ON EFFICIENT ESTIMATION IN CONTINUOUS MODELS BASED ON FINITELY QUANTIZED OBSERVATIONS

Domingo Morales

Operation Research Center

Miguel Hernández University of Elche

03202 Elche, Spain

Leandro Pardo

Department of Statistics and Operations Research

Complutense University of Madrid

28040 Madrid, Spain

Igor Vajda

Institute of Information Theory

Academy of Sciences of the Czech Republic

CZ–18208 Prague, Czech Republic

Key Words: Minimum disparity estimators, Quantized observations, Consistency, Asymptotic normality, Efficiency.

ABSTRACT

We consider minimum ϕ-divergence estimators $\widehat{\theta}_n^\phi$ of parameters θ of arbitrary dominated models $\mu_\theta \ll \lambda$ on the real line, based on finite quantizations of i.i.d. observations $X_1, \ldots, X_n$ from these models. The quantizations are represented by finite interval partitions $\mathcal{P}_n = (A_{n1}, \ldots, A_{nm_n})$ of the real line where m_n is allowed to increase to infinity for $n \to \infty$. The models with densities $f_\theta = d\mu_\theta/d\lambda$ are assumed to be regular in the sense that they admit finite Fisher informations $\mathcal{J}_\theta$. At the first place we have in mind continuous models dominated by the Lebesgue measure λ. Due to the quantizations, $\widehat{\theta}_n^\phi$ are discrete–model

estimators for which the desirable properties (computation complexity, robustness, etc.) can be controlled by a suitable choice of functions ϕ. We formulate conditions under which these estimators are consistent and efficient in the original models μ_θ in the sense that $\sqrt{n}(\widehat{\theta}_n^\phi - \theta) \xrightarrow{\mathcal{L}} N(0, \mathcal{J}_\theta^{-1})$ as $n \to \infty$.

1. INTRODUCTION

It is well known that quantizations of continuous data by fixed partitions of observation spaces reduce in all nontrivial situations the efficiency of statistical inference achievable in the original continuous model. In case of point estimators, the loss of efficiency can be measured by the asymptotic relative inefficiency (the ratio of asymptotic variances achieved in the original and quantized model, see e.g. p. 51 in Serfling (1980)). Many authors (see e.g. Ghurye and Johnson (1981), Zografos et al. (1986), Ryu (1993) and Tsairidis et al. (1997)) studied quantizations by fixed infinite partitions, where the loss of efficiency can be sufficiently small provided the partition of the whole observation space is sufficiently fine.

The situation with finite partitions is different - they cannot be uniformly fine on all parts of the observation space. For example, a finite partition of the real line always leaves two infinite intervals unpartitioned. Therefore the asymptotic relative inefficiency of estimators based on fixed finite partitions of any given size is typically unbounded. However, in Menéndez et al. (2001) we have shown that the adaptive (sample-dependent) partitions of the real line of size m defined by $m-1$ sample quantiles lead to bounded inefficiencies of a wide class of estimators, and that the losses of efficiency can be sufficiently small if m is sufficiently large. The estimators were of a minimum distance type, minimizing ϕ-divergence between theoretical and empirical distributions obtained in the quantized discrete models. The maximum likelihood estimator (MLE) based on the quantized data is included in this class.

Remind that the classes of point estimators minimizing the $\phi-$divergence between theoretical and empirical distributions were systematically studied by Vajda (1984 a, b, c, d). Special attention was payed to the estimators which minimize the so-called power divergences introduced independently by Cressie and Read (1984). For special divergences or special models the minimum $\phi-$divergence estimators were studied later more deeply by Liese and Vajda (1987), Read and Cressie (1988), Vajda (1989), Voss (1992) and other authors. The asymptotic theory for the models obtained by quantizations of continuous models of a fixed size was summarized in the above cited paper of Menéndez et al. (2001).

Computations based on sample-dependent partitions may be complicated if we work repeatedly with different samples of the same sample size n. Moreover, independently of how large m is, the full efficiency in the original continuous model cannot be unachieved in this manner. In this paper we are interested in the problem whether partitions of finite sizes m_n depending on the sample size n but not on the sample itself, and increasing to infinity with $n \to \infty$, can guarantee for the estimators studied in Menéndez et al. (2001) the efficiency in the original continuous model. More precisely, we are interested in whether the corresponding estimators $\widehat{\theta}_n$ of the true parameters $\theta \in \Theta \subset R^d$ are consistent in the sense

$$\widehat{\theta}_n \xrightarrow{P} \theta \qquad \text{as } n \to \infty \tag{1}$$

and asymptotically normal in the sense

$$\sqrt{n}\left(\widehat{\theta}_n - \theta\right) \xrightarrow{L} N\left(0, \mathcal{J}_\theta^{-1}\right) \qquad \text{as } n \to \infty \tag{2}$$

where $\mathcal{J}_\theta$ is the Fisher information matrix in the underlying continuous model. As to the partitions, we are interested in the sizes m_n growing slowly in the sense $m_n^2/n \to 0$ as $n \to \infty$.

We present relatively simple and intuitively appealing conditions on the model under which we prove (1) and (2) for all the minimum $\phi-$divergence estimators studied in Menéndez et al. (2001), and also for the estimators minimizing the more general $\phi-$disparity introduced in Meneńdez et al. (1998). Verification of the conditions may not be as simple as the conditions themselves, but it is possible as we illustrate by examples. Our results thus enable efficient estimation in the continuous models by using discrete-models estimators which may be computationally simpler, or to which we are restricted if we are facing a data compression based on finite quantization.

As to the applications of our results, let us mention that any of the estimators $\widehat{\theta}_n$ under consideration can replace the MLE based on the nonquantized observations $X_1, ..., X_n$ from the original continuous model, used in Watson (1959) and other references cited there for efficient estimation of nuisance parameters θ in the classical Pearson goodness-of-fit statistics with a fixed number of cells. These estimators can also be used for efficient estimation of nuisance parameters θ in the class of Moore-Spruil goodness-of-fit test statistics based on observations quantized into m_n cells assumed in the present paper, e.g. in the Rao-Robson-Nikulin statistics or the Dzhaparidze-Nikulin statistics (see Rao and Robson (1974), Dzhaparidze and Nikulin (1974) and Moore and Spruil (1975)).

As far as we know, general results similar to those in the present paper have not been established in the previous literature. Some results are known for particular estimators in particular models. For example, Drost (1989) studied the Moore-Spruil statistics with the location and scale nuisance parameters $\theta = (\mu, \sigma) \in \Theta = R \times (0, \infty)$, and he mentions on p. 1289 that the results (1) and (2) were proved by Bickel (1982) for the MLE $\widehat{\theta}_n = (\widehat{\mu}_n, \widehat{\sigma}_n)$ based on the observations quantized into increasing numbers m_n of cells similarly as in the present paper.

The exposition is divided into three sections. In Section 2 we introduce continuous models of parametric estimation and their discrete counterparts obtained by quatization of the observation space. To avoid unnecessary technicalities, we restrict ourselves to real observations and parameters. We introduce a class of estimators of a minimum divergence type (minimum ϕ-disparity estimators) for quantized continuous models and study the conditions on models and quantizations needed in Section 3. In Section 3 we prove the consistency (1.1) and the asymptotic normality (1.2) for the estimators introduced in Section 2. Section 4 contains examples illustrating practical applicability of the results of previous sections.

2. BASIC CONCEPTS AND AUXILIARY RESULTS

We consider a continuous bowl shaped function $\phi : [0, \infty) \longmapsto R$, decreasing on the subdomain $[0, 1]$ in the nonstrict sense $\phi(t_1) \geq \phi(t_2)$ if $0 \leq t_1 \leq t_2 \leq 1$, and increasing in a similar nonstrict sense on $[1, \infty)$. We assume that $\phi(1) = 0$ and that $\phi(t)$ is twice continuously differentiable in a neighborhood of $t = 1$ with the second derivative $\phi''(1) \neq 0$. Obviously, under these assumptions ϕ is nonnegative and $\phi''(1) > 0$.

If $\psi : [0, \infty) \mapsto R$ is convex, twice continuously differentiable in a neighborhood of 1 with $\psi''(1) \neq 0$, then

$$\phi(t) = \psi(t) - \psi'(1)(t-1) - \psi(1) \tag{3}$$

satisfies the assumptions of the present paper. For example, the convex functions $\psi_a(t) = t^\alpha/\alpha(\alpha-1)$, $\alpha \in (0,\infty) - \{1\}$, lead to the class

$$\phi_\alpha(t) = \frac{t^\alpha - \alpha(t-1) - 1}{\alpha(\alpha-1)}, \ \alpha \in (0,\infty) - \{1\}, \tag{4}$$

with the corresponding limits

$$\phi_1(t) = t\ln t - t + 1 \quad \text{and} \quad \phi_0(t) = -\ln t + t - 1, \tag{5}$$

assumed to be continuously extended to $t = 0$. We see that all ϕ_α with $\alpha > 0$ satisfy the assumptions of the paper while ϕ_0 does not so because $\phi_0(t) = \infty$ at $t = 0$. Examples of nonconvex $\phi : [0,\infty) \rightarrow R$ satisfying our assumptions are

$$\phi(t) = 1 - \frac{2t}{1+t^2} \tag{6}$$

or

$$\phi_\alpha(t) = 1 - \exp\left\{-\alpha(t-1)^2\right\}, \ \alpha > 0. \tag{7}$$

For the functions ϕ under consideration and discrete probability distributions $\boldsymbol{p} = (p_1, ..., p_m)$, $\boldsymbol{q} = (q_1, ..., q_m)$, where $\boldsymbol{q}$ dominates $\boldsymbol{p}$ in the usual sense $q_j = 0 \Rightarrow p_j = 0$, and where $1 < m < \infty$, we consider the $\phi-$*disparities*

$$D_\phi(\boldsymbol{p}; \boldsymbol{q}) = \sum_{j=1}^{m} q_j \phi\left(\frac{p_j}{q_j}\right) \quad \text{with } 0\phi\left(\frac{0}{0}\right) = 0. \tag{8}$$

Such measures of disparity between discrete distributions have been suggested by Lindsay (1994). In a rigorous systematic manner, they were introduced and studied by Menéndez et al. (1998). Since $\phi(t) \geq 0$, the $\phi-$disparities are well defined by (8) and finite for all $\boldsymbol{p}$, $\boldsymbol{q}$ under consideration. Obviously, $D_\phi(\boldsymbol{p}; \boldsymbol{q}) \geq 0$ with the equality if and only if $\boldsymbol{p} = \boldsymbol{q}$.

If ϕ is convex then the $\phi-$disparities $D_\phi(\boldsymbol{p}; \boldsymbol{q})$ are the $\phi-$divergences of Csiszár (1963, 1967). Let us mention that the $\phi-$divergences are defined for arbitrary convex functions

$\phi : (0, \infty) \to R$ which are strictly convex at $t = 1$ (i.e. not necessarily twice differentiable at $t = 1$ with $\phi''(1) > 0$) and satisfy the norming condition $\phi(1) = 0$. For example,

$$V(\boldsymbol{p};\boldsymbol{q}) = \sum_{j=1}^{m} |p_j - q_j| \tag{9}$$

is such a divergence, called *total variation* of $\boldsymbol{p}$ and $\boldsymbol{q}$ and defined by $\phi(t) = |t-1|$. The $\phi-$divergences have been systematically studied by Liese and Vajda (1987). We shall refer there for the details needed in this paper. The basic property of all $\phi-$divergences, the reflexivity $D_\phi(\boldsymbol{p};\boldsymbol{q}) \geq 0$ with the equality if and only if $\boldsymbol{p} = \boldsymbol{q}$, is well known.

The functions $\phi_\alpha, \alpha \in R$, specified by the formulas of (4), (5) are convex, thus defining $\phi_\alpha-$divergences denoted in the present paper by $I_\alpha(\boldsymbol{p};\boldsymbol{q})$, i.e.,

$$I_\alpha(\boldsymbol{p};\boldsymbol{q}) = \begin{cases} \sum_{j=1}^{m} p_j \ln(p_j/q_j) & \text{if } \alpha = 1, \\ \left(\sum_{j=1}^{m} p_j^\alpha q_j^{1-\alpha} - 1\right) / [\alpha(\alpha-1)] & \text{if } \alpha \neq 0, \alpha \neq 1, \\ \sum_{j=1}^{m} q_j \ln(q_j/p_j) & \text{if } \alpha = 0, \end{cases} \tag{10}$$

where $q/0$ is assumed to be ∞ for $q > 0$. As stated above, $I_\alpha(\boldsymbol{p};\boldsymbol{q})$ are $\phi_\alpha-$disparities for $\alpha > 0$.

Let us now consider two arbitrary divergence measures $D(\boldsymbol{p};\boldsymbol{q})$, $\widetilde{D}(\boldsymbol{p};\boldsymbol{q})$ which are reflexive in the previously mentioned sense. We say that the $\widetilde{D}-$divergence measure is *topologically stronger* than the $D-$divergence measure if for arbitrary sequences of discrete distributions $\boldsymbol{p}_n = (p_{n1}, ..., p_{nm_n})$ and $\boldsymbol{q}_n = (q_{n1}, ..., q_{nm_n})$, $1 < m_n < \infty$,

$$\lim_{n\to\infty} \widetilde{D}(\boldsymbol{p}_n;\boldsymbol{q}_n) = 0 \quad \text{implies} \quad \lim_{n\to\infty} D(\boldsymbol{p}_n;\boldsymbol{q}_n) = 0. \tag{11}$$

In what follows we shall need the following result, where it is taken into account that there are $\phi-$divergences which are not ϕ-disparities and vice versa.

Proposition 2.1. All $\phi-$divergences $D_\phi(\boldsymbol{p};\boldsymbol{q})$ are topologically stronger than the total variation $V(\boldsymbol{p};\boldsymbol{q})$, and also all ϕ-disparities $D_\phi(\boldsymbol{p};\boldsymbol{q})$ are topologically stronger than the total variation.

Proof. The first assertion has been proved in Vajda (1972) (see also Proposition 9.49 in Vajda (1989)). To prove the second assertion, take into account that, by the definition of $\phi-$disparity, there exists $\varepsilon > 0$ such that

$$\inf_{t\in[1-\varepsilon,1+\varepsilon]} \phi''(t) = \widetilde{c} > 0 \quad \text{and} \quad \inf_{t\notin(1-\varepsilon,1+\varepsilon)} \phi(t) = \widetilde{\widetilde{c}} > 0.$$

Let $c = \min\left\{\widetilde{c}, \widetilde{\widetilde{c}}\right\} > 0$ and suppose that

$$\lim_{n\to\infty} D_\phi(\boldsymbol{p}_n; \boldsymbol{q}_n) = 0. \tag{12}$$

Put

$$A_n = \{1 \le j \le m_n : |1 - p_{nj}/q_{nj}| \, \boldsymbol{I}(q_{nj} > 0) > \varepsilon\}$$

and

$$B_n = \{1 \le j \le m_n : |1 - p_{nj}/q_{nj}| \, \boldsymbol{I}(q_{nj} > 0) < \varepsilon\}$$

where $\boldsymbol{I}$ denotes the indicator function. Then (12) and the Taylor theorem for $\phi(t)$ with $\phi(1) = \phi'(1) = 0$ imply for all $\boldsymbol{p}_n$ and $\boldsymbol{q}_n$ under consideration,

$$\begin{aligned}
D_\phi(\boldsymbol{p}_n; \boldsymbol{q}_n) = & \sum_{j\in A_n} q_{nj}\phi\left(\tfrac{p_{nj}}{q_{nj}}\right) + \sum_{j\in B_n} q_{nj}\phi\left(\tfrac{p_{nj}}{q_{nj}}\right) \\
\ge & \; c\left[\sum_{j\in A_n} q_{nj} + \tfrac{1}{2}\sum_{j\in B_n} q_{nj}\left(\tfrac{p_{nj}}{q_{nj}} - 1\right)^2\right] \\
= & \; c\left[1 - Q_n + \tfrac{1}{2}\sum_{j\in B_n} q_{nj}\left(1 - \tfrac{p_{nj}}{q_{nj}}\right)^2\right],
\end{aligned}$$

where

$$Q_n = \sum_{j\in B_n} q_{nj}.$$

If Q_n is zero then $D_\phi(\boldsymbol{p}_n; \boldsymbol{q}_n) \ge c$. Hence, by (12), Q_n is positive for all n large enough with $\lim_{n\to\infty} Q_n = 1$ and, by Jensen's inequality for the convex function $\psi(t) = t^2$,

$$\begin{aligned}
\sum_{j\in B_n} q_{nj}\left(1 - \tfrac{p_{nj}}{q_{nj}}\right)^2 \ge & \; \tfrac{1}{Q_n}\left(\sum_{j\in B_n} q_{nj}\left|1 - \tfrac{p_{nj}}{q_{nj}}\right|\right)^2 \\
\ge & \left(\sum_{j\in B_n} |p_{nj} - q_{nj}|\right)^2.
\end{aligned}$$

Therefore, taking into account (12), we obtain that

$$\beta_n \triangleq \sum_{j\in B_n} |p_{nj} - q_{nj}| \to 0 \quad \text{ as } n \to \infty.$$

This means that if we put

$$P_n = \sum_{j\in B_n} p_{nj}$$

then $\lim_{n\to\infty} Q_n = 1$, together with the obvious inequality $Q_n \le \beta_n + P_n$, implies that also $\lim_{n\to\infty} P_n = 1$. Further, the inequality obtained above under the assumption $Q_n > 0$ can be written in the form

$$D_\phi(\boldsymbol{p}_n; \boldsymbol{q}_n) \ge c\left[1 - Q_n + \left(V(\boldsymbol{p}_n; \boldsymbol{q}_n) - \alpha_n\right)^2/2\right], \tag{13}$$

where $0 \le \alpha_n \le V(\boldsymbol{p}_n; \boldsymbol{q}_n)$ is defined by the equation $\alpha_n + \beta_n = V(\boldsymbol{p}_n; \boldsymbol{q}_n)$. Since

$$0 = \sum_{j\notin A_n\cup B_n} q_{nj} = \sum_{j\notin A_n\cup B_n} p_{nj},$$

it holds

$$\alpha_n = \sum_{j\in A_n} |p_{nj} - q_{nj}| \le \sum_{j\in A_n} p_{nj} + \sum_{j\in A_n} q_{nj} = 2 - P_n - Q_n,$$

i.e. the limit relations established under (12) for P_n and Q_n imply that $\lim_{n\to\infty}\alpha_n = 0$. Therefore we see from (13) that (12) implies the desired relation

$$\lim_{n\to\infty} V(\boldsymbol{p}_n; \boldsymbol{q}_n) = 0.$$

□

In this paper we study the *classical statistical model* with i.i.d. observations $X_1, ..., X_n$ distributed by μ_{θ_0} from a family $\{\mu_\theta : \theta \in \Theta\}$ of probability measures on Borel subsets of the real line R. The family is supposed to be dominated by a $\sigma-$finite measure λ on R, leading to the densities

$$f_\theta = \frac{d\mu_\theta}{d\lambda}, \ \theta \in \Theta, \tag{14}$$

on R. For simplicity we assume that Θ is an open subset of R. We also assume that the family $\{\mu_\theta : \theta \in \Theta\}$ is *regular* in the sense that all measures are concentrated on a common support

$$S = \{x \in R : f_\theta(x) > 0\}, \ \theta \in \Theta, \tag{15}$$

i.e. that they are measure-theoretically equivalent. Further, the densities $f_\theta(x)$ are assumed to be differentiable in θ for $\lambda-$almost all $x \in S$, with the derivatives

$$\dot{f}_\theta(x) = \frac{d}{d\theta} f_\theta(x) \tag{16}$$

satisfying the *Fisher information* regularity condition

$$0 < \mathcal{J}_\theta \triangleq \int \frac{\dot{f}_\theta^2}{f_\theta} d\lambda < \infty, \ \theta \in \Theta. \tag{17}$$

We are interested in the estimation of the unknown true parameter θ_0 on the basis of observations $\boldsymbol{X}_n = (X_1, ..., X_n)$ for increasing sample sizes $n = 1, 2, ...$ Our approach is specific in that it is restricted to nonsufficient statistics $\widehat{\boldsymbol{p}}_n$ of the observations, obtained by quantizations of the observation components X_i, $1 \leq i \leq n$. The quantizations are admitted to depend on sample sizes n, and they are represented by measurable partitions

$$\mathcal{P}_n = \{A_{nj} : 1 \leq j \leq m_n\}, \ n \in N, \tag{18}$$

of the convex envelope of the support S defined in (15). It is assumed that $m_1 \leq m_2 \leq ...$ are finite and

$$\lim_{n \to \infty} m_n = \infty. \tag{19}$$

The support S as well as the partition sets A_{nj} are usually intervals.

The partitions (18) lead to *discrete statistical models* $\{\boldsymbol{p}_n(\theta) : \theta \in \Theta\}$ where

$$\boldsymbol{p}_n(\theta) = \left(\boldsymbol{p}_{nj}(\theta) \triangleq \mu_\theta(A_{nj}) = \int_{A_{nj}} f_\theta d\lambda, \ 1 \leq j \leq m_n\right), \tag{20}$$

and to *discrete empirical distributions*

$$\widehat{\boldsymbol{p}}_n = \left(\widehat{p}_{nj} \triangleq \widehat{\mu}_n(A_{nj}) : 1 \leq j \leq m_n\right), \ \widehat{\mu}_n = \frac{1}{n}\sum_{i=1}^{n} \delta_{X_i}, \tag{21}$$

obtained from the empirical probability measure $\widehat{\mu}_n$ defined on the Borel subsets of R. These distributions represent the empirical evidence available in the discrete models, resulting from the original empirical evidence represented by $\widehat{\mu}_n$.

We see that $\widehat{\boldsymbol{p}}_n$ is dominated by $\boldsymbol{p}_n(\theta_0)$ (and, due to the measure - theoretic equivalence of μ_{θ_0} and μ_θ, by any $\boldsymbol{p}_n(\theta)$, $\theta \in \Theta$) in the sense specified in the definition of ϕ-disparity. Hence the ϕ-disparities $D_\phi(\widehat{\boldsymbol{p}}_n, \boldsymbol{p}_n(\theta))$ are well-defined by (8) and finite for all $\theta \in \Theta$. We shall assume without loss of generality that

$$p_{nj}(\theta) > 0 \quad \text{for all } 1 \leq j \leq m_n \text{ and } \theta \in \Theta. \tag{22}$$

It is clear from the definition (8) that $D_\phi(\widehat{\boldsymbol{p}}_n, \boldsymbol{p}_n(\theta))$ is continuous (or differentiable) in θ if all $p_{nj}(\theta)$ are continuous (differentiable) in θ.

Transformations $\widehat{\mu}_n \mapsto \widehat{\boldsymbol{p}}_n$ defined by the quantizations $\mathcal{P}_n$ are usually not statistically sufficient for the original model $\{\mu_\theta : \theta \in \Theta\}$. Consequently, the estimation procedures available in the discrete model $\{\boldsymbol{p}_n(\theta) : \theta \in \Theta\}$ are less efficient than those available in the original model. Possible exceptions are the cases where λ is a counting measure supported by a set of isolated points $s_j \in R$, $1 \leq j \leq m$, and each set $A_{nj} \in \mathcal{P}_n$, $1 \leq j \leq m \leq m_n$, covers only one support point s_j. In this case the original and the transformed model are statistically equivalent. We are interested only in the cases where this is not true. For example, if $\{\mu_\theta : \theta \in \Theta\}$ is discrete as presented above, then we are interested only in the situations where $m = \infty$ (e.g., the original model is geometric or Poisson) and $\mathcal{P}_n$ is a finite partition with the last set A_{nm_n} covering all points s_j, $j \geq m_n$. Then $\{\boldsymbol{p}_n(\theta) : \theta \in \Theta\}$ consists of m_n-dimensional reductions of the infinitely dimensional probability vectors from the original family $\{\mu_\theta : \theta \in \Theta\}$. But in the typical situation which we have in mind, the dominating λ is the Lebesgue measure (or absolutely continuous with respect to the Lebesgue measure), so that the original model is continuous in the common sense.

We study the ϕ-*disparity estimators* $\widehat{\theta}_n = \widehat{\theta}_n^\phi$ of the true parameter $\theta_0 \in \Theta$, defined as measurable functions of the empirical evidence $\widehat{\boldsymbol{p}}_n$ satisfying a.s. the condition

$$D_\phi\left(\widehat{\boldsymbol{p}}_n; \boldsymbol{p}_n\left(\widehat{\theta}_n\right)\right) = \min_{\theta \in \Theta} D_\phi(\widehat{\boldsymbol{p}}_n; \boldsymbol{p}_n(\theta)). \tag{23}$$

In the rest of this section we formulate assumptions about the original model $\{\mu_\theta : \theta \in \Theta\}$ $\equiv \{f_\theta : \theta \in \Theta\}$ and partitions $\mathcal{P}_n$ which are useful for the characterization of properties of ϕ-disparity estimators in the next section.

(A1) The components of discrete distributions $\boldsymbol{p}_n(\theta)$ are twice continuously differentiable in θ, with the first derivatives

$$\dot{\boldsymbol{p}}_n(\theta) = \left(\dot{p}_{nj}(\theta) \triangleq \frac{d}{d\theta} p_{nj}(\theta) : 1 \le j \le m_n\right)$$

satisfying the relation

$$\dot{p}_{nj}(\theta) = \int_{A_{nj}} \dot{f}_\theta d\lambda, \ \theta \in \Theta.$$

In what follows we need also the Fisher informations $\mathcal{J}_{\theta,n}$ for the discrete models $\{\boldsymbol{p}_n(\theta) : \theta \in \Theta\}$,

$$\mathcal{J}_{n,\theta} = \sum_{j=1}^{m_n} \frac{(\dot{p}_{nj}(\theta))^2}{p_{nj}(\theta)}, \ \theta \in \Theta. \tag{24}$$

Proposition 2.2. The assumption $\mathcal{J}_\theta < \infty$ introduced by (17) implies the absolute $\lambda-$integrability of the derivatives $\dot{f}_\theta$, so that the integrals considered in (A1) are well defined and finite. Further, under (A1) the conditions

$$\sum_{j=1}^{m_n} \dot{p}_{nj}(\theta) = 0 \quad \text{and} \quad \int_S \dot{f}_\theta d\lambda = 0 \tag{25}$$

are equivalent and

$$0 \le \mathcal{J}_{n,\theta} \le \mathcal{J}_\theta. \tag{26}$$

Proof. By the Schwarz inequality,

$$\int_S \left|\dot{f}_\theta\right| d\lambda = \int_S \sqrt{f_\theta}\left(\left|\dot{f}_\theta\right| / \sqrt{f_\theta}\right) d\lambda \le \sqrt{\mathcal{J}_\theta}$$

which proves the absolute integrability. Under the absolute integrability, (A1) implies

$$\sum_{j=1}^{m_n} \dot{p}_{nj}(\theta) = \int_S \dot{f}_\theta d\lambda$$

irrespectively of whether m_n is finite or not. The right-hand inequality in (26) follows under (A1) by using the Jensen inequality, see e.g. Vajda (1973). $\square$

(A2) The discrete distributions $\boldsymbol{p}_n(\theta)$ satisfy the limit relation

$$\lim_{n\to\infty} \mathcal{J}_{n,\theta_0} = \mathcal{J}_{\theta_0}$$

for the Fisher informations $\mathcal{J}_{n,\theta_0}$ and $\mathcal{J}_{\theta_0}$ defined by (24) and (17).

Next we give a sufficient condition for (A2) jointly with a condition for an analogous property of the $\phi-$divergences. Note that the $\phi-$divergences satisfy a similar inequality

$$0 \le D_\phi(\boldsymbol{p}_n(\theta); \boldsymbol{p}_n(\theta_0)) \le D_\phi(\mu_\theta, \mu_{\theta_0}) \triangleq \int_S f_{\theta_0}\phi\left(\frac{f_\theta}{f_{\theta_0}}\right) d\lambda \tag{27}$$

as the Fisher informations in (26) (cf. Liese and Vajda (1987)).

Proposition 2.3. Let all sets A_{nj} in the partitions $\mathcal{P}_n$ be intervals and, for every $\theta \in \Theta$,

$$\lim_{n\to\infty} \max_{1\le j\le m_n} p_{nj}(\theta) = 0. \tag{28}$$

Then all $\phi-$divergences satisfy the relation

$$\lim_{n\to\infty} D_\phi(\boldsymbol{p}_n(\theta); \boldsymbol{p}_n(\theta_0)) = D_\phi(\mu_\theta, \mu_{\theta_0}) \tag{29}$$

and (A1) implies (A2).

Proof. See Theorem 3 and Corollary 1 in Vajda (2002). □

Remark 2.1. The condition (28) of Proposition 2.3 holds for all $\theta \in \Theta$ if the dominating measure λ satisfies the relation

$$\lim_{n\to\infty} \max_{1\le j\le m_n} \lambda(A_{nj}) = 0.$$

In particular, it holds if

$$\lambda(S) < \infty \quad \text{and } \lambda_{nj} = \frac{\lambda(A_{nj})}{\lambda(S)} = \frac{1}{m_n} \quad \text{(cf. (19))}. \tag{30}$$

Since all measures $\mu_\theta, \theta \in \Theta$, are assumed to be measure-theoretically equivalent, we shall consider also the special case of (30) where

$$\lambda = \mu_{\theta_0}, \quad p_{nj}(\theta_0) = \frac{1}{m_n}. \tag{31}$$

Now we formulate an assumption about partitions $\mathcal{P}_n$ additional to (19).

(A3) The partitions satisfy the conditions

$$\lim_{n\to\infty} m_n = \infty, \qquad \lim_{n\to\infty} \frac{m_n^2}{n} = 0 \tag{32}$$

and

$$0 < \text{liminf}_{n\to\infty} m_n \min_{1\le j\le m_n} p_{nj}(\theta_0) \le \text{limsup}_{n\to\infty} m_n \max_{1\le j\le m_n} p_{nj}(\theta_0) < \infty.$$

This assumption has important consequences. In the following proposition, and in the rest of the paper, we say that the second derivative $\phi''(t)$ of a $\phi-$function is *Lipschitz at* $t = 1$ if, for all t from a neighborhood of 1, $|\phi''(t) - \phi''(1)| \le$ const. $|t-1|$. Obviously, $\phi''(t)$ is Lipschitz at $t = 1$ if the third derivative $\phi'''(t)$ exists in a neighborhood of $t = 1$. This condition is satisfied by all the concrete examples of ϕ-functions given above such that $\phi''(t)$ at $t = 1$ exists.

Proposition 2.4. If (A3) holds then, asymptotically for $n \to \infty$,

$$\|\widehat{\boldsymbol{p}}_n - \boldsymbol{p}_n(\theta_0)\| = O_p\left(1/\sqrt{n}\right) \tag{33}$$

and, for all $\phi-$disparities with $\phi''(t)$ Lipschitz at $t = 1$,

$$D_\phi(\widehat{\boldsymbol{p}}_n; \boldsymbol{p}_n(\theta_0)) = O_p\left(\frac{m_n}{n}\right). \tag{34}$$

Proof. By Theorem 2 of Györfi and Vajda (2001), under (A3) (without the restriction on $\max p_{nj}(\theta_0)$) the mentioned $\phi-$disparities satisfy the asymptotic law

$$\frac{n}{2\sqrt{m_n}}\left(\frac{D_\phi(\widehat{\boldsymbol{p}}_n; \boldsymbol{p}_n(\theta_0))}{\phi''(1)} - \frac{m_n}{n}\right) \xrightarrow{L} N(0,1) \quad \text{as } n \to \infty.$$

Thus (A3) implies (34). For the disparity $I_{2n} \triangleq I_2(\widehat{\boldsymbol{p}}_n, \boldsymbol{p}_n(\theta_0))$ defined by (10) for $\alpha = 2$, this means that

$$2I_{2n} = \sum_{j=1}^{m_n} \frac{(\hat{p}_{nj} - p_{nj}(\theta_0))^2}{p_{nj}(\theta_0)} = O_p\left(\frac{m_n}{n}\right).$$

But the restriction on $\max p_{nj}(\theta_0)$ in (A3) implies that there exists $C < \infty$ such that, for all sufficiently large n,

$$2I_{2n} > \frac{m_n}{C} \left\| \boldsymbol{p}_n - \boldsymbol{p}_n(\theta_0) \right\|^2 .$$

This together with (34) implies (33). □

3 CONSISTENCY AND ASYMPTOTIC NORMALITY

Let us start with the consistency of ϕ-disparity estimators $\widehat{\theta}_n = \widehat{\theta}_n^{\phi}$. Our proof of consistency is based on the following assumption employing the total variation distance (9).

(A4) The true parameter θ_0 is identifiable in the family of models $\{\boldsymbol{p}_n(\theta) : \theta \in \Theta\}$, $n \in N$, in the sense that a sequence $\theta_n \in \Theta$ satisfies the relation

$$\lim_{n\to\infty} V\left(\boldsymbol{p}_n(\theta_n), \boldsymbol{p}_n(\theta_0)\right) = 0$$

only if $\lim_{n\to\infty} \theta_n = \theta_0$.

In discrete models $\{\boldsymbol{p}(\theta) : \theta \in \Theta\}$ not depending on the sample size n, and continuous in the sense that $\theta \to \theta_0$ implies $V(\boldsymbol{p}(\theta), \boldsymbol{p}(\theta_0)) \to 0$, the true parameter is usually considered to be identifiable if for all sufficiently small $\varepsilon > 0$

$$\inf_{|\theta-\theta_0|>\varepsilon} V\left(\boldsymbol{p}(\theta), \boldsymbol{p}(\theta_0)\right) > 0.$$

Then, in particular, $V(\boldsymbol{p}(\theta), \boldsymbol{p}(\theta_0)) > 0$ for $\theta \neq \theta_0$, i.e. the mapping $\theta \mapsto \boldsymbol{p}(\theta)$ is one-to-one and continuous (in the Euclidean and total variation topology). Further, $V(\boldsymbol{p}(\theta_n), \boldsymbol{p}(\theta_0)) \to 0$ implies $|\theta_n - \theta_0| < \varepsilon$ for small $\varepsilon > 0$ and all sufficiently large n. If $\theta_{n_k} \to \theta_* \in \Theta$ then the triangle inequality implies that $V(\boldsymbol{p}(\theta_*), \boldsymbol{p}(\theta_0))$ is bounded above by a sequence tending to zero, which contradicts the assumption $\theta_* \neq \theta_0$. This means that also the inverse mapping $\boldsymbol{p}(\theta) \mapsto \theta$ is continuous.

Let us now consider the original model $\{\mu_\theta : \theta \in \Theta\}$ where $\Theta = (a, b) \subseteq R$. Further, consider a sequence of interval partitions $\mathcal{P}_n$ of the support S of $\{\mu_\theta : \theta \in \Theta\}$ such that for all $\theta \in \Theta$ and

$$V(\mu_\theta, \mu_{\theta_0}) = \int |f_\theta - f_{\theta_0}| \, d\lambda \qquad \text{(cf. (27) for } f(t) = |t-1|)$$

it holds

$$\rho_n(\theta,\theta_0) \triangleq V(\boldsymbol{p}_n(\theta), \boldsymbol{p}_n(\theta_0)) \to \rho(\theta,\theta_0) \triangleq V(\mu_\theta, \mu_{\theta_0}) \tag{35}$$

as $n \to \infty$ (see Proposition 2.3 for sufficient conditions), and suppose that $\rho(\theta,\theta_0) \neq 0$ for $\theta \neq 0$ and $\rho(\theta,\theta_0) \to 0$ if $\theta \to \theta_0$. Similarly as in the discrete case, the identifiability condition

$$\inf_{|\theta-\theta_0|>\varepsilon} \rho(\theta,\theta_0) > 0 \tag{36}$$

for all sufficiently small $\varepsilon > 0$ implies that the mapping $\theta \mapsto \mu_\theta$ is one-to-one and continuous, and also that the inverse mapping $\mu_\theta \mapsto \theta$ is continuous (in the same topologies as considered above). Unfortunately, if $n \to \infty$ then, in general, $\rho_n(\theta_n,\theta_0) \to 0$ does not imply $\rho(\theta_n,\theta_0) \to 0$ so that (35) and (36) are not sufficient conditions for (A4). The problem is that if (35) and (36) hold then $\rho_n(\theta_n,\theta_0) \to 0$ does not necessarily mean that all but finitely many θ_n are in a neighborhood of θ_0. The next two propositions present two possible solutions of this problem.

In the following proposition, we consider the standard two-point compactifications of intervals and the functions ρ and ρ_n introduced in (35).

Proposition 3.1. Let $[a,b] \subseteq [-\infty,\infty]$ be the two-point compactification of $\Theta = (a,b) \subseteq R$, and let there exist probability measures μ_a, μ_b such that the extended family $\{\mu_\theta : \theta \in [a,b]\}$ is continuous in the sense that

$$\lim_{\theta\to\theta_*} \rho(\theta,\theta_*) = 0 \qquad \text{for every } \theta_* \in [a,b].$$

If $\rho(\theta,\theta_*) \neq 0$ for all pairs $\theta \neq \theta_*$ from $[a,b]$, and the functions ρ_n satisfy (35), then (A4) holds.

Proof. Let θ_n be a sequence for which $\rho_n(\theta_n,\theta_0) \to 0$, let $\theta_* \in [a,b]$ be its condensation point and θ_{n_k} a subsequence tending to θ_*. By the triangle inequality and the inequality (27) for the total variation,

$$\begin{aligned}\rho_n(\theta_*,\theta_0) \leq\ & \rho_n(\theta_*,\theta_{n_k}) + \rho_n(\theta_{n_k},\theta_0) \\ \leq\ & \rho(\theta_*,\theta_{n_k}) + \rho_n(\theta_{n_k},\theta_0).\end{aligned}$$

The first term tends to zero due to the assumed continuity of the extended family of probability measures and the second term tends to zero by assumption. This together with (35) implies $\rho(\theta_*, \theta_0) = 0$, i.e. the identity $\theta_* = \theta_0$. $\square$

In concrete examples, the functions $\rho(\theta, \theta_0)$ of (35) are usually nonincreasing (decreasing) on the interval $(a, \theta_0]$ and nondecreasing (increasing) on $[\theta_0, b)$, with $\rho(\theta_0, \theta_0) = 0$. The next proposition assumes that the functions $\rho_n(\theta, \theta_0)$ of (35) are nonincreasing on $(a, \theta_0]$ and nondecreasing on $[\theta_0, b)$. Then for every $\theta_1 \in (\theta_0, b)$, $\theta_2 \in (a, \theta_0)$

$$\min\{\rho_n(\theta_1, \theta_0), \rho_n(\theta_2, \theta_0)\} \geq \rho_n(\theta_0, \theta_0) = 0$$

and

$$\min\{\rho(\theta_1, \theta_0), \rho(\theta_2, \theta_0)\} \geq \rho(\theta_0, \theta_0) = 0. \tag{37}$$

Obviously, under (35) and the assumption $\rho(\theta, \theta_0) \neq 0$ for $\theta \neq \theta_0$, this piecewise monotonicity of $\rho_n(\theta, \theta_0)$ implies (36).

Proposition 3.2. Let the above defined function $\rho(\theta, \theta_0)$ be continuous in $\theta \in (a, b) \subseteq S$, and let $\rho(\theta, \theta_0) \neq 0$ for $\theta \neq \theta_0$. If the functions $\rho_n(\theta, \theta_0)$ satisfy (35) and are nondecreasing on the subinterval $[\theta_0, b)$ and nonincreasing on $(a, \theta_0]$, then (A4) holds.

Proof. Let θ_{n_1} be the first element of the sequence θ_n which falls into the interval (θ_0, b) and, similarly, θ_{n_2} the first element from (a, θ_0). A simplication for the case that the whole sequence θ_n is in the intervals $(a, \theta_0]$ or $(\theta_0, b]$ is obvious. By (35) and (27) for the total variation,

$$\rho_n(\theta_{n_i}, \theta_0) \longrightarrow \rho(\theta_{n_i}, \theta_0) \geq \rho_{n_i}(\theta_{n_i}, \theta_0), \quad i \in \{1, 2\}.$$

Therefore (37) and the assumed monotonicity of the functions ρ_n imply the existence of n_0 such that for $n > n_0$

$$\rho_n(\theta, \theta_0) \geq \frac{1}{2}\left(\rho_{n_1}(\theta_{n_1}, \theta_0) + \rho(\theta_{n_1}, \theta_0)\right) \triangleq \rho_* > 0$$

if $\theta \in [\theta_{n_1}, b)$ and

$$\rho_n(\theta, \theta_0) \geq \frac{1}{2}\left(\rho_{n_2}(\theta_{n_2}, \theta_0) + \rho(\theta_{n_2}, \theta_0)\right) \triangleq \rho^* > 0$$

if $\theta \in (a, \theta_{n_2}]$. Since at the same time there exists $\widetilde{n}_0$ such that for $n > \widetilde{n}_0$

$$\rho_n(\theta_n, \theta_0) < \min\{\rho_*, \rho^*\},$$

it follows that $\theta_n \in [\theta_{n_1}, \theta_{n_2}]$ for $n > \max\{n_0,\ \widetilde{n}_0\}$. The convergence $\theta_n \to \theta_0$ now follows from the assumption $\rho_n(\theta_n, \theta_0) \to 0$ by using the compactness of $[\theta_{n_1}, \theta_{n_2}]$ and the triangle inequality, similarly as demonstrated above. □

Remark 3.1. Propositions 3.1 and 3.2 used only the metric properties of functions ρ_n and ρ (reflexivity, symmetry and triangle inequality), and also the inequalities $\rho_n \leq \rho$ and convergence $\rho_n \to \rho$ as $n \to \infty$, guaranteed for the total variation by (27) and (29). All these properties are shared by the divergence measure $H(\boldsymbol{p}, \boldsymbol{q}) = \sqrt{I_{1/2}(\boldsymbol{p}, \boldsymbol{q})}/2$ or $H(\mu_\theta, \mu_{\theta_0}) = \sqrt{I_{1/2}(\mu_\theta, \mu_{\theta_0})}/2$ (see (10) and (27), called *Hellinger distance.* Thus all what was so far said in this section remains valid with the total variation replaced by the Hellinger (or any other metric) distance. This observation is useful because, e.g., the monotonicity (unimodality) of functions $\rho_n(\theta, \theta_0)$ assumed in Proposition 3.1 is usually more easily verified for distances different from the total variation. Note that the Hellinger distance H is not one-one related with the total variation V : The bounds

$$1 - \sqrt{1 - (V/2)^2} \leq H^2 \leq V/2$$

obtained in Remark 2.38 of Liese and Vajda (1987) are attained in the whole range $0 \leq V \leq 2$. □

In the rest of this section, all asymptotic relations are considered for $n \to \infty$. Next follows our first main result.

Theorem 3.1. If the assumptions (A3) and (A4) hold then all $\phi-$disparity estimators $\widehat{\theta}_n = \widehat{\theta}_n^\phi$ are consistent in the sense that $\widehat{\theta}_n - \theta_0 = o_p(1)$ as $n \to \infty$.

Proof. By Proposition 2.4 and (A3), $D_\phi(\widehat{\boldsymbol{p}}_n; \boldsymbol{p}_n(\theta_0)) = o_p(1)$. By the definition of $\widehat{\theta}_n$ in (23),

$$D_\phi(\widehat{\boldsymbol{p}}_n; \boldsymbol{p}_n(\widehat{\theta}_n)) \leq D_\phi(\widehat{\boldsymbol{p}}_n; \boldsymbol{p}_n(\theta_0))$$

so that also $D_\phi(\widehat{\boldsymbol{p}}_n;\boldsymbol{p}_n(\theta_n)) = o_p(1)$. By Proposition 2.1, this implies that

$$V(\widehat{\boldsymbol{p}}_n;\boldsymbol{p}_n(\theta_0)) + V(\widehat{\boldsymbol{p}}_n;\boldsymbol{p}_n(\widehat{\theta}_n)) = o_p(1).$$

Hence, by the triangle inequality and symmetry for the total variation, $V(\boldsymbol{p}_n(\widehat{\theta}_n),\boldsymbol{p}_n(\theta_0)) = o_p(1)$. Now the desired relation $\widehat{\theta}_n = \theta_0 + o_p(1)$ follows from (A4). $\square$

Our next aim is the asymptotic normality of $\phi-$disparity estimators. The following proposition studies the model $\{\boldsymbol{p}_n(\theta) : \theta \in \Theta\}$ with fixed n. Thus we drop for a while the subscript n, i.e. we consider $\{\boldsymbol{p}(\theta) = (p_1(\theta), ..., p_m(\theta)) : \theta \in \Theta\}$, where $\boldsymbol{p}(\theta)$ are discrete distributions from the simplex

$$\Delta_m = \left\{\boldsymbol{p} = (p_1, ..., p_m) : p_j \geq 0,\ \sum p_j = 1\right\} \subset \square_m$$

and

$$\square_m = \{\widetilde{\boldsymbol{p}} = (\widetilde{p}_1, ..., \widetilde{p}_m) : 0 \leq \widetilde{p}_j \leq 1\}$$

is the unit cube in R^m. The Euclidean norm $\|.\|$ defines a topology on $\square_m$ and a relative topology on Δ_m. The relative interior Δ_m^o of Δ_m consists of the vectors $\boldsymbol{p} \in \Delta_m$ with all coordinates p_j positive. It is contained in the interior $\square_m^o$ of $\square_m$ (more precisely, $\Delta_m^o = \Delta_m \cap \square_m^o$).

In the proposition that follows we consider for $\boldsymbol{p}(\theta) = \boldsymbol{p}_n(\theta)$ with a fixed n and $\widetilde{\boldsymbol{p}} \in \square_m$ the $\phi-$disparities $D_\phi(\widetilde{\boldsymbol{p}};\boldsymbol{p}(\theta))$ defined in accordance with (2.6), and their derivatives

$$\dot{D}_\phi(\widetilde{\boldsymbol{p}};\boldsymbol{p}(\theta)) = \frac{d}{d\theta} D_\phi(\widetilde{\boldsymbol{p}};\boldsymbol{p}(\theta)),\ \ddot{D}_\phi(\widetilde{\boldsymbol{p}};\boldsymbol{p}(\theta)) = \frac{d}{d\theta}\dot{D}_\phi(\widetilde{\boldsymbol{p}};\boldsymbol{p}(\theta)). \tag{38}$$

If (A1) holds and $\phi(t)$ is differentiable at all $t > 0$ then $\dot{D}_\phi(\widetilde{\boldsymbol{p}};\boldsymbol{p}(\theta))$ is well defined for all $\theta \in \Theta$ (cf.(22)) and satisfies the formula

$$\dot{D}_\phi(\widetilde{\boldsymbol{p}};\boldsymbol{p}(\theta)) = \sum_{j=1}^{m}\left[\phi\left(\frac{\widetilde{p}_j}{p_j(\theta)}\right) - \frac{\widetilde{p}_j}{p_j(\theta)}\,\phi'\left(\frac{\widetilde{p}_j}{p_j(\theta)}\right)\right]\dot{p}_j(\theta). \tag{39}$$

If, moreover, ϕ is twice differentiable at all $t > 0$ then also $\ddot{D}_\phi(\widetilde{\boldsymbol{p}};\boldsymbol{p}(\theta))$ is well defined for all $\theta \in \Theta$.

Proposition 3.3. Let $\phi(t)$ be twice continuously differentiable at all $t > 0$ and let (A1), (A2) hold. If n is large enough then the Fisher information

$$\mathcal{J}_{n,\theta_0} = \sum_{j=1}^{m} \frac{\dot{p}_j(\theta_0)^2}{p_j(\theta_0)}$$

is positive and there exists a positive radius $r = r_n$ such that the open ball

$$\mathcal{B} = \mathcal{B}_n = \left\{ \widetilde{\boldsymbol{p}} \in \square_m^0 : \|\widetilde{\boldsymbol{p}} - \boldsymbol{p}(\theta_0)\| < r \right\}$$

supports a Θ-valued function $\psi(\widetilde{\boldsymbol{p}}) = \psi_n(\widetilde{\boldsymbol{p}})$ satisfying for all $\widetilde{\boldsymbol{p}} \in \mathcal{B}$ the relations

$$\dot{D}_\phi(\widetilde{\boldsymbol{p}}; \boldsymbol{p}(\psi(\widetilde{\boldsymbol{p}}))) = 0 \tag{40}$$

and

$$\ddot{D}_\phi(\widetilde{\boldsymbol{p}}; \boldsymbol{p}(\psi(\widetilde{\boldsymbol{p}}))) > 0. \tag{41}$$

Furthermore, the function $\psi : \mathcal{B} \mapsto \Theta$ is unique and continuously partially differentiable on $\mathcal{B}$.

Proof. By (2.15) and (A2), $\mathcal{J}_{n,\theta_0}$ is positive for all sufficiently large n. Put for brevity

$$\boldsymbol{p}_0 = (p_{10}, ..., p_{m0}) \stackrel{\triangle}{=} \boldsymbol{p}(\theta_0).$$

The function defined for $\widetilde{\boldsymbol{p}} = (\widetilde{p}_1, ..., \widetilde{p}_m) \in \square_m$ and $\theta \in \Theta$ by

$$\Phi(\widetilde{\boldsymbol{p}}, \theta) = \Phi(\widetilde{p}_1, ..., \widetilde{p}_m, \theta) \stackrel{\triangle}{=} \dot{D}_\phi(\widetilde{\boldsymbol{p}}; \boldsymbol{p}(\theta))$$

is coordinatewise differentiable on the open set $\square_m^o \times \Theta$. Denote the corresponding derivatives by $\dot{\Phi}_1, ..., \dot{\Phi}_m, \dot{\Phi}_{m+1}$, i.e. let

$$\dot{\Phi}_j(\widetilde{\boldsymbol{p}}, \theta) = \frac{\partial}{\partial \widetilde{p}_j} \Phi(\widetilde{\boldsymbol{p}}, \theta), \ \dot{\Phi}_{m+1}(\widetilde{\boldsymbol{p}}, \theta) = \frac{\partial}{\partial \theta} \Phi(\widetilde{\boldsymbol{p}}, \theta) = \ddot{D}_\phi(\widetilde{\boldsymbol{p}}; \boldsymbol{p}(\theta)) \ \text{(cf. (38))}. \tag{42}$$

Since $\phi(1) = \phi'(1) = 0$ (cf. the definition of the disparity function $\phi(t)$), it follows from (39) that $\Phi(\boldsymbol{p}_0, \theta_0) = 0$. Differentiating (39) by θ and substituting $\widetilde{\boldsymbol{p}} = \boldsymbol{p}_0$, $\theta = \theta_0$, we obtain that

$$\dot{\Phi}_{m+1}(\boldsymbol{p}_0, \theta_0) = \phi''(1) \mathcal{J}_{n,\theta_0} > 0, \tag{43}$$

where the inequality follows from the assumptions $\phi''(1) > 0$ and $\mathcal{J}_{n,\theta_0} > 0$. Hence we proved that $\dot{D}_\phi(\boldsymbol{p}_0; \boldsymbol{p}(\theta_0)) = 0$, $\ddot{D}_\phi(\boldsymbol{p}_0; \boldsymbol{p}(\theta_0)) > 0$, so that the existence of the desired $r > 0$ and the existence, uniqueness and the continuous differentiability of $\psi : \mathcal{B} \to \Theta$ follow from the *implicit function theorem* (see e.g. p. 148 in Fleming (1977)). $\square$

In addition to the assumptions (A1)-(A4) of Section 2, we shall need the following important assumption.

(A5) The ϕ-disparity estimator $\widehat{\theta}_n$ is for $n \to \infty$ asymptotically linear in the sense that

$$\widehat{\theta}_n = \theta_0 + \frac{1}{\mathcal{J}_{n,\theta_0}} \sum_{j=1}^{m_n} \frac{\dot{p}_{nj}(\theta_0)}{p_{nj}(\theta_0)} \left(\widehat{p}_{nj} - p_{nj}(\theta_0)\right) + o_p\left(1/\sqrt{n}\right).$$

We see that if (A1), (A2) hold then the expressions in (A5) are well defined for all sufficiently large n so that this assumption is meaningful.

In order to find a sufficient condition for (A5), we apply Proposition 3.3 under the condition

$$\lim_{n\to\infty} P\left(\|\widehat{\boldsymbol{p}}_n - \boldsymbol{p}_n(\theta_0)\| \geq r_n\right) = 0 \tag{44}$$

imposed on the empirical distribution $\widehat{\boldsymbol{p}}_n$ and the radius $r = r_n$. By Proposition 2.4, this condition holds under (A3) if

$$\lim_{n\to\infty} \sqrt{n} r_n = \infty. \tag{45}$$

Proposition 3.4. Let (A4) and the assumptions of Proposition 3.3 hold. If the radius $r = r_n$ of Proposition 3.3 and the empirical distribution $\widehat{\boldsymbol{p}}_n$ fulfil the asymptotic condition (44) then (A5) holds for the $\phi-$disparity estimator $\widehat{\theta}_n$.

Proof. By (44), the event $\widehat{\boldsymbol{p}}_n \in \mathcal{B}_n$ takes place with a probability tending to 1 for $n \to \infty$. Hence it suffices to prove the asymptotic linearity formula under the condition $\widehat{\boldsymbol{p}}_n \in \mathcal{B}_n$. The proof is based on the observation that, by (40) and (41), $\psi(\boldsymbol{p}_n(\theta_0))$ is an argmin of $\Psi_{1n}(\theta) \triangleq D_\phi(\boldsymbol{p}_n(\theta_0), \boldsymbol{p}_n(\theta))$ on Θ, and $\psi(\widehat{\boldsymbol{p}}_n)$ is an argmin of $\Psi_{2n}(\theta) \triangleq D_\phi(\widehat{\boldsymbol{p}}_n, \boldsymbol{p}_n(\theta))$.

This means that we can assume that

$$\psi\left(\boldsymbol{p}_n\left(\theta_0\right)\right)=\theta_0 \quad \text{and} \quad \psi\left(\widehat{\boldsymbol{p}}_n\right)=\widehat{\theta}_n. \tag{46}$$

Indeed, it follows from (A4) that $\theta \neq \theta_0$ implies $\boldsymbol{p}_n(\theta) \neq \boldsymbol{p}(\theta_0)$ for all but finitely many n. This means that θ_0 is the unique argmin of $\Psi_{1n}(\theta)$ for all but finitely many n, and $\widehat{\theta}_n$ is by definition an argmin of $\Psi_{2n}(\theta)$ on Θ for all n. Therefore the equalities of (3.12) hold. Now, with (46) in mind, we expand for a fixed n the function $\psi(\widetilde{\boldsymbol{p}})$ of the variable $\widetilde{\boldsymbol{p}}=(\widetilde{p}_1,...,\widetilde{p}_m) \in \mathcal{B}_n$ around the point $\boldsymbol{q}=(q_1,...,q_m) \stackrel{\triangle}{=} \boldsymbol{p}_n(\theta_0)$. To this end are needed the continuous partial derivatives

$$\dot{\psi}_j\left(\widetilde{\boldsymbol{p}}\right)=\frac{\partial}{\partial \widetilde{p}_j}\psi\left(\widetilde{\boldsymbol{p}}\right),\ 1 \leq j \leq m_n$$

we obtain the asymptotic formula

$$\psi\left(\widetilde{\boldsymbol{p}}\right)=\psi\left(\boldsymbol{q}\right)+\sum_{j=1}^{m_n}\dot{\psi}_j\left(\boldsymbol{q}\right)\left(p_j-q_j\right)+o\left(\|\widetilde{\boldsymbol{p}}-\boldsymbol{q}\|\right) \tag{47}$$

for $\|\widetilde{\boldsymbol{p}}-\boldsymbol{q}\| \to 0$. The derivatives can be evaluated by means of $\dot{\Phi}_j(\widetilde{\boldsymbol{p}},\theta)$, $1 \leq j \leq m_n+1$, defined by (42), using the identity (40). Namely, by applying the derivatives $\partial/\partial\widetilde{p}_j$ on both sides of (40), we get the equations

$$\dot{\Phi}_j\left(\widetilde{\boldsymbol{p}},\psi\left(\widetilde{\boldsymbol{p}}\right)\right)+\Phi_{m_{n+1}}\left(\widetilde{\boldsymbol{p}},\psi\left(\widetilde{\boldsymbol{p}}\right)\right)\dot{\psi}_j\left(\widetilde{\boldsymbol{p}}\right)=0, \quad 1 \leq j \leq m_n,$$

valid for all $\widetilde{\boldsymbol{p}} \in \mathcal{B}_n$. Hence for all j under consideration in (3.13),

$$\dot{\psi}_j\left(\boldsymbol{q}\right)=-\frac{\dot{\Phi}_j\left(\boldsymbol{q},\psi\left(\boldsymbol{q}\right)\right)}{\dot{\Phi}_{m_{n+1}}\left(\boldsymbol{q},\psi\left(\boldsymbol{q}\right)\right)}=-\frac{\dot{\Phi}_j\left(\boldsymbol{q},\theta_0\right)}{\dot{\Phi}_{m_{n+1}}\left(\boldsymbol{q},\theta_0\right)} \qquad \text{(cf. (46)}$$

where, by (43),

$$\dot{\Phi}_{m_{n+1}}\left(\boldsymbol{q},\theta_0\right)=-\phi''(1)\,\mathcal{J}_{n,\theta_0}.$$

By (42),

$$\dot{\Phi}_j\left(\boldsymbol{q},\theta_0\right)=\left(\frac{\partial}{\partial p_j}\Phi\left(\widetilde{\boldsymbol{p}},\theta_0\right)\right)_{\widetilde{\boldsymbol{p}}=\boldsymbol{q}}$$

where $\Phi\left(\widetilde{\boldsymbol{p}},\theta\right)$ are given for all $(\widetilde{\boldsymbol{p}},\theta)\in\mathcal{B}_n\times\Theta$ by (39). Evaluating the derivatives, we find that

$$\dot{\Phi}_j\left(\boldsymbol{q},\theta_0\right)=-\phi''\left(1\right)\dot{p}_{nj}\left(\theta_0\right)q_j.$$

By (17), the assumption (A2) implies $\mathcal{J}_{n,\theta_0}>0$ for all sufficiently large n. Therefore, for all these n,

$$\dot{\psi}_j\left(\boldsymbol{q}\right)=\frac{1}{\mathcal{J}_{n,\theta_0}}\frac{\dot{p}_{nj}\left(\theta_0\right)}{q_j}=\frac{1}{\mathcal{J}_{n,\theta_0}}\frac{\dot{p}_{nj}\left(\theta_0\right)}{p_{nj}\left(\theta_0\right)}.$$

Inserting these derivatives in (47), and substituting there $\boldsymbol{q}$ by $\boldsymbol{p}_n\left(\theta_0\right)$ and $\widetilde{\boldsymbol{p}}$ by $\widehat{\boldsymbol{p}}_n$, we obtain with the help of (46) the asymptotic linear formula of (A5). By (33), (A3) guarantees that this substitution satisfies for $n\to\infty$ the convergence $\|\widehat{\boldsymbol{p}}_n-\boldsymbol{p}_n\left(\theta_0\right)\|\to 0$ assumed in (47). In fact, it guarantees more, namely $\|\widehat{\boldsymbol{p}}_n-\boldsymbol{p}_n\left(\theta_0\right)\|=O_p\left(1/\sqrt{n}\right)$, which gives (A5) its final form. □

The method used above to establish the asymptotic linearity formula, based on the implicit function theorem, is more easily applicable in situations where the discrete model $\{\boldsymbol{p}\left(\theta\right):\theta\in\Theta\}$ does not depend on the sample size n. Indeed, in this case the condition (44) needs no verification, as $d_n=d>0$ is constant and $\|\widehat{\boldsymbol{p}}_n-\boldsymbol{q}\left(\theta_0\right)\|=o_p\left(1\right)$. Cox (1984) was probably the first who used this method in such a situation. He studied the MLE, i.e. the minimum ϕ_0-divergence estimator where ϕ_0 is given by (5). In Chapter 5 of Pardo (1997), this method was applied to all minimum $\phi-$divergence estimators in discrete models $\{\boldsymbol{p}\left(\theta\right):\theta\in\Theta\}$.

The asymptotic linearity formula of (A5) enables to prove the asymptotic normality of all $\phi-$disparity estimators. To this end we need to add the following condition on partitions to those considered in (A3). This condition is based on the assumption (A1).

(A6) The partitions $\mathcal{P}_n$ satisfy the relation

$$\lim_{n\to\infty}\sum_{j=1}^{m_n}\left(\frac{\left|\dot{p}_{nj}\left(\theta_0\right)\right|}{\sqrt{p_{nj}\left(\theta_0\right)}}\right)^3=0.$$

We present a sufficient condition for (A6).

Proposition 3.5. If (A1) is satisfied and

$$\lim_{n\to\infty} \max_{1\le j\le m_n} \frac{\dot{p}_{nj}(\theta_0)^2}{p_{nj}(\theta_0)} = 0 \tag{48}$$

then (A6) holds.

Proof. Under (A1),

$$\left(\frac{|\dot{p}_{nj}(\theta_0)|}{\sqrt{p_{nj}(\theta_0)}}\right)^3 \le \frac{\dot{p}_{nj}(\theta_0)^2}{p_{nj}(\theta_0)} \max_{1\le j\le m_n} \frac{|\dot{p}_{nj}(\theta_0)|}{\sqrt{p_{nj}(\theta_0)}},$$

so that, by (26),

$$\sum_{j=1}^{m_n}\left(\frac{|\dot{p}_{nj}(\theta_0)|}{\sqrt{p_{nj}(\theta_0)}}\right)^3 \le \mathcal{J}_{\theta_0} \max_{1\le j\le m_n} \frac{|\dot{p}_{nj}(\theta_0)|}{\sqrt{p_{nj}(\theta_0)}},$$

where $\mathcal{J}_{\theta_0}$ was assumed in (17) to be finite. $\square$

Theorem 3.2. Let the assumptions (A1) - (A3) and (A5), (A6) be satisfied. Then for all $\phi-$disparity functions with the second derivative $\phi''(t)$ Lipschitz at $t=1$, the $\phi-$disparity estimators $\widehat{\theta}_n = \widehat{\theta}_n^{\phi}$ are asymptotically normal in the sense that

$$\sqrt{n}\left(\widehat{\theta}_n - \theta_0\right) \stackrel{\mathcal{L}}{\to} N\left(0, \mathcal{J}_{\theta_0}^{-1}\right) \quad \text{as } n\to\infty,$$

where $\mathcal{J}_{\theta_0}$ is the Fisher information (17).

Proof. Put for brevity $\boldsymbol{p}_n = (p_{n1},...,p_{nm_n}) \triangleq \boldsymbol{p}_n(\theta_0)$, $\dot{\boldsymbol{p}}_n = (\dot{p}_{n1},...,\dot{p}_{nm_n}) \triangleq \dot{\boldsymbol{p}}_n(\theta_0)$ and $\mathcal{J} = \mathcal{J}_{\theta_0}$. By (A2) and (A5), it suffices to prove that

$$\sqrt{n}\sum_{j=1}^{m_n} \frac{\dot{p}_{nj}}{p_{nj}}\left(\widehat{p}_{nj} - p_{nj}\right) \stackrel{\mathcal{L}}{\to} N(0,\mathcal{J}) \qquad \text{as } n\to\infty. \tag{49}$$

By the Proposition on pp. 311-313 of Beirlant et al. (1994), (49) holds if the independent random variables

$$w_{nj} = \frac{Y_{nj}}{n} \triangleq \frac{\text{Poisson}(np_{nj})}{n}, \quad 1\le j\le m_n,$$

lead to the sums

$$Z_n \triangleq \sqrt{n}\sum_{j=1}^{m_n} \frac{\dot{p}_{nj}}{p_{nj}} (w_{nj} - p_{nj}) \quad \text{and} \quad \frac{N_n - n}{\sqrt{n}} \triangleq \sqrt{n}\sum_{j=1}^{m_n} (w_{nj} - p_{nj})$$

satisfying for every t, $u \in R$ the limit law

$$tZ_n + u\frac{N_n - n}{\sqrt{n}} \xrightarrow{\mathcal{L}} N\left(0, t^2\mathcal{J}+u^2\right) \quad \text{as } n \to \infty,$$

i.e. if

$$\sqrt{n}\sum_{j=1}^{m_n} \left(\frac{\dot{p}_{nj}}{p_{nj}}t + u\right)(w_{nj} - p_{nj}) \xrightarrow{\mathcal{L}} N\left(0, t^2\mathcal{J}+u^2\right) \quad \text{as } n \to \infty. \tag{50}$$

By the Liapunov theorem (p. 127 in Rao (1973)), (50) holds if for $n \to \infty$

$$\sigma_n^2 \triangleq var\left[\sqrt{n}\sum_{j=1}^{m_n} \left(\frac{\dot{p}_{nj}}{p_{nj}}t + u\right)(w_{nj} - p_{nj})\right] \to t^2\mathcal{J}+u^2 \tag{51}$$

and

$$\mu_n^3 \triangleq \sum_{j=1}^{m_n} E\left|\sqrt{n}\left(\frac{\dot{p}_{nj}}{p_{nj}}t + u\right)(w_{nj} - p_{nj})\right|^3 \to 0. \tag{52}$$

As is easy to verify, $\sigma_n^2 = t^2\mathcal{J}_{n,\theta_0}+u^2$, so that (51) follows from (A2). Further,

$$\mu_n^3 = n^{-3/2}\sum_{j=1}^{m_n} \frac{\left|\dot{p}_{nj}t + p_{nj}u\right|^3}{p_{nj}^3} E\left|Y_{nj} - np_{nj}\right|^3$$

where

$$\begin{aligned} E\left|Y_{nj} - np_j\right|^3 &\leq \left[E\left(Y_{nj} - np_{nj}\right)^2 E\left(Y_{nj} - np_{nj}\right)^4\right]^{1/2} \\ &= \left[np_{nj}\left(3n^2p_{nj}^2 + np_{nj}\right)\right]^{1/2} \\ &\leq \sqrt{3}\left(np_{nj}\right)^{3/2}. \end{aligned}$$

Consequently,

$$\begin{aligned} \mu_n^3 &\leq \sqrt{3}\sum_{j=1}^{m_n}\left|\tfrac{\dot{p}_{nj}t+p_{nj}u}{\sqrt{p_{nj}}}\right|^3 \\ &\leq k\sum_{j=1}^{m_n}\left[\left(\tfrac{|\dot{p}_{nj}|}{\sqrt{p_{nj}}}\right)^3 + \left(\sqrt{p_{nj}}\right)^3\right], \end{aligned}$$

where we used the inequality $|x+y|^3 \le 4\left(|x|^3+|y|^3\right)$ obtained from Jensen's inequality for the convex $\psi(t)=|t|^3$, and $k=4\sqrt{3}\max\left\{|t|^3,|u|^3\right\}$. The desired relation (52) thus follows from (A6) and from the inequality

$$\sum_{j=1}^{m_n}\left(\sqrt{p_{nj}}\right)^3 \le \left(\max_{1\le j\le m_n} p_{nj}\right)^{1/2}\sum_{j=1}^{m_n} p_{nj},$$

by taking into account that the condition imposed on $\max_{1\le j\le m_n} p_{nj}$ in (A3) and the assumption (9) imply that $\max_{1\le j\le m_n} p_{nj}\to 0$ for $n\to\infty$. □

4. EXAMPLES

Assumptions (A1)-(A6) can be satisfied by appropriate sequences of partitions in common statistical models where an a priori knowledge enables to localize the true parameter θ_0 to a compact set $\Theta_c \subset \Theta$. It usually suffices to the consider interval partitions of R by the cutpoints $x_{n,j}$; $1\le j\le m_n$ which are quantiles of the distribution $F_{\theta_c}(x)=\mu_{\theta_c}((-\infty,x))$ of orders m_n for a fixed $\theta_c\in\Theta_c$. For example, let us look at the quite restrictive condition (A3). If the parameter θ_0 coincides with θ_c then $\boldsymbol{p}_n(\theta_0)$ defined by (20) is uniform so that $m_n p_{nj}(\theta_0)=1$ for all $1\le j\le m_n$ and (A3) holds. If θ_0 is not too far away from θ_c then in most models (A3) still holds. As an illustration consider $\theta_c=0$ in the logistic location model where

$$F_\theta(x)=\frac{1}{1+e^{-x+\theta}} \qquad \text{and} \qquad x_{nj}=\ln\frac{j}{m_n-j}.$$

Then

$$x_{n,1}=\ln\frac{1}{m_n-1} \qquad \text{and} \qquad p_{n,1}(\theta_0)=\frac{1}{1+(m_n-1)\,e^{\theta_0}}$$

so that for $m_n\to\infty$

$$\lim_{n\to\infty} m_n p_{n,1}(\theta_0)=e^{-\theta_0} \qquad \text{for every } \theta_0\in R.$$

Similarly

$$x_{n,m_n-1}=\ln(m_n-1) \qquad \text{and} \qquad p_{n,m_n}(\theta_0)=\frac{e^{\theta_0}}{m_n-1+e^{\theta_0}}$$

so that for $m_n\to\infty$

$$\lim_{n\to\infty} m_n p_{n,m_n}(\theta_0)=e^{\theta_0} \text{ for every } \theta_0\in R.$$

More generally, for every $\theta_0 \in R$ it holds

$$e^{-|\theta_0|} \leq \lim_{n\to\infty} m_n \min_{1\leq j\leq m_n} p_{nj}(\theta_0) \leq \lim_{n\to\infty} m_n \max_{1\leq j\leq m_n} p_{nj}(\theta_0) \leq e^{|\theta_0|}$$

when m_n satisfies (32) so that (A3) holds for θ_0 from any bounded subset $\Theta \subset R$.

Verification of the assumptions (A1), (A2) and (A6) is usually easy but verification of (A4) and (A5) involves extensive technicalities. The next example is selected with the aim to reduce these technicalities to the level allowing to demonstrate verification of all assumptions (A1)-(A6) on a reasonably limited space. It deals with a parameter estimation of spectral density based on independent observations of frequencies distributed by this density.

Let $\{\mu_\theta : \theta \in \Theta\}$ with $\Theta = (-1, 1)$ be the statistical model with distribution functions and densities

$$F_\theta(x) = \frac{1}{2\pi}\boldsymbol{I}_{[0,2\pi)}(x)\,(x + \theta \sin x) \ \text{ and } \ f_\theta(x) = \frac{1}{2\pi}\boldsymbol{I}_{[0,2\pi)}(x)\,(1 + \theta\cos x) \tag{53}$$

where $\boldsymbol{I}$ is the indicator function. Further, let us consider for m_n assumed in (A3) the uniform partitions $\mathcal{P}_n$ of the observation space $[0, 2\pi)$ into the intervals $A_{nj} = [x_{n,j-1}, x_{nj})$, $1 \leq j \leq m_n$ defined by the cutpoints

$$0 = x_{n,0} < x_{nj} = \frac{2\pi j}{m_n} < x_{nm_n} = 2\pi \tag{54}$$

of the observation space $[0, 2\pi)$. The resulting discrete models $\boldsymbol{p}_n(\theta) = (p_{nj}(\theta) : 1 \leq j \leq m_n)$ are given by the formula

$$p_{nj}(\theta) = \frac{1}{m_n} + \theta c_{nj} \ \text{ for } \ c_{nj} = \frac{\sin x_{nj} - \sin x_{n,j-1}}{2\pi} = \int_{A_{nj}} \frac{d}{d\theta} f_\theta(x)\, dx. \tag{55}$$

Finally, let $\widehat{\theta}_n$ be the MLE which minimizes the I_1-divergence

$$I_1(\widehat{\boldsymbol{p}}_n; \boldsymbol{p}_n(\theta)) = \sum_{j=1}^{m_n} \widehat{p}_{nj} \ln \widehat{p}_{nj} - \sum_{j=1}^{m_n} \widehat{p}_{nj} \ln p_{nj}(\theta), \ \ \theta \in (-1, 1) \tag{56}$$

(cf.(10)) when the vector of observed relative frequencies is $\widehat{\boldsymbol{p}}_n = (\widehat{p}_{nj} : 1 \leq j \leq m_n)$.

By the mean value theorem and the monotonicity of $\cos x$, we get from (55)

$$\cos x_{nj} < m_n c_{nj} < \cos x_{n,j-1} \qquad \text{for } x_{nj}, x_{n,j-1} \in [0, \pi] \tag{57}$$

and

$$\cos x_{n,j-1} < m_n c_{nj} < \cos x_{nj} \qquad \text{for } x_{nj}, x_{n,j-1} \in [\pi, 2\pi]\,. \tag{58}$$

By differentiating with respect to θ, we obtain from (55) and (56)

$$\dot{p}_{nj}(\theta) = c_{nj},\;\; \ddot{p}_{nj}(\theta) = 0 \tag{59}$$

and

$$\dot{I}_1(\widehat{\boldsymbol{p}}_n; \boldsymbol{p}_n(\theta)) = -\sum_{j=1}^{m_n} \widehat{p}_{nj} \frac{m_n c_{nj}}{1+\theta m_n c_{nj}},\;\; \ddot{I}_1(\widehat{\boldsymbol{p}}_n; \boldsymbol{p}_n(\theta)) = \sum_{j=1}^{m_n} \widehat{p}_{nj} \left(\frac{m_n c_{nj}}{1+\theta m_n c_{nj}}\right)^2. \tag{60}$$

Finally, from (53) and (59) we obtain the Fisher informations

$$\mathcal{J}_\theta = \frac{1}{2\pi}\int_0^{2\pi} \frac{\cos^2 x}{1+\theta\cos x}dx = \begin{cases} \frac{1-\sqrt{1-\theta^2}}{\theta^2\sqrt{1-\theta^2}} & \text{if } \theta \neq 0 \\ \frac{1}{2} & \text{if } \theta = 0 \end{cases} \tag{61}$$

where $\mathcal{J}_\theta \geq \mathcal{J}_0 = 1/2$ for all $\theta \in (-1,1)$, and

$$\mathcal{J}_{n,\theta} = m_n \sum_{j=1}^{m_n} \frac{c_{nj}^2}{1+\theta m_n c_{nj}} = \frac{1}{m_n}\sum_{j=1}^{m_n} \frac{(m_n c_{nj})^2}{1+\theta m_n c_{nj}}. \tag{62}$$

From (59) and (55) we see that (A1) holds. By (61), $\mathcal{J}_\theta$ is an integral of a function continuous on the integration domain $[0, 2\pi]$. By (62) and (57), (58), $\mathcal{J}_{n,\theta}$ are Riemann sums for this integral so that (A2) holds for every $\theta_0 \in (-1,1)$. By (55) and (57), (58),

$$m_n p_{nj}(\theta) = 1 + \theta c_{nj} \in (1 - |\theta|\,, 1 + |\theta|) \tag{63}$$

for every $\theta \in (-1,1)$, every $1 \leq j \leq m_n$ and every n. This implies the validity of the assumption (A3) for all $\theta_0 \in (-1,1)$. Let us now turn attention to (A4). By (9) and (55), for every $\theta_n, \theta_0 \in (-1,1)$

$$V_n \triangleq V(\boldsymbol{p}_n(\theta_n); \boldsymbol{p}_n(\theta_0)) = \sum_{j=1}^{m_n} |(\theta_n - \theta_0)\, c_{nj}|\,.$$

Suppose for simplicity that $m_n = 4k_n$ for an integer k_n. Then $c_{nj} > 0$ for $1 \leq j \leq k_n$ so that

$$V_n = 4\,|\theta_n - \theta_0| \sum_{j=1}^{k_n} c_{nj} \geq 4\,|\theta_n - \theta_0|\,\frac{1}{m_n}\sum_{j=1}^{k_n} \cos x_{nj}$$

where the inequality follows from (57). By (54), $x_{n,k_n} = \pi/2$, so that on the right hand side are Riemann sums of the integral $\int\limits_0^{\pi/2} \cos x dx = 1$. Therefore,

$$V_n = 4\left|\theta_n - \theta_0\right|(1 + o(1)) \text{ as } n \to \infty$$

which implies (A4) for all $\theta_0 \in (-1,1)$. Further, by (55) and (59),

$$\sum_{j=1}^{m_n}\left(\frac{|\dot{p}_{nj}(\theta)|}{\sqrt{p_{nj}(\theta)}}\right)^3 = \sum_{j=1}^{m_n}\left(\frac{\sqrt{m_n}\,|c_{nj}|}{\sqrt{1+\theta m_n c_{nj}}}\right)^3 = \frac{R_n(\theta)}{\sqrt{m_n}}$$

where, by (57) and (58),

$$R_n(\theta) = \frac{1}{m_n}\sum_{j=1}^{m_n}\left(\frac{|m_n c_{nj}|}{\sqrt{1+\theta m_n c_{nj}}}\right)^3$$

are Riemann sums of the Riemann integral

$$\frac{1}{2\pi}\int_0^{2\pi}\left(\frac{|\cos x|}{\sqrt{1+\theta\cos x}}\right)^3 dx \le \frac{1}{2\pi}\frac{1}{(1-|\theta|)^{3/2}}\int_0^{2\pi}|\cos x|^3\,dx < \infty.$$

Therefore (A6) holds for all $\theta_0 \in (-1,1)$.

It remains to prove (A5). Let θ, $\theta_0 \in (-1,1)$ be arbitrary and θ_0 fixed. By (33) and (A3), the empirical distributions $\widehat{\boldsymbol{p}}_n$ satisfy the relation

$$\lim_{n\to\infty} P\left(\|\widehat{\boldsymbol{p}}_n - \boldsymbol{p}_n(\theta_0)\| \ge r_n\right) = 0 \qquad \text{for } r_n = \frac{1-|\theta_0|}{2m_n}. \tag{64}$$

In other words, the sequence r_n and the empirical distribution $\widehat{\boldsymbol{p}}_n$ fulfil (44) as assumes Proposition 3.4. Since (A4) was proved above and the function $\phi_1(t) = t\ln t - t - 1$ defining the I_1- divergence is twice continuously differentiable at any $t > 0$, the remaining assumptions of Proposition 3.3 hold as well. Thus, by Propositions 3.3 and 3.4, it suffices to prove for each sufficiently large n the existence of a function $\psi = \psi_n$ for which (40) and (41) hold on the ball $\mathcal{B}_n$ of radius r_n defined by (64). If $\widetilde{\boldsymbol{p}}_n = (\widetilde{p}_{nj} : 1 \le j \le m_n) \in \mathcal{B}_n$, i.e. if $\|\widetilde{\boldsymbol{p}}_n - \boldsymbol{p}_n(\theta_0)\| < r_n$, then (63) implies for $\alpha = (1-|\theta_0|)/2$, $\beta = 3/2 + |\theta_0|/2$ and all $1 \le j \le m_n$ that

$$\frac{\alpha}{m_n} \le \widetilde{p}_{nj} \le \frac{\beta}{m_n}. \tag{65}$$

Hence, by putting $\widehat{\boldsymbol{p}}_n = \widetilde{\boldsymbol{p}}_n$ in the definition of $I_1(\widehat{\boldsymbol{p}}_n; \boldsymbol{p}_n(\theta))$ in (60), and by (57), (58), we obtain that each $\widetilde{\boldsymbol{p}}_n \in \mathcal{B}_n$ fulfils the relations

$$\begin{aligned} \ddot{I}_1(\widetilde{\boldsymbol{p}}_n; \boldsymbol{p}_n(\theta)) \geq & \ \tfrac{\alpha}{(1+|\theta_0|)^2} \tfrac{1}{m_n} \sum_{j=1}^{m_n} (m_n c_n)^2 \\ = & \ \tfrac{\alpha}{(1+|\theta_0|)^2} \left(\tfrac{1}{2\pi} \int_0^{2\pi} \cos^2 x dx + o(1) \right) \text{ as } n \to \infty \\ = & \ \tfrac{\alpha}{2(1+|\theta_0|)^2} + o(1) \text{ as } n \to \infty. \end{aligned} \tag{66}$$

This means that if n is large enough then $\dot{I}_1(\widetilde{\boldsymbol{p}}_n; \boldsymbol{p}_n(\theta))$ is strictly increasing in the variable $\theta \in (-1, 1)$. Next we prove that for each $\widetilde{\boldsymbol{p}}_n \in \mathcal{B}_n$ it holds

$$\lim_{\theta \downarrow -1} \lim_{n \to \infty} \dot{I}_1(\widetilde{\boldsymbol{p}}_n; \boldsymbol{p}_n(\theta)) = -\infty, \tag{67}$$

$$\lim_{\theta \uparrow 1} \lim_{n \to \infty} \dot{I}_1(\widetilde{\boldsymbol{p}}_n; \boldsymbol{p}_n(\theta)) = \infty. \tag{68}$$

Indeed, by (60) and (65), $\widetilde{\boldsymbol{p}}_n \in \mathcal{B}_n$ implies

$$-\frac{\beta}{m_n} \sum_{j=1}^{m_n} \frac{m_n c_{nj}}{1 + \theta m_n c_{nj}} \leq \dot{I}_1(\widetilde{\boldsymbol{p}}_n; \boldsymbol{p}_n(\theta)) \leq -\frac{\alpha}{m_n} \sum_{j=1}^{m_n} \frac{m_n c_{nj}}{1 + \theta m_n c_{nj}}$$

where, by (57) and (58),

$$-\frac{1}{m_n} \sum_{j=1}^{m_n} \frac{m_n c_{nj}}{1 + \theta m_n c_{nj}} = -\frac{1}{2\pi} \int_0^{2\pi} \frac{\cos x}{1 + \theta \cos x} dx + o(1) \text{ as } n \to \infty.$$

It is easy to verify that the last integral as a function of $\theta \in (-1, 1)$ tends to ∞ or $-\infty$ if $\theta \downarrow -1$ or $\theta \uparrow 1$. Thus (67) and (68) follow from the last formula. The strict monotonicity of $\dot{D}_{\phi_1}(\widetilde{\boldsymbol{p}}_n; \boldsymbol{p}_n(\theta)) = \dot{I}_1(\widetilde{\boldsymbol{p}}_n; \boldsymbol{p}_n(\theta))$ together with (66), (67) implies that, for all sufficiently large n and all $\widetilde{\boldsymbol{p}}_n \in \mathcal{B}_n$, there exist unique solutions $\psi_n(\widetilde{\boldsymbol{p}}_n)$ of the likelihood equation $\dot{I}_1(\widetilde{\boldsymbol{p}}_n; \boldsymbol{p}_n(\theta)) = 0$ in the variable θ, i.e. that the functions ψ_n satisfy (40) for all $\widetilde{\boldsymbol{p}} = \widetilde{\boldsymbol{p}}_n \in \mathcal{B}_n$ and $\boldsymbol{p}(\theta) = \boldsymbol{p}_n(\theta)$. By (66), these functions satisfy in the same sense also (41). This completes the proof of (A5).

Now, using Theorems 3.1 and 3.2, we can close the analysis of the continuous model (53) by the statement that the MLE $\widehat{\theta}_n$ of the true parameter $\theta_0 \in (-1, -1)$, based on data finitely quantized by the uniformly distributed values $x_{n,1},...,x_{n,m_n-1}$ given by (54) is consistent, and also efficient and asymptotically normal in this model in the sense

$$\sqrt{n}\left(\widehat{\theta}_n-\theta_0\right)\xrightarrow{L}N\left(0,1/\mathcal{J}_{\theta_0}\right)\text{ as }n\to\infty,$$

where $\mathcal{J}_{\theta_0}$ is given by (61).

ACKNOWLEDGEMENTS

This work has been supported by the Spanish grants BMF2003-00892, BMF2003-04820, GV04B-670 and the Czech grants ASCR A1075403 and MSMTV 1M 0572.

BIBLIOGRAPHY

Beirlant, J., Györfi, L. and Lugosi, G. (1994). On the asymptoic normality of the L_1- and L_2- errors in histogram density estimation. *The Canadian Journal of Statistics* **22,** 309-318.

Bickel, P. J. (1982). On adaptive estimation. *The Annals of Statistics,* **10**, 647-671.

Cox, C. (1984). An elementary introduction to maximum likelihood estimation for multinomial models: Birch's theorem and delta method. *The American Statistician,* **38**, 283-287.

Cressie, N. A. C. and Read, R.C. (1984): Multinomial goodness of fit tests. *J. Royal Statist. Soc., Ser. B,* **46**, 440-464.

Csiszár, I. (1963). Eine Informationstheoretische Ungleichung und ihre Anwendung auf den Beweis der Ergodizität von Markoffschen Ketten. *Publications of the Mathematical Institute of Hungarian Academy of Sciences* **8**, Series A, 85–108.

Csiszár, I. (1967). Information type measures of difference of probability distributions and indirect observations. *Studia Scientiam Mathematica Hungarica*, **2**, 299–318.

Drost, D. C. (1989). Generalized chi-square goodness-of-fit tests for location-scale models when the number of classes tends to infinity. *The Annals of Statistics,* **17,** 1285-1300.

Dzhaparidze, K. O. and Nikulin, M. S. (1974). On a modification of the standard statistics Pearson. *Theory of Probability and Applications,* **19,** 851-851.

Fleming, W. (1977). *Functions of Several Variables,* Second Edition, Springer-Verlag. New York.

Györfi, L. and Vajda, I. (2002). Asymptotic distributions for goodness-of-fit statistics in a sequence of multinomial models. *Statistics and Probability Letters,* **56**, 57-67.

Ghurye, S. G. and Johnson, B. R. (1981). Discrete approximations to the information integral. *The Canadian Journal of Statistics*, **9**, 27–37.

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

Lindsay, B. G. (1994). Efficiency versus robustness: The case for minimum Hellinger distance and other methods. *Annals of Statistics,* **22**, 1081-1114.

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

Menéndez, M. L., Morales, D., Pardo, L. and Vajda, I. (2001). Minimum divergence estimators based on grouped data. *Annals of the Institute of Statistical Mathematics*, **53**, 277–288.

Moore, D. S. and Spruil, M. C. (1975). Unified large-sample theory of general chi-squared statistics for tests of fit. *Annals of Statistics,* **3**, 599-616.

Pardo, L. (1997). *Statistical Information Theory (In Spanish)*, Hesperides, Spain.

Rao, C. R. (1973). *Linear Statistical inference and its applications,* Second Edition, John Wiley, New York.

Rao, K. C. and Robson, D. S. (1974). A chi-square statistic for goodness-of-fit tests within the exponential family. *Communications in Statistics - Theory and Methods*, **3,** 1139-1153.

Read, R.C. and Cressie, N.A.C. (1988): *Goodness-of-fit Statistics for Discrete Multivariate data*, Springer Verlag, New York.

Ryu, K. (1993). Monotonicity of the Fisher Information and Kullback–Leibler divergence measure. *Economics Letters*, **42**, 121–128.

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

Tsairidis, Ch., Zografos, K. and Ferentinos, K. (1997). Fisher's information matrix and divergence for finite and optimal partitions of the sample space. *Communications in Statistics–Theory and Methods*, **26**, 2271–2289.

Vajda, I. (1972): On the f-divergence and singularity of probability measures. *Periodica Mathematica Hungarica,* **6**, 317-324.

Vajda, I. (1973). χ^{α}–divergence and generalized Fisher information. *Transactions of the Sixth Prague conference on Information Theory, Statistical Decision Functions and Random Processes*, 873-886, Academia, Prague.

Vajda, I. (1984 a): Minimum divergence principle in statistical estimation. *Statistics and Decisions,* Supplement. Issue No. 1, 239-261.

Vajda, I. (1984 b): Motivation, existence and equivariance of D-estimators. *Kybernetika* **20**, 189-208.

Vajda, I. (1984 c): Consistency of D-estimators. *Kybernetika* **20**, 283-303.

Vajda, I. (1984 d): Asymptotic efficiency and robustness of D-estimators. *Kybernetika* **20**, 358-375.

Vajda, I. (1989): *Theory of Statistical Inference and Information.* Boston, Kluwer.

Vajda, I. (2002): On convergence of information contained in quantized observations. *Transactions of IEEE on Information Theory*, **48**, 2163-2172.

Vos, P. V. (1992): Minimum f-divergence estimators and quasi-likelihood functions. *Annals of the Institute of Statistical Mathematics*, **44**, 261–279.

Watson, G. S. (1959). Some recent results in chi-square goodness-of-fit tests. *Biometrics* **15**, 440-468.

Zografos K., Ferentinos K. and Papaioannou T. (1986). Discrete approximations to the Csiszár, Rényi and Fisher measures of information. *Canadian Journal of Statistics* **14**, 355-366.