Akademie věd České republiky
Ústav teorie informace a automatizace

Academy of Sciences of the Czech Republic
Institute of Information Theory and Automation

RESEARCH REPORT

I. Vajda and E. C. van der Meulen:

Asymptotic properties of spacings–based divergence statistics

No. 2241 November 2008

ÚTIA AV ČR, P. O. Box 18, 182 08 Prague, Czech Republic
Telex: 122018 atom c, Fax: (+42) (2) 688 4903
E-mail: utia@utia.cas.cz

This report constitutes an unrefereed manuscript which is intended to be submitted for publication. Any opinions and conclusions expressed in this report are those of the author(s) and do not necessarily represent the views of the Institute.

On spacings-based divergence statistics and their asymptotic equivalence

I. Vajda[1] and E. C. van der Meulen[2]

[1] *Institute of Information Theory and Automation, Academy of Sciences of the Czech Republic, Pod vodárenskou věží 4, 182 08 Prague, Czech Republic. E-mail: vajda@utia.cas.cz*

[2] *Department of Mathematics, Katholieke Universiteit Leuven, Celestijnenlaan 200B, B-3001 Heverlee, Belgium. E-mail: edward.vandermeulen@wis.kuleuven.ac.be*

This paper introduces two types of goodness-of-fit statistics $T_\phi^{(m)}$ and $\tilde{T}_\phi^{(m)}$ based on m-spacings which are appropriately scaled divergences (ϕ-divergences or ϕ-disparities) of quantized hypothetical and empirical distributions. The goodness-of-fit statistics based on m-spacings known from the literature are systemized into several distinct types of statistics $U_\phi^{(m)}$ and compared with the two types of the divergence statistics $T_\phi^{(m)}$ and $\tilde{T}_\phi^{(m)}$. Mutual asymptotic equivalence between all these types of statistics is established. This equivalence helps to understand why many ad hoc defined spacings-based statistics exhibit desirable asymptotic properties. The results of the paper are applicable e.g. to the class of power functions ϕ of arbitrary real orders leading to the power divergences $T_\phi^{(m)}$ and $\tilde{T}_\phi^{(m)}$ of arbitrary orders.

Key words: Goodness-of-fit statistics, Spacings, Spacings-based statistics, Divergences, ϕ-divergences, Power divergences, ϕ-disparities, Robust disparities, Spacings-based divergence statistics, Asymptotic equivalence.

1 Divergence statistics

We consider real-valued independent identically distributed observations $X_1, \dots, X_n$ with a distribution function $F(x)$ and the problem of testing the hypothesis $\mathcal{H}_0$ that F is a given continuous increasing distribution function F_0. As is well known, we can then assume without loss of generality that the observation space is the interval $\mathcal{X} = (0, 1)$ and $F_0(x) = x$ on $\mathcal{X}$. Further, we can restrict ourselves to test statistics T_n which are functions of sufficient statistics. Examples of sufficient statistics are the empirical distribution function

$$F_n(x) = \frac{1}{n} \sum_{i=1}^{n} \boldsymbol{I}(x \geq X_i), \quad x \in \mathcal{X} \tag{1.1}$$

where $\boldsymbol{I}$ is the indicator function, and the order statistics

$$0 = Y_0 \leq Y_1 = X_{n:1} \leq \dots \leq Y_n = X_{n:n} \leq Y_{n+1} = 1 \tag{1.2}$$

where Y_0 and Y_{n+1} are dummy variables and the inequalities are typically strict with probability one.

It is natural to consider test statistics of the form $T = T_n = c_n D(F_0, F_n)$ where c_n is an appropriate scaling constant and $D(F, G)$ is a nonnegative measure of divergence between two distribution functions F and G on $\mathcal{X} = (0, 1)$. We shall deal with the divergence measures

$$D(F, G) = D_\phi(\boldsymbol{p}, \boldsymbol{q}) = \sum_{j=1}^{k} q_j \phi\left(\frac{p_j}{q_j}\right) \tag{1.3}$$

corresponding to quantizations $\boldsymbol{p} = (p_1, p_2, ..., p_k)$, $\boldsymbol{q} = (q_1, q_2, ..., q_k)$ of F, G by an interval partition $\mathcal{P}$ of $\mathcal{X} = (0, 1)$ using certain cutpoints

$$0 = a_0 < a_1 \cdots < a_{k-1} < a_k = 1 \quad \text{for} \quad k > 1 \tag{1.4}$$

and to ϕ from the class $\boldsymbol{\Phi}$ of convex functions $\phi : (0, \infty) \mapsto \mathbb{R}$ which are twice continuously differentiable in a neighborhood of 1 with $\phi''(1) > 0$ and $\phi(1) = 0$.

If $\phi \in \boldsymbol{\Phi}$ is convex then the divergences (1.3) are *ϕ-divergences* introduced for arbitrary probability distributions by Csiszár (1963) (for details about the definition of ϕ-divergences and their properties see Liese and Vajda (1987, 2006)). If $\phi \in \boldsymbol{\Phi}$ is differentiable everywhere on $(0, \infty)$ and the derivative satisfies for all $t > 0$ the condition

$$\phi'(t)\, \mathrm{sign}(t - 1) > 0 \tag{1.5}$$

then $\phi(t)$ is is said *disparity function* and the divergence (1.3) is said *ϕ-disparity*. For differentiable convex $\phi \in \boldsymbol{\Phi}$ relation (1.5) automatically holds so that the concept of ϕ-disparity in some sense generalizes the concept of ϕ-divergence. This generalization was motivated in the papers of Lindsay (1994) and Menéndes et al. (1998) introducing the concept of *ϕ-disparity* by the need to robustify the statistical estimation and testing based on the minimum divergence between distributions resulting from the accepted *statistical model* and the empirical distribution resulting from the observed *statistical reality*. It was argued that differentiable convex functions $\phi \in \boldsymbol{\Phi}$ with unbounded derivatives lead to the methods which are are non-robust (too sensitive to contaminations of statistical observations). Since derivatives of typical convex functions are not bounded (see e.g. the next example), the authors of those papers proposed to replace the convexity of ϕ by the technically less appealing but practically often more desirable disparity property (1.5) combined with the boundedness of the derivative,

$$\sup_{t>0} |\, \phi'(t)|, < \infty.$$

In other words, in robust statistics are applicable the divergences (1.3) for functions $\phi \in \boldsymbol{\Phi}$ with the property

$$0 < \phi'(t)\, \mathrm{sign}(t - 1) < \infty. \tag{1.6}$$

Example 1.1. As examples of convex functions $\phi \in \mathbf{\Phi}$ leading to ϕ-divergences $D_\phi(\boldsymbol{p}, \boldsymbol{q})$ may serve the *power functions* $\{\phi_\alpha : \alpha \in \mathbb{R}\}$ defined on $(0, \infty)$ by

$$\phi_\alpha(t) = \frac{t^\alpha - \alpha(t-1) - 1}{\alpha(\alpha - 1)} \tag{1.7}$$

if $\alpha \neq 0$, $\alpha \neq 1$, and by the continuous extensions

$$\phi_1(t) = t \ln t - t + 1 \quad \text{and} \quad \phi_0(t) = -\ln t + t - 1 \tag{1.8}$$

of these functions to $\alpha = 0$, $\alpha = 1$ in the opposite case. The corresponding divergences $D_{\phi_\alpha}(\boldsymbol{p}, \boldsymbol{q})$ are called *power divergences* and denoted simply by $D_\alpha(\boldsymbol{p}, \boldsymbol{q})$. The class of power divergences $D_\alpha(\boldsymbol{p}, \boldsymbol{q})$ contains the following classical divergences: the *quadratic divergence*

$$D_2(\boldsymbol{p}, \boldsymbol{q}) = \frac{1}{2}\chi^2(\boldsymbol{p}, \boldsymbol{q}) = \frac{1}{2}\sum_{j=1}^{k} \frac{(p_j - q_j)^2}{q_j} \tag{1.9}$$

where $\chi^2(\boldsymbol{p}, \boldsymbol{q})$ is also known as *Pearson divergence*, the *harmonic divergence*

$$D_{-1}(\boldsymbol{p}, \boldsymbol{q}) = D_2(\boldsymbol{q}, \boldsymbol{p}) = \frac{1}{2}\sum_{j=1}^{k} \frac{(p_j - q_j)^2}{q_j}, \tag{1.10}$$

the *logarithmic divergences*

$$D_0(\boldsymbol{p}, \boldsymbol{q}) = D_1(\boldsymbol{q}, \boldsymbol{p}) \quad \text{and} \quad D_1(\boldsymbol{p}, \boldsymbol{q}) = I(\boldsymbol{p}, \boldsymbol{q}) = \sum_{j=1}^{k} p_j \ln \frac{p_j}{q_j} \tag{1.11}$$

where $I(\boldsymbol{p}, \boldsymbol{q})$ is known as the *information divergence* (often denoted also as $D(\boldsymbol{p} \parallel \boldsymbol{q})$), and the *square root divergence*

$$D_{1/2}(\boldsymbol{p}, \boldsymbol{q}) = 4H^2(\boldsymbol{p}, \boldsymbol{q}) = 4\sum_{j=1}^{k} \left(\sqrt{p_j} - \sqrt{q_j}\right)^2 \tag{1.12}$$

where $H(\boldsymbol{p}, \boldsymbol{q})$ is known as *Hellinger distance*.

Example 1.2. Functions $\phi \in \mathbf{\Phi}$ with the disparity property (1.5) are obtained by the shift

$$\phi(t) = \rho(t-1), \quad t > 0 \tag{1.13}$$

of typical ρ-functions used to define robust statistical M-estimators (see e.g. Hampel et al. (1986) or Jurečková and Sen (1996)). The robustness means that the correspoding sensitivity function $\psi(t) = \rho'(t), t \in \mathbb{R}$ is bounded. Classical example is the so-called Huber class $\{\rho_\alpha : \alpha > 0\}$ with the bounded sensitivity functions

$$\psi_\alpha(t) = \rho'_\alpha(t) = \boldsymbol{I}(|\,t\,| \leq \alpha)t + \boldsymbol{I}(|\,t\,| > \alpha)\alpha, \quad t \in \mathbb{R}.$$

Another example is the class $\{\rho_\alpha : \alpha > 0\}$ of bounded functions

$$\rho_\alpha(t) = \left(1 - \exp\{-\alpha t^2 / 2\}\right) / \alpha, \quad t \in \mathbb{R} \tag{1.14}$$

with bounded (redescending) sensitivity functins $\psi_\alpha(t) = \rho'_\alpha(t) = t\exp\{-\alpha t^2\}$.

We admit that the size $k = k_n$ of the interval partition $\mathcal{P} = \{(a_{j-1}, a_j] : 1 \le j \le k\}$ of $\mathcal{X}$ introduced in (1.4), and also the cutpoints $a_1, \ldots, a_{k-1}$ themselves, may in general depend on the sample size n, but this dependence is not always explicitly denoted in this paper.

Quantizations of the special distributions F_0 (hypothetical) and F_n (empirical) by means of partitions defined by cutpoints (1.4) lead to discrete *hypothetical* and *empirical* distributions

$$\boldsymbol{p}_0 = (p_{0j} : 1 \le j \le k) \qquad \text{and} \qquad \boldsymbol{p}_n = (p_{nj} : 1 \le j \le k) \tag{1.15}$$

where

$$p_{0j} = F_0(a_j) - F_0(a_{j-1}) = a_j - a_{j-1} > 0 \tag{1.16}$$

and

$$p_{nj} = F_n(a_j) - F_n(a_{j-1}) > 0 \quad \text{a.s.} \tag{1.17}$$

These distributions can serve as arguments of the divergences D_ϕ in (1.3), yielding $D_\phi(\boldsymbol{p}_0, \boldsymbol{p}_n)$, and of the corresponding *divergence statistics*

$$T_\phi = T_{\phi,n} = n\, D_\phi(\boldsymbol{p}_0, \boldsymbol{p}_n) = n \sum_{j=1}^{k} p_{nj}\phi\left(\frac{p_{0j}}{p_{nj}}\right). \tag{1.18}$$

In this paper we restrict ourselves to the simplest divergence statistics T_ϕ, which are obtained when one of the distributions $\boldsymbol{p}_0, \boldsymbol{p}_n$ in (1.18) is uniform, that is, equal to

$$\boldsymbol{u}_k = (u_{kj} = 1/k : 1 \le j \le k)\,. \tag{1.19}$$

This takes place when the cutpoints a_j of (1.16) or (1.17) are the quantiles

$$a_j = G^{-1}(j/k) = \inf\{x \in (0,1] : G(x) \ge j/k\} \tag{1.20}$$

of the distribution functions $G = F_0$ or $G = F_n$, respectively. Proceeding this way we obtain two versions of divergences $D_\phi(\boldsymbol{p}_0, \boldsymbol{p}_n)$ and divergence statistics T_ϕ.

Version I. Applying the rule (1.20) to $G = F_0$ we get the *hypothetical quantiles*

$$a_j = F_0^{-1}(j/k) = j/k, \quad 1 \le j \le k-1 \tag{1.21}$$

leading, according to (1.16) and (1.18), to the *uniform hypothetical distribution* $\boldsymbol{p}_0 = \boldsymbol{u}_k$ and the *frequency-based divergence statistics*

$$T_\phi := n\, D_\phi(\boldsymbol{u}_k, \boldsymbol{p}_n) = n \sum_{j=1}^{k} p_{nj}\phi\left(\frac{1}{k\, p_{nj}}\right) \tag{1.22}$$

where the p_{nj} *'s* are given by (1.17) for the a_j of (1.21) and $a_0 = 0$, $a_k = 1$. We denote the corresponding partition by $\mathcal{P}^{(\mathrm{I})}$. In other words

$$p_{nj} = \frac{1}{n} \sum_{i=1}^{n} \boldsymbol{I}_{(a_{j-1},a_j]}(Y_i) \tag{1.23}$$

is the relative frequency of the observations $Y_1, \dots, Y_n$ in the cell $(a_{j-1}, a_j] = ((j-1)/k, j/k)]$, $j = 1, \cdots, k$, of the partition $\mathcal{P}^{(\mathrm{I})}$.

Example 1.3. A well-known class of the frequency-based statistics (1.22) consists of the power divergence statistics

$$T_\alpha = n\, D_\alpha(\boldsymbol{u}_k, \boldsymbol{p}_n) = n \sum_{j=1}^{k} p_{nj}\, \phi_\alpha \left(\frac{1}{kp_{nj}} \right), \quad \alpha \in \mathbb{R} \tag{1.24}$$

systematically studied in Read and Cressie (1988). Classical examples of such statistics are the Neyman statistic $T_2 = n\, D_2(\boldsymbol{u}_k, \boldsymbol{p}_n)$, the Pearson statistic $T_{-1} = n\, D_{-1}(\boldsymbol{u}_k, \boldsymbol{p}_n)$, the log-likelihood ratio statistic $T_0 = n\, D_0(\boldsymbol{u}_k, \boldsymbol{p}_n)$, the reversed log-likelihood ratio statistic $T_1 = n\, D_1(\boldsymbol{u}_k, \boldsymbol{p}_n)$, and the Freeman–Tukey statistic $T_{1/2} = n\, D_{1/2}(\boldsymbol{u}_k, \boldsymbol{p}_n)$. In these statistics it is admitted that the size $k = k_n$ of the interval partition $\mathcal{P} = \{(a_{j-1}, a_j] : 1 \leq j \leq k\}$ of $\mathcal{X}$ introduced in (1.4), and also the cutpoints $a_1, \dots, a_{k-1}$ themselves depend on the sample size n. These statistics were extensively studied in the literature, see e.g. Györfi and Vajda (2002) and references therein.

Version II. In this paper we study the disparity statistics T_ϕ obtained from (1.18) when rule (1.20) is applied to the empirical distribution $G = F_n$, leading to the *empirical quantiles* $a_j = F_n^{-1}(j/k)$. For simplicity we restrict ourselves to the sample sizes n divisible by k. Then, using the integers $m = n/k \geq 1$, we get the $k-1$ empirical quantiles

$$a_j = F_n^{-1}(j/k) = Y_{mj}, \quad 1 \leq j \leq k-1 \tag{1.25}$$

and the partition $\mathcal{P}^{(m)} = \mathcal{P}_n^{(m)}$ consisting of the k cells

$$(a_{j-1}, a_j] = (Y_{m(j-1)}, Y_{mj}], \quad 1 \leq j \leq k-1, \qquad (a_{k-1}, a_k] = (Y_{m(k-1)}, 1] \tag{1.26}$$

where $a_0 = Y_{m0} = Y_0 = 0$ (cf (1.2) and (1.4)), leading to the hypothetical distribution $\boldsymbol{p}_0 = (p_{0j} : 1 \leq j \leq k)$ with

$$p_{0j} = Y_{mj} - Y_{m(j-1)} \quad \text{for} \quad 1 \leq j \leq k-1, \quad \text{and} \quad p_{0k} = 1 - Y_{m(k-1)}. \tag{1.27}$$

(Note that here and in the sequel mj, $m(j-1)$ and so forth denote the products of integers and not the pairs of integers as in (1.15) – (1.18) and elsewhere. We believe that the correct meaning of mj can always be recognized. Also note that the order statistic Y_{mk} does not occur as an endpoint in the definition (1.26) of the cells $(a_{j-1}, a_j], 1 \leq j \leq k$.)

Since all cells $(a_{j-1}, a_j]$, $1 \leq j \leq k$, in (1.26) contain exactly m of the observations $Y_1, \dots, Y_n$, formulas (1.17) and (1.18) lead to the *uniform empirical distribution* $\boldsymbol{p}_n = \boldsymbol{u}_k$

and to the *spacings-nased divergence statistics*

$$T_\phi^{(m)} = n\, D_\phi(\boldsymbol{p}_0, \boldsymbol{u}_k) = m \sum_{j=1}^{k} \phi(kp_{0j}) \tag{1.28}$$

where the p_{0j} *'s* are given by (1.27) with $Y_0 = 0$. The use of the spacings terminology is justified by the fact that, since $k = n/m$, formula (1.28) can be given the form

$$T_\phi^{(m)} = m \sum_{j=1}^{k-1} \phi\left(\frac{n}{m}(Y_{mj} - Y_{m(j-1)})\right) + m\,\phi\left(\frac{n}{m}(1 - Y_{m(k-1)})\right) \tag{1.29}$$

where $Y_{mj} - Y_{m(j-1)}$ are m-spacings. It differs from the simpler expression

$$m \sum_{j=1}^{k} \phi\left(\frac{n}{m}(Y_{mj} - Y_{m(j-1)})\right)$$

by the modification of the last term

$$m\phi\left(\frac{n}{m}(Y_{mk} - Y_{m(k-1)})\right) \longmapsto m\phi\left(\frac{n}{m}(1 - Y_{m(k-1)})\right)$$

resulting from the definition of $\boldsymbol{p}_0$ in (1.27).

The first cell $(a_0, a_1] = (Y_0, Y_m]$ of the partition (1.26) can be extended and the last cell $(a_{k-1}, a_k] = (Y_{m(k-1)}, 1]$ of this partition reduced as follows

$$(a_0, a_1] \longmapsto (Y_0, Y_m] \cup (Y_{mk}, 1] \quad \text{and} \quad (a_{k-1}, a_k] \longmapsto (Y_{m(k-1)}, Y_{mk}] \tag{1.30}$$

where

$$(Y_{mk}, 1] = (Y_n, Y_{n+1}] \quad \text{and} \quad (Y_{m(k-1)}, Y_{mk}] = (Y_{n-m}, Y_n]. \tag{1.31}$$

This rearranging of cells leads to a modified partition $\tilde{\mathcal{P}}^{(m)}$ and a modified hypothetical distribution $\tilde{\boldsymbol{p}}_0$ with components

$$\tilde{p}_{0j} = Y_{mj} - Y_{m(j-1)} \quad \text{for} \quad 2 \le j \le k, \quad \text{and} \quad \tilde{p}_{01} = Y_m + 1 - Y_{mk} \quad (\text{cf. } (1.27)). \tag{1.32}$$

The corresponding divergence statistic $\tilde{T}_\phi^{(m)} = n\, D_\phi(\tilde{\boldsymbol{p}}_0, \boldsymbol{p}_k)$ is of the form

$$\tilde{T}_\phi^{(m)} = m \sum_{j=2}^{k} \phi\left(\frac{n}{m}(Y_{mj} - Y_{m(j-1)})\right) + m\,\phi\left(\frac{n}{m}(Y_m + 1 - Y_{m(k-1)})\right). \tag{1.33}$$

In this paper we compare the spacings-based divergence statistics $T_\phi^{(m)}$ and $\tilde{T}_\phi^{(m)}$ with the classical goodness-of-fit statistics based on m-spacings. The comparison is asymptotic, carried out for $m \ge 1$ fixed and $n = km$ increasing to infinity.

For $m = 1$ the partition $\mathcal{P}^{(m)}$ of (1.26) reduces to $\mathcal{P}^{(1)}$ consisting of the intervals

$$(a_{j-1}, a_j] = (Y_{j-1}, Y_j], \quad 1 \le j \le n-1, \qquad (a_{n-1}, a_n] = (Y_{n-1}, 1], \tag{1.34}$$

where $a_0 = Y_0 = 0$ and $a_n = 1 = Y_{n+1}$. Further, the components (1.27) of the distribution $\boldsymbol{p}_0$ reduce to

$$p_{0j} = Y_j - Y_{j-1} \quad \text{for} \quad 1 \le j \le n-1, \quad \text{and} \quad p_{0n} = 1 - Y_{n-1}, \tag{1.35}$$

and the m-spacings statistic $T_\phi^{(m)}$ of (1.29) reduces to the simple-spacings-formula

$$T_\phi = \sum_{j=1}^{n-1} \phi\left(n(Y_j - Y_{j-1})\right) + \phi\left(n(1 - Y_{n-1})\right). \tag{1.36}$$

Similarly, the partition $\tilde{\mathcal{P}}^{(m)}$ of (1.30) reduces to $\tilde{\mathcal{P}}^{(1)}$ consisting of the intervals

$$(a_{j-1}, a_j] = (Y_{j-1}, Y_j], \quad 2 \le j \le n, \qquad (a_0, a_1] = (Y_0, Y_1] \cup (Y_n, 1], \tag{1.37}$$

the components (1.32) of the distribution $\tilde{\boldsymbol{p}}_0$ reduce to

$$p_{0j} = Y_j - Y_{j-1} \quad \text{for} \quad 2 \le j \le n, \quad \text{and} \quad p_{0n} = Y_1 + 1 - Y_n, \tag{1.38}$$

and the m-spacings statistic $\tilde{T}_\phi^{(m)}$ of (1.33) reduces to the simple-spacings-formula

$$\tilde{T}_\phi = \sum_{j=2}^{n} \phi\left(n(Y_j - Y_{j-1})\right) + \phi\left(n(Y_1 + 1 - Y_{n-1})\right). \tag{1.39}$$

Remark 1.1. The formulas above employ the dummy observations $Y_0 = 0$ and $Y_{n+1} = 1$ introduced in (1.2). Unless otherwise explicitly stated, these dummy observations are also assumed in the formulas below, notably in (1.40) and (1.42).

It seems that the first attempt to compare the frequency-based divergence statistics (1.22) and (1.24) with the spacings-based statistics (1.36) and (1.39) was undertaken by Morales et al. (2003), Vajda and van der Meulen (2006a, 2006b) and Vajda (2007). As it was already observed there, the spacings-based statistics given in the previous literature lacked the motivation based on the notion of divergence between hypothetical and empirical distributions $\boldsymbol{p}_0$ and $\boldsymbol{p}_n$. This contrasts with the goodness-of-fit statistics based on deterministic partitions derived from the a_j in (1.21) and the related frequency counts (1.23), where the typical statistics, including the most classical Pearson statistic T_1 and likelihood ratio statistic T_0, can easily be recognized as appropriately scaled power divergences between $\boldsymbol{p}_0$ and $\boldsymbol{p}_n$. The classical spacings-based statistics, however, appear to have been motivated rather by other considerations such as the analytic simplicity of formulas and the possibility to achieve desired asymptotic properties. In fact, as pointed out by Pyke (1965) in his landmark paper, most of the classical spacings-based statistics were proposed within the context of testing the randomness of events in time, in which differences between successive order statistics (spacings) were considered to play an important role. Also, in the period 1946-1953, when most of the classical tests based on spacings were proposed, research focused mostly on studying the behavior of these tests under the null-hypothesis, rather than under an alternative, making it unnecessary to motivate the test statistic from the point of view of divergence or disparity. Although the concept of dispersion of spacings around the uniform distribution has been mentioned as a motivation for a test statistic by some authors, no known spacings-based statistic happens to be the divergence statistic $T_\phi^{(m)}$ of (1.29) or T_ϕ of (1.34) for some $\phi \in \mathbf{\Phi}$. This situation is illustrated in the next two examples for the simple-spacings statistics where $m = 1$. Then

$$T_\phi = R_\phi + \Delta_\phi \quad \text{for} \quad R_\phi = \sum_{j=1}^{n+1} \phi\left(n\left(Y_j - Y_{j-1}\right)\right) \tag{1.40}$$

and

$$\Delta_\phi = \phi\left(n(1 - Y_{n-1})\right) - \phi\left(n(Y_n - Y_{n-1}) - \phi\left(n(1 - Y_n)\right)\right), \tag{1.41}$$

while the classical simple-spacings statistics are of the form

$$S_\phi = \sum_{j=1}^{n+1} \phi\left((n+1)\right)\left(Y_j - Y_{j-1}\right)). \tag{1.42}$$

With reference to the above discussion, we mention here that Pyke (1965) writes that it is more convenient to weight the spacings by $n+1$ instead of n if one is concerned entirely with uniform observations.

Example 1.4. A first statistic of the type (1.42) is

$$S_\psi = \sum_{j=1}^{n+1} \left((n+1)\left(Y_j - Y_{j-1}\right)\right)^2, \tag{1.43}$$

which is $(n+1)^2$ times the statistic $\mathcal{G}$ introduced by Greenwood (1946). Both statistics are based on the function $\psi(t) = t^2$, which is not in the class $\mathbf{\Phi}$, but the closely related function $\phi_2(t)$ of (1.7) is. Greenwood devised his test statistic in connection with the problem of testing whether intervals between successive events in epidemiology were exponentially distributed. Irwin, in the discussion of Greenwood (1946), and Kimball (1947) suggested to use

$$\mathcal{K} \;=\; \sum_{j=1}^{n+1} \left(Y_j - Y_{j-1} - \frac{1}{n+1}\right)^2$$

instead of $\mathcal{G}$. It so happens that $\mathcal{K}$ equals $(2/(n+1))$-times the quadratic divergence $D_{\phi_2}(\hat{\boldsymbol{p}}_0, \boldsymbol{u}_{n+1}) = D_2(\hat{\boldsymbol{p}}_0, \boldsymbol{u}_{n+1})$, defined as in (1.9) with $k = n+1$, for the hypothetical distribution

$$\hat{\boldsymbol{p}}_0 = \left(\hat{p}_{01} = Y_1 - Y_0,\, \hat{p}_{02} = Y_2 - Y_1, \ldots, \hat{p}_{0,n+1} = Y_{n+1} - Y_n\right), \tag{1.44}$$

obtained from (1.2) by using the cutpoints $a_j = Y_j$ for $0 \le j \le n+1$ from the extended sample (1.2), and the uniform empirical distribution $\boldsymbol{u}_{n+1}$ obtained by putting $k = n+1$ in (1.19). Clearly, $\hat{\boldsymbol{p}}_0$ is an $(n+1)$-component distribution on the partition $\hat{\mathcal{P}}^{(1)}$ consisting of the $n+1$ cells

$$(a_{j-1}, a_j] = (Y_{j-1}, Y_j], \quad 1 \leq j \leq n+1, \tag{1.45}$$

and $\boldsymbol{u}_{n+1}$ is the uniform distribution on $\hat{\mathcal{P}}^{(1)}$. But instead of using the divergence argument, Kimball argued that $1/(n+1)$ is the common expectation of the random variables $Y_j - Y_{j-1}$ under $\mathcal{H}_0 : F = F_0$, so that $\mathcal{H}_0$ minimizes the expectation of $\mathcal{K}$. Nevertheless, $\mathcal{K}$ cannot be obtained from the class of spacings-type disparity statistics T_ϕ derived according to **Version II** by quantizing F_0 and F_n by means of the partition $\hat{\mathcal{P}}^{(1)}$ and putting $\phi = \phi_2$. Indeed, such quantization would yield the hypothetical distribution $\hat{\boldsymbol{p}}_0$, and, since the cell $(a_n, a_{n+1}] = (Y_n, 1]$ is empty, the empirical distribution $\hat{\boldsymbol{p}}_n = (1/n, \ldots, 1/n, 0)$, which is not the desired uniform distribution $\boldsymbol{u}_{n+1}$. The corresponding ϕ-disparity statistic T_ϕ (c.f. (1.18)) would be based on $D_2(\hat{\boldsymbol{p}}_0, \hat{\boldsymbol{p}}_n) \neq D_2(\hat{\boldsymbol{p}}_0, \boldsymbol{u}_{n+1})$. Nevertheless, we shall prove in Section 3 an asymptotic equivalence between $\mathcal{K}$ and T_{ϕ_2} from (1.34).

Example 1.5. A second well-known statistic from the family (1.42) is

$$\mathcal{M} = S_{\tilde{\phi}_0} = -\sum_{j=1}^{n+1} \ln\left((n+1)\left(Y_j - Y_{j-1}\right)\right), \tag{1.46}$$

which was introduced by Moran (1951) for the function $\tilde{\phi}_0(t) = -\ln t$ which belongs to $\boldsymbol{\Phi}$. One can verify that $\mathcal{M}$ is $(n+1)$-times the logarithmic divergence $D_{\tilde{\phi}_0}(\hat{\boldsymbol{p}}_0, \boldsymbol{u}_{n+1}) = D_{\phi_0}(\hat{\boldsymbol{p}}_0, \boldsymbol{u}_{n+1}) = D_0(\hat{\boldsymbol{p}}_0, \boldsymbol{u}_{n+1})$, defined as in (1.11) with $k = n+1$, for the same hypothetical and empirical distributions $\hat{\boldsymbol{p}}_0$ and $\boldsymbol{u}_{n+1}$ as in Kimball's statistic above. However, for the same reason as given in Example 1.2, $\mathcal{M}$ cannot be derived from the class of our spacings-type disparity statistics T_ϕ obtained through quantization of F_0 and F_n by $\hat{\mathcal{P}}^1$ and setting $\phi = \phi_0$, which procedure would yield $(n+1)D_0(\hat{\boldsymbol{p}}_0, \hat{\boldsymbol{p}}_n) \neq \mathcal{M}$. But in Section 3 we shall prove an asymptotic equivalence between $\mathcal{M}$ and T_{ϕ_0} of the form (1.34).

Let us now turn to comparing our m-spacings–based disparity statistics $T_\phi^{(m)}$ from (1.28) - (1.29) and the m-spacings–based statistics known from the literature for general $m \geq 1$. We shall start with Del Pino's (1979) class of statistics of the form

$$S_\phi^{(m)} = m \sum_{j=1}^{k} \phi\left(\frac{n+1}{m}(Y_{mj} - Y_{m(j-1)})\right) \tag{1.47}$$

where it is assumed that $n+1$ is divisible by k and that $m = (n+1)/k \geq 1$. Hence the notation in our paper is consistent in the sense that (1.47) reduces for $m = 1$ to the formula for S_ϕ in (1.42). Del Pino found $\phi(t) = t^2$ to be optimal among the functions ϕ considered by him. The class (1.47) was later investigated by Jammalamadaka *et al.* (1989) and many others. Jammalamadaka *et al* studied the asymptotics of $S_\phi^{(m)}$ for m tending slowly to infinity as $n \to \infty$. In such case these asymptotics depend only on the local properties of $\phi(t)$ in the neighborhood of $t = 1$ and a wide class of functions ϕ can be admitted including those with $\phi''(1) = 0$. However, as we have seen in the examples above, even for ϕ from the above introduced ϕ-divergence class $\boldsymbol{\Phi}$, the statistics (1.47) differ from those in (1.28) or (1.29). Other examples of well-known spacings-based statistics which differ from our spacings-type ϕ-disparity statistics (1.28) and (1.29) will be given in the next section. Therefore it is important to look at the problem whether the classical spacings-based statistics and our spacings-type disparity statistics are asymptotically equivalent for $n \to \infty$, and, if yes, then in what precise sense.

The objective of the present paper is to prove the mutual asymptotic equivalence of the statistics of the two mentioned origins. This equivalence helps to understand why many ad hoc defined spacings-based statistics exhibit desirable asymptotic properties. Let us describe briefly how the paper is organized. Sofar we have defined for $\phi \in \Phi$ and general $m \geq 1$ the class $\mathcal{U}_\phi^{(m)} = \left\{T_\phi^{(m)}, \tilde{T}_\phi^{(m)}\right\}$ of two spacings-based statistics and for $m = 1$ the class $\mathcal{U}_\phi = \left\{R_\phi, S_\phi, T_\phi, \tilde{T}_\phi\right\}$ of four different spacings-based statistics. In Section 2 the structure of the new spacings-type divergence statistic $T_\phi^{(m)}, \tilde{T}_\phi^{(m)}$ or $T_\phi, \tilde{T}_\phi$ is compared with that of the spacings-based statistics known from the literature, and the classes $\mathcal{U}_\phi^{(m)}$ and $\mathcal{U}_\phi$ are appropriately extended to cover all known types of spacings-based statistics. Section 3 establishes asymptotic equivalence of the statistics in the extended classes $\mathcal{U}_\phi^{(m)}$ and $\mathcal{U}_\phi$.

2 Spacings-based statistics

This section reviews various types of spacings-based goodness-of-fit statistics known from the literature. As before, $0 \leq Y_1 \leq \cdots \leq Y_n \leq 1$ are the ordered observations. Unless otherwise explicitly stated, we use also the dummy observations $Y_0 = 0$ and $Y_{n+1} = 1$.

Let us start with our spacings-type divergence statistics $T_\phi^{(m)}$ and $\tilde{T}_\phi^{(m)}$ introduced in (1.29) and (1.33). These statistics are not efficient if $m > 1$ because then they ignore the observations Y_{mj+r} for $1 \leq j \leq k-1$ and $1 \leq r \leq m-1$. Shifting the orders j/k of the quantiles in (1.25) by a quantity depending on r, we obtain the additional quantiles

$$a_j^{(r)} = F_n^{-1}\left(\frac{mj+r}{n}\right) = Y_{mj+r}, \quad 1 \leq j \leq k-1, \quad 1 \leq r \leq m-1 \tag{2.1}$$

and the corresponding shifted hypothetical probabilities $p_{0j}^{(r)} = Y_{mj+r} - Y_{m(j-1)+r}$ as alternatives to the previously considered quantiles $a_j = a_j^{(0)}$ and probabilities $p_{0j} = Y_{mj} - Y_{m(j-1)} = p_{0j}^{(0)}$. At the same time, this operation preserves the uniform shifted empirical probabilities $p_{nj}^{(r)} = 1/k = m/n$ on the cells $(a_{j-1}^{(r)}, a_j^{(r)}]$, $1 \leq r \leq m-1$. Replacing each term $\phi(\frac{n}{m}(Y_{mj} - Y_{m(j-1)}))$ in (1.29) by the average

$$\frac{1}{m}\sum_{r=0}^{m-1} \phi\left(\frac{n}{m}p_{0j}^{(r)}\right) = \frac{1}{m}\sum_{r=0}^{m-1} \phi\left(\frac{n}{m}(Y_{mj+r} - Y_{m(j-1)+r})\right) \tag{2.2}$$

of all m alternatives get a more efficient versions of $T_\phi^{(m)}$ and $\tilde{T}_\phi^{(m)}$, namely

$$\boldsymbol{T}_\phi^{(m)} = \sum_{j=0}^{n-m-1} \phi\left(\frac{n}{m}(Y_{j+m} - Y_j)\right) + m\phi\left(\frac{n}{m}(1 - Y_{n-m})\right) \tag{2.3}$$

and

$$\tilde{\boldsymbol{T}}_{\phi}^{(m)} = \sum_{j=1}^{n-m} \phi\left(\frac{n}{m}(Y_{j+m} - Y_j)\right) + m\phi\left(\frac{n}{m}(Y_m + 1 - Y_{n-m})\right). \tag{2.4}$$

These statistics for $m = 1$ reduce to the T_ϕ and $\tilde{T}_\phi$ of (1.34) and (1.39), so that the new notation is consistent with the old.

The same procedure can be applied also to Del Pino's statistic $S_\phi^{(m)}$ of (1.47) which, similarly as $T_\phi^{(m)}$, involves for $j = 1, \cdots, k$ the observations Y_{mj+r} only for $r = 0$ and ignores those for $1 \leq r \leq m-1$. Applying the averaging and substitution from the previous paragraph, with the quantiles $a_j^{(r)}$ of (2.1) replaced by

$$a_j^{(r)} = F_n^{-1}\left(\frac{mj+r}{n+1}\right)$$

and excluding the undefined observations Y_k for $k > n+1$, we get the more efficient version

$$\boldsymbol{S}_{\phi}^{(m)} = \sum_{j=0}^{n-m+1} \phi\left(\frac{n+1}{m}(Y_{j+m} - Y_j)\right) \tag{2.5}$$

of $S_\phi^{(m)}$ of (1.47). Notice that if $m = 1$, then $\boldsymbol{S}_\phi^{(m)}$ of (2.5) reduces to S_ϕ of (1.42) above, so that our notation is in this sense still consistent.

The statistics (2.5) are formally well defined for all $1 \leq m \leq n$, and not only for $m = (n+1)/k \geq 1$ corresponding to the integers $1 < k \leq n+1$. Cressie (1976, 1979), Hall (1986), and Ekström (1999) are among the authors dealing with the statistics (2.5) for fixed $m \geq 1$ and eventually also for m slowly tending to ∞ when $n \to \infty$.

If $m > 1$, and in particular if $m \to \infty$, then the statistics (2.5) assign more weight to central spacings than to those in the tails. To avoid this, Hall (1986) proposed to wrap the observations $Y_1, Y_2, ..., Y_n$ around the circle of unit circumference and to define the m-spacings $Y_{m+j} - Y_j$ for arbitrary $1 \leq m \leq n$ and j as the distance between observations Y_j and Y_{j+m} on this circle. This leads either to the extension of the ordered observations $Y_1, \ldots, Y_n$ by the formula

$$Y_{n+j} = 1 + Y_j \quad \text{for} \quad j = 1, 2, ..., m \tag{2.6}$$

where the previous dummy observation $Y_0 = 0$ is suppressed and the other dummy observation $Y_{n+1} = 1$ is redefined in accordance with (2.6) by $Y_{n+1} = 1 + Y_1$, leading to the m-spacing $Y_{j+m} - Y_j$ to be equal to $1 + Y_{m+j-n} - Y_j$ if $n+1-m \leq j \leq n$, or to the extension by the alternative formula

$$Y_{n+j} = 1 + Y_{j-1} \quad \text{for} \quad j = 1, 2, \cdots, m \tag{2.7}$$

where the dummy observations $Y_0 = 0$ and $Y_{n+1} = 1$ are placed on the circle as well, resulting in the m-spacing $Y_{j+m} - Y_j$ to be defined as $1 + Y_{m+j-n-1} - Y_j$ for $n+2-m \leq j \leq n$. These extensions of the ordered observations Y_j beyond $j > n$ allow to add in

(2.5) the tail evidence missing there, namely by adding to the substituted averages (2.2) also the previously excluded terms. Depending on whether we use the extension (2.6) or the alternative extension (2.7), we get in this manner two different extensions of (2.5), namely

$$\hat{\boldsymbol{S}}_{\phi}^{(m)} = \sum_{j=0}^{n} \phi\left(\frac{n+1}{m}\left(Y_{j+m} - Y_j\right)\right) \quad \text{for } \ Y_{n+j} \ \text{ given by (2.6)} \tag{2.8}$$

and

$$\tilde{\boldsymbol{S}}_{\phi}^{(m)} = \sum_{j=1}^{n} \phi\left(\frac{n+1}{m}\left(Y_{j+m} - Y_j\right)\right) \quad \text{for } \ Y_{n+j} \ \text{ given by (2.7).} \tag{2.9}$$

The statistics from the class (2.8) were studied for example by Cressie (1978), Rao and Kuo (1984), Ekström (1999) and Misra and van der Meulen (2001), while those from the class (2.9) were investigated among others by Hall (1986) and Morales *et al.* (2003).

Hall (1984) studied the statistics

$$\tilde{R}_{\phi}^{(m)} = \sum_{j=1}^{n-m} h\left(n\left(Y_{j+m} - Y_j\right)\right) = \sum_{j=1}^{n-m} \phi\left(\frac{n}{m}\left(Y_{j+m} - Y_j\right)\right) \tag{2.10}$$

where $h(t) = \phi(t/m)$ for $m \geq 1$ fixed and variable $t > 0$. These statistics neglect the tail data $Y_1, Y_2, ..., Y_{m-1}$ and $Y_{n-1}, Y_{n-2}, ..., Y_{n-m+1}$.They are closely related to our divergence statistics $\tilde{\boldsymbol{T}}_{\phi}^{(m)}$ of (2.4),

$$\tilde{R}_{\phi}^{(m)} = \tilde{\boldsymbol{T}}_{\phi}^{(m)} - m\phi\left(\frac{n}{m}\left(Y_m + 1 - Y_{n-m}\right)\right). \tag{2.11}$$

We consider also the extensions

$$R_{\phi}^{(m)} = \sum_{j=0}^{n-m+1} \phi\left(\frac{n}{m}\left(Y_{j+m} - Y_j\right)\right) \tag{2.12}$$

of the Hall's statistics $\tilde{R}_{\phi}^{(m)}$,

$$R_{\phi}^{(m)} = \tilde{R}_{\phi}^{(m)} + \phi\left(\frac{n}{m}\left(Y_m\right)\right) + \phi\left(\frac{n}{m}\left(Y_{n+1} - Y_{n-m+1}\right)\right). \tag{2.13}$$

Thus we have introduced the collection

$$\mathcal{U}_{\phi}^{(m)} = \left\{\boldsymbol{T}_{\phi}^{(m)}, \tilde{\boldsymbol{T}}_{\phi}^{(m)}, \boldsymbol{S}_{\phi}^{(m)}, \hat{\boldsymbol{S}}_{\phi}^{(m)}, \tilde{\boldsymbol{S}}_{\phi}^{(m)}, R_{\phi}^{(m)}, \tilde{R}_{\phi}^{(m)}\right\} \tag{2.14}$$

of m-spacings-based statistics (c.f. (2.3), (2.4), (2.5), (2.8), (2.9), (2.12), (2.10)) where $\boldsymbol{T}_{\phi}^{(m)}, \tilde{\boldsymbol{T}}_{\phi}^{(m)}$ are spacings-based divergence statistics, $\tilde{R}_{\phi}^{(m)}, \boldsymbol{S}_{\phi}^{(m)}, \boldsymbol{S}_{0,\phi}^{(m)}, \boldsymbol{S}_{1,\phi}^{(m)}$ represent four types of spacings-based goodness-of-fit statistics considered in the literature and $R_{\phi}^{(m)}$ is an auxiliary statistic.

If $m = 1$, then both $\boldsymbol{T}_\phi^{(m)}$ and $\tilde{\boldsymbol{T}}_\phi^{(m)}$ reduce to

$$T_\phi = \sum_{j=0}^{n-2} \phi\left(n\left(Y_{j+1} - Y_j\right)\right) + \phi\left(n\left(Y_{n+1} - Y_{n-1}\right)\right)$$

previously introduced in (1.36) and both $\boldsymbol{S}_\phi^{(m)}$ and $\hat{\boldsymbol{S}}_\phi^{(m)}$ reduce to the statistic

$$S_\phi = \sum_{j=0}^{n} \phi\left(\left(n+1\right)\left(Y_{j+1} - Y_j\right)\right) \tag{2.15}$$

previously introduced in (1.42). However, $\tilde{\boldsymbol{S}}_\phi^{(m)}$ does so only if ϕ is linear. Indeed, if $m = 1$ then $\tilde{\boldsymbol{S}}_\phi^{(m)}$ reduces to

$$\tilde{S}_\phi = \sum_{j=}^{n-1} \phi\left(\left(n+1\right)\left(Y_{j+1} - Y_j\right)\right) + \phi\left(\left(n+1\right)\left(Y_1 + 1 - Y_n\right)\right) \tag{2.16}$$

which *a.s.* coincides with S_ϕ only if

$$\phi\left(\left(n+1\right)Y_1\right) + \phi\left(\left(n+1\right)\left(1 - Y_n\right)\right) = \phi\left(\left(n+1\right)\left(Y_1 + 1 - Y_n\right)\right) \quad \{\text{a.s.}$$

It is easy to see that this takes place only for linear ϕ. Finally, if $m = 1$, then $R_\phi^{(m)}$ reduces to R_ϕ introduced previously in (1.40) and $\tilde{R}_\phi^{(m)}$ reduces to

$$\tilde{R}_\phi = \sum_{j=1}^{n-1} \phi\left(n(Y_{j+1} - Y_j)\right) = R_\phi - \phi(nY_1) - \phi\left(n(1 - Y_n)\right). \tag{2.17}$$

Hence in the case $m = 1$ the class of statistics (2.14) reduces to

$$\mathcal{U}_\phi = \left\{T_\phi, \tilde{T}_\phi, S_\phi, \tilde{S}_\phi, R_\phi, \tilde{R}_\phi\right\}, \tag{2.18}$$

(c.f. (1.36), (1.39), (1.42), (2.16), (1.40), (2.17)) where $T_\phi, \tilde{T}_\phi$ are spacings-based divergence statistics, $\tilde{R}_\phi, S_\phi, \tilde{S}_\phi$ represent three types of simple-spacings-based goodness-of-fit statistics considered in the literature and R_ϕ is an auxiliary statistic.

3 Asymptotic equivalence

In this section we consider the class of functions $\boldsymbol{\Phi}$ introduced in Section 1 and its subclasses $\boldsymbol{\Phi}_{\mathrm{am}} \subset \boldsymbol{\Phi}$ and $\boldsymbol{\Phi}_{\mathrm{L}} \subset \boldsymbol{\Phi}$. The subclass $\boldsymbol{\Phi}_{\mathrm{am}}$ contains all functions $\phi \in \boldsymbol{\Phi}$ with an *additive-multiplicative structure*, i.e. those for which there exist functions ξ, $\eta : (0, \infty) \mapsto \mathbb{R}$ satisfying for all s, $t \in (0, \infty)$ the functional equation

$$\phi(st) = \xi(s)\,\phi(t) + \phi(s) + \eta(s)\,(t - 1). \tag{3.1}$$

The subclass $\boldsymbol{\Phi}_{\mathrm{L}}$ contains all functions $\phi \in \boldsymbol{\Phi}$ which are *Lipschitz*, i.e. those for which there exit constants $c_\phi > 0$ such that for all $s, t > 0$

$$|\phi(t) - \phi(s)| \ \le c_\phi\,|t - s|. \tag{3.2}$$

Example 3.1. The function $\phi(t) = (1-t)^2/t$, $t > 0$ belongs to $\mathbf{\Phi}$ and satisfies (3.1) for $\xi(t) = 1/t$ and $\eta(t) = t - 1/t$. Therefore it belongs to $\mathbf{\Phi}_{\mathrm{am}}$. Further, all functions $\phi = \phi_\alpha$, $\alpha \in \mathbb{R}$ of Example 1.1 belong to $\mathbf{\Phi}$ as well and satisfy relation (3.1) with

$$\xi(t) = \xi_\alpha(t) = t^\alpha \quad \text{and} \quad \eta(t) = \eta_\alpha(t) = \begin{cases} \frac{t^\alpha - t}{\alpha - 1} & \text{if} \quad \alpha \neq 1 \\ \lim\limits_{\alpha \to 1} \frac{t^\alpha - t}{\alpha - 1} = t \ln t & \text{if} \quad \alpha = 1. \end{cases} \tag{3.3}$$

In other words, it holds

$$\phi_\alpha(st) = s^\alpha \phi_\alpha(t) + \phi_\alpha(s) + (t-1) \cdot \begin{cases} \frac{s^\alpha - s}{\alpha - 1} & \text{if} \quad \alpha \neq 1 \\ s \ln s & \text{if} \quad \alpha = 1 \end{cases} \tag{3.4}$$

for all $\alpha \in \mathbb{R}$ and s, $t > 0$. This means that the functions $\phi = \phi_\alpha$ belong to $\mathbf{\Phi}_{\mathrm{am}}$.

Example 3.2. The family of smooth (analytic) functions defined for all $\alpha > 0$ by the formula

$$\phi_\alpha(t) = \left(1 - \exp\{-\alpha(t-1)^2 \big/ 2\}\right) \big/ \alpha, \quad t > 0 \tag{3.5}$$

belongs to $\mathbf{\Phi}$. The derivatives ϕ'_α are for every $\alpha > 0$ bounded as follows

$$\sup_{t>0} |\phi'_\alpha(t)| = \sup_{t>0} |t-1| \exp\{-\alpha(t-1)^2 \big/ 2\} = \exp\{-\alpha/2\}. \tag{3.6}$$

Therefore the functions ϕ_α are Lipschitz with the Lipschitz constants $c_{\phi_\alpha} = \exp\{-\alpha/2\}$, and thus they belong to the class $\mathbf{\Phi}_{\mathrm{L}}$. Further functions belonging to this class can be found in Example 1.2.

The next two auxiliary statements analyze properties of the functions from $\mathbf{\Phi}_{\mathrm{am}}$.

Lemma 3.1. The functions ξ and η appearing in (3.1) are continuous on $(0, \infty)$ and satisfy the relations

$$\xi(1) = 1 \quad \text{and} \quad \eta(1) = 0. \tag{3.7}$$

Proof. By assumption, $\phi(1) = 0$. Since $\phi(t)$ is convex in the neighborhood of $t = 1$ and strictly convex at $t = 1$, there exists a constant $c > 0$ such that

$$c\phi\left(1 + \frac{c}{2}\right) < \frac{c}{2}\left(\phi\left(1\right) + \phi\left(1 + c\right)\right) = \frac{c}{2}\phi\left(1 + c\right).$$

This implies that the continuous function

$$\psi(s) = \frac{c}{2}\phi\left(t\ \left(1 + c\right)\right) - c\phi\left(t\ \left(1 + \frac{c}{2}\right)\right)$$

satisfies the condition $\psi(1) > 0$. Define

$$\delta(s,t) = \phi(st) - \ \phi(t) + \phi(s) + \eta(s)\left(t - 1\right)$$

for ϕ, ξ, η appearing in (3.1). By (3.1),

$$0 = \frac{c}{2}\,\delta(s, 1+c) - c\,\delta\left(s, 1+\frac{c}{2}\right) = \psi(s) - \xi(s)\,\psi(1) + \frac{c}{2}\,\phi(s).$$

Since both $\phi(s)$ and $\psi(s)$ are continuous in $s > 0$, we see that $\xi(s)$ must be continuous too. Further, by putting $t = 2$ in (3.1) we see that $\eta(s)$ must be continuous when $\xi(s)$ is continuous. Finally, if we put $s = 1$ in (3.1) and use the assumption $\phi(1) = 0$, then we obtain

$$(\xi(1) - 1)\,\phi(t) + \eta(1)\,(t-1) = 0 \quad \text{for all } t > 0.$$

This contradicts the assumption $\phi''(1) > 0$, unless $\xi(1) = 1$ which implies also $\eta(1) = 0$. □

Lemma 3.2. Every $\phi \in \mathbf{\Phi}_{\mathrm{am}}$ is differentiable on the whole domain $(0, \infty)$, the corresponding functions ξ and η are differentiable at 1 and

$$\phi'(t) = \xi'(1)\,\frac{\phi(t)}{t} + \frac{\phi'(1)}{t} + \eta'(1)\,\frac{t-1}{t} \quad \text{for all } t > 0. \tag{3.8}$$

Proof. Putting $s = 1 + \varepsilon$ and

$$\xi^*(\varepsilon) = \frac{\xi(1+\varepsilon) - \xi(1)}{\varepsilon}, \quad \eta^*(\varepsilon) = \frac{\eta(1+\varepsilon) - \eta(1)}{\varepsilon}$$

we obtain from (3.1) for every $t > 0$ and ε close to 0

$$t\,\frac{\phi(t+\varepsilon t) - \phi(t)}{\varepsilon t} = \xi^*(\varepsilon)\,\phi(t) + \frac{\phi(1+\varepsilon) - \phi(1)}{\varepsilon} + \eta^*(\varepsilon)\,(t-1). \tag{3.9}$$

Since ϕ is differentiable in a neighborhood of 1, we have for t close to 1

$$\xi^*(\varepsilon)\,\phi(t) + \eta^*(\varepsilon)\,(t-1) = t\,\phi'(t) - \phi'(1) + o(\varepsilon) \quad \text{as } \varepsilon \to 0.$$

By assumptions, $\phi(t)$ is strictly convex at $t = 1$ and thus not linear in the neighborhood of $t = 1$. Therefore the last relation implies that the limits of $\xi^*(\varepsilon)$ and $\eta^*(\varepsilon)$ for $\varepsilon \to 0$ exist, that is,

$$\xi^*(\varepsilon) = \xi'(1) + o(\varepsilon) \quad \text{and} \quad \eta^*(\varepsilon) = \eta'(1) + o(\varepsilon) \quad \text{as } \varepsilon \to 0.$$

Now (3.8) follows from (3.9). □

In the remainder of this section the observations of our statistical model are assumed to be distributed on $(0, 1]$ in two possible ways:

(i) *under a fixed alternative,*

(ii) *under local alternatives.*

Case **(i)** means that the observations are distributed by a fixed distribution function $F \sim f$ with f positive and continuous on $[0,1]$. This means that in particular $f(0) > 0$.

Case **(ii)** means that the observations from samples of sizes $n = 1, 2, \ldots$ are distributed by distribution functions

$$F^{(n)}(x) = F_0(x) + \frac{L_n(x)}{\sqrt[4]{n}} = x + \frac{L_n(x)}{\sqrt[4]{n}} \tag{3.10}$$

on $[0,1]$, where the functions $L_n : \mathbb{R} \mapsto \mathbb{R}$ are continuously differentiable, with $L_n(0) = L_n(1) = 0$, and with derivatives $\ell_n(x) = L_n'(x)$ tending on $[0,1]$ to a continuously differentiable function $\ell : \mathbb{R} \mapsto \mathbb{R}$ uniformly in the sense that

$$\sup_{0 \le x \le 1} |\ell_n(x) - \ell(x)| = o(1) \quad \text{as } n \to \infty. \tag{3.11}$$

The two cases **(i)** and **(ii)** are not mutually exclusive: their conjunction is "under the hypothesis $\mathcal{H}_0$ " where $F(x) = F_0(x)$, $f(x) = f_0(x) = \boldsymbol{I}_{[0,1]}(x)$ and $L_n(x) \equiv 0$ on $\mathbb{R}$ for all n. This means that the asymptotic results obtained under local alternatives for $\ell(x)$ of (3.11) being identically equal to 0 must coincide with the results obtained under the fixed alternative for $F(x) = F_0(x)$.

Lemma 3.3. Let $m \ge 1$ be fixed. Independently of whether the order statistics are extended by the rule (2.6) or (2.7), the tails

$$\Delta_{\phi,n}^{(m)} = \frac{1}{n} \sum_{i=n-m+1}^{n} \phi\left(\frac{n}{m}\left(Y_{i+m} - Y_i\right)\right) \tag{3.12}$$

and

$$\delta_{\phi,n}^{(m)} = \frac{1}{n} \sum_{i=0}^{m} \phi\left(\frac{n}{m}\left(Y_{i+m} - Y_i\right)\right) \tag{3.13}$$

of all disparity statistics under consideration are in both cases **(i)** and **(ii)** of the asymptotic order $O_p(1)$ as $n \to \infty$.

Proof. We shall prove $\Delta_{\phi,n}^{(m)} = O_p(1)$ in the case **(i)** under the extension rule (2.7). Modification of this proof for the other case, and the other rule and/or tail, is easy. Since m is fixed, it suffices to prove for every $i = n - m + r$ with $1 \le r \le m$

$$\phi\left(\frac{n}{m}(Y_{i+m} - Y_i)\right) = O_p(1). \tag{3.14}$$

It is known that

$$Y_{i+m} - Y_i = Y_r + 1 - Y_{n-m+r} = F^{-1}(W_r) + 1 - F^{-1}(W_{n-m+r}) \tag{3.15}$$

where

$$W_s = \frac{Z_1 + \cdots + Z_s}{Z_1 + \cdots + Z_{n+1}}, \quad 1 \le s \le n$$

and Z_i are independent standard exponential random variables (see, e. g. Hall (1984), p. 208). Since for all fixed integers r and s

$$W_r = o_p(1), \quad W_{n-s} = 1 + o_p(1) \quad \text{and} \quad nW_s = O_p(1),$$

it follows from the assumption $f(x) > 0$ for all $0 \le x \le 1$ and from the law of large numbers for the standard exponential Z_i

$$nF^{-1}(W_r) = nW_r \frac{F^{-1}(W_r)}{W_r} = O_p(1),$$

and similarly

$$n\left(1 - F^{-1}(W_{n-m+r})\right) = O_p(1).$$

Therefore (3.15) implies

$$\frac{n}{m}(Y_{i+m} - Y_i) = O_p(1)$$

and the desired relation (3.14) follows from here. □

The theorem below demonstrates that if $\phi \in \mathbf{\Phi}$ a convex or disparity function then in spite of that the statistics S_ϕ, $\tilde{S}_\phi$, R_ϕ and $\tilde{R}_\phi$ are not ϕ-divergences or ϕ-disparities of the hypothetical and empirical distributions F_0 and F_n, they still share the most important statistical properties with the statistics T_ϕ and $\tilde{T}_\phi$, which are divergences or disparities. Therefore this theorem provides a key argument for the thesis of the present paper formulated in Section 2, that the spacings-based goodness-of-fit statistics considered in the previous literature actually measure a disparity between the hypothetical and empirical distributions F_0 and F_n, although this was possibly not so intended by the various authors. In other words, this theorem demonstrates that the small modifications distinguishing these statistics from one another are asymptotically negligible, thus opening a possibility to develop a unified asymptotic theory for the whole class of statistics $\mathcal{U}_\phi^{(m)}$ defined in (2.14). Such a theory will presented in a subsequent paper.

Theorem 3.1. Consider the set of disparity statistics $\mathcal{U}_\phi^{(m)}$ defined in (2.14). If $\phi \in \mathbf{\Phi}$ then under both fixed and local alternatives the statistics $U_\phi^{(m)} \in \{R_\phi^{(m)}, \boldsymbol{T}_\phi^{(m)}, \tilde{\boldsymbol{T}}_\phi^{(m)}\} \subset \mathcal{U}_\phi^{(m)}$ satisfy the relation

$$\frac{U_\phi^{(m)} - \tilde{R}_\phi^{(m)}}{a_n} = o_p(1) \quad \text{as } n \to \infty. \tag{3.16}$$

for any positive normalization sequence a_n increasing to infinity. If $\phi \in \mathbf{\Phi}_{\text{am}} \cup \mathbf{\Phi}_{\text{L}}$ then this relation can be extended also to the statistics $U_\phi^{(m)} \in \{\boldsymbol{S}_\phi^{(m)}, \hat{\boldsymbol{S}}_\phi^{(m)}, \tilde{\boldsymbol{S}}_\phi^{(m)}\} \subset \mathcal{U}_\phi^{(m)}$. Hence in this case all statistics from $\mathcal{U}_\phi^{(m)}$ are mutually asymptotically equivalent of the order of $o_p(a_n)$.

Proof. (I) If $\phi \in \mathbf{\Phi}$ then, by inspecting the definitions of $R_\phi^{(m)}, \boldsymbol{T}_\phi^{(m)}$ and $\tilde{\boldsymbol{T}}_\phi^{(m)}$ we see that the difference between any $U_\phi^{(m)} \in \{R_\phi^{(m)}, \boldsymbol{T}_\phi^{(m)}, \tilde{\boldsymbol{T}}_\phi^{(m)}\}$ and $\tilde{R}_\phi^{(m)}$ is of the form of $\Delta_{\phi,n}^{(m)}$ or $\delta_{\phi,n}^{(m)}$ considered in Lemma 3.3. Thus in this case the desired relation (3.16) follows from Lemma 3.3.

(II) Let us now consider the second assertion for $\phi \in \mathbf{\Phi}_{\mathrm{am}}$. We shall prove for $U_\phi^{(m)} \in \{\boldsymbol{S}_\phi^{(m)}, \hat{\boldsymbol{S}}_\phi^{(m)}\}$ and $\tilde{\boldsymbol{S}}_\phi^{(m)}$ the relations

$$U_\phi^{(m)} - R_\phi^{(m)} = \varepsilon_n \, R_\phi^{(m)} + \delta_n \quad \text{and} \quad \tilde{\boldsymbol{S}}_\phi^{(m)} - \tilde{\boldsymbol{T}}_\phi^{(m)} = \varepsilon_n \, \tilde{\boldsymbol{T}}_\phi^{(m)} + \delta_n \tag{3.17}$$

for some numerical sequences ε_n, δ_n with the asymptotic properties $\varepsilon_n = o(1)$ and $\delta_n = O(1)$ as $n \to \infty$. Let us start with the additive-multiplicative decomposition (3.1) from which we get for any $p > 0$

$$\phi((n+1)\,p) = \xi\left(\frac{n+1}{n}\right)\phi(np) + \phi\left(\frac{n+1}{n}\right) + \eta\left(\frac{n+1}{n}\right)(np-1).$$

Hence

$$\phi((n+1)\,p) - \phi(np) = \varepsilon_n \phi(np) + \phi\left(\frac{n+1}{n}\right) + \eta\left(\frac{n+1}{n}\right)(np-1) \tag{3.18}$$

where $\varepsilon_n = \xi((n+1)/n) - 1 = o(1)$ as $n \to \infty$ by Lemma 3.1. Replacing p by the probabilities $p_{0j} = Y_{j+m} - Y_j$ figuring in the definitions of $\boldsymbol{S}_\phi^{(m)}$ and $R_\phi^{(m)}$ in (2.5) and (2.12), and summing over $1 \le j \le n - m + 1$, we get from (3.18) the relation

$$\boldsymbol{S}_\phi^{(m)} - R_\phi^{(m)} = \varepsilon_n \, R_\phi^{(m)} + \delta_n$$

for

$$\begin{aligned} \delta_n &= (n+1)\,\phi\left(\frac{n+1}{n}\right) - \eta\left(\frac{n+1}{n}\right) \\ &= \frac{n+1}{n}\,\frac{\phi\left(1+\frac{1}{n}\right) - \phi(1)}{\frac{1}{n}} - \eta\left(\frac{n+1}{n}\right). \end{aligned}$$

By Lemma 3.1,

$$\delta_n = \phi'(1) + o(1) \quad \text{as } n \to \infty.$$

This completes the proof of the first relation in (3.17) with $U_\phi^{(m)} = \boldsymbol{S}_\phi^{(m)}$. The proofs of the same relation with $U_\phi^{(m)} = \hat{\boldsymbol{S}}_\phi^{(m)}$ and of the second relation of (3.17) follow in a similar manner from (3.18).

(III) As the last step, consider the second assertion for $\phi \in \mathbf{\Phi}_{\mathrm{L}}$. Put

$$\mathcal{D}_n = \sum_{i=0}^{n-m-1} \left[\phi\left(\frac{n+1}{m}\,(Y_{i+m} - Y_i)\right) - \phi\left(\frac{n}{m}\,(Y_{i+m} - Y_i)\right)\right] \tag{3.19}$$

and let a_n be a positive sequence increasing to infinity. Then

$$\begin{aligned}
\boldsymbol{S}_\phi^{(m)} - \boldsymbol{T}_\phi^{(m)} &= \mathcal{D}_n - W_n^{(m)} && \text{c.f. (??) and (??)},\\
\hat{\boldsymbol{S}}_\phi^{(m)} - \boldsymbol{T}_\phi^{(m)} &= \mathcal{D}_n + Z_n^{(m)} - W_n^{(m)} && \text{c.f. (??) and (2.8)},\\
\tilde{\boldsymbol{S}}_\phi^{(m)} - \boldsymbol{T}_\phi^{(m)} &= \mathcal{D}_n + Z_n^{(m)} - \tilde{W}_n^{(m)} && \text{c.f. () and ()},
\end{aligned}$$

where

$$\begin{aligned}
W_n^{(m)} &= m\phi\left(\tfrac{n}{m}\left(1 - Y_{n-m}\right)\right)\\
Z_n^{(m)} &= \textstyle\sum_{j=n-m}^{n} \phi\left(\tfrac{n+1}{m}\left(Y_{j+m} - Y_i\right)\right)\\
\tilde{W}_n^{(m)} &= W_n^{(m)} - \phi\left(\tfrac{n+1}{m} Y_m\right)
\end{aligned}$$

and the arguments Y_{n+j} of $Z_n^{(m)}$ in the formula for $\hat{\boldsymbol{S}}_\phi^{(m)} - \boldsymbol{T}_\phi^{(m)}$ are given by (2.6) while in the formula for $\tilde{\boldsymbol{S}}_\phi^{(m)} - \boldsymbol{T}_\phi^{(m)}$ they are given by (2.7). But one can deduce from Lemma 3.3 that $W_n^{(m)} = O_p(1)$, $\tilde{W}_n^{(m)} = O_p(1)$ and $Z_n^{(m)} = O_p(1)$ irrespectively of whether the arguments Y_{n+j} of $Z_n^{(m)}$ are given by (2.6) or (2.7). Thus it suffices to prove

$$|\mathcal{D}_n| = o(a_n). \tag{3.20}$$

Since ϕ is by assumption Lipschitz function with a Lipschitz constatnt $c_\phi > 0$, (3.19) implies

$$|\mathcal{D}_n| \le \frac{c_\phi}{m} \sum_{i=0}^{n-m-1} |\, Y_{i+m} - Y_i \,| = \frac{c_\phi}{m} \sum_{i=0}^{n-m-1} (Y_{i+m} - Y_i)$$

where the equality follows from the monotonicity $Y_{i+m} \ge Y_i$ of the order statistics appearing in the last sum. Thus (3.20) follows from the obvious inequalities

$$\sum_{i=0}^{n-m-1} (Y_{i+m} - Y_i) \le Y_{n-m-1} - Y_0 \le 1.$$

□

Next we precise the result of Theorem 3.1 for the important special case of simple spacings where, as was argued at the end of Section 2, $\mathcal{U}_\phi^{(m)}$ reduces to the class $\mathcal{U}_\phi$ given by (2.18).

Corollary 3.1. Consider the set of disparity statistics $\mathcal{U}_\phi$ defined in (2.18). If $\phi \in \boldsymbol{\Phi}$ then under both fixed and local alternatives the statistics $U_\phi \in \{R_\phi, T_\phi, \tilde{T}_\phi\}$ satisfy the relation

$$U_\phi - \tilde{R}_\phi = o_p(a_n) \quad \text{as } n \to \infty. \tag{3.21}$$

for any normalization sequence a_n increasing to infinity. If $\phi \in \boldsymbol{\Phi}_{\mathrm{am}} \cup \boldsymbol{\Phi}_{\mathrm{L}}$ then this relation can also be extended to the statistics $U_\phi \in \{S_\phi, \tilde{S}_\phi\}$. Hence in this case all statistics from $\mathcal{U}_\phi$ are mutually asymptotically equivalent of the order of $o_p(a_n)$.

4 Acknowledgements

The research for this paper was supported by GAČR grant 102/07/1131, MŠMT grant 1M 0572, and K. U. Leuven project GOA/98/06.

5 References

Cressie, N. (1976) On the logarithms of high-order spacings. *Biometrika*, **63**, 345–355.

Cressie, N. (1978) Power results for tests based on high-order gaps. *Biometrika*, **65**, 214–218.

Cressie, N. (1979) An optimal statistic based on higher order gaps. *Biometrika*, **66**, 619–627.

Csiszár, I. (1963) Informationstheoretische Ungleichung und ihre Anwendung auf den Beweis der Ergodizität von Markoffschen Ketten. *Publ. Math. Inst. Hungarian Acad. Sci., Ser. A*, **8**, 85–108.

Del Pino, G. E. (1979) On the asymptotic distribution of k-spacings with applications to goodness of fit tests. *Ann. Statist.*, **7**, 1058–1065.

Dudewicz, E. J. and van der Meulen, E. C. (1981) Entropy-based tests of uniformity. *J. Amer. Statist. Assoc.*, **76**, 967–974.

Ekström, M. (1999) Strong limit theorems for sums of logarithms of high order spacings. *Statistics*, **33**, 153–169.

Greenwood, M. (1946) The statistical study of infectious diseases. *J. Roy. Statist. Soc., Ser. A*, **109**, 85–110.

Györfi, L. and Vajda, I. (2002) Asymptotic distribution for the goodnesss-of-fit statistics in a sequence of multinomial models. *Statist. Probab. Lett.,* **56**, 57-67.

Hall, P. (1984) Limit theorems for sums of general functions of m-spacings. *Math. Proc. Cambridge Philos. Soc.*, **96**, 517–532.

Hall, P. (1986) On powerful distributional tests based on sample spacings. *J. Multivariate Anal.*, **19**, 201–224.

Hampel, F. R., Ronchetti E. M., Rousseuw P. J. and Stahel W. A. (1986) *Robust Statistics: The Approach Based on Influence Functions.* New York: Wiley.

Jammalamadaka, S. R., Zhou, X. and Tiwari, R. C. (1989) Asymptotic efficiency of spacings tests for goodness of fit. *Metrika*, **36**, 355-377.

Jurečková, J. and Sen, P. (1996) *Robust Statistical Procedures: Asymptotics and Interrelations.* New York: Wiley.

Kimball, B. F. (1947) Some basic theorems for developing tests of fit for the case of nonparametric probability distribution functions. *Ann. Math. Statist.*, **18**, 540–548.

Liese, F. and Vajda, I. (1987) *Convex Statistical Distances.* Leipzig: Teubner.

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

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

Menéndez M. L., Morales, D., Pardo, L. and Vajda, I. (1998) Two approaches to grouping of data and related disparity statistics. *Commmun. in Statist. - Theory and Methods,* **27**, 609-633.

Misra, N. and van der Meulen, E. C. (2001) A new test of uniformity based on overlapping simple spacings. *Comm. Statist. Theory Methods*, **30**, 1435–1470.

Morales, D. Pardo, M. C. and Vajda, I.(2003) Limit laws for disparities of spacings. *J. Nonparametr. Statist.*, **15**, 325–342.

Moran, P. A. P. (1951) The random division of an interval –Part II. *J. Roy. Statist. Soc. Ser. B*, **13**, 147–150.

Pyke, R. (1965) Spacings. *J. Roy. Statist. Soc. Ser. B*, **27**, 395–449.

Rao, J. S. and Kuo, M. (1984) Asymptotic results on Greenwood statistic and some of its generalizations. *J. Roy. Statist. Soc. Ser. B*, **46**, 228–237.

Read, T. R. C. and Cressie, N. (1988) *Goodness of Fit Statistics for Discrete Multivariate Data.* Berlin: Springer.

Shao, Y. and Hahn, M. G. (1995) Limit theorems for the logarithm of sample spacings. *Statist. Probab. Lett.*, **24**, 121–132.

Vajda, I. and van der Meulen, M. C. (2006a) *Goodness-of-fit Tests Based on Observations Quantized by Hypothetical and Empirical Quantiles.* Res. Rep. 2160, February 2006, Institute of Information Thgeory and Automation, Prague.

Vajda, I. and van der Meulen, M. C. (2006b) On estimation and testing based on m-spacongs. *Prague Stochastics* (eds. Hušková and Janžura), Prague: Matfyzpress.

Vajda, I. (2007) Asymptotic comparisons of divergence-base goodness-of-fit statistics. *Publ. Statist. Inst. Univ. Prais,* **51**,49-66.