



HAL
open science

Asymptotic comparisons of divergence-base: goodness-of-fit statistics

Igor Vajda

► **To cite this version:**

Igor Vajda. Asymptotic comparisons of divergence-base: goodness-of-fit statistics. *Annales de l'ISUP*, 2007, LI (1-2), pp.49-66. <hal-03635353>

HAL Id: hal-03635353

<https://hal.science/hal-03635353v1>

Submitted on 8 Apr 2022

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



HAL Authorization

Pub. Inst. Stat. Univ. Paris
LI, fasc. 1-2, 2007, 49 à 66

ASYMPTOTIC COMPARISONS OF DIVERGENCE-BASED GOODNESS-OF-FIT STATISTICS

Igor Vajda

Institute of Information Theory and Automation
18208 Prague, Czech Republic
<vajda@utia.cas.cz>

Résumé. On considère un problème de test d'ajustement basé sur des divergences entre une distribution donnée et la distribution empirique. On étudie dans cet article les lois asymptotiques de divergence test statistiques basées sur des espacements et on les compare avec les lois asymptotiques de divergence test statistiques basées sur des histogrammes. On porte une attention particulière à l'efficacité asymptotique de ces test statistiques.

Abstract. The goodness-of-fit testing based on divergences between hypothetical and empirical distributions is considered. Asymptotic laws for divergence test statistics based on spacings are presented and compared with the asymptotic laws for the divergence test statistics based on histograms. Particular attention is paid to comparisons of asymptotic efficiencies of these two types of statistics.

AMS 2000 Classification: 62G10, 62G20, 62G30.

Keywords: m -spacings, histogram-based divergence statistics, spacings-based divergence statistics, Pitman efficiency.

1. Introduction. This paper is a revised and considerably extended version of an invited lecture at the 38-th Meeting of the French Statistical Society, cf. Vajda (2006) dealing with the classical statistical problem of testing simple hypotheses in the model with real valued i.i.d. observations. It continues the research initiated by Morales et al. (2003) and Vajda and van der Meulen (2006 a,b) about the possibility to interpret the spacings-based goodness-of-fit statistics introduced by Pyke (1972), Cressie (1976), Hall (1986), Read and Cressie (1988), van Es (1992), Ekström (1999), Misra and van der Meulen (2001) and others cited there as appropriately scaled ϕ -divergences between finitely quantized empirical and hypothetical distributions. Similar interpretation is quite obvious in the case of classical goodness-of-fit statistics of the Pearson, Neyman and log-likelihood type and it was used to extend or sharpen a number of classical statistical results (cf. e.g. Menéndez et al. (1998 a,b), Györfi and Vajda (2002), Vajda (2003), Morales et al. (2005) and Harremoës and Vajda (2007)). The results of Morales et al. (2003) and

Vajda and van der Meulen (2006 a,b) indicate that extension of this interpretation to the spacings-based statistics is equally fruitful.

One can see from the papers of Menéndez et al. (1998 a), Györfi and Vajda (2002), Morales et al. (2003) and Vajda and van der Meulen (2006 a,b) that practically all goodness-of-fit test statistics considered in the literature are appropriately scaled ϕ -divergences (or more general ϕ -disparities) between the classical empirical distributions resulting from the observations and the hypothetical distributions proposed by statisticians, both of them quantized by finite interval partitions of the observation space possibly depending on the sample size. Our paper starts in Section 2 with classification of the goodness-of-fit statistics into type *I* and type *II* based on the ideas of these previous papers. The type of statistics depends on the used partition of the observation space. The type *I* statistics are ϕ -divergences or ϕ -disparities obtained by partitioning of the observation space which depends on hypothetical distribution and leads to the histogram representation of empirical evidence. These statistics are thus called *histogram-based* and classical examples are the Pearson, Neyman and log-likelihood statistics. The type *II* statistics are ϕ -divergences or ϕ -disparities obtained by partitioning of the observation space which depends on empirical distribution and leading to the representation of empirical evidence by spacings between observations. These statistics are thus called *spacings-based* and classical examples are the Greenwood, Moran and Hall statistics.

Section 3 summarizes and extends the results of Morales et al. (2003) and Vajda and van der Meulen (2006 a) about asymptotic distributions of both the types of test statistics under local alternatives. These results are used in Section 4 to compare the Pitman asymptotic efficiencies within as well as between the classes of the type *I* and type *II* statistics. New theorems of Sections 3 and 4 are proved in Section 5.

2. Two types of divergence-based statistics. Consider the classical statistical model with given i.i.d. observations X_1, \dots, X_n distributed by an unknown c.d.f. $F(x)$ defined on the real line \mathbb{R} and the problem of testing a given of hypothesis $\mathcal{H}_0 : F = F_0$ concerning the distribution F . The standard solution of this problem is based on goodness-of-fit statistics. As well known, it suffices to deal with the test statistics T_n based on sufficient statistics. Examples of the sufficient statistics are the empirical distribution function

$$F_n(x) = \frac{1}{n} \sum_{i=1}^n \mathbf{I}(x \geq X_i), \quad x \in \mathbb{R} \quad (1)$$

where $\mathbf{I}(\cdot)$ denotes the indicator function, or the order statistics

$$(Y_1, \dots, Y_n) = (X_{n:1}, \dots, X_{n:n}). \quad (2)$$

Practically all test statistics considered in the literature are of the form

$$T_n^I = D(F_n, F_0) \quad \text{or} \quad T_n^{II} = D(F_0, F_n) \quad (3)$$

where $D(F, G)$ is a nonnegative measure of disparity between c.d.f.'s $F(x)$ and $G(x)$ on the real line \mathbb{R} . We shall deal with an important class of disparity measures between discrete distributions obtained from the distributions F_n and F_0 by means of partitions

$$\mathcal{P} = \{(a_{j-1}, a_j] : 1 \leq j \leq k\}, \quad -\infty = a_0 < a_1 < \dots < a_k = \infty \quad (4)$$

of \mathbb{R} . These partitions may depend on the sample size n in the sense that both the partition size $k = k_n$ and the cutpoints a_1, \dots, a_{k-1} themselves depend on n , but this dependence is usually not explicitly denoted in the paper. Restrictions of F_0, F_n on these partitions lead to discrete probability distributions

$$\mathbf{p}_0 = (p_{0j} : 1 \leq j \leq k), \quad \mathbf{p}_n = (p_{nj} : 1 \leq j \leq k) \quad (5)$$

where

$$p_{0j} = F_0(a_j) - F_0(a_{j-1}) \quad \text{and} \quad p_{nj} = F_n(a_j) - F_n(a_{j-1}). \quad (6)$$

These distributions are then inserted as arguments of the measures of disparities between discrete probability distributions $\mathbf{p} = (p_j : 1 \leq j \leq k)$ and $\mathbf{q} = (q_j : 1 \leq j \leq k)$ called ϕ -disparities and defined by the formula

$$D_\phi(\mathbf{p}, \mathbf{q}) = \sum_{j=1}^k q_j \phi\left(\frac{p_j}{q_j}\right), \quad \phi \in \Phi \quad (7)$$

where Φ is the class of functions $\phi : (0, \infty) \mapsto \mathbb{R}$ twice continuously differentiable in the neighborhood of 1 with $\phi''(1) > 0$ and $\phi(1) = 0$, and such that the difference $\phi(t) - \sigma(t)$ from the support line $\sigma(t) = \phi'(1)(t - 1)$ is monotone on the intervals $(0, 1)$ and $(1, \infty)$. Most frequently used $\phi \in \Phi$ are convex on $(0, \infty)$ for which the corresponding ϕ -disparities (7) are traditionally called ϕ -divergences. In this paper we usually follow the divergence terminology but we have in mind the more general above defined class Φ of disparity functions.

In accordance with (3), the divergence measures (7) lead to the *divergence goodness-of-fit statistics*

$$T_{\phi, n}^I = n D_\phi(\mathbf{p}_n, \mathbf{p}_0) \quad \text{and} \quad T_{\phi, n}^{II} = n D_\phi(\mathbf{p}_0, \mathbf{p}_n) \quad \text{for } \phi \in \Phi. \quad (8)$$

We shall consider two types of partitions (4), both of them uniform in some sense, and both of them defined by means of quantile functions

$$F^{-1}(\alpha) = \inf\{x \in \mathbb{R} : F(x) \geq \alpha\}, \quad \alpha \in (0, 1)$$

for appropriate distributions F . The type I partitions are *hypothetically uniform*, defined by the cutpoints

$$a_j = F_0^{-1}(j/k), \quad 0 \leq j \leq k \quad (9)$$

corresponding to the hypothetic distribution F_0 and leading to the uniform hypothetic distributions $\mathbf{p}_0 = \mathbf{u}_k$ where

$$\mathbf{u}_k = (u_j = 1/k : 1 \leq j \leq k), \quad (10)$$

and to the empirical distributions $\mathbf{p}_n = (p_{nj} : 1 \leq j \leq k)$ where the empirical probabilities

$$p_{nj} = F_n(a_j) - F_n(a_{j-1}) \quad \text{for } 1 \leq j \leq k \quad (11)$$

are the relative frequencies of observations in the intervals between the cutpoints (9). According to (7), (8) and (10), these type *I* partitions lead to the *type I statistics*

$$T_{\phi,n}^I = n \sum_{j=1}^{k_n} p_{0j} \phi \left(\frac{p_{nj}}{p_{0j}} \right) = \frac{n}{k_n} \sum_{j=1}^{k_n} \phi(k_n p_{nj}), \quad \phi \in \Phi \quad (12)$$

for the random empirical probabilities p_{nj} given by (11).

On the other hand, the type *II* partitions are *empirically uniform*. They are defined by means of the random empirical cutpoints

$$a_j = F_n^{-1}(j/k) \equiv Y_{mj}, \quad 1 \leq j \leq k. \quad (13)$$

corresponding to the empirical distribution F_n and to a given (possibly constant) sequence of integers $m = m_n \geq 1$ specifying $k = k_n$ by the formula

$$k = k_n = \frac{n}{m}. \quad (14)$$

The type *II* partitions lead to the uniform empirical distributions $\mathbf{p}_n = \mathbf{u}_k$ and to the hypothetical distributions \mathbf{p}_0 with the probabilities

$$p_{0j} = F_0(a_j) - F_0(a_{j-1}) = \begin{cases} F_0(Y_{mj}) - F_0(Y_{m(j-1)}) & \text{for } 1 \leq j \leq k-1 \\ F_0(Y_{n+1}) - F_0(Y_{m(k-1)}) & \text{for } j = k \end{cases} \quad (15)$$

specified by the order statistics (2) and two dummy observations $Y_0 = -\infty$ and $Y_{n+1} = \infty$. By (8) and (10), the type *II* partitions lead to the *type II statistics*

$$T_{\phi,n}^{II} = n \sum_{j=1}^{k_n} p_{nj} \phi \left(\frac{p_{0j}}{p_{nj}} \right) = \frac{n}{k_n} \sum_{j=1}^k \phi(k_n p_{0j}), \quad \phi \in \Phi \quad (16)$$

for the random hypothetical probabilities p_{0j} given by (15).

REMARK 1. Notice that the symbols *I* and *II* (and the related type *I* and type *II* terminology) are used in two different meanings. In (8) they indicate the order of

hypothetical and empirical distributions as arguments of the functional D_ϕ while in (9) and (13) they indicate the hypothetical and empirical partitions. Since however in this paper the type of order always combines with the same type of partition, this ambiguity will not perhaps lead to a confusion.

We see that the formulas (12) and (16) for the type I and type II disparity statistics are formally similar, but the distributions of the vectors $\mathbf{Z}_n^I = k_n \mathbf{p}_n$ of the random arguments $Z_{nj}^I = k_n p_{nj}$ in the formula (12) differ from the distributions of the vectors $\mathbf{Z}_n^{II} = k_n \mathbf{p}_0$ of the random arguments $Z_{nj}^{II} = k_n p_{0j}$ in the formula (16).

The above derived formulas (11) and (15) for the random arguments $\mathbf{p}_0, \mathbf{p}_n$ of the disparity statistics $T_{\phi,n}^I, T_{\phi,n}^{II}$, and the related ϕ -disparity statistics themselves, can be simplified by taking into account the well known fact that in the testing problem under consideration the observation space \mathbb{R} can be reduced without loss of generality to the interval $[0, 1] \subset \mathbb{R}$ and $F_0(x)$ to the hypothetical distribution $F_0(x) = x$ on this interval.

In this simplified situation the theoretical cutpoints in the definition of empirical probabilities (11) are $a_j = F_0^{-1}(j/k) = j/k$, leading to the empirical probabilities (11) of the form

$$p_{nj} = F_n(j/k) - F_n((j-1)/k) \quad \text{for } 1 \leq j \leq k \quad (17)$$

and to the type I statistics

$$T_{\phi,n}^I = \sum_{j=1}^n \phi(np_{nj}), \quad \phi \in \Phi. \quad (18)$$

The empirical probabilities p_{nj} are relative frequencies of observations in the intervals $((j-1)/k, j/k]$ of a uniform partition of the observation space usually represented by histograms. Therefore the type I statistics are called in this paper *histogram-based statistics*.

On the other hand, in this simplified situation the hypothetical probabilities (15) reduce into the standard m -spacings, i.e., they take on the form

$$p_{0j} = \begin{cases} Y_{mj} - Y_{m(j-1)} & \text{for } 1 \leq j \leq k-1 \\ Y_{n+1} - Y_{m(k-1)} & \text{for } j = k \end{cases} \quad (19)$$

where are used the new dummy observations

$$Y_0 = 0 \quad \text{and} \quad Y_{n+1} = 1. \quad (20)$$

The simplest hypothetical probabilities (19) are obtained for $m = 1$ when $k = k_n = n$ and when they reduce to the simple spacings (1-spacings)

$$p_{0j} = \begin{cases} \Delta_j = Y_j - Y_{j-1} & \text{for } 1 \leq j \leq n-1 \\ \Delta_n = Y_{n+1} - Y_{n-1} & \text{for } j = n. \end{cases} \quad (21)$$

In this case the type *II* statistics (16) takes on the simpler form

$$T_{\phi,n}^{II} = \sum_{j=1}^n \phi(n\Delta_j), \quad \phi \in \Phi. \quad (22)$$

In the more complicated case $1 < m < n$ the empirical distributions $\mathbf{p}_n = \mathbf{u}_n$ remain the same as in the previous simple case $m = 1$ but the type *II* statistics are usually transformed to a more efficient form by replacing the basic version $\mathbf{p}_0 = (p_{01} = \Delta_1, \dots, p_{0n} = \Delta_n)$ of the hypothetical distribution in definition (16) by the averages

$$\mathbf{q}_0 = \frac{1}{m} \sum_{k=1}^m T^k \mathbf{p}_0 \quad (23)$$

of all its alternative versions $T^k \mathbf{p}_0$ obtained by cyclic shifts, i.e.,

$$T^k \mathbf{p}_0 = (\Delta_k, \dots, \Delta_n, \Delta_1, \dots, \Delta_{k-1}). \quad (24)$$

Then

$$\mathbf{q}_0 = \left(q_{0j} = \frac{\Delta_{m,j}}{m} = \frac{Y_{j+m} - Y_j}{m} : 1 \leq j \leq n \right)$$

(for the definition of Y_{n+j} when $0 < j < n$ and other details see e.g. Morales et al (2003)). This leads to the more efficient form

$$T_{\phi,n}^{II} = n \sum_{j=1}^n p_{nj} \phi \left(\frac{q_{0j}}{p_{nj}} \right) = \sum_{j=1}^n \phi \left(\frac{n\Delta_{m,j}}{m} \right), \quad \phi \in \Phi \quad (25)$$

of the type *II* statistics. Since Δ_j in (22) and $\Delta_{m,j} = Y_{j+m} - Y_j$ in the last formula are the m -spacings for $1 \leq m < n$, the type *II* statistics defined by (22) and (25) are called in this paper *spacings-based statistics*.

Our spacings-based statistics (22) and (25) formally differ from the classical spacings-based statistics known from the literature. However, in Morales et al (2003) we proved that if $m = m_n$ increases to infinity as $n \rightarrow \infty$ then for some common functions ϕ the type *II* statistics (25) asymptotically coincide with the spacings-based statistics studied by Hall (1989), Ekström (1999) and some others. In Vajda and van der Meulen (2006 a,b) we proved for $m = 1$ similar coincidence of the type *II* statistics (22) with the classical spacings-based statistics studied previously by Pyke (1972), Cressie (1976), van Es (1992), Misra and van der Meulen (2001) and other authors cited there. This helps to explain the pleasing statistical properties of the classical spacings-based statistics in spite of that their original motivation differed from the rigorous divergence motivation given in the present paper (cf. the cited references and also Read and Cressie (1988)).

3. Asymptotic distributions. In this section we present four theorems dealing with asymptotic distributions of the above introduced statistics in two different situations. The first two of these theorems assume k_n of (14) increasing for $n \rightarrow \infty$ *slowly* in the sense

$$k_n \rightarrow \infty \quad \text{but} \quad \frac{k_n^2}{n} \rightarrow 0 \quad (26)$$

and the remaining two assume k_n increasing *fastly* in the sense $k_n = n$.

The asymptotic properties are in all four theorems studied under the local alternatives

$$F = F_{(n)} \sim f_{(n)} = 1 + \left(\frac{\sqrt{k_n}}{n} \right)^{1/2} \ell \quad (27)$$

where $\ell : [0, 1] \mapsto \mathbb{R}$ is continuous (and consequently bounded) with

$$\int_0^1 \ell(x) dx = 0 \quad \text{and} \quad \|\ell\|^2 = \int_0^1 \ell^2(x) dx \quad (28)$$

where $\|\ell\|^2$ is finite because ℓ is bounded. Clearly, $F_{(n)}$ reduces to the hypothesis F_0 when ℓ is identically zero. Similarly as in (26), all the remaining asymptotic relations in the paper are considered for $n \rightarrow \infty$. Proofs of all four theorems are deferred to the last section of the paper.

THEOREM 1. If k_n increases to infinity slowly in the sense of (26) then the statistics $T_{\phi,n}^I$ defined by (12) satisfy for all $\phi \in \Phi$ the relation

$$\frac{T_{\phi,n}^I - k_n/2}{\sqrt{k_n}} \xrightarrow{\mathcal{L}} N \left(\frac{\|\ell\|^2}{2}, \frac{1}{2} \right). \quad (29)$$

We see that if k_n slowly increases to infinity in the sense of Theorem 1 then the asymptotic distribution of $T_{\phi,n}^I$ under local alternatives does not depend on $\phi \in \Phi$. The next theorem demonstrates similar invariance with respect to $\phi \in \Phi$ also for the statistics $T_{\phi,n}^{II}$.

THEOREM 2. If k_n increases to infinity slowly in the sense of (26) then the statistics $T_{\phi,n}^{II}$ defined by (25) satisfy for all $\phi \in \Phi$ which are four times differentiable in the neighborhood of 1 the relation

$$\frac{T_{\phi,n}^{II} - k_n/2}{\sqrt{k_n}} \xrightarrow{\mathcal{L}} N \left(\frac{\|\ell\|^2}{2}, \frac{1}{3} \right). \quad (30)$$

Menéndez et al. (1998 b) proved that if k_n increases to infinity with the rate of n then the asymptotic distribution of $T_{\phi,n}^I$ under local alternatives already depends on $\phi \in \Phi$.

Similar dependence for the statistics $T_{\phi,n}^{II}$ in the case when k_n increases to infinity with the rate of n was established by Vajda and van der Meulen (2006 a,b) (cf. Theorem 4 below).

The next two theorems deal with the case $k_n = n$. They are restricted to the special family of so-called power divergence statistics resulting from the family of infinitely differentiable convex functions $\{\phi_\alpha : \alpha \in \mathbb{R}\} \subset \Phi$ defined by

$$\phi_\alpha(t) = \frac{t^\alpha - 1}{\alpha(\alpha - 1)} \quad \text{if } \alpha \notin \{0, 1\} \quad (31)$$

and

$$\phi_0(t) = -\ln t \quad \text{or} \quad \phi_1(t) = t \ln t \quad \text{otherwise.} \quad (32)$$

In accordance with (7), these functions lead to the power divergences (called sometimes α -divergences) defined by

$$D_\alpha(\mathbf{p}, \mathbf{q}) = D_{\phi_\alpha}(\mathbf{p}, \mathbf{q}) = \frac{1}{\alpha(\alpha - 1)} \left(\sum_{j=1}^{k_n} q_n^{\alpha-1} p_{nj}^\alpha - 1 \right) \quad \text{if } \alpha \notin \{0, 1\} \quad (33)$$

and by

$$D_0(\mathbf{p}, \mathbf{q}) = D_{\phi_0}(\mathbf{p}, \mathbf{q}) = \sum_{j=1}^{k_n} q_n \ln \left(\frac{q_n}{p_n} \right) \quad \text{or} \quad D_1(\mathbf{p}, \mathbf{q}) = D_{\phi_1}(\mathbf{p}, \mathbf{q}) = D_0(\mathbf{q}, \mathbf{p}) \quad (34)$$

otherwise.

Using the power divergences (33) and (34), we get from (12) for any sequence of integers $k = k_n$ and empirical probabilities p_{nj} given by (17) the family of *type I power divergence statistics*

$$T_{\alpha,n}^I = nD_{\phi_\alpha}(\mathbf{p}_n, \mathbf{p}_0) = \frac{n}{\alpha(\alpha - 1)} \left(\sum_{j=1}^{k_n} k_n^{\alpha-1} p_{nj}^\alpha - 1 \right) \quad \text{if } \alpha \notin \{0, 1\} \quad (35)$$

and

$$T_{0,n}^I = nD_0(\mathbf{p}_n, \mathbf{p}_0) = -\frac{n}{k_n} \sum_{j=1}^{k_n} \ln(k_n p_{nj}), \quad T_{1,n}^I = nD_1(\mathbf{p}_n, \mathbf{p}_0) = n \sum_{j=1}^{k_n} p_{nj} \ln(k_n p_{nj}) \quad (36)$$

otherwise. Similarly, if $m = m_n \geq 1$ is a given sequence (with the subscript n suppressed) or a constant, then we get from (25) for the corresponding m -spacings $\Delta_{m,j} = Y_{m+j} - Y_j$ and sequence $k_n = mn$ the family of *type II power divergence statistics*

$$T_{\alpha,n}^{II} = T_{\phi_\alpha,n}^{II} = \frac{1}{\alpha(\alpha - 1)} \left(\sum_{j=1}^n (k_n \Delta_{m,j})^\alpha - 1 \right) \quad \text{if } \alpha \notin \{0, 1\} \quad (37)$$

and

$$T_{0,n}^{II} = T_{\phi_0,n}^{II} = - \sum_{j=1}^n \ln(k_n \Delta_{m,j}) \quad \text{or} \quad T_{1,n}^{II} = T_{\phi_1,n}^{II} = \sum_{j=1}^n (k_n \Delta_{m,j}) \ln(k_n \Delta_{m,j}) \quad (38)$$

otherwise. Notice that $k_n = n$ if and only if m is constant, $m = 1$.

The next theorem uses for each $1 \leq j \leq n$ the point $x_{n,j} \in [(j-1)/n, j/n]$ such that the local alternative function $\ell(x)$ of (27) satisfies the condition

$$\frac{\ell(x_{n,j})}{n} = \int_{(j-1)/n}^{j/n} \ell(x) d(x).$$

By the mean value theorem for integrals, such $x_{n,j}$ exists. Since we assume $k_n = n$, it follows from (17)

$$\begin{aligned} p_{nj} &= F_{(n)}(j/n) - F_{(n)}((j-1)/n) = \int_{(j-1)/n}^{j/n} f_{(n)}(x) d(x) \quad (\text{cf. (27)}) \\ &= \frac{1}{n} + \left(\frac{\sqrt{n}}{n} \right)^{1/2} \int_{(j-1)/n}^{j/n} \ell(x) d(x) = \frac{1}{n} \left(1 + \frac{\ell(x_{n,j})}{\sqrt[4]{n}} \right). \end{aligned} \quad (39)$$

Further, by (27),

$$p_{0j} = F_0(j/n) - F_0((j-1)/n) = 1/n \quad (40)$$

so that

$$\frac{p_{nj}}{p_{0j}} = 1 + \frac{\ell(x_{n,j})}{\sqrt[4]{n}}. \quad (41)$$

THEOREM 3. If $k_n = n$ then the power divergence statistics $T_{\alpha,n}^I$ defined by (35) and (36) satisfy for all $\alpha > 0$ the asymptotic relations

$$\frac{T_{\alpha,n}^I - n\mu_{\alpha,n}}{\sqrt{n\sigma_{\alpha,n}^2}} \xrightarrow{\mathcal{L}} N(0, 1) \quad (42)$$

where

$$\mu_{\alpha,n} = \frac{1}{n} \sum_{j=1}^n \mathbf{E}\phi_{\alpha}(Z_{n,j}) \quad (\text{cf. (31), (32)}), \quad (43)$$

$$\sigma_{\alpha,n}^2 = \frac{1}{n} \sum_{j=1}^n \text{var}\phi_{\alpha}(Z_{n,j}) - \left[\frac{1}{n} \sum_{j=1}^n \text{covar}(Z_{n,j}, \phi_{\alpha}(Z_{n,j})) \right]^2 \quad (44)$$

and $Z_{n,j}$ are Poisson-distributed random variables with parameters $\lambda_{nj} = p_{nj}/p_{0j}$,

$$Z_{n,j} \sim \text{Poisson}(\lambda_{nj}) = \text{Poisson} \left(1 + \frac{\ell(x_{n,j})}{\sqrt[4]{n}} \right) \quad (\text{cf. (41)}) \quad (45)$$

for $x_{n,j}$ defined by (39).

REMARK 2. It follows from the definition of the function ϕ_α in (31) that if $\alpha \notin \{0, 1\}$ then the definitions (43) and (44) can be replaced by

$$\mu_{\alpha,n} = \frac{1}{\alpha(\alpha-1)} \left(\frac{1}{n} \sum_{j=1}^n \mathbb{E}(Z_{n,j}^\alpha) - 1 \right) \quad (46)$$

and

$$\sigma_{\alpha,n}^2 = \frac{1}{\alpha^2(\alpha-1)^2} \left\{ \frac{1}{n} \sum_{j=1}^n \text{var}(Z_{n,j}^\alpha) - \left[\frac{1}{n} \sum_{j=1}^n \text{covar}(Z_{n,j}, Z_{n,j}^\alpha) \right]^2 \right\}. \quad (47)$$

The last two formulas can also be found on p. 58 in Read and Cressie (1988) where a less general result than Theorem 3 is given, namely the asymptotic relation (42) under the hypothesis, i.e., under (45) replaced by the assumption $Z_{n,j} \sim \text{Poisson}(1)$.

For the sake of completeness we state here also the symmetric theorem for the type II statistics $T_{\alpha,n}^{II}$ proved in Vajda and van der Meulen (2006 a). This theorem deals with $\alpha > -1/2$ and uses the constants

$$\mu_0 = \gamma, \quad \mu_1 = 1 - \gamma \quad \text{and} \quad \mu_\alpha = \frac{\Gamma(\alpha+1) - 1}{\alpha(\alpha-1)} \quad \text{for } \alpha \notin \{0, 1\}, \quad (48)$$

$$\sigma_0^2 = \frac{\pi^2}{6} - 1, \quad \sigma_1^2 = \frac{\pi^3}{3} - 3 \quad \text{and} \quad \sigma_\alpha^2 = \frac{\Gamma(2\alpha+1) - (\alpha^2+1)\Gamma^2(\alpha+1)}{\alpha^2(\alpha-1)^2} \quad \text{for } \alpha \notin \{0, 1\} \quad (49)$$

and the functionals

$$m_\alpha(\ell) = \frac{\|\ell\|^2}{2} \Gamma(\alpha+1) \quad (\text{cf. (27)}). \quad (50)$$

$$(51)$$

Here $\Gamma(\cdot)$ is the gamma function and the constants $\mu_\alpha, \sigma_\alpha^2$ as well as the functional $m_\alpha(\ell)$ s are continuous in the parameter $\alpha \in \mathbb{R}$. (Note that a more general form of this theorem valid for a class of statistics $T_{\phi,n}^{II}$ can be found in Vajda and van der Meulen (2006 b).)

THEOREM 4. If $k_n = n$ then the power divergence statistics $T_{\alpha,n}^{II}$ defined by (37) and (38) satisfy for all $\alpha > -1/2$ the asymptotic relations

$$\frac{T_{\alpha,n}^{II} - n\mu_\alpha}{\sqrt{n}} \xrightarrow{\mathcal{L}} N(m_\alpha(\ell), \sigma_\alpha^2). \quad (52)$$

where $\mu_\alpha, \sigma_\alpha^2$ and $m_\alpha(\ell)$ are defined by (48), (49) and (50) respectively.

4. Asymptotic efficiencies. Let us now turn to the problem of comparing asymptotic efficiencies of two sequences of test statistics $T = \{T_n\}$ and $\tilde{T} = \{\tilde{T}_n\}$ satisfying under $F = F_{(n)}$ (cf. (27)) and $n \rightarrow \infty$ and $k_n \rightarrow \infty$ two asymptotic normality laws

$$\frac{T_n - k_n \mu}{\sqrt{k_n}} \xrightarrow{\mathcal{L}} N(\mathbf{m}, \sigma^2) \quad \text{and} \quad \frac{\tilde{T}_n - k_n \tilde{\mu}}{\sqrt{k_n}} \xrightarrow{\mathcal{L}} N(\tilde{\mathbf{m}}, \tilde{\sigma}^2)$$

where $\mathbf{m} = \tilde{\mathbf{m}} = 0$ under $F = F_0$. If $|\mathbf{m}|/\sigma > |\tilde{\mathbf{m}}|/\tilde{\sigma}$ then the asymptotic power of the test based on the statistic T_n will exceed that of the test based on the statistic \tilde{T}_n . This leads us to the definition of *asymptotic relative efficiency* of the test statistics T_n and \tilde{T}_n by the formula

$$\text{ARE}(T, \tilde{T}) = (\mathbf{m}/\sigma)^4 / (\tilde{\mathbf{m}}/\tilde{\sigma})^4. \quad (53)$$

Notice that if $k_n = n$ then, in accordance with page 273 in Fraser (1957), our $\text{ARE}(T, \tilde{T})$ reduces to the well-known *Pitman's asymptotic relative efficiency* of the statistics T_n and \tilde{T}_n . In other words, $\text{ARE}(T, \tilde{T})$ is then the limit of the inverse ratios of sample sizes leading under the local alternatives (27) to the same asymptotic powers of the tests using the statistics T_n and \tilde{T}_n .

REMARK 3. If k_n increases to infinity slowly in the sense considered in Theorems 1 and 2 then we get from these theorems and definition (53) that $\text{ARE}(T_\phi, T_{\tilde{\phi}}) = 1$ for any pair $T_\phi = \{T_{\phi,n}\}$ and $T_{\tilde{\phi}} = \{T_{\tilde{\phi},n}\}$ of sequences of test statistics considered in Theorem 1 or Theorem 2 respectively. Therefore all histogram-based statistics from the family $\{T_{\phi,n}^I : \phi \in \Phi\}$ are asymptotically equally efficient and, similarly, all spacings-based statistics from the family $\{T_{\phi,n}^{II} : \phi \in \Phi\}$ are asymptotically equally efficient. Contrary to this, definition (53) together with Theorem 4 implies that the Pitman's asymptotic relative efficiency of two different sequences of test statistics $T_\alpha^{II} = \{T_{\alpha,n}^{II}\}$, $T_{\tilde{\alpha}}^{II} = \{T_{\tilde{\alpha},n}^{II}\}$ considered in that theorem differs in general from 1 and is given by the formula

$$\text{ARE}(T_\alpha^{II}, T_{\tilde{\alpha}}^{II}) = \frac{(m_\alpha(\ell)/\sigma_\alpha)^4}{(m_{\tilde{\alpha}}(\ell)/\sigma_{\tilde{\alpha}})^4} \quad \text{if } \alpha, \tilde{\alpha} > -1/2 \quad (54)$$

for the right-hand side terms given in (49) and (50). Vajda and van der Meulen (2006 a) studied the ratio $m_\alpha(\ell)/\sigma_\alpha$ as function of the variable $\alpha > -1/2$ and found that it is maximized at $\alpha = 2$. Therefore $T_{2,n}^{II}$ is asymptotically most Pitman efficient in the class of type II statistics $\{T_{\alpha,n}^{II} : \alpha > -1/2\}$ considered in Theorem 4. Unfortunately, Theorem 3 above does not allow to evaluate in a similar explicit manner $\text{ARE}(T_\alpha^I, T_{\tilde{\alpha}}^I)$ for the sequences $T_\alpha^I = \{T_{\alpha,n}^I\}$, $T_{\tilde{\alpha}}^I = \{T_{\tilde{\alpha},n}^I\}$ from the class of type I statistics $\{T_{\alpha,n}^I : \alpha > 0\}$ considered in that theorem. But, with a reference to the arguments on p. 175 in Read and Cressie (1988), we can take for granted that $T_{2,n}^I$ is asymptotically most Pitman efficient among all the statistics $T_{\alpha,n}^I$ from this class.

Next follow our last two theorems comparing the asymptotic efficiencies of the above considered histogram and spacings based statistics of type I and type II. Their proofs are deferred to the next section.

THEOREM 5. If the partition size k_n increases to infinity slowly in the sense of (26) then

$$ARE(T_\phi^I, T_\phi^{II}) = 4/9 \quad \phi \in \Phi. \quad (55)$$

Therefore the histogram-based goodness-of-fit test statistics $T_{\phi,n}^I$ are for all $\phi \in \Phi$ asymptotically less efficient than the spacings-based goodness-of-fit test statistics $T_{\phi,n}^{II}$ in the sense that for the same sample size they need more than twice larger number of partition cells to achieve the same asymptotic power.

For the reasons stated in Remark 3, we cannot give for $\alpha > 0$ direct asymptotic efficiency comparisons of the histogram-based goodness-of-fit test statistics $T_{\alpha,n}^I$ from the class $\{T_{\alpha,n}^I : \alpha > 0\}$ with their spacings-based counterparts $T_{\alpha,n}^{II}$ from the class $\{T_{\alpha,n}^{II} : \alpha > 0\}$ similar to the comparison of $T_{\phi,n}^I$ and $T_{\phi,n}^{II}$ in Theorem 5. Instead of this, with a reference to the second part of Remark 3, we can compare at least the asymptotically most efficient representatives $T_{2,n}^I$ and $T_{2,n}^{II}$ from these classes. The result is in the next theorem.

THEOREM 6. If the partition size k_n increases to infinity fast in the sense $k_n = n$ then $ARE(T_{2,n}^I, T_{2,n}^{II}) = 1/4$. In other words, the spacings-based goodness-of-fit test statistic $T_{2,n}^{II}$ is 4 times more efficient in the Pitman sense than the classical Pearson statistic $T_{2,n}^I$.

REMARK 4. The last theorem means that the Pearson's $T_{2,n}^I$ needs 4 times larger sample size than its spacings-based counterpart $T_{2,n}^{II}$ to achieve the same asymptotic power. So large asymptotic efficiency gap between these two statistics comes out as something quite surprising. However, this paradox can be explained by the fact that of the large number $k_n = n$ of partition cells defined by the theoretical cutpoints $a_j = j/n$, many are with a positive probability empty which decreases the asymptotic power of the Pearson's $T_{2,n}^I$. On the other hand, of the same number of partition cells defined by the empirical cutpoints $a_j = F_n^{-1}(j/n)$ none is empty which keeps the asymptotic power of the spacings-based statistic $T_{2,n}^{II}$ on a considerably higher level. For the dependence of the asymptotic power of the classical Pearson statistic $T_{2,n}^I$ on the rate with which increases the number of cells k_n we can refer to the recent paper by Morales et al. (2005) and earlier ones cited there.

5. Proofs of theorems. Theorem 2 in fact just restates the asymptotic relation (2.15) in Morales et al. (2003) derived as a consequence of Theorem 2.2 *ibid.* Further, as said above, Theorem 4 is due to Vajda and van der Meulen (2006 a). Therefore it suffices to prove Theorems 1, 3, 5 and 6. As before, all asymptotic relations are considered for $n \rightarrow \infty$.

Proof of Theorem 1: By the result on p. 275 in Vajda (2003), if k_n increases to infinity slowly in the sense of (26) and the statistic $T_{2,n}^I$ defined by (35) for $\alpha = 2$ satisfies the condition

$$\frac{T_{2,n}^I}{\sqrt{k_n}} \rightarrow \mu \quad (\text{cf. (11) in Vajda (2003)}) \quad (56)$$

then for all four times differentiable $\phi \in \Phi$

$$\frac{2T_{\phi,n}^I - k_n}{\sqrt{k_n}} \xrightarrow{\mathcal{L}} N(\mu, 2) \quad (\text{cf. (12) in Vajda (2003)}). \quad (57)$$

Further, by (35),

$$T_{2,n}^I = \frac{nk}{2} \sum_{j=1}^k (p_{nj} - 1/k)^2$$

where in accordance with (17), (27)

$$p_{nj} = F_{(n)}(j/k) - F_{(n)}((j-1)/k) = \frac{1}{k} + \left(\frac{\sqrt{k}}{n}\right)^{1/2} \int_{(j-1)/k}^{j/k} \ell(x) dx.$$

Similarly as in (39), we get from here for every $1 \leq j \leq k = k_n$ the existence of a point $x_{nj} \in ((j-1)/k, j/k)$ with the property

$$p_{nj} = \frac{1}{k} + \left(\frac{\sqrt{k}}{n}\right)^{1/2} \frac{1}{k} \ell(x_{nj}) = \frac{1}{k} + \left(\frac{1}{n\sqrt{k}}\right)^{1/2} \frac{1}{\sqrt{k}} \ell(x_{nj}).$$

Hence it follows from (26) and from the definition of Riemann integral

$$\begin{aligned} \frac{T_{\phi,n}^I}{\sqrt{k_n}} &= \frac{nk_n}{2\sqrt{k_n}} \sum_{j=1}^{k_n} \frac{1}{n\sqrt{k_n}} \cdot \frac{1}{k_n} \ell^2(x_{nj}) \\ &= \frac{1}{2k_n} \sum_{j=1}^{k_n} \ell^2(x_{nj}) \rightarrow \frac{1}{2} \int_0^1 \ell^2(x) dx \end{aligned}$$

so that (56) holds for $\mu = \|\ell\|^2/2$. Since (57) is equivalent to

$$\frac{T_{\phi,n}^I - k_n/2}{\sqrt{k_n}} \xrightarrow{\mathcal{L}} N(\mu, 1/2),$$

the desired relation (29) is proved.

Proof of Theorem 3: Asymptotic distributions of statistics $T_{\phi,n}^I$ for k_n tending fastly to infinity were studied by Menéndez et al. (1998 b). By their main theorem, if $k_n = n$ and

$$\limsup_{n \rightarrow \infty} \max_n \max_{1 \leq j \leq n} np_{nj} < \infty \quad (58)$$

then all $\phi \in \Phi$ with

$$\phi(0) < \infty \quad \text{and} \quad \lim_{t \rightarrow \infty} \frac{\phi(t)}{t} < \infty \quad (59)$$

satisfy the asymptotic relation

$$\frac{\sqrt{n}[D_\phi(\hat{\mathbf{p}}_n, \mathbf{p}_0) - \mu_{\phi,n}]}{\sigma_{\phi,n}} \xrightarrow{\mathcal{L}} N(0, 1) \quad (60)$$

where

$$\mu_{\phi,n} = \frac{1}{n} \sum_{j=1}^n \mathbf{E}\phi(Z_{n,j}),$$

$$\sigma_{\phi,n}^2 = \frac{1}{n} \sum_{j=1}^n \text{var}\phi(Z_{n,j}) - \left[\frac{1}{n} \sum_{j=1}^n \text{covar}(Z_{n,j}, \phi(Z_{n,j})) \right]^2$$

and

$$Z_{n,j} \sim \text{Poisson}(p_{nj}/p_{0j}).$$

This theorem is applicable here. Indeed, since $\ell(x)$ is bounded on $[0,1]$, the condition (58) holds. Further, if $\alpha > 0$ then $\phi = \phi_\alpha$ satisfies the condition (59). Finally, by (35) and (36), $T_{\alpha,n} = nD_{\phi_\alpha}(\hat{\mathbf{p}}_n, \mathbf{p}_0)$. Therefore the desired result follows from (60) by substituting $\phi = \phi_\alpha$ and using $\lambda_{nj} = p_{nj}/p_{0j}$ given in (41).

Proof of Theorem 5. By Theorems 1, 2 and (53), for all $\phi \in \Phi$

$$\text{ARE}(T_\phi^I, T_\phi^{II}) = \frac{(\mu/\sigma)^4}{(\tilde{\mu}/\tilde{\sigma})^4} \quad (61)$$

where

$$\mu = \tilde{\mu} = \frac{\|\ell\|^2}{2}, \quad \sigma = \frac{1}{\sqrt{2}} \quad \text{and} \quad \tilde{\sigma} = \frac{1}{\sqrt{3}}.$$

This simple substitution in (61) leads for all $\phi \in \Phi$ to the desired relation (55).

Proof of Theorem 6. We shall use Theorems 3 and 4. By Theorem 3,

$$\frac{T_{2,n}^I - n\mu_{2,n}}{\sqrt{n\sigma_{2,n}^2}} \xrightarrow{\mathcal{L}} N(0, 1) \quad (62)$$

where

$$\mu_{2,n} = \frac{1}{2} \left(\frac{1}{n} \sum_{j=1}^n \mathbf{E}(Z_{n,j})^2 - 1 \right) \quad (\text{cf. (46)})$$

and

$$\sigma_{2,n}^2 = \frac{1}{4} \left\{ \frac{1}{n} \sum_{j=1}^n \text{var}Z_{n,j}^2 - \left[\frac{1}{n} \sum_{j=1}^n \text{covar}(Z_{n,j}, Z_{n,j}^2) \right]^2 \right\} \quad (\text{cf. (47)})$$

for

$$Z_{n,j} \sim \text{Poisson}(\lambda_{nj}), \quad \lambda_{nj} = 1 + \frac{\ell(x_{n,j})}{\sqrt[4]{n}}. \quad (63)$$

and $x_{n,j}$ defined by (39). By definition of var and covar,

$$\text{var} Z_{n,j}^2 = \text{E}Z_{n,j}^4 - (\text{E}Z_{n,j}^2)^2$$

and

$$\text{covar}(Z_{n,j}, Z_{n,j}^2) = \text{E}Z_{n,j}^3 - (\text{E}Z_{n,j})(\text{E}Z_{n,j}^2)$$

where by the standard formulas of probability theory

$$\text{E}Z_{n,j} = \lambda_{nj}, \quad \text{E}Z_{n,j}^2 = \lambda_{nj} + \lambda_{nj}^2, \quad \text{E}Z_{n,j}^3 = \lambda_{nj} + 3\lambda_{nj}^2 + \lambda_{nj}^3$$

and

$$\text{E}Z_{n,j}^4 = \lambda_{nj} + 7\lambda_{nj}^2 + 6\lambda_{nj}^3 + \lambda_{nj}^4.$$

Using these formulas we obtain the expressions

$$\mu_{2,n} = \frac{1}{2} \left(\frac{1}{n} \sum_{j=1}^n (\lambda_{nj} + \lambda_{nj}^2) - 1 \right)$$

and

$$\sigma_{2,n}^2 = \frac{1}{4} \left\{ \frac{1}{n} \sum_{j=1}^n (\lambda_{nj} + 6\lambda_{nj}^2 + 4\lambda_{nj}^3) - \left[\frac{1}{n} \sum_{j=1}^n (\lambda_{nj} + 2\lambda_{nj}^2) \right]^2 \right\}.$$

Inserting λ_{nj} given in (63), these expressions can be given the form

$$\mu_{2,n} = \frac{1}{2} (A_n + B_n - 1)$$

and

$$\sigma_{2,n}^2 = \frac{1}{4} \{ A_n + 6B_n + 4C_n - [A_n + 2B_n]^2 \}$$

where

$$\begin{aligned} A_n &= \frac{1}{n} \sum_{j=1}^n \lambda_{nj} = 1 + \frac{1}{\sqrt[4]{n}} a_n, \\ B_n &= \frac{1}{n} \sum_{j=1}^n \lambda_{nj}^2 = 1 + \frac{1}{\sqrt[4]{n}} a_n + \frac{b_n}{\sqrt{n}}, \\ C_n &= \frac{1}{n} \sum_{j=1}^n \lambda_{nj}^3 = 1 + \frac{3}{\sqrt[4]{n}} a_n + \frac{3b_n}{\sqrt{n}} + \frac{1}{\sqrt[4]{n^3}} c_n \end{aligned}$$

for

$$\begin{aligned} a_n &= \frac{1}{n} \sum_{j=1}^n \ell(x_{n,j}) = \int_0^1 \ell(x) dx = 0 \quad (\text{cf. (28)}), \\ b_n &= \frac{1}{n} \sum_{j=1}^n \ell^2(x) \lambda_{nj} \rightarrow \|\ell\|^2 < \infty, \\ c_n &= \frac{1}{n} \sum_{j=1}^n \ell^3(x_{n,j}) \rightarrow \int_0^1 \ell^3(x) dx < \infty. \end{aligned}$$

Therefore

$$\mu_{2,n} = \frac{1}{2} \left(1 + \frac{b_n}{\sqrt{n}} \right) = \frac{1}{2} \left(1 + \frac{\|\ell\|^2}{\sqrt{n}} (1 + o(1)) \right) = \frac{1}{2} \left(1 + \frac{\|\ell\|^2}{\sqrt{n}} + o\left(\frac{1}{\sqrt{n}}\right) \right)$$

and

$$\sigma_{2,n}^2 = \frac{1}{4} \{ 11 + o(1) - [3 + o(1)]^2 \} = \frac{1}{2} (1 + o(1)).$$

Thus the limit law (62) can be rewritten to the form

$$\frac{T_{2,n}^I - n/2 - \sqrt{n} \|\ell\|^2 / 2}{\sqrt{n}/2} \xrightarrow{\mathcal{L}} N(0, 1)$$

which is equivalent to

$$\frac{T_{2,n}^I - n/2}{\sqrt{n}} \xrightarrow{\mathcal{L}} N\left(\frac{\|\ell\|^2}{2}, \frac{1}{2}\right). \quad (64)$$

Now we turn to Theorem 4. By (48) - (50),

$$\mu_2 = \frac{\Gamma(3) - 1}{2} = \frac{1}{2}, \quad \sigma_2^2 = \frac{\Gamma(5) - 5\Gamma(3)}{4} = 1 \quad \text{and} \quad m_\alpha(\ell) = \frac{\|\ell\|^2}{2} \Gamma(3) = \|\ell\|^2$$

so that this theorem implies the asymptotic relation

$$\frac{T_{2,n}^{II} - n/2}{\sqrt{n}} \xrightarrow{\mathcal{L}} N(\|\ell\|^2, 1).$$

By (62) and (54), this relation together with (64) implies

$$\text{ARE}(T_\alpha^{II}, T_{\hat{\alpha}}^{II}) = \frac{\left(\frac{\|\ell\|^2/2}{1/\sqrt{2}}\right)^4}{(\|\ell\|^2/1)^4} = \left(\frac{\sqrt{2}}{2}\right)^4 = \frac{1}{4}$$

which completes the proof.

Acknowledgement. Preparation of this paper was partly supported by the MŠMT grant 1M0572 and the GAČR grant 102/07/1131.

Bibliography

- [1] Cressie, N. (1976). On the logarithms of high-order spacings. *Biometrika* **63**, 345–355.
- [2] Ekström, M. (1999). Strong limit theorems for sums of logarithms of high order spacings. *Statistics* **33**, 153–169.
- [3] Györfi, L. and Vajda, I. (2002). Asymptotic distributions for goodness-of-fit statistics in a sequence of multinomial models. *Statistics and Probability Letters* **56**, 57–67.
- [4] Fraser, D. A. S. (1957). *Nonparametric Methods in Statistics*. John Wiley, New York.
- [5] Hall, P. (1986). On powerful distributional tests based on sample spacings. *Journal of Multivariate Analysis* **19**, 201–224.
- [6] Harremoës P. and Vajda, I. (2007). On the Bahadur-efficient testing of uniformity by means of the entropy. *IEEE Transactions on Information Theory* **53** (in print).
- [7] Misra, N. and van der Meulen, E.C. (2001). A new test of uniformity based on overlapping simple spacings. *Communications in Statistics – Theory and Methods*. **30**, 1435–1470.
- [8] Menéndez, M. L. Morales, D. Pardo, M. C. and Vajda, I. (1998 a). Two approaches to grouping of data and related disparity statisticstotics. *Communications in Statistics – Theory and Methods*. **27**, 603–633.
- [9] Menéndez, M. L. Morales, D. Pardo, M. C. and Vajda, I. (1998 b). Asymptotic distributions of f -divergences of hypothetical and observed frequencies in sparse testing schemes. *Statistica Neerlandica* **8**, 313–328.
- [10] Morales, D. Pardo, L. Pardo, M. C. and Vajda, I. (2003). Limit laws for disparities of spacings. *Nonparametric Statistics* **15**, 325–342.
- [11] Morales, D. Pardo, L. Pardo, M. C. and Vajda, I. (2005). On the optimal number of classes in the Pearson goodness-of-fit tests. *Kybernetika* **41**, 677–698.
- [12] Pyke, R. (1972). Spacings revisited. *Proceedings of the Sixth Berkeley Symposium on Mathematical Statistics and Probability* **1**, 417–427.
- [13] Read, T. R. C. and Cressie, N. (1988). *Goodness of Fit Statistics for Discrete Multivariate Data*. Springer, Berlin.
- [14] Vajda, I. and van der Meulen, E. C. (2006 a). *Goodness-of-fit Tests Based on Observations Quantized by Hypothetical and Empirical Quantiles*. Research Report No. 2160, Institute of Information Theory and Automation, Prague.

- [15] Vajda, I. and van der Meulen, E. C. (2006 b). On estimation and testing by means of ϕ -disparities based on m -spacings. *Proceedings of Prague Stochastics 2006* (Ed. by M. Hušková and M. Janžura), pp.701-708. Matfyzpress, Prague.
- [16] Vajda, I. (2003). Asymptotic laws for stochastic disparity statistics. *Tatra Mountains Mathematical Publications* **26**, 269–280.
- [17] Vajda, I. (2006). Asymptotics for Pearson-type goodness-of-fit statistics. 38-e *Journées de Statistique de la SFdS*, pp. 1-7. SFdS-EDF, Paris.
- [18] van Es, B. (1992). Estimating functionals related to density by a class of statistics based on spacings. *Scandinavian Journal of Statistics* **19**, 61–72.