, D. Morales, J. Morales, and I. Vajda: On efficiency of estimation and testing with data quantized to fixed numbers of cells. Metrika *57* (2003), 1–27.

[12] M. L. Menéndez, D. Morales, L. Pardo, and I. Vajda: Asymptotic distributions of f-divergences of hypotetical and observed frequencies in sparse testing schemes. Statist. Neerlandica *8* (1998), 313–328.

[13] M. L. Menéndez, D. Morales, L. Pardo, and I. Vajda: Approximations to powers of ϕ-disparity goodness of fit tests. Comm. Statist. – Theory Methods *8* (2001), 313–328.

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

[15] I. Vajda: On the f-divergence and singularity of probability measures. Period. Math. Hungar. *2* (1972), 223–234.
[16] I. Vajda: On convergence of information contained in quantized observations. IEEE Trans. Inform. Theory *48* (2002), 2163–2172.
[17] I. Vajda: Asymptotic laws for stochastic disparity statistics. Tatra Mount. Math. Publ. *26* (2003), 269–280.

Domingo Morales, Operations Research Center, Miguel Hernández University of Elche, Avenida del Ferrocaril s/n, 03202 Elche. Spain.
e-mail: d.morales@umh.es

Leandro Pardo, Department of Statistics and Operations Research, Faculty of Mathematics, Complutense University of Madrid, Plaza de Ciencias 3, 28040 Madrid. Spain.
e-mail: Leandro_Pardo@mat.ucm.es

Igor Vajda, Institute of Information Theory and Automation – Academy of Sciences of the Czech Republic, Pod Vodárenskou věží 4, 182 08 Praha 8. Czech Republic.
e-mail: vajda@utia.cas.cz