Akademie věd České republiky
Ústav teorie informace a automatizace

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

RESEARCH REPORT

P. Boček, I. Vajda and E. C. van der Meulen:

Goodness-of-Fit Disparity Statistics Obtained by Hypothetical and Empirical Quantizations

No. 2292 Srpen 2010

ÚTIA AV ČR, P. O. Box 18, 182 08 Prague, Czech Republic
E-mail: utia@utia.cas.cz

Goodness-of-Fit Disparity Statistics
Obtained by Hypothetical and Empirical Quantizations

P. Boček, I. Vajda and E. C. van der Meulen

Abstract

Goodness-of-fit disparity statistics are defined as appropriately scaled ϕ-disparities or ϕ-divergences of quantized hypothetical and empirical distributions. It is shown that the classical Pearson-type statistics are obtained if we quantize by means of hypothetical percentiles, and that new spacings-based disparity statistics are obtained if we quantize by means of empirical percentiles. The main attention is paid to the asymptotic properties of the new disparity statistics and their comparisons with the spacings-based statistics known from the literature. First the asymptotic equivalence between them is proved, and then for the new statistics a general law of large numbers is proved, as well as an asymptotic normality theorem both under local and fixed alternatives. Special attention is devoted to the limit laws for the power divergence statistics of orders $\alpha \in \mathbb{R}$. Parameters of these laws are evaluated for $\alpha \in (-1, \infty)$ in a closed form and their continuity in α on the subinterval $(-1/2, \infty)$ is proved. These closed form expressions are used to compare local asymptotic powers of the tests based on these statistics, which allows to extend previous asymptotic optimality results to the class of power divergence statistics. Tables of values of the asymptotic parameters are presented for selected representative orders of $\alpha > -1/2$.

Key words: asymptotic normality, asymptotic optimality, consistency, goodness-of-fit, power divergences, spacings, ϕ-disparities, ϕ-divergences.

1 Data and their statistical models

In this chapter we consider the explanation of *observed data* $x_1, x_2, ..., x_n$ statistically as a sequence of independent outcomes from a statistical model. Our aim is to review and extend the criteria of goodness-of-fit of the model and data, and to study their properties and applications in decisions about the acceptability of concrete models for concrete data.

Let us start with the example of data first studied by Pearson (1894), which represent measurements of the ratio of the forehead to body of $n = 1000$ crabs. Pearson partitioned the original domain of measurements $(a, b) = (0.5795, 0.6995)$ into intervals of equidistant size 0.004 and counted the frequency for each interval. Table 1.1 gives the measurement values v_j represented by midpoints of the intervals and the corresponding frequencies φ_j for $1 \leq j \leq 30$.

Table 1.1 Ratio of the forehead to body of 1000 crabs

Value	Freq.	Value	Freq.	Value	Freq.	Value	Freq.	Value	Freq.
0.5815	1	0.5855	3	0.5895	5	0.5935	2	0.5975	7
0.6015	10	0.6055	13	0.6095	19	0.6135	20	0.6175	25
0.6215	40	0.6255	31	0.6295	60	0.6335	62	0.6375	54
0.6415	74	0.6455	84	0.6495	86	0.6535	96	0.6575	85
0.6615	75	0.6655	47	0.6695	43	0.6735	24	0.6775	19
0.6815	9	0.6855	5	0.6895	0	0.6935	1	0.6975	0

As second example we use the data studied recently by Ning, Gao and Dudewicz (2008) which are results of measurements of cadmium concentrations in the kidney cortex of $n = 43$ horses. These are presented in Table 1.2 below.

Table 1.2 Amounts of cadmium in horse kidneys

11.9	16.7	23.4	25.8	25.9	27.5	28.5	31.1	32.5	35.4	38.3
38.5	41.8	42.9	50.7	52.3	52.5	52.6	54.5	54.7	56.6	56.7
58.0	60.8	61.8	62.3	62.5	62.6	63.0	67.7	68.5	69.7	73.1
76.0	76.9	77.7	78.2	80.3	93.7	101.0	104.5	105.4	107.0	-

Statistical models are probability distributions on data spaces $\mathcal{X}$. A general probability distribution is specified by a **probability measure** (briefly, **p.m.**) P defined on appropriate subsets of $\mathcal{X}$. The data spaces of the examples in Tables 1.1 and 1.2 are the halflines $(0, \infty)$. Throughout this chapter we restrict ourselves to real valued data with the data space being an interval $\mathcal{X} = (a, b) \subseteq \mathbb{R}$. Then each probability measure P is uniquely specified by the **distribution function** (briefly, **d.f.**)

$$F(x) = P((-\infty, x]), \quad x \in \mathbb{R}$$

We restrict ourselves to statistical models with increasing and continuously differentiable d.f.'s $F(x)$ on the data space . These are uniquely specified by the positive continuous **probability density functions** (briefly, **p.d.f.**'s)

$$f(x) = \frac{\mathrm{d}F(x)}{\mathrm{d}x} \quad x \in (a, b)$$

as well as by the increasing **percentile functions** (briefly, **p.f.**'s)

$$Q(y) = F^{-1}(y) \ \ \text{(i.e. } x \in [a, b] \text{ such that } F(x) = y), \quad y \in [0, 1]$$

where

$$Q(0) = a \quad \text{and} \quad Q(1) = b$$

due to the assumption that $F(x)$ is increasing on (a, b).Therefore the observed data $x_1, x_2, ..., x_n$ are interpreted as realizations of independent copies $X_1, X_2, ..., X_n$ of a random variable X which can be specified equivalently by a p.m. P on $\mathcal{X}$ or a d.f. $F(x)$ on $\mathbb{R}$ or a p.d.f. $f(x)$ on the data space (a, b), or by a p.f. $Q(y)$ on the percentile space $[0, 1]$.

The data $x_1, x_2, ..., x_n$ themselves can be represented by the so-called **empirical probability measure** (briefly, **e.p.m.**)

$$P_n = \frac{1}{n} \sum_{i=1}^{n} \delta_{x_i} \tag{1.1}$$

on the data space $\mathcal{X} = (a, b)$ where δ denotes the Dirac probability measure, or by the related **empirical distribution function** (briefly, **e.d.f.**)

$$F_n(x) = P_n((-\infty, x)) = \frac{1}{n} \sum_{i=1}^{n} \boldsymbol{I}(x \geq x_i) \tag{1.2}$$

on $\mathbb{R}$ where $\boldsymbol{I}$ denotes the indicator function. Both these representations are unique up to the ordering of the data. In other words, P_n as well as F_n are one-to-one related to the order statistic $(x_{n:1}, x_{n:2}, ..., x_{n:n})$ of the data vector $(x_1, x_2, ..., x_n)$. The loss of order means no loss of statistical information, because the order statistics are known to be statistically sufficient for the statistical models independent identically distributed (i.i.d.) observations (cf. Lehmann and Romano (2005)).

An alternative form of presentation of the data $x_1, x_2, ..., x_n$ is by the histogram $f_n(x)$ on the data space (a, b). This is the density of the restriction of the e.p.m. P_n on the algebra generated by the $k + 1$ intervals obtained by slicing (a, b) at the cutpoints

$$a = c_0 < c_1 < \ ... \ < c_k < c_{k+1} = b.$$

Then for every $x \in [c_j, c_{j+1})$

$$f_n(x) = \#x_i \in [c_j, c_{j+1}).$$

For example, the frequencies φ_j of Table 1.1 define the histogram

$$f_n(x) = \varphi_j \quad \text{for} \quad x \in [c_j, c_{j+1}) \tag{1.3}$$

where

$$c_j = 0.5795 + 0.004 \cdot j \quad \text{for} \;\; 0 \le j \le 30 \tag{1.4}$$

are uniform cutpoints of the interval $(a, b) = (0.5795, 0.6995)$. The histogram form of presentation of data is statistically sufficient in exceptional situations only.

Figure 1.1 contrasts the histogram (1.3) presenting the data from Table 1.1 with the p.d.f. $f(x)$ of the maximum likely normal model N(μ,σ^2) of these data where

$$\mu = \frac{1}{1000}\sum_{i=1}^{1000} x_i = 0.6447 \;\; \text{and} \;\; \sigma^2 = \frac{1}{999}\sum_{i=1}^{1000}(x_i - \mu)^2 = 0.00036. \tag{1.5}$$

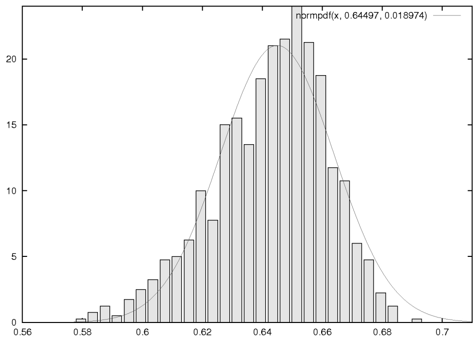
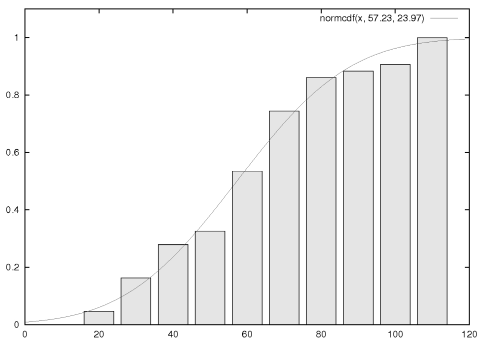


Figure 1.2 shows the e.d.f. $F_n(x)$ of the data from Table 1.2 together with the d.f. $F(x)$ of the normal model N(μ,σ^2) with sample mean and variance

$$\mu = \frac{1}{43}\sum_{i=1}^{43} x_i = 57.2326 \;\; \text{and} \;\; \sigma^2 = \frac{1}{42}\sum_{i=1}^{43}(x_i - \mu)^2 = 574.98. \tag{1.6}$$



In this chapter we deal with quality criteria for fitting various possible statistical models $F(x)$ to the data represented by e.d.f.'s $F_n(x)$ including the asymptotic properties of these criteria for $n \to \infty$. The basic concepts and notations introduced in this section are used throughout all what follows below.

2 Assessment of goodness-of-fit

Intuitively one can expect that all numerical goodness-of-fit criteria will be measures of distance, divergence or disparity between on the one hand the empirical reality represented by a data-based p.m. P or d.f. F, e.g. the e.p.m. P_n or e.d.f. F_n, and on the other the hypothetical statistical model given by a p.m. P_0 or a d.f. F_0. In this chapter the terms distance, divergence and disparity have a specific mathematical meaning which is specified in the definitions below, where we deal primarily with p.m.'s P, P_0 rather than with d.f.'s F, F_0.

Definition 2.1 (**i**) By a *distance* $D(P, P_0)$ of p.m.'s P, P_0 we mean a standard mathematical metric distance, taking values in the interval $[0, \infty)$, which is reflexive (i.e. $D(P, P_0) = 0$ if and only if $P = P_0$), symmetric (i.e. $D(P, P_0) = D(P_0, P)$) and satisfies the triangle inequality (i.e. $D(P, P_0) \leq D(P, P_1) + D(P_1, P_0)$ for all p.m.'s P, P_0, P_1).

(**ii**) A **divergence** (more precisely, an **information-theoretic divergence**) $D(P, P_0)$ is a reflexive functional taking values in the interval $[0, \infty)$ and satisfying the information processing law. To formulate this law, let $T : \mathcal{X} \to \mathcal{Y}$ be a mapping describing the results of processing the data x from the observation space $\mathcal{X}$ towards another space $\mathcal{Y}$, with $y = T(x)$. The information processing law says that no processing rule T can increase the divergence, in symbols

$$D(PT^{-1}, P_0T^{-1}) \leq D(P, P_0), \tag{2.1}$$

with the equality being valid if and only if T is statistically sufficient for P, P_0.
(**iii**) A **disparity** of p.m.'s P, P_0 is a nonnegative functional $D(P, P_0)$ which is reflexive in the above sense.

Convention 2.1 Throughout the chapter we denote by P, P_0 an arbitrary pair of probability measures on a general observation space $\mathcal{X}$. They will be represented by means of their p.d.f.'s

$$p = \frac{\mathrm{d}P}{\mathrm{d}\lambda} \quad \text{and} \quad p_0 = \frac{\mathrm{d}P_0}{\mathrm{d}\lambda} \tag{2.2}$$

with respect to (w.r.t.) a dominating measure λ on $\mathcal{X}$ (in symbols, $P \sim p, P_0 \sim p_0$). The only restriction imposed on the hypothetical model P_0 is the positivity of p_0 almost everywhere w.r.t. λ.

Continuous case. If $\mathcal{X} = (a, b) \subseteq \mathbb{R}$ then the p.m.'s (P, P_0) are in a one-to-one manner represented by the d.f.'s (F, F_0), and for absolutely continuous F, F_0 and Lebesque measure λ it holds that

$$p = f \quad \text{and} \quad p_0 = f_0 \tag{2.3}$$

where (f, f_0) are the usual p.d.f.'s of (F, F_0) (in symbols, $F \sim f, F_0 \sim f_0$). As stated above, we assume $f_0 > 0$ on (a, b).

Discrete case. If $\mathcal{X} = (1, 2, ..., k)$, then the densities (p, p_0) of (P, P_0) w.r.t. the counting measure $\lambda(1) = \lambda(2) = ... = \lambda(k) = 1$ reduce to the stochastic vectors

$$\boldsymbol{p} = (p_j \equiv P(j) : 1 \leq j \leq k), \quad \boldsymbol{p}_0 = (p_{0j} \equiv P_0(j) : 1 \leq j \leq k), \tag{2.4}$$

(in symbols, $P \sim \boldsymbol{p}, P_0 \sim \boldsymbol{p}_0$). As stated above, we assume $p_{0j} > 0$ for all $1 \leq j \leq k$.

2.1 Special distances, divergences and disparities

(a) The L_1**-distance**

$$L_1(P, P_0) = \int |p - p_0| \mathrm{d}\lambda \quad \text{for } P \sim p, P_0 \sim p_0 \tag{2.5}$$

is an example of a distance on the class of all p.m.'s which satisfies only partly the information processing law. Namely it satisfies inequality (2.1) but fails to satisfy the necessary condition for equality. To see this it suffices to consider the discrete p.m.'s $P \sim \boldsymbol{p}$ and $P_0 \sim \boldsymbol{p}_0$ given by (2.4), where

$$L_1(P, P_0) = L_1(\boldsymbol{p}, \boldsymbol{p}_0) = ||\boldsymbol{p} - \boldsymbol{p}_0||_1 = \sum_{j=1}^{k} |p_j - p_{0j}|. \tag{2.6}$$

Example 2.1 The binary coding $T : \mathcal{X} \to \mathcal{Y} = \{1, 0\}$ of the ternary observations $x \in \mathcal{X} = \{1, 2, 3\}$ defined by

$$T(1) = 1 \quad \text{and} \quad T(2) \equiv T(3) = 0 \tag{2.7}$$

significantly reduces the information for discrimination between the discrete models

$$P = (1/10,\ 5/10,\ 4/10) \quad \text{and} \quad P_0 = (9/10, 1/10, 0).$$

Indeed, the discrimination rule

$$\delta(1) = P_0 \quad \text{and} \quad \delta(2) = \delta(3) = P$$

based on the original uncoded observations from $\mathcal{X} = \{1, 2, 3\}$ is errorless if $x = 3$ and the discrimination errors for the remaining observations x take place with the probability 1/10. On the other hand, arbitrary discrimination rule $\delta : \mathcal{Y} \to \{P, P_0\}$ based on the encoded data from $\mathcal{Y} = \{1, 0\}$ admits discrimination errors with the probability 9/10. This reduction of discernibility is caused by the loss of information due to the coding which is evidently not a statistically sufficient transformation. However, this evidence is not reflected by the L_1- distance which remains unaffected by the coding, namely

$$L_1((1/10,\ 5/10,\ 4/10), (9/10, 1/10, 0)) = L_1((1/10,\ 9/10), (9/10, 1/10)) = 8/5.$$

Note that, nevertheless, the L_1-distance $L_1(f_n, f)$ is widely used as a goodness-of-fit criterion between the model p.d.f.'s f and histogram-like representations f_n of the observed data, since being introduced to the nonparametric statistics by Devroye and Györfi (1985).

(b) The L_2-**distance**

$$L_2(P, P_0) = \left(\int (p - p_0)^2 \mathrm{d}\lambda\right)^{1/2} \quad \text{for } P \sim p, P_0 \sim p_0$$

does not satisfy the information processing law in the sense that processing of the observations can increase the L_2-distance between the models P, P_0. To this end it suffices to consider the discrete p.m.'s $P \sim \boldsymbol{p}$ and $P_0 \sim \boldsymbol{p}_0$ given by (2.4) where

$$L_2(P, P_0) = L_2(\boldsymbol{p}, \boldsymbol{p}_0) = ||\boldsymbol{p} - \boldsymbol{p}_0||_2 = \left(\sum_{j=1}^{k} (p_j - p_{0j})^2\right)^{1/2}. \tag{2.8}$$

The mentioned violation of the information processing law can be verified by taking $k = 3$ and the models $P \sim \boldsymbol{p} = (0,\ 1/2,\ 1/2)$ and $P_0 \sim \boldsymbol{p}_0 = (1/2,\ 1/4,\ 1/4)$ on the observation space $\mathcal{X} = \{1, 2, 3\}$. Applying the coding (2.7) to the observations $x \in \mathcal{X}$ we obtain

$$L_2(PT^{-1}, P_0T^{-1}) = \sqrt{1/2} > L_2(P, P_0) = \sqrt{3/8}.$$

This flaw of the L_2- distance justifies the preference of the above-mentioned L_1- distance method over the L_2-method and underlines the importance of the L_1- method in statistical research.

(c) The **Kolmogorov distance**

$$K(F, F_0) = \sup_{x \in \mathbb{R}} |F(x) - F_0(x)| \tag{2.9}$$

introduced by Kolmogorov (1933) is the distance in the above stated metric sense and defines the well known Kolmogorov-Smirnov goodness-of-fit statistic $T_n = \sqrt{n} K(F_n, F_0)$ (Smirnov (1948); for more see

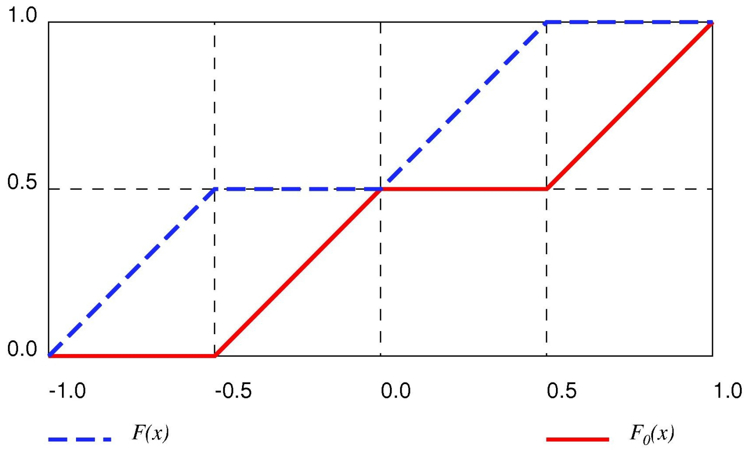


Figure 2.1 D.f.'s F, F_0 for the p.d.f.'s of (2.10) and (2.11).

Lehmann and Romano (2005)). Applicability of the Kolmogorov distance is restricted by the fact that it contradicts the information processing law in a similar way as the L_2-distance. A simple example where

$$K(FT^{-1}, F_0T^{-1}) = 1 > K(F, F_0) = 1/2$$

is obtained by using the p.d.f.'s

$$f(x) = \boldsymbol{I}(-1 < x < -1/2) + \boldsymbol{I}(0 < x < 1/2), \tag{2.10}$$
$$f_0(x) = \boldsymbol{I}(-1/2 < x < 0) + \boldsymbol{I}(1/2 < x < 1) \tag{2.11}$$

and the data processing formula

$$T(x) = x + \frac{1}{2}\left[\boldsymbol{I}(-1/2 < x < 0) - \boldsymbol{I}(0 < x < 1/2)\right] \tag{2.12}$$

which transforms the interval $(-1/2, 0)$ on $(0, 1/2)$ and vice versa. This formula is skew symmetric about $x = 0$ and thus the data processing is reversible in the sense $T^{-1} = T$. The d.f.'s F, F_0 for this example are shown in Figure 2.1 and their modifications resulting from the data processing (2.12) in Figure 2.2.

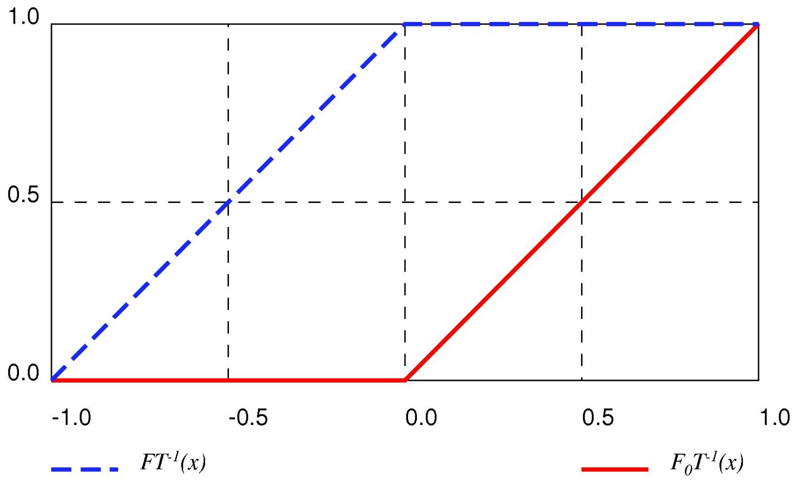


Figure 2.2 D.f.'s FT^{-1}, F_0T^{-1} modified by the data processing formula (2.12).

For the special discrete binary p.m.'s

$$P_x \sim \boldsymbol{p}_x = (F(x), 1 - F(x)) \text{ and } P_{0x} \sim \boldsymbol{p}_{0x} = (F_0(x), 1 - F_0(x)) \tag{2.13}$$

(2.9) and (2.6) imply

$$K(F,F_0)=\frac{1}{2}\sup_{x\in\mathbb{R}} L_1(P_x,P_{0x}).$$

Similar relations take place for other measures of goodness-of-fit between F, F_0 and their one-point approximations $F(x), F_0(x)$ formally represented by the binary p.m.'s (2.13). We mention the best known of them.

(d) The **Pearson divergence** is given by the formula

$$\chi^2(P,P_0)=\int\frac{(p-p_0)^2}{p_0}\mathrm{d}\lambda \quad \text{for } P\sim p, P_0\sim p_0 \tag{2.14}$$

which in the discrete case considered in (2.4) reduces to

$$\chi^2(\boldsymbol{p},\boldsymbol{p}_0)=\sum_{j=1}^{k}\frac{(p_j-p_{0j})^2}{p_{0j}}. \tag{2.15}$$

It defines the well-known Pearson goodness-of-fit test statistic

$$T_n=n\chi^2(\boldsymbol{p}_n,\boldsymbol{p}_0)=\sum_{j=1}^{k}\frac{(np_{nj}-np_{0j})^2}{np_{0j}}=\sum_{j=1}^{k}\frac{(\varphi_{nj}-np_{0j})^2}{np_{0j}} \tag{2.16}$$

for testing hypotheses $\mathcal{H}_0 : P_0$ on the basis of observations represented by the e.p.m. P_n where $\boldsymbol{p}_0$ and $\boldsymbol{p}_n$ are the restrictions of the hypothetical and empirical p.m.'s P_0 and P_n on the test cells $C_j=[c_j,c_{j+1})$ and $\varphi_{nj}=\#x_i\in C_j$ are the observed cell frequencies for $1\le j\le k$. The Pearson divergence is a divergence in the rigorous sense stated above, but it is not a distance since it is neither symmetric nor satisfies the triangle inequality.

(e) Another example of a divergence is the **LeCam divergence**

$$LC(P,P_0)=\int\frac{(p-p_0)^2}{p+p_0}\mathrm{d}\lambda \quad \text{for } P\sim p, P_0\sim p_0 \tag{2.17}$$

which in the discrete case considered in (2.4) reduces to

$$LC(\boldsymbol{p},\boldsymbol{p}_0)=\sum_{j=1}^{k}\frac{(p_j-p_{0j})^2}{p_j+p_{0j}}. \tag{2.18}$$

Since it is a divergence in the sense of Definition 2.1, all its roots are divergences in the sense of the same definition too but, as proved by Kafka, Österreicher and Vincze (1991), the square root $\sqrt{LC(P,P_0)}$ is distinguished by being a metric distance in the space of p.m.'s P, P_0.

(f) In the discrete case, by (2.15) and (2.13),

$$\chi^2(P_x,P_{0x})=\frac{(F(x)-F_0(x))^2}{F_0(x)}+\frac{(F(x)-F_0(x))^2}{1-F_0(x)}=\frac{(F(x)-F_0(x))^2}{F_0(x)(1-F_0(x))}.$$

Since $\chi^2(P_x,P_{0x})$ is a disparity of binary distributions P_x, P_{0x} for all $x\in\mathbb{R}$, the integral over $\mathbb{R}$,

$$AD(F,F_0)=\int\chi^2(P_x,P_{0x})\mathrm{d}x=\int\frac{(F(x)-F_0(x))^2}{F_0(x)(1-F_0(x))}\mathrm{d}x, \tag{2.19}$$

is a disparity of d.f.'s F, F_0 in the above stated sense. We call it the **Anderson-Darling disparity** because its scaled version

$$T_n=n\cdot AD(F_n,F_0)=n\int\frac{(F_n(x)-F_0(x))^2}{F_0(x)(1-F_0(x))}\mathrm{d}x$$

is the well known Anderson-Darling goodness-of-fit statistic for testing the hypothesis that the data represented by the e.d.f. F_n were generated by the d.f. F_0 (hypothesis $\mathcal{H}_0$).

(g) Similarly,we call

$$CM(F, F_0) = \int \chi^2(P_x, P_{0x}) F_0(x)(1 - F_0(x))\mathrm{d}x = \int (F(x) - F_0(x))^2 \mathrm{d}x \tag{2.20}$$

the **Cramér-von Mises disparity** because it is a disparity of p.d.f.'s F, F_0 in the above defined sense and

$$T_n = n \cdot CM(F_n, F_0) = n \int_{\mathcal{X}} (F_n(x) - F_0(x))^2 \mathrm{d}x$$

is the Cramér-von Mises goodness-of-fit statistics for testing the hypothesis $\mathcal{H}_0 = F_0$ under the empirical evidence F_n.

The goodness-of-fit statistics mentioned in **(f)** and **(g)** were introduced by von Mises (1947) and Anderson and Darling (1954) (see also Darling (1957)). We refer in this respect to Durbin (1973) or to pp. 58–64 in Serfling (1980).

2.2 Examples

The divergences and disparities will be more systematically studied in the next section. It the remainder of this section we apply the goodness-of-fit criteria introduced in this section to the data from Tables 1.1 and 1.2.

Example 2.2.1 Consider the discrete e.p.m. $\boldsymbol{p}_n = (p_{n1}, ..., p_{n,30})$ representing the data given by the frequencies φ_j of Table 1.1 and defined by the formula

$$p_{nj} = \varphi_j 10^{-3}, \qquad 1 \le j \le 30. \tag{2.21}$$

In addition to the normal model N(μ,σ^2) for these data given by (1.5), we consider the mixed normal model

$$\mathrm{MixN}(\pi, \mu_1, \mu_2, \sigma_1^2, \sigma_2^2) = \pi \mathrm{N}(\mu_1, \sigma_1^2) + (1 - \pi)\mathrm{N}(\mu_2, \sigma_2^2)$$

for the parameters

$$\pi = 0.5, (\mu_1, \sigma_1^2) = (0.6343, 0.000361), (\mu_2, \sigma_2^2) = (.6551, 0.00014641)$$

used by Pearson (1894). The third model considered by us is the mixed generalized lambda model

$$\mathrm{MixGLD}(\pi, \theta, \vartheta) = \pi \mathrm{GLD}(\theta) + (1 - \pi)\mathrm{GLD}(\vartheta)$$

from p. 88 of Ning, Gao and Dudewicz (2008) where the generalized lambda component models are given by the percentile functions

$$Q_\theta(y) = \theta_1 + \frac{y^{\theta_3} - (1 - y)^{\theta_4}}{\theta_2} \quad \text{and} \quad Q_\vartheta(y) = \vartheta_1 + \frac{y^{\vartheta_3} - (1 - y)^{\vartheta_4}}{\vartheta_2}.$$

for the parameters

$$\begin{aligned} \pi &= 0.802 \\ \theta &= (0.6415, 13.218, 0.135, 0.205) \\ \vartheta &= (0.6564, 11.328, 0.55, 0.15). \end{aligned} \tag{2.22}$$

Each d.f. $F_0 \in \{\mathrm{N}(\mu,\sigma^2),\ \mathrm{MixN}(\pi, \mu_1, \mu_2, \sigma_1^2, \sigma_2^2),\ \mathrm{MixGLD}(\pi, \theta, \vartheta)\}$ represents a different hypothesis about the stochastic source of the data from Table 1.1. Each one defines a hypothetical p.d.f. f_0 and a hypothetical discrete p.m. $\boldsymbol{p}_0 = (p_{01}, ..., p_{0k})$ obtained by quantization of the observation space $\mathcal{X} = (a, b) \subseteq \mathbb{R}$ into k cells. In the present example $(a, b) = (0.5795, 0.6995)$ and we consider quantization into

$k = 30$ partition intervals symmetric about the centers v_j given in Table 1.1 leding to $\boldsymbol{p}_0 = (p_{01}, ..., p_{0,30})$ with the components

$$p_{0j} = \int_{v_j-0.002}^{v_j+0.002} f_0(x)\mathrm{d}x = F_0(v_j + 0.002) - F_0(v_j - 0.002). \tag{2.23}$$

Table 2.2.1 presents the values of the distance or divergence criteria $L_1(\boldsymbol{p}, \boldsymbol{p}_0), L_2(\boldsymbol{p}, \boldsymbol{p}_0), \chi^2(\boldsymbol{p}, \boldsymbol{p}_0)$ and $LC(\boldsymbol{p}, \boldsymbol{p}_0)$ given by (2.6), (2.8), (2.15) and (2.18) for the e.p.m. $\boldsymbol{p} = \boldsymbol{p}_n$ given by (2.21) and hypothetical p.m.'s $\boldsymbol{p}_0$ given by (2.23).

Model	L_1	L_2	Pearson	LeCam
Normal	0.17476	0.047971	0.075765	0.034877
Mix Normal	0.0958	0.025351	0.020596	0.011084
Mix of Lambdas	0.24636	0.063542	0.14132	0.054918

Table 2.2.1 Values of several criteria for the models N, MixN and MixGLD of the data from Table 1.1.

Example 2.2.2 Consider the e.d.f. F_n defined by the data of Table 1.2. In addition to the normal model $\mathrm{N}(\mu,\sigma^2)$ for these data given by (1.6), we consider the generalized lambda models GLD(θ) given by the percentile function

$$Q_\theta(y) = \theta_1 + \frac{y^{\theta_3} - (1-y)^{\theta_4}}{\theta_2}$$

for the parameters

$$\theta = (41.7897, 0.01134, 0.09853, 0.3606) \tag{2.24}$$

obtained on p. 97 of Karian and Dudewicz (2000). We use also the mixed generalized lambda model

$$\mathrm{MixGLD}(\pi, \theta, \vartheta) = \pi \mathrm{GLD}(\theta) + (1-\pi)\mathrm{GLD}(\vartheta)$$

with the component generalized lambda models given by the percentile functions

$$Q_\theta(y) = \theta_1 + \frac{y^{\theta_3} - (1-y)^{\theta_4}}{\theta_2} \quad \text{and} \quad Q_\vartheta(y) = \vartheta_1 + \frac{y^{\vartheta_3} - (1-y)^{\vartheta_4}}{\vartheta_2}.$$

for the parameters

$$\begin{aligned} \pi &= 0.4 \\ \theta &= (57.8233, 0.0076, 0.1432, 0.1356) \\ \vartheta &= (56.2136, 0.0193, 0.4601, 0.4838) \end{aligned} \tag{2.25}$$

introduced on p. 91 of Ning, Gao and Dudewicz (2008). The present models $F_0 \in \{\mathrm{N}(\mu,\sigma^2)$, GLD($\theta$), MixGLD($\pi, \theta, \vartheta$)$\}$ differ from those considered in the previous example but, similarly as above, they represent three different hypotheses about the sources of data from Table 1.2. Table 2.2.2 presents values of the distance or disparity $K(F, F_0)$, $AD(F, F_0)$ and $CM(F, F_0)$ given by (2.9), (2.19) and (2.20) for the d.f. $F(x) = F_n(x)$ from Figure 1.2 and the present models $F_0(x)$.

Model	$K(F, F_0)$	$AD(F, F_0)$	$CM(F, F_0)$
Normal	0.45177	0.38039	0.05513
Lambda	0.07917	0.39040	0.06034
Mix of Lambdas	0.21747	78.4530	0.78635

Table 2.2.2 Values of several criteria for the models N, GLD and MixGLD of the data from Table 1.2.

3 Criteria of goodness-of-fit

This section studies more systematically those criteria of goodness-of-fit between the empirical evidence represented by the e.d.f. F_n on the one hand and the hypothetical model specified by its d.f. F_0 on the other, that are measures of dissimilarity $D(F_n, F_0)$ between the distribution functions F_n and F_0. Since $F_n \to F_0$ a.s. for $n \to \infty$ and, consequently, $D(F_n, F_0) \to 0$ stochastically as $n \to \infty$ for reasonable dissimilarity measures D, the statistics T_n for testing the hypothesis $\mathcal{H}_0 : F_0$ on the basis of empirical evidence represented by F_n, are considered in the form $T_n = m_n D(F_n, F_0)$. Here $m_n \to \infty$ is an appropriate scaling sequence for which T_n tends to a limit distribution. The value $Q(1-\alpha)$ of the p.f. of this distribution is then used as a critical value of T_n for the asymptotically α-level test of hypothesis $\mathcal{H}_0$. Examples were given in Section 2, e.g. the Kolmogorov distance $K(F_n, F_0)$ and corresponding Kolmogorov-Smirnov statistic $T_n = \sqrt{n} K(F_n, F_0)$.

Similarly, if the empirical evidence is represented by a discrete distribution $\boldsymbol{p}_n$ and the hypothetical model is specified by the discrete distribution $\boldsymbol{p}_0$ then the dissimilarity between the former and the sequel is $D(\boldsymbol{p}_n, \boldsymbol{p}_0)$ and the statistic for testing the hypothesis $\mathcal{H}_0 : \boldsymbol{p}_0$ is $T_n = m_n D(\boldsymbol{p}_n, \boldsymbol{p}_0)$. Examples were given in Section 2, e.g. the Pearson distance $\chi^2(\boldsymbol{p}_n, \boldsymbol{p}_0)$ and the corresponding Pearson statistic $T_n = n\chi^2(\boldsymbol{p}_n, \boldsymbol{p}_0)$.

Goodness-of-fit criteria are not only studied for e.d.f.'s F_n or related e.p.m.'s $\boldsymbol{p}_n$ simply representing the data sets $\{x_1, x_2, ..., x_n\}$ in a straightforward manner (1.2), but also for more sophisticated models F and $\boldsymbol{p}$ obtained by data-based statistical inference like e.g. those obtained by maximum likelihood estimation. Therefore we deal in the rest of this section with arbitrary d.f.'s F and p.m.'s $\boldsymbol{p}$.

3.1 Disparities, divergences and metric distances

To unify the treatment of the situation when goodness-of-fit is considered for d.f.'s (F, F_0) or p.m.'s $(\boldsymbol{p}, \boldsymbol{p}_0)$, it is convenient to represent simultaneously both hypothetical models $F_0, \boldsymbol{p}_0$ by the corresponding general p.m.'s P_0, and alternative models $F, \boldsymbol{p}$ by the corresponding general p.m.'s P. Thus we deal in this subsection primarily with dissimilarity measures $D(P, P_0)$. We respect Convention 2.1 and use the notations (2.2) - (2.4) introduced there.

We define the class of measures of **dissimilarity** of probability measures P, P_0 on $\mathcal{X}$ by

$$D_\phi(P, P_0) = \int_{\mathcal{X}} p_0 \phi\left(\frac{p}{p_0}\right) \mathrm{d}\lambda \quad \text{for } P \sim p, P_0 \sim p_0 \tag{3.1}$$

generated by continuous functions $\phi : (0, \infty) \mapsto [0, \infty)$ with continuous extension $0 \le \phi(0) \le \infty$, such that the integral (3.1) exists. In particular, for arbitrary d.f. F and hypothetical d.f. F_0 on an interval observation space $\mathcal{X} = (a, b) \subseteq \mathbb{R}$,

$$D_\phi(F, F_0) = \int_a^b f_0(x) \phi\left(\frac{f(x)}{f_0(x)}\right) \mathrm{d}x \quad \text{for } F \sim f, F_0 \sim f_0, \tag{3.2}$$

and, in the discrete case, for

$$P \sim \boldsymbol{p} = (p_1, p_2, ..., p_k) \quad \text{and} \quad P_0 \sim \boldsymbol{p}_0 = (p_{01}, p_{02}, ..., p_{0k}) \tag{3.3}$$

c.f. (2.4),

$$D_\phi(P, P_0) \equiv D_\phi(\boldsymbol{p}, \boldsymbol{p}_0) = \sum_{j=1}^{k} p_{0j} \phi\left(\frac{p_j}{p_{0j}}\right). \tag{3.4}$$

Let us clarify for which functions ϕ the dissimilarities (3.1) - (3.4) are disparities, divergences or distances in the sense of the previous section. Denote by $\boldsymbol{\Phi}$ the class of all differentiable functions $\phi : (0, \infty) \mapsto \mathbb{R}$ with continuous extension $\phi(0)$ and the property

$$(\phi'(t) - \phi'(1))\operatorname{sign}(t-1) > 0 \quad \text{for all } t \in (0, \infty). \tag{3.5}$$

Then the standardized version

$$\tilde{\phi}(t) = \phi(t) - \phi(1) - \phi'(1)(t-1) \tag{3.6}$$

of $\phi(t)$ is increasing in the domain $t \geq 0$ and decreasing in the domain $t \leq 0$, i.e. is positive for $t \in (0, \infty)$ except for $t = 1$ where $\tilde{\phi}(1) = 0$. Therefore the integral

$$D_{\tilde{\phi}}(P, P_0) = \int_{\mathcal{X}} p_0 \tilde{\phi}\left(\frac{p}{p_0}\right) \mathrm{d}\lambda = D_\phi(P, P_0) - \phi(1) \tag{3.7}$$

exists, takes on values in the closed interval $[0, \infty]$, and the dissimilarity $D_{\tilde{\phi}}(P, P_0)$ is reflexive in the sense that equality holds if and only if $P = P_0$. Hence the expressions $D_{\tilde{\phi}}(P, P_0)$ are well-defined disparities, and also the expressions $D_\phi(P, P_0)$ are disparities up to shifts $\phi(1)$. This justifies to speak about all dissimilarities $D_\phi(P, P_0), D_\phi(F, F_0)$ and $D_\phi(\boldsymbol{p}, \boldsymbol{p}_0)$ given by (3.1) - (3.4) for $\phi \in \boldsymbol{\Phi}$ as **disparities in the wide *sense*,** transformed by constant shifts $\phi(1)$ to *disparities* defined in the precise sense of the previous section. Of course, the subset

$$\boldsymbol{\Phi}_{\text{disp}} = \{\phi \in \boldsymbol{\Phi} : \phi(1) = 0\}, \quad \boldsymbol{\Phi}_{\text{disp}} \subset \boldsymbol{\Phi}$$

defines proper disparities by (3.1) - (3.4).

Let $\boldsymbol{\Phi}_{\text{div}}$ be the class of differentiable convex functions $\phi : (0, \infty) \mapsto \mathbb{R}$ with continuous extension $\phi(0)$ and strict convexity at $t = 1$. Then $\phi(1) + \phi'(1)(t-1)$ is the support straight line of the function $\phi(t)$ which is strictly smaller than $\phi(t)$ at all $t \neq 1$ due to the strict convexity of $\phi(t)$ at $t = 1$. However, the assumed convexity of $\phi(t)$ on the whole domain $(0, \infty)$ means that the function (3.6) is increasing on $(1, \infty)$ and decreasing on $(0, 1)$ so that (3.5) holds. Consequently,

$$\boldsymbol{\Phi}_{\text{div}} \subset \boldsymbol{\Phi}_{\text{disp}},$$

i.e. expression (3.7) is reflexive. As proved in Csiszár (1967) or Liese and Vajda (1987) (see also a new statistical proof in Vajda and Liese (2006)), the disparities (3.7) with $\phi \in \boldsymbol{\Phi}_{\text{div}}$ satisfy the information processing law, i.e. they are *divergences* in the sense defined in the previous section.

Example 3.1.1 The functions defined on $(0, \infty)$ by

$$\phi_0(t) = -\ln t \quad \text{and} \quad \phi_1(t) = t \ln t \tag{3.8}$$

with extensions $\phi_0(0) = \infty$ and $\phi_1(t) = 0$ satisfy condition (3.5). Thus they belong to $\boldsymbol{\Phi}$ and define the wide-sense disparities

$$D_1(P, P_0) \equiv D_{\phi_1}(P, P_0) = \int_{\mathcal{X}} p \ln\left(\frac{p}{p_0}\right) \mathrm{d}\lambda \tag{3.9}$$

and

$$D_0(P, P_0) \equiv D_{\phi_0}(P, P_0) = \int_{\mathcal{X}} p_0 \ln\left(\frac{p_0}{p}\right) \mathrm{d}\lambda = D_1(P_0, P). \tag{3.10}$$

Since $\phi_i(1) = 0$, both these functions belong to the subset $\boldsymbol{\Phi}_{\text{disp}} \subset \boldsymbol{\Phi}$, so that (3.9) and (3.10) are proper disparities. Further, both functions $\phi_i(t)$ are differentiable and strictly convex on the domain $(0, \infty)$. Therefore they belong to $\boldsymbol{\Phi}_{\text{div}}$ and (3.9) and (3.10) are divergences. In fact, $D_1(P, P_0)$ is known as the **information divergence** or **Kullback divergence**, and $D_0(P, P_0)$ is usually called the **reversed information divergence** or **reversed Kullback divergence**. Note that

$$T_n = nD_1(\boldsymbol{p}_n, \boldsymbol{p}_0) = \sum_{j=1}^{k} np_{nj} \ln \frac{np_{nj}}{np_{0j}} = \sum_{j=1}^{k} \varphi_{nj} \ln \frac{\varphi_{nj}}{np_{0j}} \tag{3.11}$$

is the well-known likelihood ratio test statistic for testing the hypothesis $\mathcal{H}_0 : P_0$ on the basis of observations represented by the e.p.m. P_n, where $\boldsymbol{p}_0$ and $\boldsymbol{p}_n$ are restrictions of the hypothetical and empirical p.m.'s P_0 and P_n to the test cells $C_j = [c_j, c_{j+1})$, and $\varphi_{nj} = \#x_i \in C_j$ are the observed cell frequencies for $1 \leq j \leq k$.

Example 3.1.2 The functions defined on $(0,\infty)$ by

$$\phi_{0,-1}(t) = \frac{1}{t} \quad \text{and} \quad \phi_{0,2}(t) = t^2$$

with extensions $\phi_{0,-1}(0) = \infty$ and $\phi_{0,2}(t) = 0$ satisfy condition (3.5). Thus they belong to $\mathbf{\Phi}$ and define the wide-sense disparities

$$D_{0,-1}(P,P_0) \equiv D_{\phi_{0,-1}}(P,P_0) = \int_{\mathcal{X}} \frac{p_0^2}{p} \mathrm{d}\lambda$$

and

$$D_{0,2}(P,P_0) \equiv D_{\phi_{0,2}}(P,P_0) = \int_{\mathcal{X}} \frac{p^2}{p_0} \mathrm{d}\lambda = D_{0,-1}(P_0,P).$$

Since $\phi_{0,-1}(1) = \phi_{0,2}(1) = 1$, the functions $\phi_{-1}(t) = \phi_{0,-1}(t) - 1$ and $\phi_2(t) = \phi_{0,2}(t) - 1$ define the proper disparities

$$\tilde{D}_{0,-1}(P,P_0) = D_{0,-1}(P,P_0) - 1 = \int_{\mathcal{X}} \frac{p_0^2}{p} \mathrm{d}\lambda - 1 = \int_{\mathcal{X}} \frac{(p-p_0)^2}{p} \mathrm{d}\lambda \tag{3.12}$$

and

$$\begin{aligned} \tilde{D}_{0,2}(P,P_0) &= D_{0,2}(P,P_0) - 1 = \int_{\mathcal{X}} \frac{p^2}{p_0} \mathrm{d}\lambda - 1 \\ &= \int_{\mathcal{X}} \frac{(p-p_0)^2}{p} \mathrm{d}\lambda = \tilde{D}_{0,-1}(P_0,P). \end{aligned} \tag{3.13}$$

The same proper disparities are defined by the standardized versions

$$\tilde{\phi}_{0,-1}(t) = \frac{1}{t} - 2 + t = \frac{(1-t)^2}{t} \quad \text{and} \quad \tilde{\phi}_{0,2}(t) = t^2 - 1 - 2t = (t-1)^2$$

belonging to $\mathbf{\Phi}_{\text{disp}} \subset \mathbf{\Phi}$. Further, both functions $\tilde{\phi}_{0,-1}(1)$ and $\tilde{\phi}_{0,2}(1)$ are differentiable and strictly convex on the domain $(0,\infty)$. Therefore they belong to $\mathbf{\Phi}_{\text{div}}$ and (3.12), (3.13) are divergences. In fact, $\tilde{D}_{0,2}(P,P_0) = \chi^2(P,P_0)$ is the Pearson divergence defined in 2.1.**(d)** and $\tilde{D}_{0,-1}(P,P_0) = \chi^2(P_0,P)$ is the reversed Pearson divergence.

Example 3.1.3 Power divergences The functions defined on $(0,\infty)$ by

$$\phi_\alpha(t) = \frac{t^\alpha - 1}{\alpha(\alpha-1)} \quad \text{for the powers } \alpha \in \mathbb{R} - \{0,1\} \tag{3.14}$$

and for the remaining powers by

$$\phi_1(t) = t\ln t \quad \text{and} \quad \phi_0(t) = -\ln t \tag{3.15}$$

belong to $\mathbf{\Phi}$ defined in Example 3.1.1. They satisfy the condition $\phi_\alpha(1) = 0$ so that they belong also to the subset $\mathbf{\Phi}_{\text{disp}} \subset \mathbf{\Phi}$. Since they are differentiable and strictly convex on $(0,\infty)$, they belong in fact also to $\mathbf{\Phi}_{\text{div}}$ and define the divergences

$$D_\alpha(P,P_0) \equiv D_{\phi_\alpha}(P,P_0) = \frac{\int_{\mathcal{X}} p^\alpha p_0^{1-\alpha} \mathrm{d}\lambda - 1}{\alpha(\alpha-1)} \quad \text{for } \alpha \in \mathbb{R} - \{0,1\}. \tag{3.16}$$

Remarks **(i)** It holds that

$$D_1(P,P_0) \equiv D_{\phi_1}(P,P_0) = \int_{\mathcal{X}} p \ln\left(\frac{p}{p_0}\right) \mathrm{d}\lambda = D_0(P_0,P) \equiv D_{\phi_0}(P_0,P) \tag{3.17}$$

since the functions (3.15) are the same as in Example 3.1.1. Hence the members $D_1(P,P_0)$ and $D_0(P,P_0)$ of the power divergence family are the **Kullback** and **reversed Kullback divergences** introduced in

Example 3.1.1.
(ii) Further,

$$D_\alpha(P,P_0) = \tilde{D}_\alpha(P,P_0)/2 \quad \text{for } \alpha = -1 \text{ and } \alpha = 2$$

where the right-hand side consists of the disparities introduced in Example 3.1.2. Therefore $2D_2(P,P_0)$ and $2D_{-1}(P,P_0)$ are the well known *Pearson* and *reversed Pearson divergences* of Example 3.1.2, introduced already in Section 2.1, part **(d)** (briefly, 2.1**(d)**).
(iii) Another well-known member of the family of power divergences not mentioned before is the **Hellinger divergence**

$$D_{1/2}(P,P_0) = 4H(P,P_0) = 4\int_{\mathcal{X}} (\sqrt{p} - \sqrt{p_0})^2 \, \mathrm{d}\lambda, \tag{3.18}$$

which in the discrete case (2.4) reduces to

$$D_{1/2}(\boldsymbol{p},\boldsymbol{p}_0) = 4 = 4\sum_{j=1}^{k} \left(\sqrt{p_j} - \sqrt{p_{0j}}\right)^2. \tag{3.19}$$

The divergence $H(\boldsymbol{p},\boldsymbol{p}_0)$ was introduced by Matusita (1956) but it is better known as **squared Hellinger distance**. Indeed,

$$\sqrt{H(P,P_0)} = ||\sqrt{p} - \sqrt{p_0}||_2$$

is the L_2-distance of the square roots of p.d.f.'s and, as such, is a metric distance. Of course, $\sqrt{D_{1/2}(\boldsymbol{p},\boldsymbol{p}_0)}$ is a metric distance too.
(iv) The expression

$$T_n = nD_{1/2}(\boldsymbol{p}_n,\boldsymbol{p}_0)/2 = 2n\sum_{j=1}^{k} \left(\sqrt{p_j} - \sqrt{p_{0j}}\right)^2$$

is known as the **Freeman-Tukey statistic** for testing the hypothesis $\mathcal{H}_0 : P_0$ on the basis of observations represented by the e.p.m. P_n, where $\boldsymbol{p}_0$ and $\boldsymbol{p}_n$ are the restrictions of the hypothetical and empirical p.m.'s P_0 and P_n on the test cells $C_j = [c_j, c_{j+1})$ for $1 \le j \le k$.
(v) The square root $\sqrt{D_{1/2}(\boldsymbol{p},\boldsymbol{p}_0)}$ is the only member among all powers of all power divergences which is a distance in the sense introduced above (cf Definition 2.1) because $D_{1/2}(P,P_0)$ is the only power divergence which is symmetric in the variables P, P_0.
(vi) The nonnegative standardized versions of the functions (3.14) are

$$\tilde{\phi}_\alpha(t) = \frac{t^\alpha - \alpha(t-1) - 1}{\alpha(\alpha-1)} \quad \text{for } \alpha \in \mathbb{R} - \{0,1\} \tag{3.20}$$

and those corresponding to (3.15) are the limits

$$\tilde{\phi}_1(t) = t\ln t - t + 1 \quad \text{and} \quad \tilde{\phi}_0(t) = -\ln t + t - 1 \tag{3.21}$$

of $\tilde{\phi}_\alpha(t)$ for $\alpha \to 0$ and $\alpha \to 1$. Obviously

$$D_{\tilde{\phi}_\alpha}(P,P_0) = D_{\phi_\alpha}(P,P_0) = D_\alpha(P,P_0) \quad \text{for all } \alpha \in \mathbb{R}. \tag{3.22}$$

(vii) The classes of modified power divergences $\alpha D_\alpha(P,P_0), \alpha > 0$ are one-to-one related to the logarithmic power divergences

$$R_\alpha(P,P_0) = \frac{\ln[\alpha(\alpha-1)D_\alpha(P,P_0)+1]}{\alpha-1} = \frac{\ln\int_{\mathcal{X}} p^\alpha p_0^{1-\alpha}\mathrm{d}\lambda}{\alpha-1}, \quad \alpha > 0 \tag{3.23}$$

of Rényi (1961) where $R_1(P,P_0) = \lim_{\alpha\to 1} R_\alpha(P,P_0) = D_1(P,P_0)$. They were studied e.g. by Perez (1967). The more significant modified versions

$$H_\alpha(P,P_0) = \alpha(\alpha-1)D_\alpha(P,P_0) + 1 = \int_{\mathcal{X}} p^\alpha p_0^{1-\alpha}\mathrm{d}\lambda, \quad \alpha \in \mathbb{R}, \tag{3.24}$$

sometimes called **Hellinger integrals**, were studied e.g. by Chernoff (1952), Vajda (1971) and Liese (1982). The Hellinger integrals are skew symmetric about $\alpha = 1/2$ in the sense that

$$H_\alpha(P, P_0) = H_{1-\alpha}(P_0, P), \quad \alpha \in \mathbb{R}. \tag{3.25}$$

The power divergences (3.16),(3.17) are skew symmetric in the same sense, i.e.

$$D_\alpha(P, P_0) = D_{1-\alpha}(P_0, P), \quad \alpha \in \mathbb{R}. \tag{3.26}$$

The skew-symmetrization of the formerly used power divergences $\alpha D_\alpha(P, P_0)$, $\alpha > 0$, was proposed by Cressie and Read (1984), and the skew symmetric divergences (3.16),(3.17) are used as standard representatives of the whole class of divergences $D_\phi(P, P_0), \phi \in \mathbf{\Phi}_{\mathrm{div}}$. A similar skew-symmetrization

$$R_\alpha(P, P_0) = \frac{\ln \int_{\mathcal{X}} p^\alpha p_0^{1-\alpha} \mathrm{d}\lambda}{\alpha(\alpha - 1)}, \quad \alpha \in \mathbb{R} \tag{3.27}$$

of the Rényi divergences was introduced by Liese and Vajda (1987) who used them to establish a number of properties of the power divergences (3.26).

3.2 Metricity and robustness

The power divergences $D_\alpha(P, P_0), \alpha \in \mathbb{R}$, do not represent all aspects of the class of all divergences $D_\phi(P, P_0), \phi \in \mathbf{\Phi}_{\mathrm{div}}$. For example, $\sqrt{D_{1/2}(P, P_0)}$ is the only power divergence which is a metric distance. This might suggest that the metricity of a divergence is a rare property. But, in fact, the class of all divergences contains uncountably many of them with this property. For example, all functions

$$\varphi_\beta(t) = \frac{\operatorname{sign} \beta}{1 - \beta} \left[(t^{1/\beta} + 1)^\beta - 2^{\beta - 1}(t + 1) \right] \quad \text{for } -\infty < \beta \leq 2 \tag{3.28}$$

with the terms for $\beta = 0$ and $\beta = 1$ obtained by the continuous extension rule as

$$\varphi_0(t) = |t - 1|/2 \ \text{ and } \ \varphi_1(t) = t \ln t + (t + 1) \ln \frac{2}{t + 1} \tag{3.29}$$

belong to $\mathbf{\Phi}_{\mathrm{div}}$. The square roots $\sqrt{D_{\phi_\beta}(P, P_0)}$ of the corresponding divergences are metric distances (see Vajda (2009)). LeCam's divergence of 2.1.**(g)** is among them, since

$$D_{\varphi_{-1}}(P, P_0) = LC(P, P_0).$$

The total variation function $\varphi_0(t)$ formally does not belong to $\mathbf{\Phi}_{\mathrm{div}}$ because it is not differentiable at $t = 1$, but this may be cured by defining this derivative as the mean of the left- and right-hand derivatives,

$$\varphi_0'(1) = \frac{\varphi_0'(1+) + \varphi_0'(1-)}{2} = \frac{1}{2} - \frac{1}{2} = 0. \tag{3.30}$$

Let us now look at the robustness of the testing or estimation based on the disparity statistics

$$T_{\phi,n} = n D_\phi(P_n, P_0), \quad \phi \in \mathbf{\Phi}_{\mathrm{disp}}$$

reflecting the proximity of the hypothetical p.m. P_0 and e.p.m. P_n, or their special forms

$$T_{\phi,n} = n D_\phi(\boldsymbol{p}_n, \boldsymbol{p}_0) = \sum_{j=1}^k n p_{0j} \phi\left(\frac{n p_{nj}}{n p_{0j}}\right) = \sum_{j=1}^k n p_{0j} \phi\left(\frac{\varphi_{nj}}{n p_{0j}}\right), \quad \phi \in \mathbf{\Phi}_{\mathrm{disp}}$$

using restrictions $\boldsymbol{p}_0$ and $\boldsymbol{p}_n$ of P_0 and P_n on the test cells $C_j = [c_j, c_{j+1})$ where $\varphi_{nj} = \# x_i \in C_j$ are the observed cell frequencies for $1 \leq j \leq k$.

If $\phi \in \mathbf{\Phi}_{\mathrm{div}}$, i.e. if $D_\phi(P, P_0)$ is a divergence then $\phi(t)$ is always unbounded and its derivative $\phi'(t)$ is usually unbounded on the domain $(0, \infty)$. For example, all power divergence functions $\phi_\alpha(t)$ as well as their derivatives $\phi_\alpha'(t)$ are unbounded. The statistical disparity measures were introduced by Lindsay

(1994) and more systematically investigated by Menéndes et al. (1998). In these papers it is argued that the robustness of statistical inference based on minimization of disparities between models requires bounded generating functions $\phi \in \mathbf{\Phi}_{\text{disp}}$, or at least bounded derivatives $\phi\prime$, because these two functions represent the criterion function and the influence function of robust statistics. Thus, from the point of robustness of statistical decisions based on disparity statistics, the attention is concentrated on the functions $\rho(t-1) \in \mathbf{\Phi}_{\text{disp}}$ for the classical criterial ρ-functions of robust statistics leading to bounded influence functions $\psi(t)$ proportional to the derivatives $\rho'(t)$.

Example 3.2.1 A classical example is the family of **Huber ρ-functions**

$$\rho_k(t) = \boldsymbol{I}(|t| \le k)t^2, \quad k > 0$$

smoothly extended as linear functions in the domain $\{t \in \mathbb{R} : |t| > k\}$ with the constantly bounded influences $\psi_k(t) = \rho_k'(t) = 2k$ of the observations t from this domain (see Hampel et al. (1986) or Jurečková and Sen (1996)). The Huber functions define the family

$$\phi_k(t) = \rho_k(t-1), \quad k > 0 \tag{3.31}$$

of functions from $\mathbf{\Phi}_{\text{div}}$ with the bounded derivatives

$$\psi_k(t) = \phi_k'(t) = 2[(t-1)\boldsymbol{I}(|t-1| \le k) + k\boldsymbol{I}(|t-1| > k), \quad k > 0. \tag{3.32}$$

These functions generate the family of robust divergences $D_{\phi_k}(P, P_0), k > 0$.

Example 3.2.2 Another classical example is the family of ρ-functions

$$\rho_\alpha(t) = (\alpha - 1)t\boldsymbol{I}(t < 0) + \alpha t\boldsymbol{I}(t > 0), \quad 0 < \alpha < 1$$

leading to the robust statistical inference based on the so-called **regression quantiles** (see Jurečková and Sen (1996)). The total variation generating function is the special case

$$\varphi_0(t) = \rho_{1/2}(t-1).$$

We put

$$\phi_\alpha(t) = \rho_\alpha(t-1), \quad 0 < \alpha < 1 \tag{3.33a}$$

and similarly as in (3.30), we use the generalized form

$$\phi_\alpha'(1) = \frac{\phi_\alpha'(1+) + \phi_\alpha'(1-)}{2} = \alpha - \frac{1}{2} \tag{3.34}$$

to extend the derivative $\phi_\alpha'(t)$ of the family of robust divergence generating functions to $t = 1$ in order to achieve the formal validity for including this family in $\mathbf{\Phi}_{\text{div}}$. The derivatives

$$\psi_\alpha(t) = \phi_\alpha'(t) = (\alpha - 1)\boldsymbol{I}(t < 1) + \alpha\boldsymbol{I}(t > 1), \quad 0 < \alpha < 1 \tag{3.35}$$

of these functions are bounded on the domain $(0, \infty)$ so that the functions (3.33a) generate the family of robust divergences $D_{\phi_\alpha}(P, P_0)$, $0 < \alpha < 1$.

4 Disparities based on partitions

In the previous section we assessed goodness-of-fit between two statistical models given by p.m.'s P and P_0 by the disparity, divergence or distance

$$D_\phi(P, P_0) = \int_{\mathcal{X}} p_0 \phi\left(\frac{p}{p_0}\right) \mathrm{d}\lambda \tag{4.1}$$

where the concrete type depends on the extended real valued function $\phi \in \mathbf{\Phi}$ and, according to Convention 2.1, p, p_0 are densities of P, P_0 given by (2.2) and p_0 is positive on $\mathcal{X}$. Definition (2.2) of the densities p, p_0 means that for every $A \subset \mathcal{X}$

$$P(A) = \int_A p \mathrm{d}\lambda \quad \text{and} \quad P_0(A) = \int_A p_0 \mathrm{d}\lambda \quad \text{for} \quad \lambda = P + P_0. \tag{4.2}$$

In some situations it is necessary to restrict the p.m.'s P and P_0 on a partition $\mathcal{P} = \{C_1, ..., C_k\}$ of $\mathcal{X}$ into disjoint cells C_j , resulting into the quantizations

$$\boldsymbol{p} = (p_j \equiv P(C_j) : 1 \le j \le k) \quad \text{and} \quad \boldsymbol{p}_0 = (p_{0j} \equiv P_0(C_j) : 1 \le j \le k) \tag{4.3}$$

of these p.m.'s and to the reduced disparity, divergence or distance

$$D_\phi(\boldsymbol{p}, \boldsymbol{p}_0) = \sum_{j=1}^{k} p_{0j} \phi\left(\frac{p_j}{p_{0j}}\right). \tag{4.4}$$

Partitioning of the observation space means that observations $x \in \mathcal{X}$ are replaced by the indices of the partition sets containing these observations, i.e. by

$$T(x) \in \{1, 2, ..., k\} \quad \text{where} \quad T^{-1}(j) = C_j \quad \text{for} \quad 1 \le j \le k. \tag{4.5}$$

Among other this means that if ϕ generates the divergence $D_\phi(P, P_0)$, then the information processing law implies that

$$D_\phi(\boldsymbol{p}, \boldsymbol{p}_0) \le D_\phi(P, P_0) \tag{4.6}$$

where the equality holds if and only if the partition $\mathcal{P}$ is statistically sufficient for (P, P_0).

The situation described above takes place when the first of the p.m.'s is an e.p.m., i.e. when it is uniform on the observation support set $S_n = \{x_1, x_2, ..., x_n\}$ according to (1.1). Then (4.2) holds for

$$p(x) = p_n(x) \equiv \boldsymbol{I}(x \in S_n) \quad \text{and} \quad p_0(x) \equiv \boldsymbol{I}(x \notin S_n)$$

because

$$(P_n + P_0)(S_n \cap A) = P_n(A) \quad \text{and} \quad (P_n + P_0)((\mathcal{X} - S_n) \cap A) = P_n(A) \quad \text{for every } A \subset \mathcal{X}.$$

This, together with the fact that P_n is supported by S_n and P_0 is supported by $\mathcal{X} - S_n$ implies that

$$\begin{aligned} D_\phi(P_n, P_0) &= \int_{\mathcal{X}} p_0 \, \phi\left(\frac{p_n}{p_0}\right) \mathrm{d}(P_n + P_0) \\ &= \int_{S_n} \boldsymbol{I}(x \notin S_n) \, \phi\left(\frac{\boldsymbol{I}(x \in S_n)}{\boldsymbol{I}(x \notin S_n)}\right) \mathrm{d}P_n \\ &+ \int_{\mathcal{X} - S_n} \boldsymbol{I}(x \notin S_n) \, \phi\left(\frac{\boldsymbol{I}(x \in S_n)}{\boldsymbol{I}(x \notin S_n)}\right) \mathrm{d}P_0 \\ &= \int_{S_n} 0 \, \phi\left(\frac{1}{0}\right) \mathrm{d}P_n + \int_{\mathcal{X} - S_n} 1 \, \phi\left(\frac{0}{1}\right) \mathrm{d}P_0 \\ &= \lim_{t \to \infty} \frac{\phi(t)}{t} + \phi(0), \end{aligned} \tag{4.7}$$

where in the last line we replaced the undefined expression by the limit on the basis of the principle of continuous extension. The existence of the limit is guaranteed only for convex ϕ, i.e. for divergences, but even then the value (4.7) is constant, often infinite (for the power divergences it is finite only for the powers $0 < \alpha < 1$). Thus, without any further specification or restriction, the disparities, divergences or distances $D_\phi(P_n, P_0)$ are meaningless for statistical inference. Hence for the rest of the chapter we adopt the following convention.

Convention 4.1 The disparities, divergences or distances $D_\phi(P_n, P_0)$ are replaced by their discrete versions (4.4) resulting from finite partitions $\mathcal{P} = \{C_1, ..., C_k\}$ of the observation space $\mathcal{X}$ by the quantization rule (4.3). Further, we restrict ourselves to **real valued observations** and assume interval observation spaces $\mathcal{X} = (a, b) \subseteq \mathbb{R}$ on which statistical models P_n, P, P_0 are given by d.f.'s F_n, F, F_0 where both F and F_0 are assumed to have positive densities $f > 0$ and $f_0 > 0$. Moreover, we restrict ourselves to the interval partitions

$$\mathcal{P} = \{C_j \equiv (c_{j-1}, c_j] : 1 \le j \le k\} \quad \text{for} \quad a = c_0 < c_1 < ... < c_{k-1} < c_k = b \tag{4.8}$$

where the quantization rule (4.3) takes on the form

$$\boldsymbol{p} = (p_j \equiv F(c_j) - F(c_{j-1}) : 1 \le j \le k), \tag{4.9}$$

$$\boldsymbol{p}_n = (p_{nj} \equiv F_n(c_j) - F_n(c_{j-1})) : 1 \le j \le k) \tag{4.10}$$

$$\boldsymbol{p}_0 = (p_{0j} \equiv F_0(c_j) - F_0(c_{j-1}) : 1 \le j \le k). \tag{4.11}$$

By (4.4), the disparities, divergences or distances $D_\phi(\boldsymbol{p}_n, \boldsymbol{p}_0)$ depend on the frequencies of the observations $x_1, x_2, ..., x_n$ in the intervals of the partition $\mathcal{P}$ but not on the order of these observations. Therefore the vector $(x_1, x_2, ..., x_n)$ of observations can be replaced by the order statistics $(x_{n:1}, x_{n:2}, ..., x_{n:n})$. Further, the hypothetical model F_0, which is compared with the alternative model F or the empirical model F_n, is known and by the assumptions the function $F_0(y)$ is strictly increasing on the observation space $\mathcal{X} = (a, b)$. Hence it can be used to transform in a one-to-one manner this observation space into the simple standardized observation space $\mathcal{Y} = (F_0(a), F_0(b)) = (0, 1)$ commonly used in the literature dealing with testing hypotheses $\mathcal{H}_0 : F_0$ against alternatives $\mathcal{A} : F$. The d.f. F governs in an i.i.d. manner the random observations $X_1, X_2, ..., X_n$ generating the realizations $x_1, x_2, ..., x_n$. In this new observation space we deal with the ordered observations

$$F_0(a) \equiv 0 = Y_0 < Y_1 \equiv F_0(X_{n:1}) \le ... \le Y_n \equiv F_0(X_{n:n}) < 1 = Y_{n+1} \equiv F_0(b) \tag{4.12}$$

and with the hypotheses $\mathcal{H}_0 : F_0(Q_0)$, alternatives $\mathcal{A} : F(Q_0)$ and e.d.f.'s $F_n(Q_0)$, all defined on $[0, 1]$ by means of the increasing **hypothetical percentile function** (briefly, **h.p.f.**)

$$Q_0(y) = F_0^{-1}(y) \quad \text{on } [0, 1]. \tag{4.13}$$

This motivates the next convention which will also hold for the remainder of this chapter.

Convention 4.2 We assume without loss of generality that we test the hypothesis of uniformity $\mathcal{H}_0 : F_0$ with constant p.d.f. $f_0(y) = 1$ and linear h.p.f. $Q_0(y) = F_0(y) = y$ on $(0, 1)$ against the alternative $\mathcal{A} : F$ with a p.d.f. $f(y)$ positive on $(0, 1)$ and the percentile function $Q(y) = F^{-1}(y)$ increasing on $[0, 1]$. The testing is based on the e.d.f. $F_n(y)$ on $[0, 1]$ with jumps $1/n$ at the points

$$0 < Y_1 \le ... \le Y_n < 1 \tag{4.14}$$

obtained by ordering the random observations

$$X_1, X_2, ..., X_n \quad \text{i.i.d. by the p.d.f. } f(y) \text{ on the observation space } (0, 1). \tag{4.15}$$

Alternatively, the testing can be based on the one-to-one related **empirical percentile function** (briefly, **e.p.f.**)

$$Q_n(y) = F_n^{-1}(y) = \inf\{z \in (0, 1] : F_n(z) \ge y\} \text{ on } [0, 1]. \tag{4.16}$$

Consequently

$$Q_n\left(\frac{j}{n}\right) = Y_j, \quad 0 \le j \le n \tag{4.17}$$

and the disparities, divergences or distances $D_\phi(P, P_0)$ and $D_\phi(P_0, P)$ are given by the formulas

$$D_\phi(P, P_0) = \int_0^1 \phi(p)\,\mathrm{d}y \quad \text{and} \quad D_\phi(P_0, P) = \int_0^1 p\phi\left(\frac{1}{p}\right)\mathrm{d}y. \tag{4.18}$$

This is the the basic conceptual framework for the rest of the chapter.

The information available to the statistician when he faces the problem of testing hypothesis $\mathcal{H}_0 : F_0$ is represented by the d.f.'s F_n and F_0. It is used by him to calculate the decision tool $D_\phi(\boldsymbol{p}_n, \boldsymbol{p}_0)$ or $T_n = m_n D_\phi(\boldsymbol{p}_n, \boldsymbol{p}_0)$ using the rules (4.4), (4.10) and (4.11). Naturally, this procedure is simplified if either the distribution $\boldsymbol{p}_0$ or $\boldsymbol{p}_n$ is uniform,

$$\boldsymbol{p}_0 = \left(\frac{1}{k}, \frac{1}{k}, ..., \frac{1}{k}\right) \quad \text{or} \quad \boldsymbol{p}_n = \left(\frac{1}{k}, \frac{1}{k}, ..., \frac{1}{k}\right) \quad \text{for some } k > 1 \tag{4.19}$$

where k may increase with the sample size n, i.e. the dependence

$$k = k_n \tag{4.20}$$

is admitted but the subscript n is suppressed unless it plays an explicit role. These two possibilities are mutually exclusive for large k and are studied separately in the next two subsections.

4.1 Partitioning by hypothetical percentiles

Let us start with the first case considered in (4.19). It takes place if the partition (4.8) is defined by the $k+1$ cutpoints

$$c_j = Q_0\left(\frac{j}{k}\right) = \frac{j}{k}, \quad 0 \le j \le k \quad \text{for the h.p.f. } Q_0 \text{ (c.f. (4.13))}, \tag{4.21}$$

which are the **hypothetical percentiles** of the uniformly distributed orders $0, 1/k, ..., (k-1)/k, 1$. Thus, the partitioning by uniformly distributed hypothetical percentiles leads to the discrete alternative, empirical and hypothetical distributions

$$\boldsymbol{p} = \left(p_j \equiv F\left(\frac{j}{k}\right) - F\left(\frac{j-1}{k}\right) : 1 \le j \le k\right), \tag{4.22}$$

$$\boldsymbol{p}_n = \left(F_n\left(\frac{j}{k}\right) - F_n\left(\frac{j-1}{k}\right)) : 1 \le j \le k\right), \tag{4.23}$$

$$\boldsymbol{p}_0 = \left(p_{0j} \equiv F_0\left(\frac{j}{k}\right) - F_0\left(\frac{j-1}{k}\right) = \frac{1}{k} : 1 \le j \le k\right) \tag{4.24}$$

respectively. The disparities, divergences or distances $D_\phi(\boldsymbol{p}_n, \boldsymbol{p}_0)$ and $D_\phi(\boldsymbol{p}_0, \boldsymbol{p}_n)$ are given by the formulas

$$D_\phi(\boldsymbol{p}_n, \boldsymbol{p}_0) = k \sum_{j=1}^{k} \phi\left(k p_{nj}\right) \quad \text{for} \quad p_{nj} \equiv F_n\left(\frac{j}{k}\right) - F_n\left(\frac{j-1}{k}\right) \tag{4.25}$$

and

$$D_\phi(\boldsymbol{p}_0, \boldsymbol{p}_n) = \sum_{j=1}^{k} p_{nj} \phi\left(\frac{1}{k p_{nj}}\right) \quad \text{for} \quad p_{nj} \equiv F_n\left(\frac{j}{k}\right) - F_n\left(\frac{j-1}{k}\right) \tag{4.26}$$

where the latter is meaningful only if all p_{nj} are positive.

The probability that all p_{nj} are positive decreases when k increases and vanishes for $k \ge n+1$, since there are only n observations, i.e. F_n has at most n jumps. Therefore the version (4.26) is not used in the sequel and attention is restricted to the goodness-of-fit criteria (4.25) or their scaled versions $T_{\phi,n} = m_n D_\phi(\boldsymbol{p}_n, \boldsymbol{p}_0)$. All examples of statistics $T_{\phi,n} = m_n D_\phi(\boldsymbol{p}_n, \boldsymbol{p}_0)$ given in previous sections are of this type if their cells C_j are specified by the cutpoints (4.21).

4.2 Partitioning by empirical percentiles.

In this subsection we study the second possibility considered in (4.19). It takes place if the partition (4.8) is defined by the same number of $k+1$ cutpoints

$$0 = c_0 < c_1 < ... < c_{k-1} < c_k = 1 \tag{4.27}$$

as in Subsection 4.1, but the first k of them are the **empirical percentiles** of the uniformly distributed orders $0, 1/k, ..., (k-1)/k$, i.e.,

$$c_j = \begin{cases} Q_n\left(\frac{j}{k}\right), \quad 0 \le j \le k-1 \quad \text{for the e.p.f. } Q_n \text{ (c.f. (4.16))} \\ \\ 1 \quad \text{for } j = k \end{cases} \tag{4.28}$$

This formula is not a complete parallel to (4.21) of the previous subsection, because here the last cutpoint is $c_k = 1$ and not $Q_n(k/k) = Q_n(1) = Y_n < 1$. (See in this respect Remark 4.2.1 below.) The cutpoints formula simplifies when k divides n. Unless otherwise explicitly stated, we assume that

$$k = k_n = \frac{n}{m} \quad \text{for } m = m_n = 1, 2, ... \qquad \text{(cf. (4.20))}. \tag{4.29}$$

In accordance with the agreement above, the subscript n is suppressed in k_n, m_n and all expressions involving k_n, m_n unless it is explicitly needed to display.

By the definitions of F_n, F_0, under the assumption (4.29) the cutpoints formula (4.28) implies

$$c_j = Q_n\left(\frac{jm}{n}\right) = Y_{jm}, \quad \text{so that, by the hypothesis of uniformity,} \quad F_0(c_j) = Y_{jm}, \quad 0 \le j \le k-1. \tag{4.30}$$

Hence from (4.10) and (4.27) we get

$$\boldsymbol{p}_n = \left(p_{nj} \equiv F_n(Y_{jm}) - F_n\left(Y_{(j-1)m}\right) = \frac{1}{k} : 1 \le j \le k+1\right) \tag{4.31}$$

because $F_n(Y_{km}) \equiv F_n(Y_n) = F_n(c_k) \equiv 1$, while from (4.11) and (4.27) we get

$$\boldsymbol{p}_0 = \begin{cases} \left(p_{0j} \equiv Y_{jm} - Y_{(j-1)m} : 1 \le j \le k\right) \\ \text{with} \\ Y_{km} \equiv Y_n \text{ replaced by } Y_{n+1} = 1. \end{cases} \tag{4.32}$$

The replacement of $Y_{km} \equiv Y_n$ by $Y_{n+1} = 1$ in (4.32) is necessary, because $c_{k-1} = Y_{(k-1)m}$ but the last cutpoint of (4.27) is $c_k = 1$ and not $Y_{km} \equiv Y_n < 1$, so that

$$F_0(Y_{km}) \equiv F_0(Y_n) < F_0(c_k) \equiv F_0(Y_{n+1}) \equiv 1. \tag{4.33}$$

Notice that under the assumption (4.29) the cutpoints formula (4.28) implies that each cell of the partition of the interval contains exactly m observations.

The disparities, divergences or distances $D_\phi(\boldsymbol{p}_n, \boldsymbol{p}_0)$ and $D_\phi(\boldsymbol{p}_0, \boldsymbol{p}_n)$ of the distributions defined by (4.31) and (4.32) are given by the formulas

$$D_\phi(\boldsymbol{p}_n, \boldsymbol{p}_0) = D_{\phi^*}(\boldsymbol{p}_0, \boldsymbol{p}_n) \quad \text{for} \quad \phi^*(t) = t\phi\left(\frac{1}{t}\right) \quad \text{on } (0, \infty) \tag{4.34}$$

where

$$\begin{aligned} D_\phi(\boldsymbol{p}_0, \boldsymbol{p}_n) &= \sum_{j=1}^{k-1} \frac{1}{k}\phi\left(k(Y_{jm} - Y_{(j-1)m})\right) + \frac{1}{k}\phi\left(k(1 - Y_{(k-1)m})\right) \\ &= \frac{m}{n}\sum_{j=1}^{k} \phi\left(\frac{n}{m}(Y_{jm} - Y_{(j-1)m})\right) + \frac{m}{n}\phi\left(\frac{n}{m}(1 - Y_{(j-1)m})\right). \end{aligned} \tag{4.35}$$

The version (4.34) leads to a too complicated formula in terms of the original function ϕ. Thus for theoretical analysis as well as for practical applications it is more convenient to work with version (4.35). The stochastic differences $Y_{jm} - Y_{(j-1)m}$ are generally referred to as the m-spacings.

The disparities, divergences or distances (4.35) define the m-spacings based goodness-of-fit statistics

$$\begin{aligned} T_\phi^{(m)} &= T_{\phi,n}^{(m)} = nD_\phi(\boldsymbol{p}_0, \boldsymbol{p}_n) \\ &= m\sum_{j=1}^{k-1} \phi\left(\frac{n}{m}(Y_{jm} - Y_{(j-1)m})\right) + m\phi\left(\frac{n}{m}(1 - Y_{(k-1)m})\right). \end{aligned} \tag{4.36}$$

Remark 4.2.1 If in full analogy with (4.21) the cutpoint scheme (4.27, (4.28) is replaced by

$$c_j = Q_n\left(\frac{j}{k}\right), \quad 0 \le j \le k \quad \text{for the e.p.f. } Q_n \text{ (c.f. (4.16)}$$

then $c_k = Y_n < 1$, so that the components $p_{0j} = F_0(c_j) - F_0(c_{j-1}) \equiv c_j - c_{j-1}$ of the hypothetical distribution $\boldsymbol{p}_0$ satisfy the strict inequality

$$\sum_{j=1}^{k} p_{0j} = c_k = Y_n < 1$$

i.e. they cannot be normalized to 1. This can be solved by adding the cutpoint $c_{k+1} = 1$ when the collection of cutpoints

$$0 = c_0 < c_1 < ... < c_{k-1} < c_k < c_{k+1} = 1$$

generates hypothetical and empirical distributions $\boldsymbol{p}_0$ and $\boldsymbol{p}_n$ with $k+1$ components, both normalized to 1, but the last component of the empirical distribution is then

$$p_{n,k+1} = F_n(c_{k+1}) - F_0(c_k) = F_n(1) - F_0(Y_n) \equiv 0,$$

so that the empirical distribution is not uniform. Nevertheless, for $\phi \in \boldsymbol{\Phi}_{\text{div}}$ the ϕ-divergences $D_\phi(\boldsymbol{p}_0, \boldsymbol{p}_n)$ are well defined by the formula

$$D_\phi(\boldsymbol{p}_0, \boldsymbol{p}_n) = D_\phi(\boldsymbol{p}_0, \boldsymbol{p}_n) = \sum_{j=1}^{k} \frac{1}{k}\phi\left(k(Y_{jm} - Y_{(j-1)m})\right) + (1 - Y_n)\phi^*(0) \tag{4.37}$$

for the function $\phi^*(t)$ defined above (for details see e.g. Liese and Vajda (2006)). Thus the ϕ-divergences with finite limit $\phi^*(0) = \lim_{t\to\infty} \phi(t)/t$ define meaningful spacings-based divergence statistics

$$T_\phi^{*(m)} = T_{\phi,n}^{*(m)} = m\sum_{j=1}^{k} \phi\left(\frac{n}{m}(Y_{jm} - Y_{(j-1)m})\right) + (1 - Y_n)\phi^*(0). \tag{4.38}$$

In what follows we would like to deal with more general statistics than just the divergence statistics, so that as starting point we prefer the more universal statistics (4.36).

In the remaining sections we study the properties and applications of the simple-spacings based variants

$$T_\phi = T_{\phi,n} = \sum_{j=1}^{n-1} \phi\left(n(Y_j - Y_{j-1})\right) + \phi\left(n(1 - Y_{n-1})\right) \tag{4.39}$$

of the statistics $T_\phi^{(1)}$ in (4.36), i.e. $T_\phi^{(m)}$ when $m = 1$. We decompose them into **representative parts** R_ϕ and an **asymptotically vanishing parts** V_ϕ as follows:

$$T_\phi = R_\phi + V_\phi \quad \text{with} \quad R_\phi = \sum_{j=1}^{n+1} \phi\left(n\left(Y_j - Y_{j-1}\right)\right), \quad \phi \in \boldsymbol{\Phi} \tag{4.40}$$

and

$$V_\phi = \phi\left(n(1 - Y_{n-1})\right) - \phi\left(n(Y_n - Y_{n-1})\right) - \phi\left(n(1 - Y_n)\right) \tag{4.41}$$

where we put as before $Y_{n+1} = 1$.

Example 4.2.1 The power divergences $D_\alpha(\boldsymbol{p}_0, \boldsymbol{p}_n)$ of orders $\alpha \in \mathbb{R}$ from Example 3.1.3 define spacings-based statistic $T_\alpha = T_{\alpha,n}$ obtained by inserting in (4.39) the generating power functions ϕ_α or $\tilde{\phi}_\alpha$. For example, the power divergence of order 2 with the generating function $\tilde{\phi}_2(t) = (t-1)^2/2$ given in (3.20) defines the spacings-based statistic $T_2 = T_{2,n}$ with the representative part

$$R_2 = R_{2,n} = \frac{1}{2}\sum_{j=1}^{n+1}(n(Y_j - Y_{j-1}) - 1)^2 = \frac{n^2}{2}\sum_{j=1}^{n+1}\left(Y_j - Y_{j-1} - \frac{1}{n}\right)^2. \tag{4.42}$$

Obviously, this is a spacings-based version of the classical Pearson goodness-of-fit statistic

$$T_n = nD_2(\boldsymbol{p}_n, \boldsymbol{p}_0) = \frac{1}{2}n\chi^2(\boldsymbol{p}_n, \boldsymbol{p}_0) \quad \text{(see (2.16))}.$$

Similarly, the power divergence of order 0 with the generating function $\phi_0(t) = -\ln t$ from (3.15) defines the spacings-based statistic $T_0 = T_{0,n}$ with the representative part

$$R_0 = R_{0,n} = -\sum_{j=1}^{n+1}\ln\left(n\left(Y_j - Y_{j-1}\right)\right), \tag{4.43}$$

which is nothing but a spacings-based version of the classical likelihood ratio goodness-of-fit statistic

$$T_n = nD_0(\boldsymbol{p}_0, \boldsymbol{p}_n) = nD_1(\boldsymbol{p}_n, \boldsymbol{p}_0) \quad \text{(see (3.11))}.$$

5 Goodness-of-fit statistics based on spacings

5.1 Objectives of the following sections

This chapter is devoted to the systematic analysis of the disparity and divergence spacings statistics and to their comparisons with the spacings statistics studied in the previous literature. The primary aim is to show that while the motivation of the latter is not based on the concept of similarity between empirical and hypothetical distributions, this idea is in fact hidden somewhere behind because they asymptotically coincide with the former. Therefore the *first objective* is to prove the mutual asymptotic equivalence of the disparity and divergence spacings statistics introduced in this chapter and the spacings statistics known from the literature. This equivalence helps to understand why many ad hoc defined spacings-based statistics exhibit desirable asymptotic properties. The secondary aim is to present in a relatively simple unified manner the asymptotic properties of the many various types of spacings statistics studied in the previous literature. Thus the *second objective* of this chapter is to prove the consistency and asymptotic normality under fixed and local alternatives for a sufficiently wide variety of our spacings-type disparity or divergence statistics. These results are important for applications of the spacings statistics in the testing of goodness-of-fit, and they may also be useful in the estimation of functionals of the type of ϕ-disparity or ϕ-divergence. The last aim is to apply this asymptotic theory to the spacings-based power divergence statistics and compare their asymptotic parameters and properties for various divergence orders $\alpha \in \mathbb{R}$. Therefore the *third objective* is the explicit evaluation of the asymptotic parameters of spacings-based power divergence statistics and an analysis of their properties including their continuity in the parameter $\alpha \in \mathbb{R}$. To achieve all these objectives within a reasonably limited space, we pay the main attention to the simple spacings with $m = 1$ and, starting with Section 6, we deal exclusively with simple spacings.

It seems that the spacings-based goodness-of-fit test statistics given in the literature lacked sofar the motivation of taking into account the notion of disparity between hypothetical and empirical distributions $\boldsymbol{p}_0$ and $\boldsymbol{p}_n$. This contrasts with the goodness-of-fit statistics based on deterministic partitions specified by the uniformly distributed constant cutpoints c_j given in (4.21) and by the related random frequency counts (4.23), 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, all known spacings-based statistic are close to the divergence statistic $T_\phi^{(m)}$ of (4.36) or T_ϕ of (4.39) for some ϕ in $\mathbf{\Phi}_{\rm div}$, but none of them happens to be precisely equal to this divergence statistic. This situation is illustrated in the next examples for the simple-spacings statistics $T_\phi = T_{\phi,n}$ with $\phi \in \mathbf{\Phi}_0$ given by (4.40) as the sum $R_\phi + V_\phi$ where the representative terms

$$R_\phi = \sum_{j=1}^{n+1} \phi\left(n\left(Y_j - Y_{j-1}\right)\right), \quad \phi \in \mathbf{\Phi} \tag{5.1}$$

slightly differ from the statistics known from the literature, which are of the form

$$S_\phi = \sum_{j=1}^{n+1} \phi\left((n+1)\right)\left(Y_j - Y_{j-1}\right)), \quad \phi \in \mathbf{\Phi} \tag{5.2}$$

where ϕ is often from the divergence subclass $\mathbf{\Phi}_{\rm div} \subset \mathbf{\Phi}$. Hence the departure from the divergence statistics is mainly the scaling of the spacings by $n+1$ instead of n. A possible explanation for this is suggested in the next example. We prove in the next section that the departure from the divergence statistics as such is asymptotically negligible in the sense that the so-called asymptotically vanishing term

$$V_\phi = \phi\left(n(1 - Y_{n-1})\right) - \phi\left(n(Y_n - Y_{n-1})\right) - \phi\left(n(1 - Y_n)\right) \tag{5.3}$$

really vanishes asymptotically and the modification of the scaling factor by $n/(n+1)$ is asymptotically negligible.

Example 5.1.1 The first known statistic of the type (5.2) is

$$\mathcal{G} = \sum_{j=1}^{n+1} (Y_j - Y_{j-1})^2 \tag{5.4}$$

of Greenwood (1946) who devised it for testing the hypothesis that the intervals between successive events in epidemiology are exponentially distributed. Obviously,

$$(n+1)^2 \mathcal{G} = S_\phi \equiv \sum_{j=1}^{n+1} \phi\left((n+1)\right)\left(Y_j - Y_{j-1}\right)) \tag{5.5}$$

for $\phi(t) = t^2$ from $\mathbf{\Phi}$. Therefore the Greenwood proposal was neither the divergence nor the disparity spacing statistic. However, Irwin in the discussion of Greenwood (1946), and Kimball (1947) suggested to replace $\mathcal{G}$ by the modification of the power divergence spacing statistic (4.42) defined by

$$\mathcal{K} = \sum_{j=1}^{n+1} \left(Y_j - Y_{j-1} - \frac{1}{n+1}\right)^2 = \frac{2}{(n+1)^2} \sum_{j=1}^{n+1} \tilde{\phi}_2\left((n+1)\right)\left(Y_j - Y_{j-1}\right)) \tag{5.6}$$

for $\tilde{\phi}_2(t) = (t-1)^2/2$ from $\mathbf{\Phi}_{\rm div}$ generating the quadratic spacing statistics R_2 in (4.42). The motivation of the Irwin and Kimball statistic $\mathcal{K}$ may be deduced from the fact that for any real $x_1, x_2, ..., x_{n+1}$

$$E \sum_{j=1}^{n+1} (Y_j - Y_{j-1} - x_j)^2 \geq E \sum_{j=1}^{n+1} \left(Y_j - Y_{j-1} - \frac{1}{n+1}\right)^2$$

and that the inequality is strict unless

$$E\left(Y_j - Y_{j-1}\right) = \frac{1}{n+1} \quad \text{for} \quad 1 \leq j \leq n+1$$

which in turn takes place if and only if $F = F_0$. Therefore the minimal expected values of the Kimball criterion $\mathcal{K}$ characterize the hypothesis F_0 and the larger expected values are reserved for the alternatives $F \neq F_0$.

Example 5.1.2 Another classical spacing statistic which is a slight modification of the power divergence spacings statistic (4.43) was defined by Moran (1951) as

$$\mathcal{M} = S_{\phi_0} = -\sum_{j=1}^{n+1} \ln\left((n+1)\left(Y_j - Y_{j-1}\right)\right) = \sum_{j=1}^{n+1} \phi_0\left((n+1)\left(Y_j - Y_{j-1}\right)\right) \tag{5.7}$$

where $\phi_0(t) = -\ln t$ from $\mathbf{\Phi}_{\text{div}}$ generated the logarithmic spacing statistics R_0 in (4.43).

5.2 Types of statistics studied

As stated above, the analysis of the spacings-based disparity or divergence statistics generated by functions $\phi \in \mathbf{\Phi}$ is in the rest of this chapter restricted to the case of the simple spacings with $m = 1$. In the previous subsection we defined for these spacings three different statistics, viz. T_ϕ of (4.39), R_ϕ of (5.1), and S_ϕ of (5.2). The first of these was obtained by application of the partition of the observation space by n empirical percentiles of equidistant orders to the disparity or divergence $D_\phi(F_n, F_0)$ of the empirical and hypothetical distribution. The remaining two were modifications of T_ϕ representing the spacing statistics from the pioneering work of Greenwood, Kimball and Moran. In the present subsection some other modifications are introduced, which represent the spacing statistics known from the literature subsequent to the mentioned pioneering work. Since those statistics generally used the m-spacings, we return temporarily in this subsection to our disparity or divergence statistics $T_\phi^{(m)}$ from (4.36) based on m-spacings, in order to make the comparisons more realistic and credible.

Let us start with Del Pino's (1979) class of statistics of the form

$$S_\phi^{(m)} = m \sum_{j=1}^{k} \phi\left(\frac{n+1}{m}\left(Y_{mj} - Y_{m(j-1)}\right)\right) \tag{5.8}$$

where it is assumed that $n+1$ is divisible by k and that $m = (n+1)/k \geq 1$. Here the notation in our chapter is consistent in the sense that (5.8) reduces for $m = 1$ to the formula for S_ϕ in (5.2). Del Pino found $\phi(t) = t^2$ to be optimal among the functions ϕ considered by him. The class (5.8) 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 in this regard a wide class of functions ϕ can be admitted, including those with $\phi''(1) = 0$, so that they can be used for functions which generate disparities or divergences. However, as we have seen in Examples 5.1.1 and 5.1.2 for some $\phi \in \mathbf{\Phi}_{\text{div}}$, the statistics (5.8) differ from those in (4.36). Other examples of well-known spacings-based statistics which differ from our spacings-type ϕ-disparity statistics (4.36) 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.

Let us return to our spacings-type ϕ-disparity statistic $T_\phi^{(m)}$ introduced in (4.36). Notice that $T_\phi^{(m)}$ cannot be efficient when $m > 1$, because then it ignores 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 percentiles in (4.30) by a quantity depending on r, we obtain the shifted empirical percentiles

$$c_j^{(r)} = F_n^{-1}\left(\frac{mj+r}{n}\right) = Y_{mj+r}, \quad 1 \leq j \leq k-1, \quad 1 \leq r \leq m-1 \tag{5.9}$$

as cutpoints and, instead of $p_{0j} = Y_{mj} - Y_{m(j-1)} = p_{0j}^{(0)}$, the shifted hypothetical probabilities $p_{0j}^{(r)} = Y_{mj+r} - Y_{m(j-1)+r}$, while still preserving the uniform shifted empirical probabilities $p_{nj}^{(r)} = 1/k = m/n$ on the cells $(c_{j-1}^{(r)}, c_j^{(r)}]$, $1 \leq r \leq m-1$. Replacing each term $\phi(\frac{n}{m}(Y_{mj} - Y_{m(j-1)}))$ in (4.36) by the average

$$\frac{1}{m}\sum_{r=0}^{m-1} \phi\left(\frac{n}{m}(Y_{mj+r} - Y_{m(j-1)+r})\right) \tag{5.10}$$

of all $\phi(np_{0j}^{(r)}/m)$ for $0 \leq r \leq m-1$, we get a potentially more efficient version of $T_\phi^{(m)}$, namely

$$\hat{T}_\phi^{(m)} = \sum_{j=0}^{n-m-1} \phi\left(\frac{n}{m}(Y_{j+m} - Y_j)\right) + m\phi\left(\frac{n}{m}(1 - Y_{n-m})\right) \tag{5.11}$$

which for $m = 1$ reduces to T_ϕ of (4.39), so that the notation of our chapter is again consistent.

A similar procedure can be carried out for $S_\phi^{(m)}$ of (5.8), which involves the observations Y_{mj} for $1 \leq j \leq k$, but ignores the observations Y_{mj+r} for $0 \leq j \leq k-1$ and $1 \leq r \leq m-1$. Applying the averaging and substitution from the previous paragraph, with n replaced by $n+1$ in (5.8), and excluding the terms containing undefined expressions (that is, the terms $Y_{mk+r} - Y_{m(k-1)+r}, 1 \leq r \leq m-1$, where $mk + r > n+1$), we get a similar possibly more efficient version

$$\hat{S}_\phi^{(m)} = \sum_{j=0}^{n-m+1} \phi\left(\frac{n+1}{m}(Y_{j+m} - Y_j)\right) \tag{5.12}$$

of Del Pino's statistic $S_\phi^{(m)}$ of (5.8). Notice that if $m = 1$, then $\hat{S}_\phi^{(m)}$ of (5.12) reduces to S_ϕ of (5.2) above, so that our notation is in this sense still consistent. The statistics (5.12) 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 (5.12) for fixed $m \geq 1$ and eventually also for m slowly tending to ∞ when $n \to \infty$.

In spite of the fact that, when carrying out our analysis above, we went through several important papers (and many other ones listed in these as references), covering altogether four decades of research on spacings-based statistics, we did not in this literature come across the framework of the statistics S_ϕ and its modification R_ϕ when restricting ourselves to the simple spacings with $m = 1$. To make this connection, take into account that if $m > 1$, and in particular if $m \to \infty$, then the statistics (5.12) assign more weight to the 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 to the following two possible extensions of the ordered observations $Y_1, \ldots, Y_n$.

(i) By the formula

$$Y_{n+j} = 1 + Y_j \quad \text{for} \quad j = 1, 2, ..., n \tag{5.13}$$

where the previous dummy observation $Y_0 = 0$ is suppressed and the other dummy observation $Y_{n+1} = 1$ is redefined in accordance with (5.13) 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 $n + 1 - m \leq j \leq n$.

(ii) By the alternative formula

$$Y_{n+j} = 1 + Y_{j-1} \quad \text{for} \quad j = 0, 1, \cdots, n \tag{5.14}$$

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 $1 + Y_{m+j-n-1} - Y_j$ if $n + 2 - m \leq j \leq n$.

Both these extensions of the ordered observations Y_j beyond $j > n$ allow to add in (5.12) the tail evidence missing there by adding to the substituted averages (5.10) also the previously excluded terms. Depending on whether we use (5.13) or the alternative extension (5.14), we get in this manner two different extensions of (5.12), namely

$$\tilde{S}_\phi^{(m)} = \sum_{j=1}^{n} \phi\left((n+1)\,\frac{Y_{j+m}-Y_j}{m}\right) \quad \text{where} \quad Y_{j+m} = 1 + Y_{j+m-n} \tag{5.15}$$

if $j = n+1-m, \cdots, n$, or

$$\tilde{\tilde{S}}_\phi^{(m)} = \sum_{j=0}^{n} \phi\left((n+1)\,\frac{Y_{j+m}-Y_j}{m}\right) \quad \text{where} \quad Y_{j+m} = 1 + Y_{j+m-n-1} \tag{5.16}$$

if $j = n+2-m, \cdots, n$, and $Y_0 = 0$ (cf (5.14)).

The statistics from the class (5.15) were studied for example by Hall (1986) and Morales *et al.* (2003), while those from the class (5.16) were investigated among others by Cressie (1978), Rao and Kuo (1984), Ekström (1999) and Misra and van der Meulen (2001) and others cited there.

Recently Jimenez and Shao (2009) studied for convex functions ϕ the statistics

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

for m, n such that $n+1$ is divisible by k and $m = (n+1)/k \geq 1$. Under the hypothetic p.d.f.'s $F(y) = y$ considered in this chapter these statistics reduce to the statistic $S_\phi^{(m)}$ of Del Pino's (5.8).

As said above, this chapter deals only with the ordinary spacings where $m = 1$. We have seen that then the statistic T_ϕ takes on the form presented in (4.39) and both $S_\phi^{(m)}$ of (5.8) and $\hat{S}_\phi^{(m)}$ of (5.12) reduce to the statistic

$$S_\phi = \sum_{j=1}^{n+1} \phi\left((n+1)\left(Y_j - Y_{j-1}\right)\right), \quad \text{where } Y_{n+1} = 1 \text{ and } Y_0 = 0 \tag{5.17}$$

introduced in (5.2). Consequently, $JS_\phi^{(m)}$ reduces for $m = 1$ to S_ϕ too. It is easy to see that in this case also $\tilde{\tilde{S}}_\phi^{(m)}$ of (5.16) reduces to S_ϕ. However, $\tilde{S}_\phi^{(m)}$ of (5.15) does not do so unless ϕ is linear. Indeed, if $m = 1$, $\tilde{S}_\phi^{(m)}$ reduces to

$$\tilde{S}_\phi = \sum_{j=1}^{n-1} \phi\left((n+1)\left(Y_{j+1} - Y_j\right)\right) + \phi\left((n+1)\left(Y_1 + 1 - Y_n\right)\right) \tag{5.18}$$

which coincides with

$$S_\phi = \sum_{j=1}^{n-1} \phi\left((n+1)\left(Y_{j+1} - Y_j\right)\right) + \phi\left((n+1)\,Y_1\right) + \phi\left((n+1)\left(1 - Y_n\right)\right) \tag{5.19}$$

only if

$$\phi\left((n+1)\,Y_1\right) + \phi\left((n+1)\left(1 - Y_n\right)\right) = \phi\left((n+1)\left(Y_1 + 1 - Y_n\right)\right)$$

which takes place with a positive probability only for linear ϕ. It is to be noted that some of the results described in this chapter follow from the papers dealing with general m-spacing statistics cited in this and the following sections. Our simple proofs are to some extent based on the arguments established in these papers.

In what follows we use the functions

$$\phi^{(n)}(t) = \phi\left(\frac{n+1}{n}\,t\right) \tag{5.20}$$

and, in addition to T_ϕ, S_ϕ, and R_ϕ (introduced in (5.1)), also the statistic $\tilde{S}_\phi$ of (5.18). Moreover, we study another new type of spacings-type disparity statistic denoted by $\tilde{T}_\phi$. To obtain it, we redefine the partition $\mathcal{Q} = \{(0, Y_1], \ldots, (Y_{n-2}, Y_{n-1}], (Y_{n-1}, 1)\}$ of $(0,1)$ given by (4.32) when $m = 1$. The new partition $\mathcal{Q}'$ of $(0,1)$ is obtained by rearranging the n intervals of the partition $\mathcal{Q}$ into n new intervals by the rule

$$(0, Y_1] \mapsto (0, Y_1] \cup (Y_n, 1) \quad \text{and} \quad (Y_{n-1}, 1) \mapsto (Y_{n-1}, Y_n] \tag{5.21}$$

while keeping the remaining intervals $(Y_{j-1}, Y_j], \quad 2 \le j \le n-1$, unaltered. This new partition $\mathcal{Q}'$ leads to the **modified hypothetical distribution**

$$\tilde{\boldsymbol{p}}_0 = (\tilde{p}_{01} = Y_1 + 1 - Y_n,\, \tilde{p}_{02} = Y_2 - Y_1, \ldots, \tilde{p}_{0n} = Y_n - Y_{n-1})$$

but preserves the original uniform empirical distribution $\boldsymbol{p}_n$ on the cells, as each of the new n intervals still contains exactly one of the observations $Y_1, \ldots, Y_n$. Therefore the new partition leads to the new spacings-type disparity statistic

$$\begin{aligned} \tilde{T}_\phi &= n\, D_\phi\left(\tilde{\boldsymbol{p}}_0, \boldsymbol{p}_n\right) = \sum_{j=1}^{n} \phi(n\tilde{p}_{0j}) \\ &= \sum_{j=2}^{n} \phi\left(n(Y_j - Y_{j-1})\right) + \phi\left(n(Y_1 + 1 - Y_n)\right) \end{aligned} \tag{5.22}$$

which differs from T_ϕ of (4.39). Applying (5.20), we obtain the useful relations

$$\tilde{S}_\phi = \tilde{T}_{\phi^{(n)}} \quad \text{and} \quad S_\phi = R_{\phi^{(n)}}. \tag{5.23}$$

In addition to the statistics $R_\phi, S_\phi, \tilde{S}_\phi, T_\phi, \tilde{T}_\phi$, defined above in (5.1), (5.2) , (5.18), (4.39), and (5.22), respectively, we use in this chapter also the auxiliary spacings-based statistics

$$\tilde{R}_\phi = \sum_{j=1}^{n-1} \phi\left(n(Y_{j+1} - Y_j)\right) = R_\phi - \phi(nY_1) - \phi\left(n(1 - Y_n)\right) \tag{5.24}$$

investigated previously by authors neglecting the tail probabilities such as for example Hall (1984). Thus we can conclude this subsection by introducing the sets

$$\mathcal{U}_\phi = \left\{R_\phi, \tilde{R}_\phi, S_\phi, \tilde{S}_\phi, T_\phi, \tilde{T}_\phi\right\}, \quad \phi \in \boldsymbol{\Phi} \tag{5.25}$$

of the spacings-based statistics of the special types introduced here and studied in the following sections. The statistics from $\mathcal{U}_\phi$ are representative in the sense that they cover all known spacings-based statistics studied in the previous literature as special cases when the attention is restricted to the simple spacings.

The set $\boldsymbol{\Phi}$ of differentiable functions $\phi : (0, \infty) \to \mathbb{R}$ was introduced in the Subsection 3.1 by mild additional conditions guaranteeing the existence of the integrals

$$D_\phi(P, P_0) = \int p_0 \phi\left(\frac{p}{p_0}\right) \mathrm{d}\lambda \quad \text{(cf. (3.7))}$$

called disparities in the wide sense, which were justified as measures of disparity, divergence or distance only for ϕ from the subclasses $\boldsymbol{\Phi}_{\text{disp}} \supset \boldsymbol{\Phi}_{\text{div}}$ of $\boldsymbol{\Phi}$. On the other hand, the functions ϕ which defined the statistics $U_\phi \in \mathcal{U}_\phi$ considered in the cited literature imposed on the functions $\phi : (0, \infty) \to \mathbb{R}$ usually somewhat different additional conditions, namely the continuity and the continuous second order differentiability in the neighborhood of 1 with $\phi''(1) \ne 0$ and $\phi(1) = 0$. Therefore in the sequel we study the class of statistics

$$\mathcal{U}_\phi = \left\{R_\phi, \tilde{R}_\phi, S_\phi, \tilde{S}_\phi, T_\phi, \tilde{T}_\phi\right\}, \quad \phi \in \boldsymbol{\Phi}_0 \tag{5.26}$$

where $\boldsymbol{\Phi}_0$ is the set of all continuous functions $\phi : (0, \infty) \mapsto R$ which are twice continuously differentiable in a neighborhood of 1 with $\phi''(1) > 0$ and $\phi(1) = 0$. The replacement of $\boldsymbol{\Phi}$ by $\boldsymbol{\Phi}_0$ means no loss of generality and guarantees that the class (5.26) contains all disparity and divergence statistics $R_\phi, S_\phi, \tilde{S}_\phi$, T_ϕ, $\tilde{T}_\phi$, defined above or, more generally, that

$$\mathbf{\Phi}_{\text{div}} \subset \mathbf{\Phi}_{\text{disp}} \subset \mathbf{\Phi}_0. \tag{5.27}$$

For references later, we summarize the definition formulas

$$R_\phi = R_{\phi,n} = \sum_{j=1}^{n+1} \phi\left(n\left(Y_j - Y_{j-1}\right)\right) \tag{5.28}$$

$$S_\phi = S_{\phi,n} = \sum_{j=1}^{n+1} \phi\left(\left(n+1\right)\left(Y_j - Y_{j-1}\right)\right) \tag{5.29}$$

(e.g. Jammalamadaka et al. (1989), Jiménez and Shao (2009))

$$T_\phi = T_{\phi,n} = \sum_{j=1}^{n-1} \phi\left(n(Y_j - Y_{j-1})\right) + \phi\left(n(1 - Y_{n-1})\right) = nD_\phi(\boldsymbol{p}_0, \boldsymbol{p}_n) \quad \text{(cf. (4.36), (4.39))} \tag{5.30}$$

$$\tilde{R}_\phi = \tilde{R}_{\phi,n} = \sum_{j=2}^{n} \phi\left(n(Y_j - Y_{j-1})\right) \tag{5.31}$$

(e.g. Hall (1984), Cressie (1976) - (1979))

$$\tilde{S}_\phi = \tilde{S}_{\phi,n} = \sum_{j=2}^{n} \phi\left(\left(n+1\right)\left(Y_j - Y_{j-1}\right)\right) + \phi\left(\left(n+1\right)\left(Y_1 + 1 - Y_n\right)\right) \tag{5.32}$$

(e.g. Hall (1986), Morales et al. (2003))

$$\tilde{T}_\phi = \tilde{T}_{\phi,n} = \sum_{j=2}^{n} \phi\left(n(Y_j - Y_{j-1})\right) + \phi\left(n(Y_1 + 1 - Y_n)\right) = nD_\phi(\tilde{\boldsymbol{p}}_0, \boldsymbol{p}_n) \quad \text{(cf. (5.22))} \tag{5.33}$$

where in all these formulas $Y_0 = 0, Y_{n+1} = 1$ and

$$Y_j = F_0(X_{n:j}) \quad \text{for } X_{n:j} \sim F,\ 1 \le j \le n \quad \text{and } \mathcal{H} : F_0. \tag{5.34}$$

5.3 Structural spacings statistics

In this section and in the rest of this chapter we study the subclasses

$$\mathcal{U}_\phi = \left\{R_\phi, \tilde{R}_\phi, S_\phi, \tilde{S}_\phi, T_\phi, \tilde{T}_\phi\right\} \quad \text{for } \phi \in \mathbf{\Phi}_2 \text{ or } \phi \in \mathbf{\Phi}_1 \quad \text{where } \mathbf{\Phi}_2 \subset \mathbf{\Phi}_1 \subset \mathbf{\Phi}_0 \tag{5.35}$$

of the class (5.26) of statistics defined by (5.28) - (5.34). Here $\mathbf{\Phi}_1 \subset \mathbf{\Phi}_0$ is the subset of functions $\phi \in \mathbf{\Phi}_0$ for which there exist functions $\xi, \eta, \zeta : (0, \infty) \mapsto \mathbb{R}$ satisfying the functional equation

$$\phi(st) = \xi(s)\,\phi(t) + \zeta(t)\,\phi(s) + \eta(s)\,(t-1) \quad \text{for all } s,\, t \in (0, \infty). \tag{5.36}$$

The narrower class $\mathbf{\Phi}_2$ consists of all $\phi \in \mathbf{\Phi}_1$ which admit functions $\xi, \eta, \zeta : (0, \infty) \mapsto \mathbb{R}$ satisfying the stronger functional equation

$$\phi(st) = \xi(s)\,\phi(t) + \phi(s) + \eta(s)\,(t-1). \tag{5.37}$$

Assertion 5.3.1 The functions ξ, ζ and η are continuous on $(0, \infty)$ and satisfy the relations

$$\xi(1) = \zeta(1) = 1 \quad \text{and} \quad \eta(1) = 0. \tag{5.38}$$

Proof The continuity of ξ and η can be obtained by putting $t = 2$ and $t = 3$, and that of ζ by putting $s = 2$ in (5.36). If we put $s = 1$ in (5.36) or (5.37) and use the assumption $\phi(1) = 0$, then we obtain that for all $t \in (0, \infty)$

$$(\xi(1) - 1)\,\phi(t) + \eta(1)\,(t - 1) = 0.$$

This contradicts the assumption $\phi''(1) > 0$, unless $\xi(1) = 1$ which implies also $\eta(1) = 0$. By putting $t = 1$ in (5.36) we find that $\zeta(1) = 1$. □

Assertion 5.3.2 Every $\phi \in \mathbf{\Phi}_1$ is differentiable on $(0, \infty)$, the corresponding functions ξ and η are differentiable at 1, and for every $t > 0$

$$\phi'(t) = \xi'(1)\,\frac{\phi(t)}{t} + \phi'(1)\,\frac{\zeta(t)}{t} + \eta'(1)\,\frac{t-1}{t}. \tag{5.39}$$

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 (5.36) for every $t > 0$ and ε close to 0

$$t\,\frac{\phi(t+\varepsilon t) - \phi(t)}{\varepsilon t} = \xi^*(\varepsilon)\,\phi(t) + \frac{\phi(1+\varepsilon) - \phi(1)}{\varepsilon}\,\zeta(t) + \eta^*(\varepsilon)\,(t-1). \tag{5.40}$$

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

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

By the assumptions concerning $\mathbf{\Phi}$, $\phi(t)$ is not linear in a 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 (5.39) follows from (5.40) for all $t > 0$. □

Example 5.3.1 The function $\phi(t) = (1-t)/t$, $t > 0$, belongs to $\mathbf{\Phi}$ and satisfies (5.37) for $\xi(t) = 1/t$ and $\eta(t) \equiv 0$. Therefore it belongs to $\mathbf{\Phi}_2 \subset \mathbf{\Phi}$. The function $\phi(t) = (1-t)^2/t$, $t > 0$, belongs to $\mathbf{\Phi}$ too and satisfies (5.37) for the same $\xi(t)$ as above and $\eta(t) = t - 1/t$. Therefore it belongs to $\mathbf{\Phi}_2$. The functions defined on $(0, \infty)$ by

$$\phi_\alpha(t) = \frac{t^\alpha \ln t}{(2\alpha - 1)}, \quad \alpha \in \mathbb{R} - \{\tfrac{1}{2}\}$$

belong to $\mathbf{\Phi}$ and satisfy (5.36) for $\xi(t) = \zeta(t) = t^\alpha$ and $\eta(t) \equiv 0$. Therefore

$$\{\phi_\alpha : \alpha \in \mathbb{R} - \{\tfrac{1}{2}\}\} \subset \mathbf{\Phi}_1$$

and $\phi_0 \in \mathbf{\Phi}_2$. But ϕ_1 satisfies also (5.37) for $\xi(t) = t$ and $\eta(t) = t \ln t$ and therefore ϕ_1 belongs to $\mathbf{\Phi}_2$.

5.4 Organization of the following sections

The rest of the chapter deals with the asymptotic properties and applications of the classes of statistics considered in (5.35). Let us mention briefly how the following sections are organized. Section 6 establishes the asymptotic equivalence of the statistics from the class $\mathcal{U}_\phi = \left\{R_\phi, \tilde{R}_\phi, S_\phi, \tilde{S}_\phi, T_\phi, \tilde{T}_\phi\right\}$, $\phi \in \mathbf{\Phi}_1$, and presents the general asymptotic theory of the structural statistics from the class $\mathcal{U}_\phi = \left\{R_\phi, \tilde{R}_\phi, S_\phi, \tilde{S}_\phi, T_\phi, \tilde{T}_\phi\right\}$, $\phi \in \mathbf{\Phi}_2$. Section 7 applies this theory to and makes it precise for the power divergence statistics of Example 3.1.3 and comments on comparable results in the previous literature. Section 8 presents a universal program for evaluation of power divergence spacings statistics and their applications in testing the goodness-of-fit. Finally, Section 9 contains proofs of the assertions of Sections 6 and 7.

6 Asymptotic properties of structural statistics

In the remainder of this chapter the observations 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 $[0, 1]$. Case (ii) means that the observations from samples of sizes $n = 1, 2, \dots$ are distributed by distribution functions

$$F^{(n)}(x) = F_0(x) + \frac{L_n(x)}{\sqrt[4]{n}} = x + \frac{L_n(x)}{\sqrt[4]{n}} \tag{6.1}$$

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

$$\sup_{0 \le x \le 1} |\ell_n(x) - \ell(x)| = o(1) \quad \text{as } n \to \infty. \tag{6.2}$$

The two possibilities (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 (6.2) being identically equal to 0 must coincide with the results obtained under the fixed alternative for $F(x) = F_0(x)$.

6.1 Asymptotic equivalence

The theorems below demonstrate that if $\phi \in \boldsymbol{\Phi}_2$ defines a ϕ-divergence or ϕ-disparity, then the statistics S_ϕ, $\tilde{S}_\phi$, R_ϕ and $\tilde{R}_\phi$, which are formally not scaled ϕ-divergences or ϕ-disparities of the hypothetical and empirical distributions F_0 and F_n, share the most important statistical properties with the statistics T_ϕ and $\tilde{T}_\phi$, which are scaled ϕ-divergences or ϕ-disparities of this type . Therefore they provide a key argument for the thesis of the present chapter 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. But the main purpose of the following theorems is to present a systematic asymptotic theory for the whole set of statistics (5.35) and to demonstrate that the small modifications distinguishing these statistics from one another are asymptotically negligible. The restriction to the functions from $\boldsymbol{\Phi}_2$ or even $\boldsymbol{\Phi}_1$ is not essential – it only simplifies the proof of the next theorem.

Assertion 6.1.1. Consider the observations under fixed or local alternatives, and the set of statistics $\{R_\phi, \tilde{R}_\phi, S_\phi, \tilde{S}_\phi, T_\phi, \tilde{T}_\phi\}$ defined in (5.1), (5.24), (5.2), (5.18), (4.39), and (5.22). If $\phi \in \boldsymbol{\Phi}_1$, then for any statistic $U_\phi \in \{R_\phi, S_\phi, \tilde{S}_\phi, T_\phi\}$

$$U_\phi - \tilde{R}_\phi = O_p(1) \quad \text{as } n \to \infty, \tag{6.3}$$

and, if $\phi \in \boldsymbol{\Phi}_2$, then

$$S_\phi - R_\phi = \varepsilon_n R_\phi + \delta_n \quad \text{and} \quad \tilde{S}_\phi - \tilde{T}_\phi = \varepsilon_n \tilde{T}_\phi + \delta_n \tag{6.4}$$

where $\varepsilon_n = o(1)$ and $\delta_n = \phi'(1) + o(1)$ as $n \to \infty$. □

The proofs of this assertion and of the remaining ones of the chapter are deferred to Section 9.

6.2 Assumptions and notations

In this subsection we study the same spacings-type ϕ-disparity statistics $R_\phi, \tilde{R}_\phi, S_\phi, \tilde{S}_\phi, T_\phi$ and $\tilde{T}_\phi$, defined by (5.1), (5.24), (5.2), (5.18), (4.39), and (5.22), for ϕ from $\boldsymbol{\Phi}_2$ or $\boldsymbol{\Phi}_1$ as in the previous subsection. Unless otherwise explicitly stated, these statistics are assumed to be distributed under the fixed or local alternatives introduced as case (i) and case (ii) in the beginning of this section .

For every continuous function $\psi : (0,\infty) \mapsto \mathbb{R}$ we define the condition

$$\lim_{t\to\infty} t^{-\alpha}|\psi(t)| = \lim_{t\downarrow 0} t^{\beta}|\psi(t)| = 0 \quad \text{for some } \alpha \geq 0 \text{ and } \beta < 1 \tag{6.5}$$

and the integral

$$\langle \psi \rangle = \langle \psi(t) \rangle = \int_0^\infty \psi(t)\, e^{-t}\, \mathrm{d}t. \tag{6.6}$$

Obviously, if (6.5) holds then $\langle \psi \rangle$ exists and is finite.

Let $\phi \in \mathbf{\Phi}_1$ satisfy (6.5) and let

$$\xi = \xi_\phi, \quad \zeta = \zeta_\phi \quad \text{and} \quad \eta = \eta_\phi \tag{6.7}$$

be the corresponding functions satisfying the functional equation (5.36). Then all functions

$$\psi(t) = \phi(ts) - \phi(t)\,\zeta(s), \quad s > 0,$$

satisfy (6.5) too, and by (5.36) the linear combinations

$$\psi(t) = \xi(t)\,\phi(s) + \eta(t)\,(s-1), \quad s > 0,$$

of functions $\xi(t)$ and $\eta(t)$ also satisfy (6.5). Since $\phi(s)$ is not linear in the neighborhood of $s = 1$, it follows from here that $\xi(t)$ and $\eta(t)$ themselves satisfy (6.5). Therefore the integrals $\langle \xi \rangle$ and $\langle \eta \rangle$ exist and are finite.

For the fixed alternatives $F \sim f$ we shall consider the linear combinations

$$\mu_\phi(f) = \langle \xi \rangle\, D_\phi(F_0, F) + \langle \phi \rangle\, D_\zeta(F_0, F) \tag{6.8}$$

where

$$D_\phi(F_0, F) = \int_0^1 f(x)\phi\left(\frac{1}{f(x)}\right) \mathrm{d}x \tag{6.9}$$

and

$$D_\zeta(F_0, F) = \int_0^1 f(x)\zeta\left(\frac{1}{f(x)}\right) \mathrm{d}x \tag{6.10}$$

are disparities of the distributions F_0 and F, well defined by (4.1) under the present assumptions about the densities f_0 and f, and are finite. If $\phi(t)$ is convex on $(0,\infty)$, or $\phi(t) - \phi'(1)\,(t-1)$ is monotone on $(0,1)$ and $(1,\infty)$, then $D_\phi(F_0, F)$ is a nonnegative ϕ-divergence or ϕ-disparity of F_0 and F. Similarly, if $\zeta(t)$ is convex on $(0,\infty)$, or $\zeta(t) - \zeta(1) - \zeta'(1)\,(t-1)$ is monotone on $(0,1)$ and $(1,\infty)$, then the ϕ^*-divergence or ϕ^*-disparity of F_0 and F for

$$\phi^*(t) = \zeta(t) - \zeta(1) = \zeta(t) - 1 \quad \text{(cf (5.38))}$$

satisfies the relation $D_{\phi^*}(F_0, F) = D_\zeta(F_0, F) - 1$. Hence the formula for $\mu_\phi(f)$ can be written for every $\phi \in \mathbf{\Phi}_1$ in the more intuitive form

$$\mu_\phi(f) = \langle \xi \rangle\, D_\phi(F_0, F) + \langle \phi \rangle\, [D_{\phi^*}(F_0, F) + 1] \tag{6.11}$$

where ξ and ϕ^* depend on ϕ as specified above, and $D_\phi(F_0, F)$, $D_{\phi^*}(F_0, F)$ are divergences or disparities between the hypothesis F_0 and the alternative F for typical $\phi \in \mathbf{\Phi}_1$. For $\phi \in \mathbf{\Phi}_2 \subset \mathbf{\Phi}_1$ it holds that $\zeta \equiv 1$, so that (6.11) then simplifies to

$$\mu_\phi(f) = \langle \xi \rangle\, D_\phi(F_0, F) + \langle \phi \rangle. \tag{6.12}$$

In particular for $\phi \in \mathbf{\Phi}_2$

$$\mu_\phi(f_0) = \langle \phi \rangle. \tag{6.13}$$

6.3 Consistency under $\mathcal{H}_0$ and fixed alternatives

Assertion 6.3.1 Consider the observations under a fixed alternative $F \sim f$ with f positive and continuous on $[0,1]$, and denote by U_ϕ any statistic from the class $\{R_\phi, \tilde{R}_\phi, T_\phi, \tilde{T}_\phi\}$. If $\phi \in \mathbf{\Phi}_1$ satisfies (6.5), then

$$\frac{U_\phi}{n} \xrightarrow{p} \mu_\phi(f) \quad \text{for } n \to \infty \tag{6.14}$$

where $\mu_\phi(f)$ is given by (6.11). If $\phi \in \mathbf{\Phi}_2$ satisfies (6.5), then the asymptotic relation (6.14) remains valid also for $U_\phi = \tilde{S}_\phi$ and $U_\phi = S_\phi$, and $\mu_\phi(f)$ is given by the simpler formula (6.12). □

Corollary 6.3.1 Under $\mathcal{H}_0$ (6.14)) reduces to

$$\frac{U_\phi}{n} \xrightarrow{p} \mu_\phi(f_0) = \langle \phi \rangle = \int_0^\infty \phi(t)\, e^{-t}\, \mathrm{d}t \qquad \text{for } n \to \infty \tag{6.15}$$

□

In the sequel we use the L_2-norm

$$\|\ell\| = \left(\int_0^1 \ell^2(x)\, \mathrm{d}x \right)^{1/2}$$

and we denote the integral (6.6) usually by $\langle \psi(t) \rangle$ rather than $\langle \psi \rangle$.

6.4 Asymptotic normality under local alternatives

Assertion 6.4.1 Consider the observations under the local alternatives (6.1) with a limit function $\ell(x)$ introduced in (6.2), and denote by U_ϕ any statistic from the set $\{R_\phi, \tilde{R}_\phi, S_\phi, \tilde{S}_\phi, T_\phi, \tilde{T}_\phi\}$. If $\phi \in \mathbf{\Phi}_2$ satisfies the stronger version of (6.5) with $\beta < 1/2$ then

$$\frac{1}{\sqrt{n}}(U_\phi - n\mu_\phi) \xrightarrow{\mathcal{D}} N(m_\phi(\ell), \sigma_\phi^2) \quad \text{as } n \to \infty \tag{6.16}$$

where

$$\mu_\phi = \langle \phi(t) \rangle, \quad \sigma_\phi^2 = \langle \phi^2(t) \rangle - \langle \phi(t) \rangle^2 - (\langle t\phi(t) \rangle - \langle \phi(t) \rangle)^2 \tag{6.17}$$

and

$$m_\phi(\ell) = \frac{\|\ell\|^2}{2} \left(\langle t^2 \phi(t) \rangle - 4\langle t\phi(t) \rangle + 2\langle \phi(t) \rangle \right). \tag{6.18}$$

□

6.5 Asymptotic normality under fixed alternatives

Let us now consider the fixed alternative $F \sim f$ defined at the beginning of this section under (i), and $\phi \in \mathbf{\Phi}_2$ with $\xi = \xi_\phi$, $\eta = \eta_\phi$, satisfying the functional equation (5.37), and denote by ϕ', ξ', η' the derivatives of ϕ, ξ, η as in Assertion 5.3.2. To express the asymptotic normality under this alternative, we need auxiliary functions $\Psi_i = \Psi_{i,\phi}$ of the variable $x \in (0,1)$:

$$\begin{aligned} \Psi_1(x) &= \xi'(1)\, \langle \phi(t) \rangle\, f(x)\, \xi\left(\frac{1}{f(x)}\right) + \xi'(1)\, f(x)\, \phi\left(\frac{1}{f(x)}\right) \\ &\quad + [\phi'(1) - \eta'(1)]\, f(x) + \eta'(1) \end{aligned} \tag{6.19}$$

$$\begin{aligned} \Psi_2(x) &= \left(\langle \phi^2(t) \rangle - \langle \phi(t) \rangle^2 \right) f(x)\, \xi^2\left(\frac{1}{f(x)}\right) + f(x)\, \eta^2\left(\frac{1}{f(x)}\right) \\ &\quad + 2(\langle t\phi(t) \rangle - \langle \phi(t) \rangle) f(x)\, \xi\left(\frac{1}{f(x)}\right) \eta\left(\frac{1}{f(x)}\right), \end{aligned} \tag{6.20}$$

$$\Psi_3(x) = (\langle t\phi(t) \rangle - \langle \phi(t) \rangle)\, \sqrt{f(x)}\, \xi\left(\frac{1}{f(x)}\right) + \sqrt{f(x)}\, \eta\left(\frac{1}{f(x)}\right), \tag{6.21}$$

and also

$$\Psi_4(x) = \frac{\sqrt{f(x)}}{F(x)} \int_0^x \left(1 - \frac{F(y)\, f'(y)}{f^2(y)}\right) \Psi_1(y)\,\mathrm{d}y \tag{6.22}$$

when the alternative density has a continuous derivative $f'(x)$ on $(0,1)$.

Assertion 6.5.1 Consider the observations under the fixed alternative $F \sim f$ with f positive and continuous on $[0,1]$ and continuously differentiable on $(0,1)$ with the derivative f' bounded. If U_ϕ is a statistic from the set $\{R_\phi, \tilde{R}_\phi, S_\phi, \tilde{S}_\phi, T_\phi, \tilde{T}_\phi\}$, and $\phi \in \mathbf{\Phi}_2$ satisfies the stronger version of (6.5) with $\beta < 1/2$, then

$$\frac{1}{\sqrt{n}}(U_\phi - n\mu_\phi(f)) \xrightarrow{\mathcal{D}} N(0, \sigma_\phi^2(f)) \quad \text{as} \quad n \to \infty \tag{6.23}$$

where $\mu_\phi(f)$ is given by (6.12) and

$$\sigma_\phi^2(f) = \int_0^1 \Psi_2(x)\,\mathrm{d}x - 2\int_0^1 \Psi_3(x)\,\Psi_4(x)\,\mathrm{d}x + \int_0^1 \Psi_4^2(x)\,\mathrm{d}x \tag{6.24}$$

for $\Psi_2(x)$, $\Psi_3(x)$ and $\Psi_4(x)$ defined by (6.20) – (6.22). □

Remark 6.5.1 Under the hypothesis $F_0 \sim f_0 \equiv 1$ both Assertions 6.4.1 and 6.5.1 deal with the same statistical model. Therefore, if $f = f_0$, the asymptotic parameters $(\mu_\phi, \sigma_\phi^2)$ from (6.17) and $(\mu_\phi(f_0), \sigma_\phi^2(f_0))$ from (6.12) and (6.24) must be the same, that is, the equalities

$$\mu_\phi(f_0) = \langle\phi\rangle \quad \text{and} \quad \sigma_\phi^2(f_0) = \langle\phi^2\rangle - \langle\phi\rangle^2 - (\langle t\phi(t)\rangle - \langle\phi\rangle)^2$$

must hold. The first equality is clear from (6.13). For $f = f_0$ we get from (9.62) by partial integration

$$\Psi_1(y) = \langle t\phi'(t)\rangle = \langle t\phi(t)\rangle - \langle\phi\rangle \quad \text{for all } y \in (0,1).$$

Thus, by (6.22), $\Psi_4(x)$ is under the hypothesis constant, equal to $\langle t\phi(t)\rangle - \langle\phi\rangle$. Similarly, by (6.20), (6.21) and Assertion 5.3.1, $\Psi_2(x) = \langle\phi^2\rangle - \langle\phi\rangle^2$ and $\Psi_3(x) = \Psi_4(x)$. Hence (6.24) implies the desired result

$$\sigma_\phi^2(f_0) = \Psi_2(x) - 2\Psi_4^2(x) + \Psi_4^2(x) = \sigma_\phi^2.$$

Remark 6.5.2 The expressions μ_ϕ, σ_ϕ^2 are well defined by (6.17) for every continuous function $\phi : (0,\infty) \mapsto \mathbb{R}$ satisfying the condition (6.5) with $\beta < 1/2$. If this condition holds for some function $\psi : (0,\infty) \mapsto \mathbb{R}$, then it holds also for all linear transformations $\phi(t) = a\psi(t) + b(t-1) + c$ and

$$\mu_\phi = a\mu_\psi + c, \quad \sigma_\phi^2 = a^2\sigma_\psi^2. \tag{6.25}$$

Let us now consider a fixed alternative $F \sim f$ with the density continuously differentiable on $(0,1)$. Then, using expression (9.52) for $\mu_\phi(f)$, and (9.54)–(9.56) for $s_i^2(f)$, the formulas

$$\mu_\phi(f) = \int_0^1 f(x) \left\langle \phi\left(\frac{t}{f(x)}\right)\right\rangle \mathrm{d}x \quad \text{and} \quad \sigma_\phi^2(f) = s_1^2(f) + s_2^2(f) + s_3^2(f) \tag{6.26}$$

define $\mu_\phi(f)$ and $\sigma_\phi^2(f)$ for all continuously differentiable functions $\phi : (0,\infty) \mapsto \mathbb{R}$ such that both $\phi(t)$ and $\tilde{\phi}(t) = t\phi'(t)$ satisfy (6.5) with $\beta < 1/2$. If ψ is one of the functions satisfying all these conditions then all linear transformations $\phi(t) = a\psi(t) + b(t-1) + c$ satisfy these conditions too and

$$\mu_\phi(f) = a\mu_\psi(f) + c, \quad \sigma_\phi^2(f) = a^2\sigma_\psi^2(f). \tag{6.27}$$

Formulas (6.25) and (6.27) are verifiable from the definitions mentioned in this remark and are useful for the evaluation of asymptotic means and variances.

Remark 6.5.3 We observe that the asymptotic results of Assertions 6.3.1, 6.4.1 and 6.5.1 are in each case for a fixed ϕ the same for *any* statistic U_ϕ from the class of statistics considered, confirming the asymptotic equivalence of these statistics.

7 Asymptotic properties of power spacings statistics

7.1 Power spacing statistics

In the rest of this chapter we deal with and frequently refer to the statistics generated by the power functions $\phi = \tilde{\phi}_\alpha$ introduced in (3.20), (3.21) as nonnegative linear transforms of the simpler functions $\phi = \phi_\alpha$ defined by (3.14), (3.15). In order to simplify the notations, we intrchange the symbols $\phi_\alpha \longleftrightarrow \tilde{\phi}_\alpha$, i.e. we use the functions defined by

$$\phi_\alpha(t) = \frac{t^\alpha - \alpha(t-1) - 1}{\alpha(\alpha-1)} \quad \text{if } \alpha \notin \{0,1\}, \tag{7.1}$$

and otherwise by the corresponding limits

$$\phi_1(t) = t\ln t - t + 1 \quad \text{and} \quad \phi_0(t) = -\ln t + t - 1, \tag{7.2}$$

and their simpler alternatives

$$\tilde{\phi}_\alpha(t) = \frac{t^\alpha - 1}{\alpha(\alpha-1)} \quad \text{for} \quad \alpha \notin \{0,1\}, \quad \tilde{\phi}_1(t) = t\ln t, \quad \tilde{\phi}_0(t) = -\ln t. \tag{7.3}$$

The rest of this chapter pays special attention to the subclass

$$\mathcal{U}_\alpha = \{R_{\phi_\alpha}, \tilde{R}_{\phi_\alpha}, S_{\phi_\alpha}, \tilde{S}_{\phi_\alpha}, T_{\phi_\alpha}, \tilde{T}_{\phi_\alpha}\}, \quad \alpha \in \mathbb{R} \tag{7.4}$$

of the spacings-based structural disparity statistics studied in the previous section which are generated by the power functions $\phi = \phi_{\ \alpha} : (0,\infty) \mapsto \mathbb{R}$ defined for all powers $\alpha \in \mathbb{R}$ byas in (3.20), (3.21). It is easy to verify that these functions belong to the subset $\mathbf{\Phi}_2$, that is, they satisfy the functional equation (5.37) with

$$\xi(t) = \xi_\alpha(t) = t^\alpha \quad \text{and} \quad \eta(t) = \eta_\alpha(t) = \begin{cases} \frac{t^\alpha - t}{\alpha - 1} & \text{if} \quad \alpha \neq 1 \\ \lim\limits_{\alpha\to 1} \frac{t^\alpha - t}{\alpha-1} = t\ln t & \text{if} \quad \alpha = 1 \end{cases} \tag{7.5}$$

In other words, if $\alpha \in \mathbb{R}$ then

$$\phi_\alpha(st) = s^\alpha\phi_\alpha(t) + \phi_\alpha(s) + (t-1)\cdot \begin{cases} \frac{s^\alpha - s}{\alpha-1} & \text{if} \quad \alpha \neq 1 \\ s\ln s & \text{if} \quad \alpha = 1 \end{cases} \tag{7.6}$$

for all s, $t > 0$.

It is also easy to verify that the functions $\phi_{\ \alpha}$, $\alpha \in \mathbb{R}$ are convex, belong to $\mathbf{\Phi}_{\text{div}}$ and define $\phi_{\ \alpha}$-divergences (or briefly, α-divergences). Referring to definitions (3.1) and (3.2), we introduce the following simplified notation for these $\phi_{\ \alpha}$-divergences:

$$D_\alpha(\boldsymbol{p}, \boldsymbol{p}_0) = D_{\phi_\alpha}(\boldsymbol{p}, \boldsymbol{p}_0) = \frac{1}{\alpha(\alpha-1)}\left(\sum_{j=1}^k p_j^\alpha p_{0j}^{1-\alpha} - 1\right) = D_{1-\alpha}(\boldsymbol{p}_0, \boldsymbol{p}) \quad (\text{cf. } (3.4))$$

if $\alpha \notin \{0,1\}$, and

$$D_1(\boldsymbol{p}, \boldsymbol{p}_0) = D_{\phi_1}(\boldsymbol{p}, \boldsymbol{p}_0) = \sum_{j=1}^k p_j \ln\frac{p_j}{p_{0j}} = D_0(\boldsymbol{p}_0, \boldsymbol{p})$$

otherwise. Similarly (cf. (3.2)),

$$D_0(F_0, F) = D_{\phi_0}(F_0, F) = \int_0^1 f\ln\frac{f}{f_0}\,\mathrm{d}x = \int_0^1 f(x)\ln f(x)\,\mathrm{d}x, \tag{7.7}$$

$$D_1(F_0, F) = D_{\phi_1}(F_0, F) = \int_0^1 f_0\ln\frac{f_0}{f}\,\mathrm{d}x = -\int_0^1 \ln f(x)\,\mathrm{d}x, \tag{7.8}$$

$$\begin{aligned} D_\alpha(F_0, F) &= D_{\phi_\alpha}(F_0, F) = \frac{1}{\alpha(\alpha-1)}\left(\int_0^1 f\left(\frac{f_0}{f}\right)^\alpha \mathrm{d}x - 1\right) \\ &= \frac{1}{\alpha(\alpha-1)}\left(\int_0^1 f(x)^{1-\alpha}\mathrm{d}x - 1\right) \quad \text{if} \quad \alpha \notin \{0,1\}. \end{aligned} \tag{7.9}$$

Similar to the corresponding ϕ_{α}-divergences themselves, the ϕ_{α}-divergence statistics T_{ϕ_α}, $\tilde{T}_{\phi_\alpha}$ and S_{ϕ_α} are not altered if the nonnegative convex functions $\phi_{\alpha} \in \mathbf{\Phi}_2$ are replaced by the convex functions $\phi_\alpha(t)$ from $\mathbf{\Phi}_2$ given by (7.3).

For references later we present formulas for selected statistis from the class (7.4). In the first set are our true divergence alternatives

$$T_{\phi_\alpha} = T_{\phi_\alpha,n} = \frac{1}{\alpha(\alpha-1)}\left[n^\alpha\left(\sum_{j=1}^{n-1}(Y_j - Y_{j-1})^\alpha + (1 - Y_{n-1})^\alpha\right) - n\right] \tag{7.10}$$

$$T_{\phi_1} = T_{\phi_1,n} = \sum_{j=1}^{n-1} n(Y_j - Y_{j-1}) \ln\left[n(Y_j - Y_{j-1})\right] + n(1 - Y_{n-1}) \ln\left[n(1 - Y_{n-1})\right] \tag{7.11}$$

$$T_{\phi_0} = T_{\phi_0,n} = -\sum_{j=1}^{n-1} \ln\left[n(Y_j - Y_{j-1})\right] - \ln\left[(1 - Y_{n-1})\right] \tag{7.12}$$

(cf. (5.30)). In the second set are the modified divergence statistics

$$S_{\phi_\alpha} = S_{\phi_\alpha,n} = \frac{1}{\alpha(\alpha-1)}\left[(n+1)^\alpha \sum_{j=1}^{n+1}(Y_j - Y_{j-1})^\alpha - n - 1\right] \tag{7.13}$$

$$S_{\phi_1} = S_{\phi_1,n} = \sum_{j=1}^{n+1}(n+1)(Y_j - Y_{j-1}) \ln\left[n(Y_j - Y_{j-1})\right] \tag{7.14}$$

$$S_{\phi_0} = S_{\phi_0,n} = -\sum_{j=1}^{n+1} \ln\left[(n+1)(Y_j - Y_{j-1})\right] \tag{7.15}$$

(cf. (5.29)) extensively used in the literature (cf. Jammalamadaka et al. (1986), (1989), Misra and van der Meulen (2001), Jiménez and Shao (2009) and others cited there). In the third set are again the true divergence statistics

$$\tilde{T}_{\phi_\alpha} = \tilde{T}_{\phi_\alpha,n} = \frac{1}{\alpha(\alpha-1)}\left[n^\alpha\left(\sum_{j=2}^{n}(Y_j - Y_{j-1})^\alpha + (Y_1 + 1 - Y_n)^\alpha\right) - n\right] \tag{7.16}$$

$$\tilde{T}_{\phi_1} = \tilde{T}_{\phi_1,n} = \sum_{j=2}^{n} n(Y_j - Y_{j-1}) \ln\left[n(Y_j - Y_{j-1})\right] + n(Y_1 + 1 - Y_n) \ln\left[n(Y_1 + 1 - Y_n)\right] \tag{7.17}$$

$$\tilde{T}_{\phi_0} = \tilde{T}_{\phi_0,n} = -\sum_{j=2}^{n} \ln\left[n(Y_j - Y_{j-1})\right] - \ln\left[n(Y_1 + 1 - Y_n)\right] \tag{7.18}$$

(cf. (5.33)) slightly different from those used by Hall (1986), Morales et al. (2003), Vajda and van der Meulen (2006), Vajda (2007) and others cited there (they differ by the normalizing constant n instead of $n+1$).

Example 7.1.1 For $\alpha = 2$ we obtain the statistic

$$S_{\phi_2} = \frac{1}{2}\left[(n+1)^2 \sum_{j=1}^{n+1}(Y_j - Y_{j-1})^2 - (n+1)\right] = \frac{n+1}{2}\left[(n+1)\mathcal{G} - 1\right] \tag{7.19}$$

where $\mathcal{G}$ is the Greenwood statistic of Example 5.1.1.

Since the general asymptotic theory of the statistics $U_\alpha \in \mathcal{U}_\alpha$ specified by (7.4) is covered by Assertion 5.3.2 and Assertions 6.3.1–6.5.1, the assertions that follow in this section are basically their corollaries.

However, they bring explicit formulas and additional important new results, the proofs of which are not trivial. These proofs are partly based on a continuity theory for the asymptotic parameters

$$\mu_\alpha(f) = \mu_{\phi_\alpha}(f),\ \sigma_\alpha^2(f) = \sigma_{\phi_\alpha}^2(f),\ \mu_\alpha = \mu_{\phi_\alpha},\ \sigma_\alpha^2 = \sigma_{\phi_\alpha}^2 \text{ and } m_\alpha(\ell) = m_{\phi_\alpha}(\ell), \tag{7.20}$$

defined by (6.26), (6.17) and (6.18), as functions of the structural parameter $\alpha \in \mathbb{R}$. Such a theory enables us to avoid a direct calculation of the asymptotic parameters at some $\alpha_0 \in \mathbb{R}$, if these calculations are tedious and the asymptotic parameters are known at the neighboring parameters α. This theory is summarized in Assertion 7.1.2 below using Assertion 7.1.1. In Assertion 7.1.2 we take the representations (6.12) and (6.24) for $\mu_{\phi_\alpha}(f)$ and $\sigma_{\phi_\alpha}^2(f)$ rather than (6.26).

Assertion 7.1.1 Let $g(y)$ be a continuous positive function on a compact interval $[a, b] \subset \mathbb{R}$ and $\Phi(u, v)$ a continuous function of variables $u, v \in \mathbb{R}$. Furthermore let, for all α from an interval $(c, d) \subset \mathbb{R}$, $\psi_\alpha : (0, \infty) \mapsto \mathbb{R}$ be convex or concave functions differentiable at some point $t_* \in (0, \infty)$. If the values $\psi_\alpha(t)$, $t \in (0, \infty)$, and the derivatives $\psi'_\alpha(t_*)$ depend continuously on $\alpha \in (c, d)$, then for every $\alpha_0 \in (c, d)$

$$\lim_{\alpha \to \alpha_0} \int_a^b \Phi(g, \psi_\alpha(g))\, \mathrm{d}y = \int_a^b \Phi(g, \psi_{\alpha_0}(g))\, \mathrm{d}y. \tag{7.21}$$

□

Assertion 7.1.2 The asymptotic parameters μ_α, σ_α^2 and $m_\alpha(\ell)$, specified by (7.20), (6.17) and(6.18), are continuous in the variable $\alpha \in (-1/2, \infty)$. If the density f satisfies the assumptions of Assertion 6.3.1, then the asymptotic mean $\mu_\alpha(f)$ specified by (7.20) and (6.12) is continuous in the variable $\alpha \in (-1, \infty)$. If f satisfies the stronger assumptions of Assertion 6.5.1, then the asymptotic variance $\sigma_\alpha^2(f)$ specified by (7.20) and (6.24) is continuous in the variable $\alpha \in (-1/2, \infty)$. □

7.2 Consistency

In the assertion below and in the rest of the chapter, we use the gamma function of the variable $\alpha \in \mathbb{R}$ and the Euler constant,

$$\Gamma(\alpha) = \int_0^\infty t^{\alpha-1} e^{-t}\, \mathrm{d}t \quad \text{and} \quad \gamma = 0.577\ldots. \tag{7.22}$$

Assertion 7.2.1 Consider the observations under the fixed alternative $F \sim f$ assumed in Assertion 6.3.1 and denote by U_α any statistic from the class $\mathcal{U}_\alpha$ of (7.4). If $\alpha > -1$, then

$$\frac{U_\alpha}{n} \xrightarrow{p} \mu_\alpha(f) \quad \text{as } n \to \infty \tag{7.23}$$

for

$$\mu_\alpha(f) = D_\alpha(F_0, F)\, \Gamma(\alpha + 1) + \mu_\alpha, \tag{7.24}$$

where

$$\mu_0 = \gamma, \quad \mu_1 = 1 - \gamma, \quad \text{and} \quad \mu_\alpha = \frac{\Gamma(\alpha+1) - \Gamma(1)}{\alpha(\alpha-1)} \quad \text{for } \alpha \notin \{0, 1\}, \tag{7.25}$$

and $D_\alpha(F_0, F)$ are the ϕ_α-divergences (7.7)-(7.9). The ϕ_α-divergences are zero if and only if $F = F_0$, so that under the hypothesis $\mathcal{H}_0 : F = F_0$

$$\mu_\alpha(f_0) = \mu_\alpha, \quad \alpha \in \mathbb{R}. \tag{7.26}$$

Both parameters μ_α and $\mu_\alpha(f)$ are continuous in the variable $\alpha \in (-1, \infty)$ and satisfy the inequality $\mu_\alpha(f) \geq \mu_\alpha$, which is strict unless $F = F_0$. □

Since $\Gamma(\alpha + 1) = \alpha(\alpha - 1)\,\Gamma(\alpha - 1)$, (7.25) and (7.24) can be replaced for $\alpha \notin \{0, 1\}$ by

$$\mu_\alpha = \Gamma(\alpha - 1) - \frac{1}{\alpha(\alpha-1)} \quad \text{and} \quad \mu_\alpha(f) = \Gamma(\alpha - 1) \int_0^1 f^{1-\alpha}\, \mathrm{d}x - \frac{1}{\alpha(\alpha-1)}. \tag{7.27}$$

Assertion 7.2.1 can be illustrated by Table 7.2.1 , in which actual values of the parameters μ_α and $\mu_\alpha(f)$ are presented for selected parameters α. In this table, f denotes any density considered in Assertion 6.3.1, and the expressions for $D_\alpha(F_0, F)$, $H(F_0, F)$, and $\chi^2(F_0, F)$ can be easily discerned from those used in Examples 3.1.1 and 3.1.3, thereby replacing P and P_0 by F_0 and F and sums by integrals.

Table 7.2.1 Values of μ_α and $\mu_\alpha(f)$ for selected $\alpha > -1$.

α	μ_α	$\mu_\alpha(f)$
$-\frac{1}{2}$	$\frac{4}{3}(\sqrt{\pi}-1) \doteq 1.030$	$\sqrt{\pi}\, D_{-1/2}(F_0, F) + \mu_{-1/2} = \frac{4\sqrt{\pi}}{3} \int_0^1 f^{3/2}\, \mathrm{d}x - \frac{4}{3}$
0	$\gamma \doteq 0.577$	$D_0(F_0, F) + \mu_0 = \int_0^1 f \ln f\, \mathrm{d}x + \gamma$
$\frac{1}{2}$	$4 - 2\sqrt{\pi} \doteq 0.455$	$2\sqrt{\pi}\, H(F_0, F) + \mu_{1/2} = 4 - 2\sqrt{\pi} \int_0^1 \sqrt{f} \mathrm{d}x$
1	$1 - \gamma \doteq 0.423$	$D_1(F_0, F) + \mu_1 = 1 - \gamma - \int_0^1 \ln f\, \mathrm{d}x$
$\frac{3}{2}$	$\sqrt{\pi} - \frac{4}{3} \doteq 0.439$	$\frac{3\sqrt{\pi}}{4} D_{3/2}(F_0, F) + \mu_{3/2} = \sqrt{\pi} \int_0^1 \frac{\mathrm{d}x}{\sqrt{f}} - \frac{4}{3}$
2	$\frac{1}{2} = 0.500$	$\chi^2(F_0, F) + \mu_2 = \int_0^1 \frac{\mathrm{d}x}{f} - \frac{1}{2}$
$\frac{5}{2}$	$\frac{\sqrt{\pi}}{2} - \frac{4}{15} \doteq 0.620$	$\frac{15\sqrt{\pi}}{8} D_{5/2}(F_0, F) + \mu_{5/2} = \frac{\sqrt{\pi}}{2} \int_0^1 \frac{\mathrm{d}x}{f^{3/2}} - \frac{4}{15}$
3	$\frac{5}{6} \doteq 0.833$	$6D_3(F_0, F) + \mu_3 = \int_0^1 \frac{\mathrm{d}x}{f^2} - \frac{1}{6}$
4	$\frac{23}{12} \doteq 1.917$	$24D_4(F_0, F) + \mu_4 = 2\int_0^1 \frac{\mathrm{d}x}{f^3} - \frac{1}{12}$

7.3 Asymptotic normality under local alternatives

Assertion 7.3.1 Consider the observations under the local alternatives (6.1) with the limit function $\ell(x)$ introduced in (6.2), and denote by U_α any statistic from the class $\mathcal{U}_\alpha$ of (7.4). If $\alpha > -1/2$, then

$$\frac{1}{\sqrt{n}}(U_\alpha - n\mu_\alpha) \xrightarrow{\mathcal{D}} N(m_\alpha(\ell), \sigma_\alpha^2) \qquad \text{as } n \to \infty \tag{7.28}$$

where the parameters μ_α, $m_\alpha(\ell)$, and σ_α^2 are continuous in the variable $\alpha \in (-1/2, \infty)$, and are given by (7.25) and the formulas

$$m_\alpha(\ell) = \frac{\|\ell\|^2}{2}\,\Gamma(\alpha+1) \tag{7.29}$$

$$\sigma_\alpha^2 = \frac{\Gamma(2\alpha+1) - (\alpha^2+1)\,\Gamma^2(\alpha+1)}{\alpha^2(\alpha-1)^2} \qquad \text{for } \alpha \notin \{0, 1\} \tag{7.30}$$

and

$$\sigma_0^2 = \frac{\pi^2}{6} - 1, \qquad \sigma_1^2 = \frac{\pi^3}{3} - 3. \tag{7.31}$$

□

Assertion 7.3.1 provides the possibility to compute and compare asymptotic relative efficiencies of tests of the hypothesis $\mathcal{H}_0 : F_0 \sim f_0$ based on the statistics $U_\alpha \in \mathcal{U}_\alpha$, $\alpha > -1/2$, for various values of α. The Pitman asymptotic relative efficiency (ARE) of one test relative to another is defined as the limit of the inverse ratio of sample sizes required to obtain the same limiting power at the sequence of alternatives converging to the null hypothesis. If we define the "efficacies" of the statistics $U_\alpha \in \mathcal{U}_\alpha$ of Assertion 7.3.1 by

$$\mathrm{eff}(U_\alpha) = \frac{\Gamma^2(\alpha+1)}{\sigma_\alpha^2} = \frac{(m_\alpha(\ell))^2}{\sigma_\alpha^2}\left(\frac{2}{\|\ell\|^2}\right)^2 \qquad \text{for } \|\ell\|^2 \neq 0$$

then under the assumptions of Assertion 7.3.1 we get in accordance with Section 4 in Del Pino (1979)

$$\mathrm{ARE}(U_{\alpha_1}, U_{\alpha_2}) = \frac{\mathrm{eff}(U_{\alpha_1})}{\mathrm{eff}(U_{\alpha_2})}$$

where U_{α_1} and U_{α_2} are arbitrary statistics from $\mathcal{U}_{\alpha_1}$ and $\mathcal{U}_{\alpha_2}$. Notice that arbitrary statistics U_α from the set $\mathcal{U}_\alpha$, α fixed, all have the same efficacy (cf. also Remark 6.5.3). In Table 7.3.1 we present the parameters $m_\alpha(\ell)$, σ_α^2 and $\Gamma^2(\alpha+1)/\sigma_\alpha^2$ for selected values of $\alpha > -1/2$. This table indicates that the statistics $U_2 \in \{R_{\phi_2}, \tilde{R}_{\phi_2}, S_{\phi_2}, \tilde{S}_{\phi_2}, T_{\phi_2}, \tilde{T}_{\phi_2}\}$ are most asymptotically efficient in the Pitman sense among all statistics U_α, $\alpha > -1/2$. This extends the result on p. 1457 in Rao and Kuo (1984) about the asymptotic efficiency of the Greenwood statistic $\mathcal{G} = (2S_{\phi_2} + n + 1)/(n+1)^2$ (cf. Example 5.1.1 (formula (5.4)), Example 7.1.1 (formula (7.19)), and formula (7.46) below).

Table 7.3.1 The asymptotic parameters $m_\alpha(\ell)$, σ_α^2 and eff(U_α) for selected statistics U_α of Assertion 7.3.1.

α	$m_\alpha(\ell)$	σ_α^2	eff(U_α)
0	$\frac{\|\ell\|^2}{2}$	$\frac{\pi^2}{6} - 1 \doteq 0.645$	1.550
$\frac{1}{2}$	$\|\ell\|^2 \frac{\sqrt{\pi}}{4} \doteq \frac{\|\ell\|^2}{2} \times 0.886$	$16 - 5\pi \doteq 0.292$	2.690
1	$\frac{\|\ell\|^2}{2}$	$\frac{\pi^2}{3} - 3 \doteq 0.290$	3.448
$\frac{3}{2}$	$\|\ell\|^2 \frac{3\sqrt{\pi}}{8} \doteq 1.329$	$\frac{32}{3} - \frac{13\pi}{4} \doteq 0.457$	3.871
2	$\|\ell\|^2 = \frac{\|\ell\|^2}{2} \times 2$	1	4.000
$\frac{5}{2}$	$\|\ell\|^2 \frac{15\sqrt{\pi}}{16} \doteq \frac{\|\ell\|^2}{2} \times 3.323$	$\frac{128}{15} - \frac{29\pi}{16} \doteq 2.839$	3.890
3	$\|\ell\|^2\, 3 = \frac{\|\ell\|^2}{2} \times 6$	10	3.600
4	$\|\ell\|^2\, 12 = \frac{\|\ell\|^2}{2} \times 24$	212	2.717

The general form of the asymptotic normality (7.28), as well as the continuity of the parameters μ_α, $m_\alpha(\ell)$ and σ_α^2 in $\alpha \in (-1/2, \infty)$ established in Assertion 7.3.1 appear to be new results. The special result for $\alpha = 0$ also seems to be new. The particular result for $\alpha \in (-1/2, \infty) - \{0, 1\}$ and $U_\alpha = S_{\phi_\alpha}$ follows from the asymptotic normality obtained for the statistics

$$\sum_{j=1}^{n+1} \left((n+1)\,(Y_j - Y_{j-1})\right)^\alpha = \alpha(\alpha-1)\, S_{\phi_\alpha} + n + 1 \tag{7.32}$$

(cf. (7.47) below) by Del Pino, see p. 1062 in Del Pino (1979). The particular result for $\alpha = 1$ and the statistic $U_1 = S_{\phi_1}$ with μ_1 and σ_1^2 given in Tables 7.2.1 and 7.3.1 was obtained previously by Misra and van der Meulen (2001), who however considered m-spacings for arbitrary $m \geq 1$. They compared also the efficiency of the test statistics for $\alpha = 0$, $\alpha = 1$ and $\alpha = 2$ with a similar conclusion as in Table 7.3.1.

7.4 Asymptotic normality under fixed alternatives

In this subsection we study the asymptotic distributions of the spacings-type power divergence statistics U_α from the sets $\mathcal{U}_\alpha = \{R_{\phi_\alpha}, \tilde{R}_{\phi_\alpha}, S_{\phi_\alpha}, \tilde{S}_{\phi_\alpha}, T_{\phi_\alpha}, \tilde{T}_{\phi_\alpha}\}$ for $\alpha > -1/2$ under the assumption that the observations are distributed by a fixed alternative $F \sim f$ satisfying the assumptions of Assertion 6.5.1. If $\alpha > -1/2$ then ϕ_α satisfies the assumption of Assertion 6.5.1 too. Therefore this theorem implies that

$$\frac{1}{\sqrt{n}}(U_\alpha - n\mu_\alpha(f)) \xrightarrow{\mathcal{D}} N(0, \sigma_\alpha^2(f)) \quad \text{for } n \to \infty \tag{7.33}$$

where the asymptotic parameters $\mu_\alpha(f)$, $\sigma_\alpha^2(f)$ are given by (7.20). Similarly as in the previous section, we are interested in explicit formulas for these parameters. By Assertion 6.5.1, the asymptotic mean is for all $\alpha \in \mathbb{R}$ given by the explicit formula (7.24) presented in Assertion 7.2.1. The only problem which remains is the formula for $\sigma_\alpha^2(f)$, $\alpha \in \mathbb{R}$.

The functions $\psi_\alpha(t) = t^\alpha$ with $\alpha > -1/2$ satisfy all assumptions of Remark 6.5.2 so that we can consider the quantities

$$\tau_\alpha^2(f) \equiv \sigma_{\psi_\alpha}^2(f), \quad \alpha \in (-1/2, \infty)$$

defined there. By (6.27),

$$\sigma_\alpha^2(f) = \frac{\tau_\alpha^2(f)}{\alpha^2(\alpha-1)^2} \quad \text{for } \alpha \in (-1/2, \infty) - \{0, 1\}. \tag{7.34}$$

One can find on p. 521 of Hall (1984) an expression for $\tau_\alpha^2(f)$ for all $\alpha \in (-1/2, \infty) - \{0, 1\}$, which for the case $m = 1$ can be given the form

$$\tau_\alpha^2(f) = \alpha^2(\alpha-1)^2 \left(\sigma_\alpha^2 \int_0^1 f^{1-2\alpha} \mathrm{d}x + \Gamma^2(\alpha+1)\, \Delta_\alpha(F_0, F) \right) \tag{7.35}$$

where σ_α^2 is defined by formula (7.30) and

$$\Delta_\alpha(F_0, F) = \frac{1}{\alpha^2} \int_0^1 \left(\frac{1}{(f(x))^\alpha} - \frac{1}{F(x)} \int_0^x (f(y))^{1-\alpha} \mathrm{d}y \right)^2 f(x)\, \mathrm{d}x \quad \text{for } \alpha \in \mathbb{R} - \{0\}. \tag{7.36}$$

Since Hall (1984) gave no hint about the derivation of his formula, let us mention that (7.35) is obtained if one substitutes ψ_α for ϕ in the expressions (9.54)– (9.56) below for $s_j^2(f)$, $j \in \{1, 2, 3\}$, given in the proof of Assertion 6.5.1 (thereby employing the expression

$$\begin{aligned} G(x) &= \alpha E(Z^\alpha) \int_0^x \left(1 - \frac{F f'}{f^2} \right) \frac{1}{f^{\alpha-1}} \, \mathrm{d}y \\ &= \Gamma(\alpha+1) \left((\alpha-1) \int_0^x (f(y))^{1-\alpha} \mathrm{d}y + (f(x))^{-\alpha} F(x) \right) \end{aligned}$$

for $G(x)$ of (9.53) when ϕ is replaced by ψ_α), and then forms the sum $s_1^2(f) + s_2^2(f) + s_3^2(f)$. By (7.34) and (7.35),

$$\sigma_\alpha^2(f) = \sigma_\alpha^2 \int_0^1 f^{1-2\alpha} \mathrm{d}x + \Gamma^2(\alpha+1)\, \Delta_\alpha(F_0, F), \quad \alpha \in (-1/2, \infty) - \{0, 1\}. \tag{7.37}$$

The final, intuitively appealing, form of the asymptotic variance

$$\sigma_\alpha^2(f) = (1 + 2\alpha(2\alpha-1)\, D_{2\alpha}(F_0, F))\, \sigma_\alpha^2 + \Gamma^2(\alpha+1)\, \Delta_\alpha(F_0, F) \tag{7.38}$$

(with $\sigma_\alpha^2(f_0) = \sigma_\alpha^2$ given in (7.30)), follows for $\alpha \in (-1/2, \infty) - \{0, 1\}$ by taking into account the formula for $D_{2\alpha}(F_0, F)$ obtained from (7.9). The peculiar expression $\Delta_\alpha(F_0, F)$ figuring in (7.36) and (7.38) can be better understood if we take into account the following assertion, after which we extend (7.38) to include also the values $\alpha \in \{0, 1\}$.

Assertion 7.4.1 If the fixed alternative $F \sim f$ satisfies the assumptions of Assertion 6.5.1, then the class $\{\Delta_\alpha(F_0, F) : \alpha \in \mathbb{R} - \{0\}\}$ consists of the variances

$$\begin{aligned} \Delta_\alpha(F_0, F) &= \int_0^1 \left(\frac{f^{-\alpha}}{\alpha} - \int_0^1 \frac{f^{-\alpha}}{\alpha} f \, \mathrm{d}y \right)^2 f \, \mathrm{d}x \\ &= \int_0^1 \left(\frac{f^{-\alpha}}{\alpha} \right)^2 f \, \mathrm{d}x - \left(\int_0^1 \frac{f^{-\alpha}}{\alpha} f \, \mathrm{d}x \right)^2 \end{aligned} \tag{7.39}$$

of the functions $f^{-\alpha}(X)/\alpha$ of the random argument X distributed by F. This class is continuously extended to all $\alpha \in \mathbb{R}$ by introducing the variance

$$\begin{aligned}\Delta_0(F_0, F) &= \int_0^1 \left(\ln f - \int_0^1 (\ln f)\, f\, \mathrm{d}y\right)^2 f\, \mathrm{d}x \\ &= \int_0^1 f \ln^2 f\, \mathrm{d}x - \left(\int_0^1 f \ln f\, \mathrm{d}x\right)^2 \end{aligned} \tag{7.40}$$

of the function $\ln f(X)$ of the random argument X introduced above. All $\Delta_\alpha(F_0, F)$, $\alpha \in \mathbb{R}$, are nonnegative measures of divergence of F_0 and F, reflexive in the sense that $\Delta_\alpha(F_0, F) = 0$ if and only if $F = F_0$.
□

We are now in a position to formulate the general results obtained in this chapter regarding the asymptotic normality of spacings-type power divergence statistics U_α from the sets $\mathcal{U}_\alpha = \{R_{\phi_\alpha}, \tilde{R}_{\phi_\alpha}, S_{\phi_\alpha}, \tilde{S}_{\phi_\alpha}, T_{\phi_\alpha}, \tilde{T}_{\phi_\alpha}\}$ for $\alpha > -1/2$ under the assumption of the fixed alternative, thereby specifying the parameters $\mu_\alpha(f)$ and $\sigma_\alpha^2(f)$ in (7.33) for all $\alpha > -1/2$. Inspecting once more formula (7.38), we observe that if $\alpha > -1/2$ differs from 0 and 1, then the asymptotic variance $\sigma_\alpha^2(f)$ under the alternative f exceeds the asymptotic variance $\sigma_\alpha^2 = \sigma_\alpha^2(f_0)$ achieved under the hypothesis $F_0 \sim f_0$ by a linear function of σ_α^2 given by

$$2\alpha(2\alpha - 1)\, D_{2\alpha}(F_0, F)\sigma_\alpha^2 + \Gamma^2(\alpha + 1)\, \Delta_\alpha(F_0, F) \tag{7.41}$$

with the coefficients $D_{2\alpha}(F_0, F)$ and $\Delta_\alpha(F_0, F)$ positive unless $F = F_0$. By using Assertion 7.1.2, we can now find the formulas for $\sigma_0^2(f)$ and $\sigma_1^2(f)$ which are missing in (7.37) by taking limits in (7.38) for $\alpha \to 0$ and $\alpha \to 1$. Since the limits σ_0^2 and σ_1^2 were already calculated in Assertion 7.2.1, and the limit $\Delta_0(F_0, F)$ is clear from Assertion 7.4.1, we obtain

$$\sigma_0^2(f) = \lim_{\alpha \to 0} \sigma_\alpha^2(f) = \sigma_0^2 + \Delta_0(F_0, F) \tag{7.42}$$

and

$$\sigma_1^2(f) = \lim_{\alpha \to 1} \sigma_\alpha^2(f) = (1 + 2D_2(F_0, F))\, \sigma_1^2 + \Delta_1(F_0, F) \tag{7.43}$$

where (cf. (7.39))

$$\Delta_1(F_0, F) = \int_0^1 \frac{1}{f} \mathrm{d}x - 1. \tag{7.44}$$

Together with (7.37), (7.42) and (7.43) provide formulas for $\sigma_\alpha^2(f)$ for all $\alpha > -1/2$. It is clear that $\sigma_0^2(f)$ and $\sigma_1^2(f)$ are of the form (7.38), so that the representation (7.38) holds for all $\alpha > -1/2$. We summarize our results as follows.

Assertion 7.4.2 If the alternative $F \sim f$ satisfies the assumptions of Assertion 6.5.1, then the asymptotic formula of (7.33) is valid for all $\alpha > -1/2$. The asymptotic means $\mu_\alpha(f)$ are given by the explicit formulas (7.24) − (7.9). The asymptotic variances $\sigma_\alpha^2(f)$ are given by (7.38), where the explicit formulas for $D_{2\alpha}(F_0, F)$ can be found in (7.7) − (7.9), those for σ_α^2 in (7.30) and (7.31), and the formulas for $\Delta_\alpha(F_0, F)$ in (7.39) and (7.40). The asymptotic means and variances are continuous in the variable $\alpha \in (-1/2, \infty)$. The asymptotic means satisfy the inequality $\mu_\alpha(f) \geq \mu_\alpha$ mentioned in Assertion 7.2.1. The asymptotic variances satisfy the inequality $\sigma_\alpha^2(f) \geq \sigma_\alpha^2$. Both inequalities become equalities if and only if $F = F_0$. □

Concrete forms of $\mu_\alpha(f)$ and $\sigma_\alpha^2(f_0) = \sigma_\alpha^2$ were illustrated in Tables 7.2.1 and 7.3.1. The next table illustrates $\sigma_\alpha^2(f)$ given by (7.38) for arbitrary f satisfying the assumptions of Assertion 6.5.1 and selected values of α. In each line of Table 7.4.1 two expressions for $\sigma_\alpha^2(f)$ are given: the first one is obtained by substituting α in (7.38), the second one by actually calculating $D_{2\alpha}(F_0, F)$ and $\Delta_\alpha(F_0, F)$ in each case and putting the results in a closed form. As presumed, for $f = 1$ the illustrated values reduce to σ_α^2 from Table 7.3.1 .

Table 7.4.1: Asymptotic variances $\sigma^2_\alpha(f)$ for selected $\alpha > -1/2$.

α			$\sigma^2_\alpha(f)$
0	$\sigma_0^2 + \Delta_0(F_0, F)$	$=$	$\frac{\pi^2}{6} - 1 + \int_0^1 f \ln^2 f \mathrm{d}x \; - \; \left(\int_0^1 f \ln f \,\mathrm{d}x\right)^2$
$\frac{1}{2}$	$\sigma^2_{\frac{1}{2}} + \frac{\pi}{4}\Delta_{\frac{1}{2}}(F_0, F)$	$=$	$16 - 4\pi - \pi \left(\int_0^1 \sqrt{f}\,\mathrm{d}x\right)^2$
1	$[1 + \chi^2(F_0, F)]\,\sigma_1^2 + \Delta_1(F_0, F)$	$=$	$\int_0^1 \frac{\mathrm{d}x}{f}\left(\frac{\pi^2}{3} - 2\right) - 1$
$\frac{3}{2}$	$[1 + 6D_3(F_0, F)]\,\sigma^2_{3/2} + \frac{9\pi}{16}\Delta_{3/2}(F_0, F)$	$=$	$\int_0^1 \frac{\mathrm{d}x}{f^2}\left(\frac{32}{3} - 3\pi\right) - \frac{\pi}{4}\left(\int_0^1 \frac{\mathrm{d}x}{\sqrt{f}}\right)^2$
2	$[1 + 12D_4(F_0, F)\,\sigma_2^2 + 4\Delta_2(F_0, F)]$	$=$	$2\int_0^1 \frac{\mathrm{d}x}{f^3} - \left(\int_0^1 \frac{\mathrm{d}x}{f}\right)^2$
3	$[1 + 30D_6(F_0, F)]\,\sigma_3^2 + 36\Delta_3(F_0, F)$	$=$	$14\int_0^1 \frac{\mathrm{d}x}{f^5} - 4\left(\int_0^1 \frac{\mathrm{d}x}{f^2}\right)^2$

7.5 Discussion

The general form of the asymptotic normality (7.33) established by Assertion 7.4.2, as well as the continuity of the asymptotic means and variances $\mu_\alpha(f)$ and $\sigma^2_\alpha(f)$ in the parameter $\alpha > -1/2$ proved in Assertion 7.1.2, and the explicit formulas (7.24) and (7.38) for these parameters for general α seem to be new results. However, in the references cited in Subsections 5.1 and 5.2 one can find particular versions of these results for some of the statistics U_α from the set $\{R_{\phi_\alpha}, \tilde{R}_{\phi_\alpha}, S_{\phi_\alpha}, \tilde{S}_{\phi_\alpha}, T_{\phi_\alpha}, \tilde{T}_{\phi_\alpha}\}$ or their linear functions, and for some $\alpha > -1/2$ and some distributions $F \sim f$.

Let us start with the statistic S_{ϕ_0} proposed by Moran (1951), and denoted by $\mathcal{M}$ in Example 5.1.2 (equation(5.7)). The asymptotic normality (7.33) for $\alpha = 0$, $U_0 = S_{\phi_0}$ and $f = f_0 \equiv 1$, with the parameters $\mu_0(f_0) = \mu_0$ and $\sigma_0^2(f_0) = \sigma_0^2$ given in Tables 7.2.1 and 7.3.1, was proved by Darling (1953), yielding specifically that under $\mathcal{H}_0$

$$\frac{1}{\sqrt{n}}\left(\mathcal{M} - n\gamma\right) \xrightarrow{\mathcal{D}} N\left(0, \frac{\pi^2}{6} - 1\right) \quad \text{as } n \to \infty. \tag{7.45}$$

The result of Darling was extended to all positively valued step functions f on $[0,1]$ by Cressie (1976), who also obtained $\mu_0(f)$ and $\sigma_0^2(f)$ given in Tables 7.3.1 and 7.4.1. The result of Cressie was extended by van Es (1992) to the alternatives f considered in the present chapter which satisfy a Lipschitz condition on $[0,1]$, and to all f considered in this chapter by Shao and Hahn (1995). Cressie(1976) and van Es(1992) studied S_{ϕ_0} as the special case obtained for $m = 1$ from a more general statistic based on m-spacings with $m \geq 1$. Van Es extended ideas and methods developed for $m > 1$ by Vasicek (1976) and Dudewicz and van der Meulen (1981) for proving the consistency and asymptotic normality of a spacings-based estimator of entropy. The latter authors considered only $\phi(t) = -\ln t$.

Greenwood (1946) introduced the statistic

$$\mathcal{G} = \sum_{j=1}^{n+1}(Y_j - Y_{j-1})^2 = \frac{2S_{\phi_2} + n + 1}{(n+1)^2}, \tag{7.46}$$

discussed in Examples 5.1.1 and 7.1.1. Kimball (1950) proposed the generalization

$$\sum_{j=1}^{n+1}(Y_j - Y_{j-1})^\alpha = \frac{\alpha(\alpha-1)S_{\phi_\alpha} + n + 1}{(n+1)^\alpha}, \qquad \alpha > 0, \tag{7.47}$$

and Darling (1953) proved an asymptotic normality theorem equivalent to (7.33) for $U_\alpha = S_{\phi_\alpha}$, $\alpha \in (0,\infty) - \{1\}$, and $f = f_0 \equiv 1$. Weiss (1957) extended this result of Darling to positive piecewise constant densities f. Hall (1984) obtained the asymptotic normality

$$\frac{1}{\sqrt{n}}\left(\tilde{U}_\alpha - n\alpha(\alpha-1)\,\mu_\alpha(f) - n\right) \xrightarrow{\mathcal{D}} N(0, \alpha^2(\alpha-1)^2\sigma^2_\alpha(f)) \quad \text{as } n \to \infty \tag{7.48}$$

for all statistics

$$\begin{aligned}\tilde{U}_\alpha &= \sum_{j=2}^{n} (n(Y_j - Y_{j-1}))^\alpha \\ &= \alpha(\alpha-1)\,\tilde{R}_{\phi_\alpha} - \alpha n(1 - Y_n + Y_1) + n + \alpha - 1 \\ &= \alpha(\alpha-1)\,\tilde{R}_{\phi_\alpha} + n + O_p(1)\end{aligned}$$

with $\alpha \in (-1/2, \infty) - \{0, 1\}$ for any f considered in Assertion 7.4.2. Here $\mu_\alpha(f)$ and $\sigma^2_\alpha(f)$ are the same as in Assertion 7.4.2, with $\mu_\alpha(f)$ given by the right-hand side of (7.27) and $\sigma^2_\alpha(f)$ by (7.38), $\tilde{R}_{\phi_\alpha}$ is defined as in (5.24) with $\phi = \phi_\alpha$, and the $O_p(1)$ statement follows from the proof of Assertion 6.1.1. In fact, this result of Hall (1984) was one of the arguments used in the proof of Assertion 7.4.2.

The statistic S_{ϕ_1} was proposed by Misra and van der Meulen (2001), who proved the asymptotic normality (7.33) for $U_1 = S_{\phi_1}$ and any f considered in Assertion 7.4.2, with the parameters $\mu_1(f)$ and $\sigma^2_1(f)$ given in Tables 7.2.1 and 7.3.1, yielding the result

$$\frac{1}{\sqrt{n}}\left(S_{\phi_1} - n\left(1 - \gamma - \int_0^1 \ln f \,\mathrm{d}x\right)\right) \xrightarrow{\mathcal{D}} N\left(0, \int_0^1 \left(\frac{\pi^2}{3} - 2\right)\frac{\mathrm{d}x}{f} - 1\right) \tag{7.49}$$

as $n \to \infty$. We see that the present Assertion 7.4.2 unifies and extends the results proved separately in the literature in three different situations for two particular statistics from the set (7.4). The formulas for all asymptotic parameters $\mu_\alpha(f)$ and $\sigma^2_\alpha(f)$ of the statistics U_α are shown to follow via the asymptotic equivalence of these statistics (cf. Assertion 6.1.1) and the continuity of these parameters in α (cf. Assertion 7.1.2) from Hall's formula (cf. (7.48)) for the asymptotic parameters of $\tilde{U}_\alpha$ with $\alpha \in (-1/2, \infty)$ different from 0 and 1.

8 Program for testing by power divergence statistics

Following three MATLAB based functions was proposed for evaluation of the statistics from section 7.

The statistic $T_{\phi_\alpha,n}$ (denoted as function Tpdt) and given by (7.10) - (7.12).

```
function T = Tpdt(X, F0, alpha)
% This function compute the Goodness-of-fit statistic based on
% the power-divergences.
%
% Use: Tpdt(X, F0, {alpha})
%    x       row vector; observed data
%    F0      string: hypothetical cumulative distribution distribution of vector x
%    {alpha} number: from interval (-1/2, inf) order of power-divergence
%                    (optional, default = 2)
%
% Example: Tpdt(X, 'normal_cdf(x, 0, 1)', 1.5)
eps = 1.E-4;
if (nargin<2), error('Use: Tpdt(x, F0, {alpha})'); end
Xsize = size(X); if (Xsize(1)~=1), error('The first parameter is not a row vector'); end
n = Xsize(2);
if (n<3), error('The number of the observations is too small'); end
if (nargin<3), alpha = 2; end for (i=1:n)
   x = X(i);
%  small 'x' is required in hypothetical cdf definition string F0
   Y(i) = eval(F0);
end Y = [0 sort(Y) 1]; dY = diff(Y(1:n)); if (abs(alpha)<eps)
% equation 7.12
   T = -sum(log(n*dY))-log(n*(1-Y(n)));
```

```
elseif (abs(1-alpha)<eps)
% equation 7.11
   T = log(n*dY)*(n*dY)'+log(n*(1-Y(n)))*n*(1-Y(n));
else
% equation 7.10
   T = (sum((n*dY).^alpha)+(n*(1-Y(n)))^alpha-n)/alpha/(alpha-1);
end return
```

The statistics $S_{\phi_\alpha,n}$ (denoted as Spdt) and given by (7.13) - (7.15).

```
function S = Spdt(X, F0, alpha)
% This function compute the Goodness-of-fit statistic based on
% the power-divergences.
%
% Use: Spdt(X, F0, {alpha})
%    x       row vector; observed data
%    F0      string: hypothetical cumulative distribution distribution of vector x
%    {alpha} number: from interval (-1/2, inf) order of power-divergence
%                    (optional, default = 2)
%
% Example: Spdt(X, 'normal_cdf(x, 0, 1)', 1.5)
eps = 1.E-4;
if (nargin<2), error('Use: Spdt(X, F0, {alpha})'); end
Xsize = size(X);
if (Xsize(1)~=1), error('The first parameter is not a row vector'); end
n = Xsize(2);
if (n<3), error('The number of the observations is too small'); end
if (nargin<3), alpha = 2; end
for (i=1:n)
   x = X(i);
%  small 'x' is required in hypothetical cdf definition string F0
   Y(i) = eval(F0);
end
Y = [0 sort(Y) 1];
dY = diff(Y);
if (abs(alpha)<eps)
% equation 7.15
   S = -sum(log((n+1)*dY));
elseif (abs(1-alpha)<eps)
% equation 7.14
   S = log((n+1)*dY)*((n+1)*dY)';
else
% equation 7.13
   S = (sum(((n+1)*dY).^alpha)-n-1)/alpha/(alpha-1);
end
return
```

The statistics $\tilde{T}_{\phi_\alpha,n}$ (denoted as TWpdt) and given by (7.16) - (7.18).

```
function TW = TWpdt(X, F0, alpha)
% This function compute the Goodness-of-fit statistic based on
% the power-divergences.
%
% Use: TWpdt(X, F0, {alpha})
%    x       row vector; observed data
%    F0      string: hypothetical cumulative distribution distribution of vector x
%    {alpha} number: from interval (-1/2, inf) order of power-divergence
```

```
%                     (optional, default = 2)
%
% Example: TWpdt(X, 'normal_cdf(x, 0, 1)', 1.5)
eps = 1.E-4;
if (nargin<2), error('Use: TWpdt(X, F0, {alpha})'); end
Xsize = size(X);
if (Xsize(1)~=1), error('The first parameter is not a row vector'); end
n = Xsize(2);
if (n<3), error('The number of the observations is too small'); end
if (nargin<3), alpha = 2; end
for (i=1:n)
   x = X(i);
%  small 'x' is required in hypothetical cdf definition string F0
   Y(i) = eval(F0);
end
Y = [0 sort(Y) 1];
dY = diff(Y(2:n+1));
if (abs(alpha)<eps)
% equation 7.18
   TW = -sum(log(n*dY))-log(n*(Y(2)+1-Y(n+1)));
elseif (abs(1-alpha)<eps)
% equation 7.17
   TW = log(n*dY)*(n*dY)'+log(n*(Y(2)+1-Y(n+1)))*n*(Y(2)+1-Y(n+1));
else
% equation 7.16
   TW = (sum((n*dY).^alpha)+(n*(Y(2)+1-Y(n+1)))^alpha-n)/alpha/(alpha-1);
end
return
```

The input parameters of all functions are row vector of observed data **X**, string definition of hypothetical d.f. **F0** and power-divergence order **alpha**. The mixed generalized lambda model for crabs and horses data from section 1. was tested by all 3 statistics for **alpha** $\in (-0.4, 3)$. The results of tests are presented in figures 8.1 and 8.2 with 95% significance critical value using the asymptotic normality given in Assertion 7.3.1.

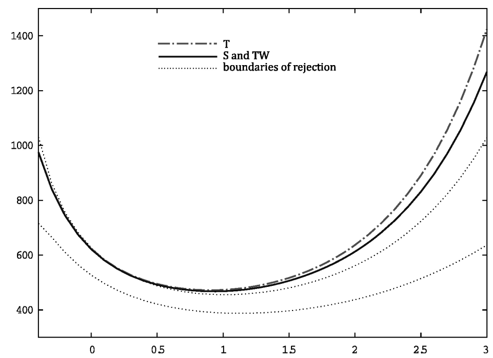
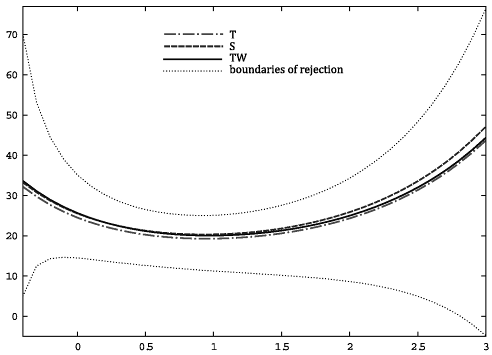


Figure 8.1 statistics $T_{\phi_\alpha,n}, S_{\phi_\alpha,n}, \tilde{T}_{\phi_\alpha,n}$, $\alpha \in (-0.4, 3)$ for data from table 1.1

9 Appendix

Proofs of the assertions stated in this chapter above can be found in Vajda and van der Meulen (2010). They are presented here for the sake of completness.



9.1 Proofs for structural spacings statistics

Proof of Assertion 6.1.1 We shall consider the fixed alternative $F(x)$ with a continuous density $f(x) > 0$ for $0 \le x \le 1$. For the local alternatives the argument is similar. By inspecting the definitions of T_ϕ, $\tilde{T}_\phi$ and R_ϕ we see that for (6.3) it suffices to prove that as $n \to \infty$

$$\phi(np_{01}) = O_p(1) \quad \text{and} \quad \phi(n(p_{01} + p_{02})) = O_p(1). \tag{9.50}$$

It is known (see for example page 208 in Hall (1986)) that $p_{01} = F^{-1}(Z_1/W_{n+1})$ and $p_{01} + p_{02} = F^{-1}((Z_1 + Z_2)/W_{n+1})$, where $Z_1, \dots, Z_{n+1}$ are independent standard exponential variables and $W_{n+1} = Z_1 + \cdots + Z_{n+1}$, so that, for $n \to \infty$,

$$\frac{W_{n+1}}{n} \xrightarrow{p} 1 \quad \text{and} \quad V_n = \frac{Z_1}{W_{n+1}} \xrightarrow{p} 0.$$

Setting

$$\Upsilon_n = \frac{F^{-1}(V_n)}{V_n} = \frac{F^{-1}(V_n) - F^{-1}(0)}{V_n}$$

and using the mean value theorem and the assumed continuity of f in the neighborhood of 0, we find that

$$\Upsilon_n \xrightarrow{p} \frac{1}{f(0)} \quad \text{as } n \to \infty$$

where, by assumptions about f, $0 < f(0) < \infty$. Thus

$$np_{01} = \frac{n}{W_{n+1}} Z_1 \Upsilon_n$$

and, by applying (5.36),

$$\phi(np_{01}) = \xi\left(\frac{n}{W_{n+1}}\right) \phi(Z_1 \Upsilon_n) + \zeta(Z_1 \Upsilon_n)\, \phi\left(\frac{n}{W_{n+1}}\right) + \eta\left(\frac{n}{W_{n+1}}\right) (Z_1 \Upsilon_n - 1).$$

Since $Z_1 \Upsilon_n = O_p(1)$ as $n \to \infty$, we obtain from Assertion 5.3.1

$$\begin{aligned} \phi(np_{01}) &= \left[\xi\left(\frac{n}{W_{n+1}}\right) + \phi\left(\frac{n}{W_{n+1}}\right) + \eta\left(\frac{n}{W_{n+1}}\right)\right] O_p(1) \\ &= [\xi(1) + \phi(1) + \eta(1) + o_p(1)]\, O_p(1) \\ &= O_p(1) \quad \text{(cf (5.38))}, \end{aligned}$$

thus proving the first relation of (9.50). Replacing $V_n = Z_1/W_{n+1}$ by $V_n = (Z_1 + Z_2)/W_{n+1}$, and using the fact that now

$$(Z_1 + Z_2)\, \Upsilon_n = (Z_1 + Z_2)\, \frac{F^{-1}(V_n) - F^{-1}(0)}{V_n} = O_p(1)$$

we obtain the second relation of (9.50). Next we prove (6.4). From (5.37) we get for any $p > 0$

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

so that

$$\phi((n+1)\,p) - \phi(np) = \varepsilon_n \phi(np) + \phi\left(\frac{n+1}{n}\right) + \eta\left(\frac{n+1}{n}\right)(np-1) \tag{9.51}$$

where $\varepsilon_n = \xi((n+1)/n) - 1 = o(1)$ as $n \to \infty$ by Assertion 5.3.1. Replacing p by the probabilities $p_{0j} = Y_j - Y_{j-1}$ figuring in the definitions of S_ϕ and R_ϕ (cf. (5.2) and (5.1)), and summing over $1 \le j \le n+1$, we get the equality

$$S_\phi - R_\phi = \varepsilon_n\, R_\phi + \delta_n$$

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 Assertion 5.3.1,

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

This completes the proof of the first relation in (6.4). The proof of the second relation is the same: we just replace p in (9.51) by the probabilities $\tilde{p}_{0j}$ figuring in the definition (5.22) of $\tilde{T}_\phi$. □

Proof of Assertion 6.3.1 By Theorem 1 of Hall (1984), the statistic $\tilde{R}_\phi$ defined by (5.24) satisfies under a fixed alternative $F \sim f$ the relation

$$\frac{\tilde{R}_\phi}{n} \xrightarrow{p} \tilde{\mu}_\phi(f) = \int_0^1 f^2(x)\left(\int_0^\infty \phi(t)\, e^{-tf(x)}\mathrm{d}t\right)\mathrm{d}x \quad \text{as } n \to \infty$$

provided $\phi : (0,\infty) \mapsto \mathbb{R}$ is continuous and exponentially bounded in the sense that $|\phi(t)| \le K(t^\alpha + t^{-\beta})$ for some $K > 0$, $\alpha \ge 0$, $\beta < 1$, and f is bounded, piecewise continuous, and bounded away from 0 (see also part (i) of Theorem 3.1 in Misra and van der Meulen (2001)). Thus (6.14) is proved for $U_\phi = \tilde{R}_\phi$ as soon as it is shown that for $\phi \in \mathbf{\Phi}_1$ the limit $\tilde{\mu}_\phi(f)$ coincides with $\mu_\phi(f)$. By substituting s for $tf(x)$ in the last integral, and using the assumption $0 < f(x) < \infty$ and the functional equation (5.36),

$$\begin{aligned}\tilde{\mu}_\phi(f) &= \int_0^1 f(x)\left(\int_0^\infty \phi\left(\frac{s}{f(x)}\right) e^{-s}\mathrm{d}s\right)\mathrm{d}x \qquad (9.52)\\ &= \int_0^1 f(x)\left(\int_0^\infty \left[\xi(s)\,\phi\left(\frac{1}{f(x)}\right) + \zeta\left(\frac{1}{f(x)}\right)\phi(s) + \eta(s)\left(\frac{1}{f(x)} - 1\right)\right] e^{-s}\mathrm{d}s\right)\mathrm{d}x\\ &= \mu_\phi(f) + \int_0^\infty \eta(s)\, e^{-s}\,\mathrm{d}s \int_0^1 (1 - f(x))\,\mathrm{d}x = \mu_\phi(f).\end{aligned}$$

The extension of (6.14) to $U_\phi \in \{T_\phi, \tilde{T}_\phi, R_\phi\}$ follows from Assertion 6.1.1. For $\phi \in \mathbf{\Phi}_2$ the extension of (6.14) to $U_\phi \in \{S_\phi, \tilde{S}_\phi\}$ follows from Assertion 6.1.1 too. □

Proof of Assertion 6.4.1 For $U_\phi = S_\phi$ the relations (6.16)−(6.18) follow from the result of Kuo and Rao (1981), cf. also Del Pino (1979) and Theorem 3.2 in Misra and van der Meulen (2001). The extension to the remaining statistics U_ϕ follows from Assertion 6.1.1. □

Proof of Assertion 6.5.1 Consider $U_\phi = \tilde{R}_\phi$ for $\phi \in \mathbf{\Phi}_2$. By Assertion 5.3.2, $\phi(t)$ has a continuous derivative $\phi'(t)$ on $(0,\infty)$. By (5.39), for every $c \in \mathbb{R}$

$$t^c|\phi'(t)| \le |\xi'(1)|\, t^{c-1}|\phi(t)| + |\phi'(1)|\, t^c + |\eta'(1)|\, t^{c-1}|t-1|.$$

Thus if ϕ satisfies (6.5) with $\beta < 1/2$ then there exists $\alpha \ge 0$ such that

$$\lim_{t\to\infty} t^{-\alpha}|\phi'(t)| = \lim_{t\downarrow 0} t^{1+\beta}|\phi'(t)| = 0.$$

This means that under the assumptions of Assertion 6.4.1 there exist $c > 0$, $K > 0$ and $b < 1/2$ such that for every $t \in (0,\infty)$

$$|\phi(t)| \le K(t^a + t^{-b}) \quad \text{and} \quad |\phi'(t)| \le K(t^a + t^{-b-1}).$$

For continuously differentiable functions ϕ satisfying these assumptions, and fixed alternatives with densities f continuously differentiable on $(0,1)$, it follows from Theorem 2 in Hall (1984) (cf. also part (ii) of Theorem 3.1 in Misra and van der Meulen (2001)) that $U_\phi = \tilde{R}_\phi$ satisfies the relation

$$\frac{1}{\sqrt{n}}(U_\phi - n\tilde{\mu}_\phi(f)) \xrightarrow{\mathcal{D}} N(0, \tilde{\sigma}^2_\phi(f)) \quad \text{for } n \to \infty$$

where: **(1)** the asymptotic mean $\tilde{\mu}_\phi(f)$ was presented and proved to be equal to $\mu_\phi(f)$ in the proof of Assertion 6.3.1 under assumptions weaker than here and, **(2)** the asymptotic variance $\tilde{\sigma}^2_\phi(f)$ can be specified by means of the standard exponential variable Z and the auxiliary function

$$G(x) = \int_0^x \left(1 - \frac{F(y)\,f'(y)}{f^2(y)}\right) E\left[Z\,\phi'\left(\frac{Z}{f(y)}\right)\right] \mathrm{d}y, \quad 0 < x < 1, \tag{9.53}$$

as the sum of

$$s_1^2(f) = \int_0^1 \left(E\phi^2\left(\frac{Z}{f(x)}\right) - \left[E\phi\left(\frac{Z}{f(x)}\right)\right]^2\right) f(x)\,\mathrm{d}x \tag{9.54}$$

$$s_2^2(f) = -2\int_0^1 E\left[(Z-1)\,\phi\left(\frac{Z}{f(x)}\right)\right] \frac{G(x)}{F(x)}\, f(x)\,\mathrm{d}x \tag{9.55}$$

and

$$s_3^2(f) = \int_0^1 \left(\frac{G(x)}{F(x)}\right)^2 f(x)\,\mathrm{d}x. \tag{9.56}$$

It remains to be proved that for every $x \in (0,1)$

$$\left(E\,\phi^2\left(\frac{Z}{f(x)}\right) - \left[E\,\phi\left(\frac{Z}{f(x)}\right)\right]^2\right) f(x) = \Psi_2(x), \tag{9.57}$$

$$E\left[(Z-1)\,\phi\left(\frac{Z}{f(x)}\right)\right]\sqrt{f(x)} = \Psi_3(x) \tag{9.58}$$

and

$$\frac{G(x)\,\sqrt{f(x)}}{F(x)} = \Psi_4(x). \tag{9.59}$$

Indeed, then $\tilde{\sigma}^2_\phi(t) = \sigma^2_\phi(f)$ so that (6.23) is proved for $U_\phi = R_\phi$, and the extension of (6.23) to the remaining statistics $U_\phi \in \{\tilde{R}_\phi, S_\phi, \tilde{S}_\phi, T_\phi, \tilde{T}_\phi\}$ follows from Assertion 6.1.1. We shall prove (9.57) – (9.59) in the reversed order. By substituting $t = Z/f(y)$ in (5.39) and taking into account that $\zeta(t) \equiv 1$ we obtain

$$\begin{aligned} E\left[Z\phi'\left(\frac{Z}{f(y)}\right)\right] &= f(y)\,E\left[\xi'(1)\,\phi\left(\frac{Z}{f(y)}\right) + \phi'(1) + \eta'(1)\left(\frac{Z}{f(y)} - 1\right)\right] \\ &= f(y)\left[\xi'(1)\,E\,\phi\left(\frac{Z}{f(y)}\right) + \phi'(1) + \eta'(1)\left(\frac{1}{f(y)} - 1\right)\right] \end{aligned}$$

and, by putting $s = 1/f(x)$ and $t = Z$ in (5.37), we get

$$\phi\left(\frac{Z}{f(x)}\right) = \phi(Z)\,\xi\left(\frac{1}{f(x)}\right) + \phi\left(\frac{1}{f(x)}\right) + \eta\left(\frac{1}{f(x)}\right)(Z-1). \tag{9.60}$$

Therefore

$$E\,\phi\left(\frac{Z}{f(x)}\right) = \langle\phi\rangle\xi\left(\frac{1}{f(x)}\right) + \phi\left(\frac{1}{f(x)}\right) \tag{9.61}$$

and, consequently,

$$E\left[Z\,\phi'\left(\frac{Z}{f(y)}\right)\right] = \Psi_1(y). \tag{9.62}$$

This, together with the definitions of $\Psi_4(x)$ and $G(x)$ in (6.22) and (9.53), implies (9.59). Further, from (9.60) and the definition of $\Psi_3(x)$ in (6.21) we get (9.58). Finally, from (9.60), (9.61) and the definition of $\Psi_2(x)$ in (6.20) we obtain (9.57) which completes the proof. □

9.2 Proofs for power spacings statistics

Proof of Assertion 7.1.1 By the assumptions about g,

$$t_0 = \min_{y\in[a,b]} g(y) > 0 \quad \text{and} \quad t_1 = \max_{y\in[a,b]} g(y) < \infty.$$

If $\psi_\alpha(t)$ is convex, then for every $t \in [t_0, t_1]$ and $\alpha \in (c,d)$

$$\psi'_\alpha(t_*)\,(t - t_*) \le \psi_\alpha(t) \le \psi_\alpha(t_0) + \psi_\alpha(t_1).$$

If $\psi_\alpha(t)$ is concave, then, similarly,

$$\psi_\alpha(t_0) + \psi_\alpha(t_1) \le \psi_\alpha(t) \le \psi'_\alpha(t_*)\,(t - t_*).$$

Therefore in both cases

$$\max_{t_0\le t\le t_1} |\psi_\alpha(t)| \le \max\left\{|\psi_\alpha(t_0) + \psi_\alpha(t_1)|,\, |\psi'_\alpha(t_*)|\cdot|t_1 - t_0|\right\}.$$

The assumed continuity of $\psi'_\alpha(t_*)$ and $\psi_\alpha(t_0) + \psi_\alpha(t_1)$ in the variable $\alpha \in (c,d)$ implies that for all compact neighborhoods $N \subset (c,d)$ of α_0 the constant

$$k = \sup_{\alpha\in N}\max_{t_0\le t\le t_1} |\psi_\alpha(t)| = \sup_{\alpha\in N}\max_{y\in[a,b]} |\psi_\alpha(g(y))|$$

is finite. Put

$$K = \max_{[t_0,t_1]\times[-k,k]} \Phi(u,v).$$

The function $|\Phi(g,\psi_\alpha(g))|$ of variables $(y,\alpha) \in [a,b]\times(c,d)$ is bounded on $[a,b]\times N$ by $K < \infty$. Since for every $y \in [a,b]$

$$\lim_{\alpha\to\alpha_0} \Phi(g,\psi_\alpha(g)) = \Phi(g,\psi_{\alpha_0}(g)),$$

the Lebesgue dominated convergence theorem for integrals implies (7.21). □

Proof of Assertion 7.1.2 Since $\mu_\alpha = \mu_\alpha(f_0)$ and $\sigma^2_\alpha = \sigma^2_\alpha(f_0)$, where the hypothetical density f_0 satisfies the assumptions of Assertions 6.3.1 and 6.5.1, the continuity of μ_α and σ^2_α follows from the continuity of $\mu_\alpha(f)$ and $\sigma^2_\alpha(f)$ proved below. By (7.20) and (6.18),

$$m_\alpha(\ell) = \frac{\|\ell\|^2}{2}\left(\langle t^2\phi_\alpha(t)\rangle - 4\langle t\phi_\alpha(t)\rangle + 2\langle\phi_\alpha(t)\rangle\right)$$

where ϕ_α is given by (3.20), (3.21), and, by (6.6),

$$\langle t^j\phi_\alpha(t)\rangle = \int_0^\infty t^j\phi_\alpha(t)\,\mathrm{d}H(t), \quad j \in \{0,1,2\} \tag{9.63}$$

for $H(t) = 1 - e^{-t}$. All integrals (9.63) are finite if and only if $\alpha \in (-1, \infty)$. Further, for every fixed $t > 0$

$$\frac{\mathrm{d}}{\mathrm{d}\alpha}\, \alpha\phi_\alpha(t) \geq 0 \quad \text{at any } \alpha \in \mathbb{R}. \tag{9.64}$$

Hence the continuity of the products $\alpha\langle t^j \phi_\alpha(t)\rangle$ in the variable $\alpha \in \mathbb{R}$ follows from the monotone convergence theorem for integrals, and this implies also the desired continuity of the integrals (9.63) at any $\alpha \in (-1, \infty) - \{0\}$. Further, for every fixed $t > 0$

$$\frac{\mathrm{d}}{\mathrm{d}\alpha}\, (\alpha - 1)\, \phi_\alpha(t) \geq 0 \quad \text{for any } \alpha \in \mathbb{R}. \tag{9.65}$$

Hence the continuity of the products $(\alpha - 1)\, \langle t^j \phi_\alpha(t)\rangle$ in the variable $\alpha \in \mathbb{R}$ follows as well from the monotone convergence theorem for integrals. Similarly as above, this implies the continuity of the integrals (9.63) at the remaining point $\alpha = 0$. Further, by (7.20) and (6.12),

$$\mu_\alpha(f) = \langle \xi_\alpha \rangle\, D_\alpha(F_0, F) + \langle \phi_\alpha \rangle$$

where, by (6.6) and (7.5)

$$\langle \xi_\alpha \rangle = \int_0^\infty t^\alpha \mathrm{d}H(t) \quad \text{and} \quad \langle \phi_\alpha \rangle = \int_0^\infty \phi_\alpha(t)\, \mathrm{d}H(t).$$

These integrals are finite if and only if $\alpha \in (-1, \infty)$. The continuity of $\langle \phi_\alpha \rangle$ at $\alpha \in (-1, \infty)$ was proved above, the continuity of $D_\alpha(F_0, F)$ at $\alpha \in \mathbb{R}$ follows from the assumptions about the densities f_0 and f and from Proposition 2.14 in Liese and Vajda (1987). The continuity of $\langle \xi_\alpha \rangle$ at $\alpha \in (-1, \infty)$ follows from the monotone convergence theorem for integrals applied separately to the integration domains $(0, 1)$ and $(1, \infty)$. Finally, let us consider $\sigma_\alpha^2(f)$ defined by (6.19) – (6.24) for $\phi = \phi_\alpha$, $\xi = \xi_\alpha$, and $\eta = \eta_\alpha$ given by (3.20), (3.21) and (7.5). The integrals $\langle t\phi_\alpha(t)\rangle$, $\langle \phi_\alpha(t)\rangle$ and $\langle \phi_\alpha^2(t)\rangle$ are finite if and only if $\alpha \in (-1/2, \infty)$, and their continuity at $\alpha \in (-1/2, \infty)$ was either proved above or can be proved similarly as above. The continuity of the integral

$$\int_0^1 \left[f\xi_\alpha^2 \left(\frac{1}{f}\right) + f\eta_\alpha^2 \left(\frac{1}{f}\right) \right] \mathrm{d}x$$

at $\alpha \in (-1/2, \infty)$ follows from Assertion 7.1.1, which establishes the continuity of the component $\int \Psi_2(x)\, \mathrm{d}x$ of $\sigma_\alpha^2(f)$ in (6.24). For the continuity of the remaining two components, we take into account that $F(x) > c_1 x$ for some $c_1 > 0$ on $[0, 1]$, because f is bounded away from zero on $[0, 1]$. Furthermore, both $f(x)$ and $f'(x)$ are bounded on $[0, 1]$, so that there exists a constant c_2 such that in (6.22)

$$\frac{\sqrt{f(x)}}{F(x)} \int_0^x \left| 1 - \frac{F(y)\, f'(y)}{f^2(y)} \right| \mathrm{d}y < c_2 \quad \text{for all } x \in [0, 1]. \tag{9.66}$$

Using the function $\varphi_\alpha(t) = \alpha\phi_\alpha(t)$, which is for every $t > 0$ continuous and monotone in $\alpha \in \mathbb{R}$ (cf. (9.64)), we obtain from (6.19)

$$\Psi_1(x) = \alpha \langle \phi_\alpha \rangle\, f(x)^{1-\alpha} + f(x)\, \varphi_\alpha\left(\frac{1}{f(x)}\right) + 1 - f(x)$$

where the right-hand side is bounded on $[0, 1]$, locally uniformly in α, and continuous at any $\alpha \in \mathbb{R}$. By (6.22) and (9.66), this implies that also $\Psi_4(x)$ is bounded on $[0, 1]$, locally uniformly in α, and continuous at any $\alpha \in \mathbb{R}$. Since the integrands in

$$\int_0^1 \left[\sqrt{f}\xi_\alpha \left(\frac{1}{f}\right) + \sqrt{f}\eta_\alpha \left(\frac{1}{f}\right) \right] \Psi_4\, \mathrm{d}x \quad \text{and} \quad \int_0^1 \Psi_4^2\, \mathrm{d}x$$

are continuous on $[0, 1]$ and locally bounded in the variable $\alpha \in \mathbb{R}$, the continuity of both these integrals in the variable $\alpha \in \mathbb{R}$ follows from the Lebesgue dominated convergence theorem for integrals. This clarifies the continuity of the second and third component of $\sigma_\alpha^2(f)$ in (6.24) and thus completes the proof. □

Proof of Assertion 7.2.1 The functions from the class $\{\phi_\alpha : \alpha \in (-1, \infty)\} \subset \mathbf{\Phi}_2$ satisfy all assumptions of Assertion 6.3.1. Hence (7.23) holds for all $\alpha > -1$ and the limit $\mu_\alpha(f)$ is given in accordance with (6.12) and (7.5) by the formula

$$\mu_\alpha(f) = \langle \xi_\alpha(t) \rangle \, D_\alpha(F_0, F) + \langle \phi_\alpha(t) \rangle = \langle t^\alpha \rangle \, D_\alpha(F_0, F) + \langle \phi_\alpha(t) \rangle$$

where $\langle t^\alpha \rangle = \Gamma(\alpha + 1)$ for all $\alpha \in \mathbb{R}$. If $\alpha \notin \{0, 1\}$ then

$$\langle \phi_\alpha(t) \rangle = \frac{1}{\alpha(\alpha - 1)} \langle t^\alpha - 1 \rangle = \frac{\Gamma(\alpha + 1) - \Gamma(1)}{\alpha(\alpha - 1)}$$

but the expressions

$$\langle \phi_0(t) \rangle = \langle -\ln t \rangle \quad \text{and} \quad \langle \phi_1(t) \rangle = \langle t \ln t \rangle$$

lead to the evaluation of unpleasant integrals. This evaluation can be avoided by employing Assertion 7.1.2. From the continuity of $\mu_\alpha = \langle \phi_{\ \alpha}(t) \rangle$, it follows that

$$\mu_j = \langle \phi_j(t) \rangle = \lim_{\alpha \to j} \frac{\Gamma(\alpha + 1) - \Gamma(1)}{\alpha(\alpha - 1)} \quad \text{for } j \in \{0, 1\},$$

where the limit on the right can be easily evaluated by using L'Hospital's rule and the known formulas $\Gamma'(1) = -\gamma$, $\Gamma'(2) = 1 - \gamma$, thus leading to the values μ_j, $j \in \{0, 1\}$, given in (7.25). The continuity and the inequality $\mu_\alpha(f) \geq \mu_\alpha$ for $\alpha \in (-1, \infty)$ follow from (7.24) and (7.25) because $D_\alpha(F_0, F)$ is nonnegative and continuous in $\alpha \in \mathbb{R}$ and $\Gamma(\alpha + 1)$ is positive and continuous in $\alpha \in (-1, \infty)$. The condition for equality follows from the fact that $D_\alpha(F_0, F)$ is positive unless $F = F_0$. □

Proof of Assertion 7.3.1 Similarly as we applied Assertion 6.3.1 in the proof of Assertion 7.2.1, (7.28) follows for all $\alpha > -1/2$ from Assertion 6.4.1. If $\alpha \notin \{0, 1\}$, then the expressions for $m_\alpha(\ell)$ and σ_α^2 given in (7.29) and (7.30) follow easily from the formulas given for $m_{\phi_\alpha}(\ell)$ and $\sigma_{\phi_\alpha}^2$ in Assertion 6.4.1, but the direct evaluation of $m_j(\ell)$ and σ_j^2 from these formulas for $j \in \{0, 1\}$ is a somewhat tedious task. However, by using the continuity of $m_\alpha(\ell)$ and σ_α^2 established in Assertion 7.1.2, we obtain $m_j(\ell)$ and σ_j^2 given in (7.29) and (7.31) as the limits

$$m_j(\ell) = \lim_{\alpha \to j} m_\alpha(\ell) \quad \text{and} \quad \sigma_j^2 = \lim_{\alpha \to j} \sigma_\alpha^2 \quad \text{for } j \in \{0, 1\},$$

which expressions can be easily evaluated by using the continuity of the right-hand side of (7.29) and L'Hospital's rule, thereby employing the formulas

$$\begin{aligned} \Gamma(\alpha + k + 1) &= (\alpha + k)\,(\alpha + k - 1) \cdots (\alpha + 1)\,\Gamma(\alpha + 1), \\ \Gamma''(\alpha + 1) &= 2\Gamma'(\alpha) + \alpha\Gamma''(\alpha) \end{aligned}$$

and

$$\Gamma''(1) = \frac{\pi^2}{6} + \gamma^2, \;\; \Gamma''(2) = \frac{\pi^2}{6} - 2\gamma + \gamma^2, \;\; \Gamma''(3) = \frac{\pi^2}{3} + 2 - 6\gamma + 2\gamma^2$$

in addition to the previously used $\Gamma'(1) = -\gamma$ and $\Gamma'(2) = 1 - \gamma$. □

Proof of Assertion 7.4.1 If $\psi : [0, 1] \mapsto \mathbb{R}$ is continuous then by the assumptions about f

$$\inf_{x \in [0,1]} f(x) > 0 \quad \text{and} \quad \sup_{x \in [0,1]} |\psi(x)\, f(x)| < \infty$$

and, consequently, the function

$$\Psi(x) = \int_0^x \psi(y)\, f(y)\, \mathrm{d}y, \quad x \in (0, 1)$$

is well defined. Since

$$\frac{\mathrm{d}}{\mathrm{d}x} \frac{\Psi^2}{F} = -\left(\frac{\Psi}{F}\right)^2 f + \frac{2\Psi\psi f}{F}$$

and

$$|\Psi(y)| \le y \sup_{x\in[0,1]} |\psi(x)\, f(x)| \quad \text{as well as} \quad F(y) \ge y \inf_{x\in[0,1]} f(x),$$

the function Ψ satisfies the relation

$$\int_0^1 (\psi - \Psi/F)^2 f \,\mathrm{d}x = \int_0^1 \psi^2 f \,\mathrm{d}x - \left(\int_0^1 \psi\, f \,\mathrm{d}x\right)^2. \tag{9.67}$$

To this end take into account the relations

$$\begin{aligned}\int_0^1 (\psi - \Psi/F)^2 f \,\mathrm{d}x &= \int_0^1 \psi^2 f \,\mathrm{d}x - \int_0^1 \frac{2\Psi\psi f}{F} \,\mathrm{d}x + \int_0^1 \left(\frac{\Psi}{F}\right)^2 f \,\mathrm{d}x \\ &= \int_0^1 \psi^2 f \,\mathrm{d}x - \left(\frac{\Psi^2(1)}{F(1)} - \lim_{y\downarrow 0} \frac{\Psi^2(y)}{F(y)}\right) \\ &= \int_0^1 \psi^2 f \,\mathrm{d}x - \frac{\Psi^2(1)}{F(1)}.\end{aligned}$$

Now, using (9.67) we obtain (7.39) from the definition (7.36). Since f is assumed to be bounded and bounded away from 0,

$$\begin{aligned}\lim_{\alpha\to 0} \Delta_\alpha(F_0, F) &= \int_0^1 \left(\lim_{\alpha\to 0} \frac{f^{-\alpha} - 1}{\alpha} - \int_0^1 \lim_{\alpha\to 0} \frac{f^{-\alpha} - 1}{\alpha} f \,\mathrm{d}y\right)^2 f \,\mathrm{d}x \\ &= \int_0^1 \left(\ln f - \int_0^1 (\ln f)\, f \,\mathrm{d}y\right)^2 f \,\mathrm{d}x \\ &= \Delta_0(F_0, F)\end{aligned}$$

which proves the continuity at $\alpha = 0$. The reflexivity is clear from (7.39) and (7.40). □

Proof of Assertion 7.4.2 The proof should be clear from what was said above. The inequality $\sigma_\alpha^2(f) \ge \sigma_\alpha^2$ and the condition for equality follow from (7.38), because $D_{2\alpha}(F_0, F)$ and $\Delta_\alpha(F_0, F)$ are nonnegative measures of divergence of F_0 and F, which are equal to zero if and only if $F = F_0$, in which case the excess function (7.41) is 0. □

Acknowledgements

This research was supported by the MŠMT grant 1M0572, GAČR grant 202/10/0618 and K.U. Leuven project GOA/98/06.

References

Anderson, T. W. and Darling, D. A. (1954) A test of goodness-of-fit. *J. Amer. Statist. Assoc.* **49**, 765-769.

Chernoff, H. (1952) A measure of asymptotic efficiency of tests of a hypoothesis based on a sum of observations. *Ann. Math. Statist.*, **23**, 493-507.

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.

Cressie, N. and Read, T. R. C. (1984) Multinomial goodness-of-fit tests. *Journal of the Royal Statistical Society Series B*, **46**, 440-464.

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.

Csiszár, I. (1967) On topological properties of f-divergence. *Studia Scientiarum Math. Hungarica*, **2**, 329-339.

Darling, D. A. (1953) On a class of problems related to the random division of an interval. *Ann. Math. Statist.*, **24**, 239–253.

Darling, D. A. (1957) The Kolmogorov-Smirnov, Cramer-von Mises tests. *Ann. Math. Statist.*, **28**, 823–838.

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

Devroye, L. and Györfi, L. (1985) *Nonparametric Density Estimation: The L1 View*. J. Wiley, New York.

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

Durbin, J. (1973) *Distribution theory for tests based on the sample distribution function*. Regional Conference Series in Applied Mathematics, Vol. 9. Philadelphia: SIAM.

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.

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., Rousseeuw, P.J. and Stahel, W.A. (1986) *Robust Statistics. The Approach Based on Influence Functions.* Wiley, Ne York.

Jammalamadaka, S. R. and Tiwari, R. C. (1986) Efficiencies of some disjoint spacings tests relative to a χ^2 test. In M.L.Puri, J.P. Vilaplana and W.Wertz (eds), *New Perspectives in Theoretical and Applied Statistics*, 311-318. 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.

Jimenez, R. and Shao, Y. (2009) Limit theorems for ϕ-divergences based on k-spacings. *Comm. Statist. Theory Methods*, **38**, 695–710.

Jureckova, J. and Sen, P.K. (1996) *Robust Statistical Procedures.* Wiley

Kafka,P., Österreicher, F. and Vincze, I. (1991) On powers of Csiszár f-divergences defining a distance. *Studia Sci. Math. Hungar*. **26**, 415–422.

Karian, Z. A. and Dudewicz, E. J. (2008) *Fitting Statistical Distributions. The Generalized Lambda Distribution and Generalized Bootstrap Methods*. Boca Raton, FL: CRC Press.

Kolmogorov A. N. (1933) Sulla determinazione empirica di una legge di distibuziane. *Giorna. Ist. Attuari.* **4**, 83–91.

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.

Kimball, B. F. (1950) On the asymptotic distribution of the sum of powers of unit frequency differences. *Ann. Math. Statist.*, **21**, 263–271.

Kuo, M. and Rao, J. S. (1981) Limit theory and efficiencies for tests based on higher order spacings. In *Statistics – Applications and New Directions.* Proc. Golden Jubilee Conference of the Indian Statistical Institute, 333–352. Calcutta: Statistical Publishing Society.

Lehmann, E. and Romano (2005) *Testing Statistical Hypotheses*, Third Edition, Springer.

Liese, F. (1982) Hellinger integrals of Gaussian processes with independent increments. *Stochastics* **6**, 81-96.

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.

Matusita, K. (1956) Distances and decision rules. *Ann. Inst. Statist. Math.*, **16**, 305-320.

Menéndez, M., Morales, D. Pardo, L. and Vajda, I. (1998) Two approaches to grouping of data and related disparity statistics. *Comm. Statist. Theory 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.

Ning, W., Gao, Y., Dudewicz, E.J. (2008) Fitting mixture distributions using generalized lambda distributions and comparison with normal mixtures. *American Journal of Mathematical and Management Sciences*, **28**, Nos. 1 & 2, 81-99.

Pearson, K., (1894) Contributions to the mathematical theory of evolution. *Phil. Trans. of the Roy. Soc. of London A.* **185**, 71-110.

Perez, A. (1967) Information-theoretic risk estimates in statistical decision. *Kybernetika,* **3**, 1-21.

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.

Rényi, A. (1961) On measures of entropy and information. *Proc. 4th Berkeley Symp. on Probab. Th. Math. Statist.*, Vol. 1, 547-561.

Serfling, R. H. (1980) *Approximation Theorems of Mathematical Statistics.* Wiley, New York.

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

Smirnov, N. V. (1948) Tables for estimating the goodness of fit of empirical distributions. *Ann. Math. Statist.*, **19**, 279-281.

van Es, B. (1992) Estimating functionals related to a density by a class of statistics based on spacings. *Scand. J. Statist.*, **19**, 61–72.

Vajda, I. (1971) Limit theorems for total variation of cartesian product measures. *Studia Sci. Math. Hungar.* **6**, 317-333.

Vajda, I. (2007) Asymptotic comparisons of divergence-based goodness-of-fit statistics. *Publ. Stastist. Inst. Univ. Paris* , **51**, 49-66.

Vajda, I. (2009) On metric divergences of probability measures. *Kybernetika,* **45**, 885-900.

Vajda, I. and van der Meulen, E. C. (2006) On estimation and testing based on m-spacings. *Prague Stochastics 2006* (Eds Hušková M., Janžura M.), Matfyzpress, Prague.

Vajda, I. and van der Meulen, E. C. (2006) *Divergences Between Models and Data under Hypothetical and Empirical Quantiles.* Res. Report No. 2275, Institute of Information Theory and Automation, Prague (*available online at* **http://simu0292.utia.cas.cz/vajda/RR2274**).

Vasicek, O. (1976) A test for normality based on sample entropy. *J. Roy. Statist. Soc. Ser. B*, **38**, 54-59.

von Mises, R. (1947) On the asymptotic distribution of differential statistical functions. *Ann. Math. Statist.*, **18**, 309–348.

Weiss, L. (1957) The asymptotic power of certain test of fit based on sample spacings. *Ann. Math. Statist.*, **28**, 783–786.