Metrika (2003) 57: 1–27

Metrika

On efficiency of estimation and testing with data quantized to fixed number of cells[1]

A. M. Mayoral[2], D. Morales[2], J. Morales[2] and I. Vajda[3]

[2] Miguel Hernández University of Elche, Avenida del Ferrocarril, 03202 Elche (Alicante), Spain (e-mail: d.morales@umh.es)
[3] Institute of Information Theory and automation, Pod vodárenskou věží 4, 182 08 Prague, Czech Republic (e-mail: vajda@utia.cas.cz)

Abstract. In continuous parametrized models with i.i.d. observations we consider finite quantizations. We study asymptotic properties of the estimators minimizing disparity between the observed and expected frequencies in the quantization cells, and asymptotic properties of the goodness of fit tests rejecting the hypotheses when the disparity is large. The disparity is measured by an appropriately generalized ϕ-divergence of probability distributions so that, by the choice of function ϕ, one can control the properties of estimators and tests. For bounded functions ϕ these procedures are robust. We show that the inefficiency of the estimators and tests can be measured by the decrease of the Fisher information due to the quantization. We investigate theoretically and numerically the convergence of the Fisher informations. The results indicate that, in the common families, the quantizations into 10–20 cells guarantees "practical efficiency" of the quantization-based procedures. These procedures can at the same time be robust and numerically considerably simpler than similar procedures using the unreduced data.

Key words: Quantization, Fisher information in quantized models, convergence of Fisher information, optimal quantization, efficient estimation, efficient testing

1 Introduction and basic concepts

The common model of statistical inference on a univariate parameter θ considers empirical data $X_1, \ldots, X_n$, where X_i are mutually independent observations distributed by $F(x, \theta)$ from a family

[1] Supported by the grants A1075101 and GV99-159-1-01.

$$(F(x,\theta):\theta\in\Theta),\quad \Theta\subset R \text{ open},\tag{1}$$

of mutually different absolutely continuous distribution functions on the real line R. The data define, for sample sizes $n=1,2,\ldots,$ the empirical distribution functions

$$F_n(x)=\frac{1}{n}\sum_{i=1}^{n} I_{[X_i,\infty)}(x)$$

on R. Almost all statistical methods are based, directly or indirectly, on distances $D(F_n(\cdot);F(\cdot,\theta))$ between the empirical and theoretical distributions (they need not necessarily be metrics). In particular, the point estimators θ_n (sequences of measurable mappings $R^n\to\Theta$) are assumed to satisfy the condition

$$D(F_n(\cdot);F(\cdot,\theta_n))=\inf_{\theta\in\Theta} D(F_n(\cdot);F(\cdot,\theta))\tag{2}$$

(eventually with the equality replaced by an approximate equality, see e.g. Millar (1984), Vajda (1995) or Vajda and Janžura (1997)). Similarly, the statistical tests of

$$\mathscr{H}:F(\cdot,\theta_0)\quad\text{versus}\quad \mathscr{A}_n: G_n(\cdot)=\left(1-\frac{1}{\sqrt{n}}\right)F(\cdot,\theta_0)+\frac{1}{\sqrt{n}}F(\cdot,\theta)\tag{3}$$

for $\theta\neq\theta_0$ are assumed to reject $\mathscr{H}$ if

$$D(F_n(\cdot);F(\cdot,\theta_0))>c_{D,n}\tag{4}$$

for some critical values $c_{D,n}>0$.

Contiguous alternatives, introduced in (3) are usually employed in the asymptotic statistics in order to analyze the behaviour of tests for the sample sizes $n\to\infty$.

The interest of the statistical theory is focused on the *asymptotically efficient estimators* (briefly, AE estimators, see e. g. Section 4.1 in Serfling (1980)), defined by the condition

$$\lim_{n\to\infty}\sqrt{n}(\theta_n-\theta_0)=N(0,1/I(\theta_0))\quad P\text{-weakly}\tag{5}$$

for the *Fisher information function*

$$I(\theta)=\int\frac{\dot f^2(x,\theta)}{f(x,\theta)}\,\mathrm{d}x,\quad \theta\in\Theta,\tag{6}$$

where $\dot f(x,\theta)=\mathrm{d}f(x,\theta)/\mathrm{d}\theta$ are derivatives of the densities $f(x,\theta)=\mathrm{d}F(x,\theta)/\mathrm{d}x$ and it is assumed $0<I(\theta)<\infty$. Further, the interest is focused on the *asymptotically efficient tests* (briefly AE tests, cf. Section 10.1 in Serfling (1980)) which, for every *asymptotic size* $\alpha\in(0,1)$ satisfying the condition

$$\alpha = \limsup_{n\to\infty} P(D(F_n(\cdot); F(\cdot,\theta_0)) > c_{D,n}), \tag{7}$$

maximize the *asymptotic power*

$$\beta = \beta(\theta,\theta_0|\alpha) = \liminf_{n\to\infty} Q_n(D(F_n(\cdot); F(\cdot,\theta_0)) > c_{D,n}) \tag{8}$$

at the local alternatives $\mathscr{A}_n$ defined in (3), for θ from the neighborhood of θ_0. Note that in (5), (7) and (8), and everywhere in the sequel, P denotes the probability measure under $\mathscr{H}$ and Q_n the probability measure under $\mathscr{A}_n$, and the maximization of the asymptotic power (8) extends over a given class of tests (distances D).

Notice that the efficiency condition (5) does not depend on the class of distances taken into account, so that the AE of estimators is an absolute concept, while the maximization of (8) extends over a given class of distances so that the AE of tests is a relative concept.

Obviously, if for appropriate norming constants $a_{D,n} > 0$

$$\lim_{n\to\infty} a_{D,n} D(F_n(\cdot); F(\cdot,\theta_0)) = \mathrm{X}_k^2 \quad P\text{-weakly}, \tag{9}$$

where X_k^2 is the chi-square distributed random variable, then (7) holds with limsup replaced by lim for $c_{D,n} = \mathrm{X}_{k,1-\alpha}^2/a_{D,n}$, where $\mathrm{X}_{k,1-\alpha}^2$ denotes the $(1-\alpha)$-fractile of X_k^2.

The distances under consideration are usually from the class of ϕ-divergences, defined by the formula

$$D(F_n(\cdot); F(\cdot,\theta)) = D_\phi(p_n(\boldsymbol{y}); p(\theta|\boldsymbol{y})), \tag{10}$$

where the right-hand side is, for an appropriate function $\phi(t)$, $t > 0$, the ϕ-disparity of the discrete empirical and theoretical distributions obtained by means of a partition $\mathscr{P}$ of R into $1 < m < \infty$ intervals $A_1, \dots, A_m$ defined by increasing coordinates of a given $(m-1)$-vector $\boldsymbol{y} = (y_1, \dots, y_{m-1})$, that is,

$$p_n(\boldsymbol{y}) = (p_{n1}(\boldsymbol{y}), \dots, p_{nm}(\boldsymbol{y})) \quad \text{where} \quad p_{nj}(\boldsymbol{y}) = F_n(y_j) - F_n(y_{j-1}) \tag{11}$$

and

$$p(\theta|\boldsymbol{y}) = (p_1(\theta|\boldsymbol{y}), \dots, p_m(\theta|\boldsymbol{y})) \quad \text{where} \quad p_j(\theta|\boldsymbol{y}) = F(y_j,\theta) - F(y_{j-1},\theta) \tag{12}$$

for $y_0 = -\infty$, $y_m = \infty$.

The *ϕ-disparity* of discrete distributions $\boldsymbol{p}$ and $\boldsymbol{q}$ has been defined in Menéndez et al. (1998, 2001b) by the same formula as the ϕ-divergence of Csiszár (see Csiszár (1963, 1967), Liese and Vajda (1987), Read and Cressie (1988) or Vajda (1989)),

$$D_\phi(\boldsymbol{p}; \boldsymbol{q}) = \sum_{j=1}^{m} q_j \phi\left(\frac{p_j}{q_j}\right), \tag{13}$$

but assuming that the function $\phi(t)$ in (13) is strictly convex only in the neighborhood of $t = 1$ (where it is assumed to be twice continuously differentiable with $\phi(1) = \phi'(1) = 0$ and $\phi''(1) > 0$) and monotone on the intervals $(0, 1)$ and $(1, \infty)$.

Note that the convex functions ϕ assumed in the definition of ϕ-divergences (10), and satisfying the condition $\phi(1) = \phi'(1) = 0$, are automatically monotone in the stated sense. But these functions cannot be bounded on $(0, \infty)$. The extension of ϕ-divergences to the ϕ-disparities has been motivated by results of Lindsay (1994), Basu and Sarkar (1994a, b) and Park et al. (1995), who have shown that a robust estimation and testing is possible only when the ϕ function figuring in (10) and (13) is bounded on $(0, \infty)$. Convex functions $\phi(t)$ with $\phi(1) = \phi'(1) = 0$ and $\phi''(1) > 0$ cannot satisfy this condition.

The most common ϕ-disparities are some ϕ-divergences, namely the *information divergence* (briefly ID), $I(\boldsymbol{p}; \boldsymbol{q})$, obtained by using $\phi(t) = t \ln t$ in (13) and leading in (2) to the maximum likelihood estimator; the reversed ID, $I(\boldsymbol{q}; \boldsymbol{p})$, obtained from $\phi(t) = -\ln t$ and leading in (2) to another classical estimator introduced by Rao (1961); the *Pearson divergence* (PD), $\chi^2(\boldsymbol{p}; \boldsymbol{q})$, obtained for $\phi(t) = (t-1)^2$ and leading in (4) to the Pearson test; and the *reversed* PD (or *Neyman divergence*), $\chi^2(\boldsymbol{q}; \boldsymbol{p})$, obtained for $\phi(t) = (t-1)^2/t$ and leading in (8) to the Neyman χ^2-test, cf. Read and Cressie (1988). In the latter reference one can find also the authors who introduced the ID-tests (known also as the likelihood ratio tests), the reversed ID-tests, and the Pearson and Neyman divergence estimators.

More generally, the functions $\phi(t) = \|t-1\|^\alpha / t^{\alpha-1}$ with $\alpha \geq 1$ define ϕ-divergences and $\phi(t) = \|t-1\|^\alpha / t^\alpha$ with $\alpha \geq 0$, or $\|t-1\|/t^\alpha$ with $0 < \alpha < 1$, define ϕ-disparities.

In Section 2 we argue, by using the results of Lindsay (1994) and Menéndez et al. (2001a), that all ϕ-disparity estimators $\theta_{m,n}^{\phi,y}$ defined by (2) for the distances (10) are asymptotically efficient (best asymptotically normal) in the discrete statistical models

$$(p(\theta|\boldsymbol{y}) : \theta \in \Theta) \tag{14}$$

when the models satisfy the standard regularity imposed on discrete models in the asymptotic statistics (see e. g. Birch (1964)). In other words, this regularity implies

$$\lim_{n\to\infty} \sqrt{n}(\theta_{m,n}^{\phi,y} - \theta_0) = N(0, 1/I_m(\boldsymbol{y}|\theta_0)) \quad P\text{-weakly} \tag{15}$$

where

$$I_m(\theta|\boldsymbol{y}) = \sum_{j=1}^{m} \frac{\dot{p}_j^2(\theta|\boldsymbol{y})}{p_j(\theta|\boldsymbol{y})}, \quad \theta \in \Theta, \quad \text{for } \dot{p}_j(\theta|\boldsymbol{y}) = \frac{\mathrm{d}p_j(\theta|\boldsymbol{y})}{\mathrm{d}\theta} \tag{16}$$

is the Fisher information function in the model (14). This property has been established for the MLE by Birch (1964), for the most common ϕ-divergence estimators by Read and Cressie (1988) and for all ϕ-divergence estimators by Morales et al. (1995).

Every $\theta_{m,n}^{\phi,\boldsymbol{y}}$ with given m, $\boldsymbol{y}$ and ϕ is one possible specification of the above considered estimator θ_n in the model (1). Therefore one can ask whether it is efficient, or how much inefficient it is in this model. It follows from (5) and (15) that

$$\Delta_m(\theta_0|\boldsymbol{y}) = I(\theta_0) - I_m(\theta_0|\boldsymbol{y}) \tag{17}$$

is a measure of *asymptotic inefficiency* of all estimators $\theta_{m,n}^{\phi,\boldsymbol{y}}$ in the continuous model (1). In particular, $\theta_{m,n}^{\phi,\boldsymbol{y}}$ is efficient in the continuous model (1) if and only if the Fisher information function $I_m(\theta|\boldsymbol{y})$ in the discrete model (14) coincides with the Fisher information function (6) in the continuous one (1). Note that the nonnegativity in (17) follows e. g. from Theorem 3 in Vajda (1973).

Similarly we argue, with a reference to Menéndez et al. (1998), that the convergence condition for tests based on distances in (9) holds for the class of ϕ-divergences (10) and for $a_{D,n} = a_{\phi,n} = 2n/\phi''(1)$ and $k = m-1$. Therefore, the test in (4) with the disparity distance (10) and the critical values $c_{D,n} = c_{\phi,n}$ given by the formula

$$c_{\phi,n} = \frac{\phi''(1)\mathrm{X}^2_{m-1,1-\alpha}}{2n}, \tag{18}$$

defines a ϕ-disparity test satisfying relation (7) for all $\alpha \in (0,1)$, but with *limsup* replaced by *lim*. Further, we argue, with a reference to the recent paper of Menéndez et al. (2001b), that $\Delta_m(\theta_0|\boldsymbol{y})$, defined in (17), is a measure of asymptotic inefficiency of all ϕ-disparity tests.

Moreover, it is easy to prove by using Theorem 4 in Vajda (1973) that, for every $\theta_0 \in \Theta$ and $\boldsymbol{y} = \boldsymbol{y}_*(\theta_0)$ defined by

$$I_m(\theta_0\,|\,\boldsymbol{y}_*(\theta_0)) = \max_{\boldsymbol{y}} I_m(\theta_0|\boldsymbol{y}) \tag{19}$$

(and thus depending on m), it holds that

$$\Delta_m(\theta_0\,|\,\boldsymbol{y}_*(\theta_0)) = o(1) \quad \text{as } m \to \infty. \tag{20}$$

Main implication of (20) is that, under quantizations $\boldsymbol{y}_*(\theta_0)$, the ϕ-disparity estimators and tests defined by (2) and (4) respectively, using the class of ϕ-divergences (10) for the discrete probability distribution defined by (12) for $\boldsymbol{y} = \boldsymbol{y}_*(\theta_0)$, are for $m \to \infty$ asymptotically efficient in the continuous model (1).

Unfortunately, the computationally simple and at the same time robust and asymptotically efficient methods for the family (1), resulting from the quantization of R by the $m-1$ components of $\boldsymbol{y}_*(\theta_0)$, and the subsequent application of the ϕ-disparity with bounded ϕ, are not practically applicable because of the dependence of $\boldsymbol{y}_*(\theta_0)$ on the unknown θ_0. Some authors found the way out of this unpleasant situation by considering the infinite uniform partitions by

$$\boldsymbol{y}_\infty^{(k)} = \left(\frac{2j \pm 1}{2k} : j = 0, \pm 1, \ldots, k = 1, 2, \ldots\right)$$

for which Theorem 3 of Vajda (2001) implies

$$\Delta_\infty(\theta_0|\boldsymbol{y}_\infty^{(k)}) = o(1) \quad \text{as } k \to \infty$$

for every $\theta_0 \in \Theta$ with $I(\theta_0) < \infty$, e. g. Ghurye and Johnson (1981), Zografos et al. (1986) and Tsaridis et al. (1997). But such infinite quantizations eliminate one of the main advantages, namely the computational simplicity of subsequent inference procedures and the effective use of all cells, which are hardly possible when there is a large number of sparsely frequented cells, typical for partitions of the type $\boldsymbol{y}_\infty^{(k)}$ in case of heavier-tailed distributions.

A different more appropriate way out of this situation can be based on the fact that for the suboptimal partition vector consisting of the j/m-fractiles of $F(x, \theta_0)$

$$\boldsymbol{y}(\theta_0) = (y_1(\theta_0), \ldots, y_{m-1}(\theta_0)) \quad \text{with} \quad F(y_j(\theta_0), \theta_0) = \frac{j}{m} \tag{21}$$

an analogue of (20) holds, namely

$$\Delta_m(\theta_0 \,|\, \boldsymbol{y}(\theta_0)) = o(1) \quad \text{as } m \to \infty \tag{22}$$

by the arguments in Vajda (2001).

The second, more interesting step, follows from the results of Menéndez et al. (1998, 2001a), where it is proved that the ϕ-disparity estimators with the random partitions defined by the j/m-fractiles of the empirical distribution $F_n(x)$, i. e. by

$$\boldsymbol{y}_n = (y_{n1}, \ldots, y_{nm}) \quad \text{with} \quad y_{nj} = F_n^{-1}(j/m), \tag{23}$$

achieve the asymptotic inefficiency $\Delta_m(\theta_0 \,|\, \boldsymbol{y}(\theta_0))$, i. e. the same asymptotic inefficiency as the corresponding estimators and tests using the partition $\boldsymbol{y}(\theta_0)$. One can say that they adapt automatically to the unknown partition $\boldsymbol{y}(\theta_0)$ which is suboptimal, but still good enough.

In Section 3 we are interested in the models of location $F(x, \theta) = F(x - \theta)$ for $\theta \in \Theta = R$ in (1), where

$$I(\theta) = I(0),\ I_m(\theta|\boldsymbol{y}) = I_m(0|\boldsymbol{y}),\ \boldsymbol{y}_*(\theta) = \boldsymbol{y}_*(0) + \theta,\ \boldsymbol{y}(\theta) = \boldsymbol{y}(0) + \theta. \tag{24}$$

We considerably extend the results of Pötzelberger and Felsenstein (1993) and Tsaridis et al. (1997), who computed $I_m(0 \,|\, \boldsymbol{y}(0))$ for the most common location families and $m = 2, 3$ and 4. We present these values for $2 \le m \le 20$ in the logistic, normal, Cauchy and double exponential location families. We show that the inefficiency $\Delta_m(\theta \,|\, \boldsymbol{y}(\theta)) = \Delta_m(0 \,|\, \boldsymbol{y}(0))$ is practically negligible in these families for $10 \le m \le 20$. Further, we show that if $I(\theta) < \infty$ then $\Delta_m(0 \,|\, \boldsymbol{y}(0)) = o(1)$ as $m \to \infty$, and that in some cases

$$\Delta_m(0 \,|\, \boldsymbol{y}(0)) = O(1/m^2) \quad \text{as } m \to \infty. \tag{25}$$

2 Fisher information and efficiency of estimation

Menéndez et al. (2001a, b) considered discrete models (14) satisfying the following variants of the regularity assumptions of Birch (1964):

(B1) All coordinates of $p(\theta_0|\boldsymbol{y})$ are positive.
(B2) In the neighborhood of θ_0, the vector of derivatives $\dot{p}(\theta|\boldsymbol{y}) = (\dot{p}_1(\theta|\boldsymbol{y}), \ldots, \dot{p}_m(\theta|\boldsymbol{y}))$ considered in (16) exists and is continuous.
(B3) The vector $\dot{p}(\theta_0|\boldsymbol{y})$ is non-zero.
(B4) The mapping $\theta \mapsto p(\theta|\boldsymbol{y})$ is one-one on Θ.

The condition (B2) has the equivalent formulation:

(B2) In the neighborhood of θ_0 the vector $\dot{F}(\theta|\boldsymbol{y}) = \mathrm{d}/\mathrm{d}\theta F(\theta|\boldsymbol{y})$ exists and is continuous, where

$$F(\boldsymbol{y}, \theta) = (F(y_1, \theta), \ldots, F(y_m, \theta)). \tag{26}$$

The condition (B4) was shown to have an equivalent formulation in terms of $F(\boldsymbol{y}, \theta)$, namely

(B4) The mapping $\theta \mapsto F(\boldsymbol{y}, \theta)$ is one-one on Θ.

In particular, the cited papers studied the ϕ-divergence and ϕ-disparity estimators $\theta_{m,n}^{\phi,\boldsymbol{y}}$ minimizing on Θ the ϕ-disparity functions

$$D_\phi(p_n(\boldsymbol{y}); p(\theta|\boldsymbol{y})) = \sum_{j=1}^{m} p_j(\theta|\boldsymbol{y})\phi\left(\frac{p_{nj}(\boldsymbol{y})}{p_j(\theta|\boldsymbol{y})}\right), \quad \theta \in \Theta, \tag{27}$$

for the empirical discrete distributions

$$p_n(\boldsymbol{y}) = (p_{nj}(\boldsymbol{y}) \triangleq F_n(y_j) - F_n(y_{j-1}) : 1 \leq j \leq m), \tag{28}$$

and the theoretical discrete distributions $p(\theta|\boldsymbol{y})$. Of course, each ϕ-disparity estimator can be interpreted as an estimator of the true parameter θ_0 in the original continuous model (1). Theorem 4.1 in Menéndez et al. (2001a) stated that under (B1)–(B4) all ϕ-disparity estimators $\theta_{m,n}^{\theta,\boldsymbol{y}}$ satisfy (15) for the Fisher information given by (16).

Results of this type, and also deeper results about the higher order efficiency of ϕ-disparity estimators can be deduced from Lindsay (1994), who also argued that the ϕ-disparity estimators are robust in the case where $\phi(t)$, $t > 0$, is bounded.

Our attention will be paid to the particular case $p(\theta_0|\boldsymbol{y}) = q$ where q denotes from now on the uniform distribution,

$$q = \left(\frac{1}{m}, \frac{1}{m}, \ldots, \frac{1}{m}\right), \tag{29}$$

i.e. to the case where $\boldsymbol{y}$ quantizes the observation space R into *statistically equivalent blocks*. It is useful to consider the following assumption:

(A0) For every $1 \leq j \leq m-1$ there exists $y_{0j} \in R$ such that $F(y_{0j}, \theta_0) = j/m$ and $F(x, \theta_0)$ is increasing in the neighborhood of $x = y_{0j}$.

Menéndez et al. (2001b) studied this particular case and, in addition, they considered the following assumptions:

(A1) ≡ (B1) for $\boldsymbol{y} = \boldsymbol{y}_0 = (y_{01}, \dots, y_{0m-1})$ figuring in (A0).
(A2) In the neighborhood of $(\boldsymbol{y}_0, \theta_0) \in R^{m-1} \times R = R^m$, $F(\boldsymbol{y}, \theta)$ given by (26) is continuous and the vector of derivatives $\dot{F}(\theta|\boldsymbol{y})$ figuring in the second version of (B2) exists and is continuous.
(A3) ≡ (B3) for $\boldsymbol{y} = \boldsymbol{y}_0$ figuring in (A0).
(A4) (B3) holds for all $\boldsymbol{y} \in R^{m-1}$ from the neighborhood of $\boldsymbol{y}_0$.

Thus (A0)–(A4) hold if (B1)–(B4) hold for $\boldsymbol{y} = \boldsymbol{y}_0$ and, moreover, (B2), (B4) hold also for $\boldsymbol{y}$ from the neighborhood of $\boldsymbol{y}_0$. Therefore (A1)–(A4) are slightly stronger than (B1)–(B4) taken at $\boldsymbol{y} = \boldsymbol{y}_0$ given in (A0).

Menéndez et al. (2001a) studied also the ϕ-disparity estimators $\theta^{\phi,\boldsymbol{y}}_{m,n}$ for the empirical quantization vectors $\boldsymbol{y} = \boldsymbol{y}_n$ given by (23). We put for brevity

$$\theta^{\phi,\boldsymbol{y}_n}_{m,n} = \tilde{\theta}^{\phi}_{m,n}, \tag{30}$$

and in (27) we replace $p_n(\boldsymbol{y}_n)$ for simplicity by q. (As follows from (28), the difference $p_n(\boldsymbol{y}_n) - q$ is asymptotically negligible.) Therefore the empirical ϕ-disparity estimator $\tilde{\theta}^{\phi}_{m,n}$ minimizes on Θ the random function

$$D_\phi(q; p(\theta|\boldsymbol{y}_n)) = \sum_{j=1}^{m} p_j(\theta|\boldsymbol{y}_n)\phi\left(\frac{1}{mp_j(\theta|\boldsymbol{y}_n)}\right), \quad \theta \in \Theta.$$

For the reversed ϕ-disparities with $\tilde{\phi}(t) = t\phi(1/t)$, the formula for the minimized function simplifies as follows

$$D_{\tilde{\phi}}(q; p(\theta|\boldsymbol{y}_n)) = D_\phi(p(\theta|\boldsymbol{y}_n); q) = \frac{1}{m}\sum_{j=1}^{m} \phi(mp_j(\theta|\boldsymbol{y}_n)). \tag{31}$$

In Theorem 5.1 of the cited paper it is proved that if (A0)–(A4) hold then $\tilde{\theta}^{\phi}_{m,n}$ is asymptotically normal in the sense of (15), with the Fisher information $I_m(\theta_0|\boldsymbol{y}) = I_m(\theta_0|\boldsymbol{y}_0)$.

The vector $\boldsymbol{y}_0$ depends on the true parameter θ_0, i. e. $\boldsymbol{y}_0 = \boldsymbol{y}_0(\theta_0)$. In order to simplify notation, we put in the sequel $\boldsymbol{y}_0 = \boldsymbol{y}(\theta_0)$, i. e. we define

$$\boldsymbol{y}(\theta_0) = (y_j(\theta_0) \triangleq F^{-1}(j/m, \theta_0) : 1 \le j \le m), \tag{32}$$

where $t \mapsto F^{-1}(t, \theta_0)$ is inverse to $x \mapsto F(x, \theta_0)$. Notice that under (A0) the inverse function $F^{-1}(t, \theta_0)$ is well defined and continuous in the neighborhoods of all points $t = j/m$, $1 \le j \le m$.

The empirical ϕ-disparity estimator $\tilde{\theta}^{\phi}_{m,n}$ is one of the possible estimators in the continuous model (1), similarly as $\theta^{\phi,\boldsymbol{y}}_{m,n}$. The inefficiencies of these two estimators in this model are given respectively by $\Delta_m(\theta_0 \mid \boldsymbol{y}(\theta_0))$ and $\Delta_m(\theta_0|\boldsymbol{y})$, defined in (17). The minimum inefficiency under the quantization of size m is achieved by $\boldsymbol{y} = \boldsymbol{y}_*(\theta_0)$ defined in (19), so that $0 \le \Delta_m(\theta_0 \mid \boldsymbol{y}_*(\theta_0)) \le \Delta_m(\theta_0 \mid \boldsymbol{y}(\theta_0))$ or, equivalently,

$$I_m(\theta_0 \,|\, \boldsymbol{y}(\theta_0)) \le I_m(\theta_0 \,|\, \boldsymbol{y}_*(\theta_0)) \le I(\theta_0) \quad \text{(cf. (17))}.$$

Unfortunately, the least inefficient estimation procedure cannot be practically used because $\boldsymbol{y}_*(\theta_0)$ depends on the true θ_0 which is to be estimated. However, as said above, the inefficiency $\Delta(\theta_0 \,|\, \boldsymbol{y}(\theta_0))$ is achieved by the empirical ϕ-disparity estimators under consideration, under the regularity of model (1) summarized in (A0)–(A4). Moreover, as follows from Lindsay (1994) and other authors cited in Section 1, these estimators are robust with respect to contaminations of the original data $X_1, \ldots, X_n$, when the function ϕ considered in (31) is bounded. This is also seen directly from (31), where the effect of arbitrary deviations of the random quantization $\boldsymbol{y}_n$ on the minimized ϕ-disparity function is clearly limited when ϕ is bounded. This motivates the practical statistical interest in the following definition.

The empirical ϕ-disparity estimators $\tilde{\theta}^{\phi}_{m,n}$ are efficient in the model (1) *asymptotically* for $m \to \infty$ if $\boldsymbol{y}(\theta_0)$ satisfies (20), i. e. if

$$I(\theta_0) - I_m(\theta_0 \,|\, \boldsymbol{y}(\theta_0)) = o(1) \quad \text{as } m \to \infty. \tag{33}$$

The problem of asymptotic efficiency of this kind is solved by the next result which follows from the arguments in Example 4 of Vajda (2001).

Theorem 1. *Let the densities $f(x,\theta)$ of distributions considered in (1) be a.e. differentiable in θ, satisfying the conditions $I(\theta_0) < \infty$ and*

$$\frac{\mathrm{d}}{\mathrm{d}\theta}\int_a^b f(x,\theta)\,\mathrm{d}x = \int_a^b \frac{\mathrm{d}}{\mathrm{d}\theta} f(x,\theta)\,\mathrm{d}x \quad \text{for all } \infty \le a < b \le \infty.$$

Then (33) holds.

3 Fisher information and efficiency of testing

The Fisher informations $I(\theta_0)$, $I_m(\theta_0|\boldsymbol{y})$ and $I_m(\theta_0 \,|\, \boldsymbol{y}_*(\theta_0))$ play a similar role in the testing of statistical hypotheses as in the statistical point estimation. The goodness of fit tests (4) with the ϕ-disparity distance (10) were among the tests studied by Menéndez et al. (1998) under the assumption that the vector $p(\theta_0|\boldsymbol{y})$ has all coordinates positive. In Theorem 4.1 ibid., the corresponding test statistics were shown to be asymptotically X^2_k-distributed in the sense of (9) with $k = m - 1$ for the norming constants $a_{\phi,n} = 2n/\phi''(1)$. (This result was proved also in part I of Theorem 3.1 of Inglot et al. (1991) for $\phi(t)$ three times continuously differentiable in the neighborhood of $t = 1$). Therefore the corresponding tests of asymptotic size $\alpha \in (0,1)$ (cf. (7)) are of the form

$$2n\sum_{j=1}^{m} p_j(\theta_0|\boldsymbol{y})\phi\left(\frac{p_{nj}(\boldsymbol{y})}{p_j(\theta_0|\boldsymbol{y})}\right) > \phi''(1)\mathrm{X}^2_{m-1,1-\alpha}. \tag{34}$$

The question is how inefficient these tests are, or which of them are efficient in the sense outlined in Section 1. To find an answer, take first into account that one can vary in these tests the used distance D, i.e. the convex functions ϕ and the quantization $\boldsymbol{y}$. Thus it is natural to consider for all $\alpha \in (0,1)$ and $m > 1$ the efficiency in the classes $\mathscr{T}_{\alpha,m}$ of the tests (34).

By the definition of Section 1, we need to maximize the power $\beta(\theta, \theta_0|\alpha)$ given by (8) over the tests from $\mathscr{T}_{\alpha,m}$, locally in the neighborhood of $\theta = \theta_0$. The next result considerably simplifies evaluation of the powers $\beta(\theta, \theta_0|\alpha)$ for the tests from $\mathscr{T}_{\alpha,m}$ at local alternatives $\mathscr{A}_n = \mathscr{A}_n(\theta)$ defined in (3) for $\theta \in \Theta$, $\theta \neq \theta_0$. In this result, and in the sequel, $F_k(x|\lambda)$ denotes the distribution function of the noncentral chi-square distributed random variable $\mathrm{X}_k^2(\lambda)$ with k degrees of freedom and a noncentrality parameter $\lambda \geq 0$, and $F_k(x) \triangleq F_k(x|0)$. Hence in particular, $F_k^{-1}(1-\alpha) = \mathrm{X}_{k,1-\alpha}^2$ in the notation introduced in Section 1 after formula (9).

Lemma 1. *For all tests (34) belonging to $\mathscr{T}_{\alpha,m}$, the asymptotic power at any local alternative $\mathscr{A}_n(\theta)$, $\theta \in \Theta$, is given by the formula*

$$\beta(\theta, \theta_0|\alpha) = 1 - F_{m-1}(\gamma|\delta) = P(\mathrm{X}_{m-1}^2(\delta) > \gamma) \tag{35}$$

for

$$\gamma = \gamma_m(\alpha) \triangleq F_{m-1}^{-1}(1-\alpha) \quad \textit{and} \quad \delta = \delta_m(\theta, \theta_0|\boldsymbol{y}) \triangleq \chi^2(p(\theta|\boldsymbol{y}); p(\theta_0|\boldsymbol{y})). \tag{36}$$

Proof. Let us consider an arbitrary test (34) from $\mathscr{T}_{\alpha,m}$. As said above,

$$\alpha = 1 - F_{m-1}(\gamma) \quad \text{for } \gamma \text{ given by (36)} \tag{37}$$

is its asymptotic size. By Corollary 3.1 in Menéndez et al. (2001b), it follows from here that (35) with γ and δ given by (36) is the asymptotic power of this test at the alternative $\theta \neq \theta_0$.

Lemma 1 reduces the problem under consideration to the maximization over $\boldsymbol{y} = (y_1, \ldots, y_{m-1})$ of

$$\pi_m(\theta, \theta_0|\alpha, \boldsymbol{y}) \triangleq 1 - F_{m-1}(\gamma_m(\alpha)\,|\,\delta_m(\theta, \theta_0|\boldsymbol{y})) \tag{38}$$

for θ from the close neighborhood of θ_0. If $\theta = \theta_0$ then the chi-square divergence $\delta_m(\theta_0, \theta_0|\boldsymbol{y})$ is zero so that (37) implies

$$\pi_m(\theta_0, \theta_0|\alpha, \boldsymbol{y}) = 1 - F_{m-1}(\gamma_m(\alpha)\,|\,0) = \alpha$$

for all $\boldsymbol{y}$ under consideration. It is easy to see that under the regularity assumed in Theorem 1,

$$\delta_m(\theta, \theta_0|\boldsymbol{y}) = I_m(\theta_0|\boldsymbol{y})(\theta - \theta_0)^2 + o((\theta - \theta_0)^2) \quad \text{for } \theta \to \theta_0, \tag{39}$$

where $I_m(\theta|\boldsymbol{y})$ is the Fisher information (16). Since for every $\gamma_m(\alpha) > 0$ the function $F_{m-1}(\gamma_m(\alpha)\,|\,\delta)$ is twice differentiable in the domain $\delta > 0$, one can expect the existence of a constant $c_m(\alpha)$ such that

$$\pi_m(\theta, \theta_0|\alpha, \boldsymbol{y}) = \alpha + c_m(\alpha) I_m(\theta_0|\boldsymbol{y})(\theta - \theta_0)^2 + o((\theta - \theta_0)^2) \quad \text{for } \theta \to \theta_0. \tag{40}$$

We see that if (40) is valid then $I_m(\theta_0|\boldsymbol{y})$ characterizes the rate of convergence of the asymptotic power $\pi_m(\theta, \theta_0|\alpha, \boldsymbol{y})$ to the asymptotic size α in the class of tests $\mathscr{T}_{\alpha,m}$, and that $c_m(\alpha)I_m(\theta_0|\boldsymbol{y})$ characterizes the same rate in the united class

$$\mathscr{T}_\alpha = \bigcup_{m>1} \mathscr{T}_{\alpha,m}.$$

Thus the efficiency in the class $\mathscr{T}_{\alpha,m}$ or $\mathscr{T}_\alpha$ is characterized by the maximal *asymptotic power ratio* at θ_0, defined by

$$\pi_{\alpha,m}(\theta_0) = \sup_{\boldsymbol{y}} \lim_{\theta\to\theta_0} \frac{[\pi_m(\theta, \theta_0|\alpha, \boldsymbol{y}) - \alpha]}{c_m(\alpha)(\theta-\theta_0)^2} = \sup_{\boldsymbol{y}} I_m(\theta_0|\boldsymbol{y}) \tag{41}$$

or

$$\pi_\alpha(\theta_0) = \sup_{m,\boldsymbol{y}} \lim_{\theta\to\theta_0} \frac{[\pi_m(\theta, \theta_0|\alpha, \boldsymbol{y}) - \alpha]}{(\theta-\theta_0)^2} = \sup_{m,\boldsymbol{y}} c_m(\alpha)I_m(\theta_0|\boldsymbol{y}), \tag{42}$$

respectively. The right-hand side supremum (41) was denoted above by $I_m(\theta_0\,|\,\boldsymbol{y}_*(\theta_0))$, i. e. the maximal power ratios are given by the formulas

$$\pi_{\alpha,m}(\theta_0) = I_m(\theta_0\,|\,\boldsymbol{y}_*(\theta_0)) \quad \text{and} \quad \pi_\alpha(\theta_0) = \sup_{m>1} c_m(\alpha)I_m(\theta_0\,|\,\boldsymbol{y}_*(\theta_0)). \tag{43}$$

We are thus interested in the validity of (40) and in evaluation of the constants $c_m(\alpha)$ figuring there. This problem is solved by the following theorem using the next lemma.

Lemma 2. *The distribution functions considered in Lemma 1 satisfy for all* $k \geq 1$ *and* $x > 0$, $\lambda > 0$ *the relation*

$$F_k(x|\lambda) - F_k(x) = \frac{\lambda}{2}[F_{k+2}(x) - F_k(x)] + R_k(x, \lambda) \tag{44}$$

where

$$R_k(0, \lambda) = R_k(\infty, \lambda) = 0 \quad \textit{and} \quad \sup_{x>0} |R_k(x, \lambda)| < 1 - e^{-\lambda/2}(1 + \lambda/2).$$

Proof. We shall start with Ferguson (1996) who on p. 62 presents for $\sqrt{\mathrm{X}_k^2(\lambda)}$ the density

$$g_k(x|\lambda) = e^{-\lambda/2} \sum_{j=0}^{\infty} \frac{(\lambda/2)^j}{j!} \frac{2xx^{k+2j-2}e^{-x^2/2}}{2^{k/2+j}\Gamma(k/2+j)}.$$

Hence the density of $\mathrm{X}_k^2(\lambda)$ is

$$f_k(x|\lambda) = e^{-\lambda/2} \sum_{j=0}^{\infty} \frac{(\lambda/2)^j}{j!} \frac{x^{(k+2j)/2-1}e^{-x/2}}{2^{(k+2j)/2}\Gamma((k+2j)/2)},$$

i.e. the distribution functions under consideration satisfy the relation

$$F_k(x|\lambda) = e^{-\lambda/2} \sum_{j=0}^{\infty} \frac{(\lambda/2)^j}{j!} F_{k+2j}(x).$$

The formula (44) with

$$R_k(x,\lambda) = (e^{-\lambda/2} - 1 + \lambda/2)F_k(x) + (e^{-\lambda/2} - 1)(\lambda/2)F_{k+2}(x) + e^{-\lambda/2} \sum_{j=2}^{\infty} \frac{(\lambda/2)^j}{j!} F_{k+2j}(x)$$

follows directly from this relation. Further, $R_k(x,\lambda) = R_k^+(x,\lambda) + R_k^-(x,\lambda)$ where

$$R_k^+(x,\lambda) = e^{-\lambda/2} \sum_{j=2}^{\infty} \frac{(\lambda/2)^j}{j!} F_{k+2j}(x)$$

and

$$R_k^-(x,\lambda) = -[(1 - \lambda/2 - e^{-\lambda/2})F_k(x) + (1 - e^{-\lambda/2})(\lambda/2)F_{k+2}(x)].$$

The rest is clear from the fact that both these functions are continuously monotone with $R_k^+(0,\lambda) = R_k^-(0,\lambda) = 0$ and

$$0 < R_k^+(\infty,\lambda) = e^{-\lambda/2}(e^{\lambda/2} - 1 - \lambda/2) = 1 - e^{-\lambda/2}(1 + \lambda/2),$$

$$0 > R_k^-(\infty,\lambda) = -[1 - \lambda/2 - e^{-\lambda/2} + (1 - e^{-\lambda/2})\lambda/2] = -R_k^+(\infty,\lambda),$$

so that $|R_k| \le \max(R_k^+, |R_k^-|)$.

Theorem 2. *If the model is regular in the sense considered in Theorem 1 then the asymptotic relation (40) with*

$$c_m(\alpha) = \frac{1 - \alpha - F_{m+1}(\gamma_m(\alpha))}{2} \quad (cf.\ (36)) \tag{45}$$

holds for every α, m and $\mathbf{y}$ considered there. The constants $c_m(\alpha)$ are positive and satisfy the inequality

$$c_m(\alpha) \ge \frac{1 - \alpha}{m + 1}. \tag{46}$$

Therefore the maximal power ratios $\pi_{\alpha,m}$ or π_α are well defined by the formulas (41)–(43) with $c_m(\alpha)$ given by (45) for all tests (34) from the classes $\mathcal{T}_{\alpha,m}$ or $\mathcal{T}_\alpha$, respectively.

Proof. Put $\gamma = \gamma_m(\alpha)$. Lemma 2 implies that

$$F_{m-1}(\gamma|\delta) - F_{m-1}(\gamma) = \frac{\delta}{2}[F_{m+1}(\gamma) - F_{m-1}(\gamma)] + o(\delta) \quad \text{as } \delta \to 0,$$

where $F_{m-1}(\gamma) = 1 - \alpha$ by (37). Hence it follows from (38) and (44) that

$$\begin{aligned}\pi_m(\theta, \theta_0|\alpha, \boldsymbol{y}) - \alpha &= 1 - F_{m-1}(\gamma\,|\,\delta_m(\theta, \theta_0|\boldsymbol{y})) - [1 - F_{m-1}(\gamma)] \\ &= -[F_{m-1}(\gamma\,|\,\delta_m(\theta, \theta_0|\boldsymbol{y})) - F_{m-1}(\gamma)] \\ &= \frac{\delta_m(\theta, \theta_0|\boldsymbol{y})}{2}[F_{m-1}(\gamma) - F_{m+1}(\gamma)] + o(\delta_m(\theta, \theta_0|\boldsymbol{y}))\end{aligned}$$

as $\delta_m(\theta, \theta_0|\boldsymbol{y}) \to 0$. This together with (39) implies (40) with $c_m(\alpha)$ given by (45). It remains to prove the inequality (46). First, observe that

$$F_{k+2}(x) = \frac{1}{k+2}\int_0^x t f_k(t)\,\mathrm{d}t$$

where $f_k(x)$ is the density of $F_k(x)$. Thus the *per partes* rule gives for every $x > 0$ and $k \geq 1$

$$xF_k(x) = \int_0^x F_k(t)\,\mathrm{d}t + (k+2)F_{k+2}(x).$$

In particular, for $m > 1$

$$F_{m+1}(x) = \frac{xF_{m-1}(x) - \int_0^x F_{m-1}(t)\,\mathrm{d}t}{m+1}$$

so that

$$F_{m+1}(\gamma) = \frac{\gamma(1-\alpha) - \int_0^\gamma F_{m-1}(t)\,\mathrm{d}t}{m+1}.$$

Now

$$\begin{aligned}\gamma(1-\alpha) - \int_0^\gamma F_{m-1}(t)\,\mathrm{d}t &= \int_0^\gamma (1 - F_{m-1}(t))\,\mathrm{d}t - \alpha\gamma \\ &= \int_0^\infty (1 - F_{m-1}(t))\,\mathrm{d}t - \left[\alpha\gamma + \int_\gamma^\infty (1 - F_{m-1}(t))\,\mathrm{d}t\right] \\ &\leq \int_0^\infty (1 - F_{m-1}(t))\,\mathrm{d}t - \alpha\int_0^\infty (1 - F_{m-1}(t))\,\mathrm{d}t \\ &= (1-\alpha)\int_0^\infty (1 - F_{m-1}(t))\,\mathrm{d}t = (1-\alpha)(m-1),\end{aligned}$$

where the inequality follows from the fact that

$$\alpha\gamma + \int_\gamma^\infty (1 - F_{m-1}(t)\,\mathrm{d}t) \geq \alpha \int_0^\gamma 1 \cdot \mathrm{d}t + \alpha \int_\gamma^\infty (1 - F_{m-1}(t))\,\mathrm{d}t$$

$$\geq \alpha \int_0^\infty (1 - F_{m-1}(t))\,\mathrm{d}t$$

and the last equality follows from the relation

$$\int_0^\infty (1 - F_{m-1}(t))\,\mathrm{d}t = \int_0^\infty t\,\mathrm{d}F_{m-1}(t) = m - 1.$$

Therefore

$$c_m(\alpha) = \frac{1 - \alpha - F_{m+1}(\gamma)}{2} \geq \frac{(1-\alpha)[1 - (m-1)/(m+1)]}{2} = \frac{1-\alpha}{m+1}$$

which completes the proof.

By Theorem 2, in all models satisfying the standard regularity assumptions, the information $I_m(\theta_0|\boldsymbol{y})$ characterizes the sensitivity of the power $\pi_m(\theta, \theta_0|\alpha, \boldsymbol{y})$ of the tests from $\mathscr{T}_{\alpha,m}$ to small deviations of alternatives θ from the hypothesis $\theta = \theta_0$. The larger is the information $I_m(\theta_0|\boldsymbol{y})$ the sharper is the increase of power from the minimal possible level α achieved at $\theta = \theta_0$. By (17), the upper bound on $I_m(\theta_0|\boldsymbol{y})$ is $I(\theta_0)$ and, by Theorem 1, this bound is achievable. Therefore the nonnegative variable $I(\theta_0) - I_m(\theta_0|\boldsymbol{y})$ can serve as a relative measure of inefficiency in the class of tests $\mathscr{T}_{\alpha,m}$. The disadvantage of this method is that even the most efficient test in the class $\mathscr{T}_{\alpha,m}$, defined by the partition $\boldsymbol{y}_*(\theta_0)$ which maximizes the information $I_m(\theta_0|\boldsymbol{y})$ (see (19)), is often characterized as inefficient, since $I(\theta_0) - I_m(\theta_0\,|\,\boldsymbol{y}_*(\theta_0))$ is not negligible unless m is large enough. Therefore a more appropriate inefficiency measure in the class $\mathscr{T}_{\alpha,m}$ seems to be the difference between the maximal asymptotic power ratio $\pi_{\alpha,m}(\theta_0)$ (see (43)) and the actual asymptotic power ratio $I_m(\theta_0|\boldsymbol{y})$, i.e. the difference $\Delta_m^*(\theta_0|\boldsymbol{y}) \triangleq I_m(\theta_0\,|\,\boldsymbol{y}_*(\theta_0)) - I_m(\theta_0|\boldsymbol{y})$. Then a test (34) is relatively efficient in $T_{\alpha,m}$ if and only if $\Delta_m^*(\theta_0|\boldsymbol{y}) = 0$.

If the sequence $c_m(\alpha)$, $m = 1, 2, \ldots$, was nondecreasing with a finite limit $c_\infty(\alpha)$, then we could obtain from Theorem 1 the formula

$$\pi_\alpha = c_\infty(\alpha) I(\theta_0)$$

for the maximal asymptotic power ratio (43), and the relative inefficiency in the whole class $\mathscr{T}_\alpha$ could be defined as the difference

$$\Delta_\alpha(\theta_0) = c_\infty(\alpha) I(\theta_0) - c_m(\alpha) I(\theta_0|\boldsymbol{y}).$$

Unfortunately, a more detailed analysis of $c_m(\alpha)$ shows that this sequence is decreasing to zero (with the rate $m^{-1/2}$). The situation is thus not as simple as one might expect. But the informations $I_m(\theta_0\,|\,\boldsymbol{y}_*(\theta_0))$ considered in (43) are increasing to $I(\theta_0)$ in typical models with the rate higher than $m^{1/2}$ so that one can expect the existence of a unique integer $m_* = m_*(\theta_0|\alpha)$ with the property

$$\pi_\alpha(\theta_0) = c_{m_*}(\alpha) I_{m_*}(\theta_0 \,|\, \boldsymbol{y}_*(\theta_0)).$$

The evaluation of $m_*(\theta_0|\alpha)$, and the possibility to investigate the inefficiency

$$c_{m_*}(\alpha) I_{m_*}(\theta_0 \,|\, \boldsymbol{y}_*(\theta_0)) - c_m(\alpha) I_m(\theta_0|\boldsymbol{y})$$

of all tests (34) from the class $\mathcal{T}_\alpha$, are left as an interesting open problem.

We can conclude the second part of this section, dealing with the role of Fisher informations in testing statistical hypotheses, by underlining the role of Fisher informations $I_m(\theta_0 \,|\, \boldsymbol{y}_*(\theta_0))$ and $I_m(\theta_0|\boldsymbol{y})$ in evaluating the relative asymptotic inefficiencies of the estimators in classes $\mathcal{T}_{\alpha,m}$ with fixed $\alpha \in (0,1)$ and $m > 1$. Note that the informations $I_m(\theta_0 \mid \boldsymbol{y}(\theta_0))$ are important too. This follows from the fact that $\boldsymbol{y}_*(\theta_0)$ cannot be expressed in a closed form even in as common models as the normal or Cauchy model of location, and from the fact observed in the previously cited papers of Menéndez et al. (1998, 2001b). Namely, that the asymptotic distribution, asymptotic size and asymptotic power of the tests (34) at the alternatives $\mathcal{A}_n$ defined in (3) remain to be preserved when the quantization $\boldsymbol{y}(\theta_0)$ is replaced by the empirical quantization $\boldsymbol{y} = \boldsymbol{y}_n$ defined in (23). The preservation of the asymptotic distribution and size was proved in Theorem 5.1 of Menéndez et al. (1998) and the preservation of the asymptotic power in Corollary 3.1 of Menéndez et al. (2001b). The inefficiency of the empirical ϕ-disparity tests tends to zero as $m \to \infty$ by Theorem 1. The asymptotic α-size versions of these tests are of the following form:

$$2n \sum_{j=1}^{m} p_j(\theta_0|\boldsymbol{y}_n)\phi\left(\frac{1}{mp_j(\theta_0|\boldsymbol{y}_n)}\right) > \phi''(1)\mathrm{X}^2_{m-1,1-\alpha}. \tag{47}$$

For some functions ϕ, the test statistic can be simplified by using $\tilde{\phi}(t) = t\phi(1/t)$ instead of $\phi(t)$, similarly as in (31).

Note also that the convergence (33) is important in the hypotheses testing too. Namely, if $I(\theta_0) - I_{m_0}(\theta_0 \mid \boldsymbol{y}(\theta_0))$ is negligible for some m_0, then the non-negative difference

$$[I(\theta_0) - I_m(\theta_0|\boldsymbol{y})] - \Delta^*_m(\theta_0|\boldsymbol{y})$$
$$= I(\theta_0) - I_m(\theta_0 \,|\, \boldsymbol{y}_*(\theta_0)) \le I(\theta_0) - I_m(\theta_0 \mid \boldsymbol{y}(\theta_0))$$

is negligible for all $m > m_0$. Therefore, in this case the upper bound $I(\theta_0) - I_m(\theta_0|\boldsymbol{y})$ on the inefficiency $\Delta^*_m(\theta_0|\boldsymbol{y})$ can be used as a relative measure of inefficiency of the tests (34) in all classes $\mathcal{T}_{\alpha,m}$ with $\alpha \in (0,1)$ and $m \ge m_0$.

4 Fisher informations in location models

In this section we restrict ourselves to location families $F(x,\theta) = F(x-\theta)$, $x, \theta \in R$ and study the Fisher information measures

$$I_m(\theta \mid \boldsymbol{y}(\theta)) = I_m(0 \mid \boldsymbol{y}(0)) \triangleq I_m \text{ and } I_m(\theta \,|\, \boldsymbol{y}_*(\theta)) = I_m(0 \,|\, \boldsymbol{y}_*(0)) \triangleq I^*_m, \tag{48}$$

and the suboptimal and optimal quantizations

$$\boldsymbol{y}(\theta) = \theta \cdot \mathbf{1} + \boldsymbol{y}(0) \quad \text{and} \quad \boldsymbol{y}_*(\theta) = \theta \cdot \mathbf{1} + \boldsymbol{y}_*(0) \quad (\mathbf{1} = (1, \dots, 1) \in R^m). \tag{49}$$

In these families, for every quantization $\boldsymbol{y}$ we have

$$F(\theta \cdot \mathbf{1} + \boldsymbol{y}, \theta) = F(\boldsymbol{y}) \triangleq \boldsymbol{\lambda} = (\lambda_1, \dots, \lambda_{m-1}), \tag{50}$$

so that the quantization $\theta \cdot \mathbf{1} + \boldsymbol{y}$ consists of the $m-1$ fractiles of the parent distribution $F(x)$ of orders $\boldsymbol{\lambda}$. In particular, $\boldsymbol{y}(0)$ consists of the fractiles of $F(x)$ of the orders uniformly spaced on $(0, 1)$ (briefly, *uniform fractile orders*) defined by

$$\boldsymbol{\lambda}_0 = (1/m, 2/m, \dots, (m-1)/m). \tag{51}$$

We denote the vector of the fractile orders leading to the optimal partitions $\boldsymbol{y}_*(0)$ by

$$\boldsymbol{\lambda}_* \triangleq F(\boldsymbol{y}_*(0)). \tag{52}$$

The components λ_{*j} of $\boldsymbol{\lambda}_*$ thus represent the *optimal fractile orders*.

Let us now consider the statistical procedures in the model (1) based on the observations randomly quantized by the empirical fractiles of orders $\boldsymbol{\lambda}$,

$$\boldsymbol{y}_n = \left(y_{nj} \triangleq F_n^{-1}(\lambda_j) : 1 \le j \le m-1 \right). \tag{53}$$

Since in this case F_n tends to $F(x - \theta_0)$ where θ_0 is the true parameter, the fractiles (53) tend to

$$\theta_0 \cdot \mathbf{1} + (F^{-1}(\lambda_j) : 1 \le j \le m-1) = \left\langle \begin{array}{ll} \boldsymbol{y}(\theta_0) & \text{if } \boldsymbol{\lambda} = \boldsymbol{\lambda}_0 \\ \boldsymbol{y}_*(\theta_0) & \text{if } \boldsymbol{\lambda} = \boldsymbol{\lambda}_*. \end{array} \right.$$

The theorems used above to argue that the inference based on data quantized by the empirical fractiles of uniform orders is characterized by the inefficiency $\Delta(\theta_0 \,|\, \boldsymbol{y}(\theta_0)) = I(\theta_0) - I_m(\theta_0 \,|\, \boldsymbol{y}(\theta_0))$ lead in the location models to a stronger conclusion. Namely, that if the data are quantized by the empirical quantiles of orders $\boldsymbol{\lambda}$, then this inefficiency is $\Delta(\theta_0|\boldsymbol{y}) = I(\theta_0) - I_m(\theta_0|\boldsymbol{y})$ for $\boldsymbol{y} = (F^{-1}(\lambda_j) : 1 \le j \le m-1)$. Thus, and this is one of the clues of this paper, the minimal inefficiency $\Delta(\theta_0 \,|\, \boldsymbol{y}_*(\theta_0))$ can be practically achieved in the location models by using the random quantization by the empirical fractiles $\boldsymbol{\lambda}_*$ of optimal orders which does not depend on the true location θ_0. This interesting result can be extended to some other (but not all) models with equivariant structure.

We shall therefore be interested not only in $\boldsymbol{y}(0)$ (or, equivalently, $\boldsymbol{\lambda}_0$) but also in $\boldsymbol{\lambda}_*$ (equivalently, $\boldsymbol{y}_*(0)$) which are practically applicable in the statistical inference about location models. Next, we use analytic as well as numerical methods to evaluate these characteristics (and also the corresponding values of I_m and I_m^* defined in (48)) in the common location families: normal, logistic, Cauchy and doubly exponential.

We start with the formula

$$I_m(0|\mathbf{y}) = \sum_{j=1}^{m} \frac{[f(y_j) - f(y_{j-1})]^2}{F(y_j) - F(y_{j-1})} \tag{54}$$

easily deducible in the location models for every $\mathbf{y} \in R^{m-1}$ under consideration from (12) and (16). This formula provides the desired values of $I_m = I_m(0 \mid \mathbf{y}(0))$ for $\mathbf{y}(0)$, and is obtained by the application of the quantile function $F^{-1}(\lambda)$ to the coordinates of $\lambda - 0$. Namely,

$$I_m = m \sum_{j=1}^{m} [f(F^{-1}(j/m)) - f(F^{-1}((j-1)/m))]^2. \tag{55}$$

Evaluation of $I_m^* = I_m(0 \mid \mathbf{y}_*(0))$ and $\mathbf{y}_*(0)$ means the maximization of the expression (54) over $\mathbf{y}$. This mathematical programming problem can in some cases be solved analytically so that I_m^* and $\mathbf{y}_*(0)$ can be expressed in a closed form. However, in most cases it will be necessary to solve this problem numerically with the aid of computers. Next we discuss some known and some new results in this area.

The following theorem sharpens Lemma 2.1 in Cheng (1973) and partly also Proposition 1 in Pötzelberger and Felsenstein (1993) and the results in Section 3 of Tsaridis et al. (1997).

Theorem 3. *Let* $\varphi(\lambda) = f(F^{-1}(\lambda))$ *be twice differentiable on* $(0,1)$. *Then* $\boldsymbol{\lambda} = (\lambda_1, \ldots, \lambda_{m-1})$ *with* $0 < \lambda_1 < \cdots < \lambda_{m-1} < 1$ *is the vector of optimal fractile orders only if it satisfies the equations*

$$\frac{\varphi(\lambda_{j+1}) - \varphi(\lambda_j)}{\lambda_{j+1} - \lambda_j} + \frac{\varphi(\lambda_j) - \varphi(\lambda_{j-1})}{\lambda_j - \lambda_{j-1}} = 2\dot{\varphi}(\lambda_j), \quad 1 \le j \le m-1, \tag{56}$$

where $\dot{\varphi}(\lambda) = \mathrm{d}\varphi(\lambda)/\mathrm{d}\lambda$ *and* $\lambda_0 = 0$, $\lambda_m = 1$. *If* $\varphi(\lambda)$ *is strictly convex or strictly concave on* $(0,1)$ *with*

$$\varphi(0) \triangleq \lim_{\lambda \downarrow 0} \varphi(\lambda) = \varphi(1) \triangleq \lim_{\lambda \uparrow 1} \varphi(\lambda) = 0 \tag{57}$$

then (56) has a unique solution $\boldsymbol{\lambda}$ *of the above considered properties and this solution represents the optimal fractile orders.*

Proof. By substituting $y_j = F^{-1}(\lambda_j)$, $0 \le j \le m$, in (54) and taking the partial derivatives of (54) with respect to λ_j, $1 \le j \le m-1$, one obtains that (56) is the stationarity condition which is necessary for the optimality of the above considered λ. Let us suppose that $\varphi(\lambda)$ is strictly concave on $(0,1)$ (otherwise we can replace it by the strictly concave $-\varphi(\lambda)$). Under the stated assumptions, $\varphi(\lambda)$ can be extended into a continuously differentiable function on $[0,1]$ with the right- and left-hand derivatives $\dot{\varphi}(0) \in (-\infty, \infty]$ and $\dot{\varphi}(1) \in [-\infty, \infty)$. The functions

$$\dot{\varphi}(\lambda) \quad \text{and} \quad \frac{\varphi(\lambda) - \varphi(\lambda_0)}{\lambda - \lambda_0}, \quad \lambda_0 \in [0, 1],$$

of variable λ are decreasing on $[0, 1]$ and $[0, 1] - \{\lambda_0\}$, respectively. By the mean value theorem, this implies that for all sufficiently small $\lambda_j \in (\lambda_{j-1}, 1)$ there exists an increasing function $\psi_j(\lambda_j) > \lambda_j$ with a range $(a_j, 1]$, where $\lambda_{j-1} < a_j < 1$, such that the equality of (56) holds if and only if $\lambda_{j+1} = \psi_j(\lambda_j)$. Let us define for sufficiently small $\lambda_1 \in (0, 1)$ the numbers $\lambda_2 = \psi_1(\lambda_1)$, $\lambda_3 = \psi_2(\lambda_2) = \psi_2(\psi_1(\lambda_1)), \ldots, \lambda_m = \psi_{m-1}(\lambda_{m-1}) = \psi_{m-1}(\psi_{m-2}(\ldots(\psi_1(\lambda_1))\ldots))$. Then there exists a unique $\lambda_1 \in (0, 1)$ leading to $\lambda_m = 1$. This proves the uniqueness of the solution of (56). The fact that this solution maximizes (54) follows from Lemma 2.1 in Cheng (1973).

The following statement follows from the arguments on pp. 130–131 of Pötzelberger and Felsenstein (1993).

Theorem 4. *If* $c = \int (f(x)[(\ln f(x))'']^2)^{1/3}\, \mathrm{d}x$ *is finite then the distribution function*

$$G(x) = c^{-1} \int_{-\infty}^{x} (f(t)[(\ln f(t))'']^2)^{1/3}\, \mathrm{d}t, \quad x \in R, \tag{58}$$

and the coordinates of the optimal partitions $\boldsymbol{y}_*(0)$ *are mutually related by the asymptotic formula*

$$y_{*j}(0) \to G^{-1}(\lambda) \quad \textit{for } \frac{j}{m} \to \lambda \in (0, 1), m \to \infty. \tag{59}$$

In other words, if m is large then $\boldsymbol{y}_*(0)$ *is close to the uniform order fractiles of the distribution* $G(x)$ *or, equivalently,*

$$G_m(x) \triangleq \frac{1}{m-1} \sum_{j=1}^{m-1} I_{[y_{*j},\, \infty)}(x) \to G(x) \quad \textit{as } m \to \infty,\ x \in R.$$

Example 1 (Logistic model). A nice illustration for applicability of Theorems 3 and 4 is provided by the logistic location family, where

$$F(x) = \frac{1}{1 + e^{-x}}, \quad F^{-1}(\lambda) = \ln \frac{\lambda}{1 - \lambda},$$

$$f(x) = \frac{e^{-x}}{(1 - e^{-x})^2} \quad \text{and} \quad \varphi(\lambda) = \lambda(1 - \lambda).$$

For this model, the optimal and sub-optimal quantizations can be assessed analytically.

As easy to verify, $I(0) = 1/3$. Further, the assumptions of Theorem 3 hold and the optimal fractile orders are the $m - 1$ uniform fractile orders $\boldsymbol{\lambda}_0$, which are the unique solution of the equations (56). Therefore

$$I_m^* = I_m = m\sum_{j=1}^{m}[\varphi(j/m) - \varphi((j-1)/m)]^2 \quad \text{(cf. (55))}$$

$$= m\sum_{j=1}^{m}\left[\frac{j}{m}\cdot\frac{m-j}{m} - \frac{j-1}{m}\cdot\frac{m-j+1}{m}\right]^2$$

$$= \frac{1}{m^3}\sum_{j=1}^{m}[m-2j+1]^2 = \frac{1}{3}\left(1-\frac{1}{m^2}\right), \tag{60}$$

where we used the formulas

$$\sum_{i=1}^{m} i = \frac{m(m+1)}{2} \quad \text{and} \quad \sum_{i=1}^{m} i^2 = \frac{m(m+1)(2m+1)}{6}.$$

It follows from (60) that the asymptotic relation (33) can now be precised nonasymptotically as follows: $I(0) - I_m = 1/(3m^2)$. Also the relation (25) holds nonasymptotically for all $m > 1$, with $O(1/m^2)$ replaced by $1/(3m^2)$. The distribution function $G(x)$ of Theorem 4 coincides in this case with $F(x)$. Therefore relation (59) of Theorem 4 holds in the stronger nonlimit form leading to the optimal quantization defined by

$$y_{*j}(0) = y_{0j}(0) = G^{-1}(j/m) = \ln\frac{j}{m-j}, \quad 1 \le j \le m-1.$$

Relative asymptotic inefficiencies (in %), $\eta_m = 100[I(0) - I_m]/I(0) = 100/m^2$ are presented in Table 1. We see that the relative asymptotic inefficiencies can be characterized as small for $m \ge 10$, and negligible when $m \ge 20$.

Example 2 (Standard normal model). Let $f(x)$ and $F(x)$ be the standard normal density and distribution function. Then $F^{-1}(\lambda)$, and therefore $\varphi(\lambda)$, cannot be presented in a closed form. However, it is easy to verify that $\dot\varphi(\lambda) = -2F^{-1}(\lambda)$ so that $\ddot\varphi(\lambda) = -2/f(F^{-1}(\lambda))$ is negative on $(0,1)$. Therefore Theorem 3 is applicable. The evaluation of the coordinates $\boldsymbol{y}(0)$, and of the corresponding values of I_m, is only a minor numerical problem – to this end it suffices to use a closed formula approximating the quantile function $F^{-1}(\lambda)$ with a guaranteed accuracy in the domain of interest $\lambda = j/m$. A more serious problem is the evaluation of the optimal partitions $\boldsymbol{y}_*(0)$, leading easily to the corresponding optimal fractile orders $\boldsymbol{\lambda}_*$ and Fisher informations I_m^*. By Theorem 3, to this end it suffices to find the solution of equations (56). This can-

Table 1. Relative asymptotic inefficiencies η_m for the logistic model (in %)

m	2	3	4	5	10	15	20
η_m	25	11.11	6.25	4	2.78	0.44	0.25

not be done analytically, as in the previous example, but it is possible to do it numerically, using the method presented in the proof of Theorem 3. Indeed, the functions $\psi_1, \psi_2, \ldots, \psi_{m-1}$ can be evaluated numerically at arbitrary arguments so that it suffices to iterate λ_1 until $\lambda_m = 1$ is achieved with a desired accuracy. Nevertheless, in the present paper we used a different approach, namely the gradient method for maximization of (54), starting with the initial estimates

$$\tilde{\boldsymbol{y}}(0) = (\tilde{y}_j(0) = G^{-1}(j/m) : 1 \leq j \leq m-1)$$

obtained by means of $G(x)$ specified in Theorem 4. This distribution function can be explicitly evaluated for the standard normal $F(x)$, namely (cf. p. 144 in Pötzelberger and Felsenstein (1993)),

$$G(x) = F(x/\sqrt{3}), \quad x \in R.$$

Therefore

$$\tilde{\boldsymbol{y}}(0) = \sqrt{3}\boldsymbol{y}(0).$$

We have used Powell's quadratically convergent method of maximization proposed by Powell (1964) to obtain the optimal partition $\boldsymbol{y}_*$ that maximizes (46). The basic idea of this method is that, starting at a point $\tilde{\boldsymbol{y}}$ in the $(m-1)$-dimensional space, we proceed in some vector direction, so that a given function of $m-1$ variables can be minimized along the line in this direction by the one-dimensional method given in Brent (1973). The tolerance of Brent's method was fixed at 10^{-12}. As an initial value we used $\tilde{\boldsymbol{y}} = \tilde{\boldsymbol{y}}(0)$. We tested the final iterations $\boldsymbol{y}_*$ by comparing them with what was obtained for the initial values $\tilde{\boldsymbol{y}} = \tilde{\boldsymbol{y}}(0)$. To calculate $F(x), F^{-1}(x)$ and $f(x)$ we used the *dcdflib* library of C functions written by Brown et al. (1994).

In Table 2 we present the coordinates of $\boldsymbol{y}(0), \tilde{\boldsymbol{y}}(0), \boldsymbol{y}_*(0)$ and the corresponding Fisher informations I_m, $\tilde{I}_m, I_m^*$ for $m = 2, 3, 4, 5, 10, 15$ and 20. Since the coordinates of $\boldsymbol{y}(0)$ and $\boldsymbol{y}_*(0)$ are symmetric around 0 (containing 0 when m is even), we present only the positive ones. As can easily be verified, in the standard normal location model under consideration $I(0) = 1$. Table 2 presents also the relative asymptotic inefficiencies (in %),

$$\eta_m = \frac{I(0) - I_m}{I(0)} 100, \;\; \tilde{\eta}_m = \frac{I(0) - \tilde{I}_m}{I(0)} 100 \quad \text{and} \quad \eta_m^* = \frac{I(0) - I_m^*}{I(0)} 100. \quad (61)$$

All presented values (except the percents) are rounded off to three decimals.

Note that the values of I_m for $m = 2, 3$ and 4 were first published in Pötzelberger and Felsenstein (1993), and those for $m = 5, 10, 15$ and 20 have been published in Menéndez et al. (2001b).

In Figure 1 we present the values of the Fisher information measures $I_m, \tilde{I}_m, I_m^*$ and the corresponding relative asymptotic inefficiencies for the uniform and optimal partitions, η_m and η_m^*, assessed for all sizes m between $m = 1$ and $m = 30$. As can be seen from Figure 1 and Table 2, the relative asymptotic inefficiencies of the three quantizations coincide for $m = 2$, but they differ when $m \geq 3$. For $m = 10$, the relative asymptotic inefficiency of the uniform

Table 2. Fisher information measures $I_m, \tilde{I}_m, I_m^*$ in the normal model, and the corresponding relative asymptotic inefficiencies for the uniform and optimal partitions, η_m and η_m^* (in %), evaluated for selected partition sizes m between $m = 2$ and $m = 20$

m	$\boldsymbol{y}(0)$	I_m	η_m	$\tilde{\boldsymbol{y}}(0)$	$\tilde{I}_m$	$\tilde{\eta}_m$	$\boldsymbol{y}_*(0)$	I_m^*	η_m^*
2	0.000	0.637	36.34	0.000	0.637	36.34	0.000	0.637	36.34
3	0.431	0.793	20.68	0.714	0.801	19.92	0.612	0.809	19.02
4	0.000 0.674	0.861	13.94	0.000 1.168	0.876	12.44	0.000 0.982	0.883	11.75
5	0.281 0.841	0.897	10.30	0.476 1.458	0.914	8.59	0.381 1.244	0.920	7.99
10	0.000 0.320 0.641 0.961 1.282	0.959	4.06	0.000 0.555 1.110 1.665 2.220	0.973	2.71	0.000 0.405 0.834 1.325 1.968	0.977	2.29
15	0.115 0.346 0.577 0.808 1.039 1.270 1.501	0.976	2.38	0.200 0.600 1.000 1.400 1.800 2.200 2.600	0.986	1.38	0.137 0.414 0.703 1.013 1.360 1.776 2.344	0.989	1.07
20	0.000 0.183 0.366 0.548 0.731 0.914 1.097 1.279 1.462 1.645	0.985	1.63	0.000 0.317 0.633 0.950 1.266 1.583 1.899 2.216 2.532 2.849	0.992	0.86	0.000 0.208 0.420 0.638 0.866 1.111 1.381 1.690 2.068 2.593	0.994	0.62

quantization $\boldsymbol{y}(0)$ doubles that of the optimal one. We see that for all three partitions $\boldsymbol{y}(0)$, $\tilde{\boldsymbol{y}}(0)$ and $\boldsymbol{y}^*(0)$ the relative asymptotic inefficiencies are reasonably small for $m \geq 10$, and they are practically negligible for $m \geq 20$.

Example 3 (Standard Cauchy model). The standard Cauchy location family is specified by

$$f(x) = \frac{1}{\pi(1+x^2)} \quad \text{and} \quad F(x) = \frac{1}{2} + \frac{1}{\pi} \operatorname{arctg} x, \quad x \in R.$$

Here

$$F^{-1}(\lambda) = \operatorname{tg}[\pi(\lambda - 1/2)] \quad \text{and} \quad \varphi(\lambda) = \frac{1}{\pi(1 + \operatorname{tg}^2[\pi(\lambda - 1/2)])}, \quad \lambda \in (0, 1).$$

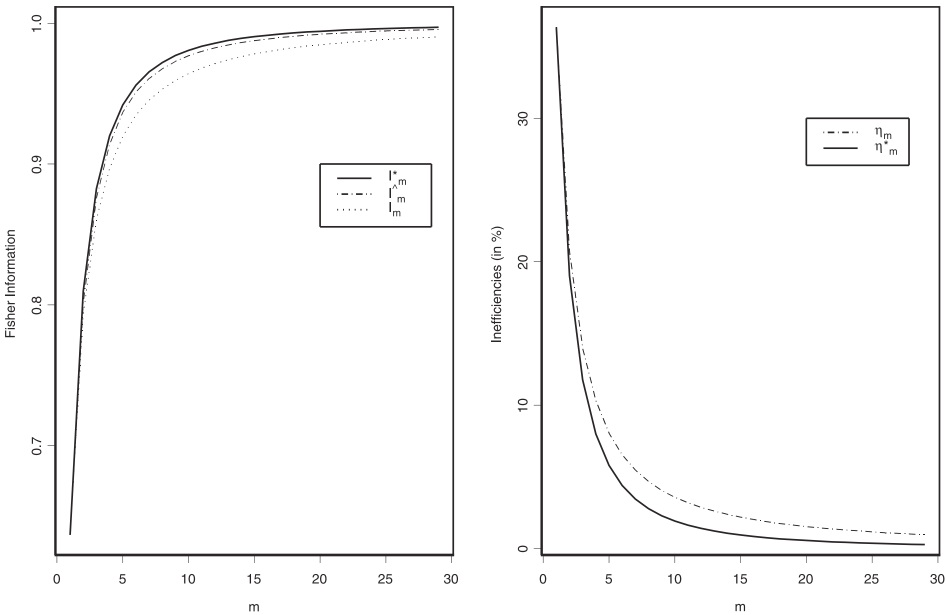

Fig. 1. Fisher information measures and relative asymptotic inefficiencies for the uniform, asymptotic and optimal partitions, evaluated for $m = 1, \ldots, 30$ in the normal location model.

Further, $I(0) = 1/2$ and the norming constant of Theorem 4 is

$$c = \int \frac{(x^2-1)^{2/3}}{(1+x^2)^{5/3}}\,\mathrm{d}x = 2\int_0^{\pi/2} |\sin^2 t - \cos^2 t|^{2/3}\,\mathrm{d}t$$
$$= 2\int_0^{\pi/2} \cos^{2/3} t\,\mathrm{d}t.$$

Thus the G-function of Theorem 4 is of the form

$$G(x) = \frac{1}{c}\int_{-\infty}^{x} \frac{(y^2-1)^{2/3}}{(1+y^2)^{5/3}}\,\mathrm{d}y = \frac{1}{2}\left[1 + \frac{\int_0^{\mathrm{artg}\,x} \cos^{2/3} t\,\mathrm{d}t}{\int_0^{\pi/2} \cos^{2/3} t\,\mathrm{d}t}\right].$$

Values of the quantile function $G^{-1}(\lambda)$ can be obtained by a numerical solution of the equation $G(x) = \lambda$. In this manner we obtained the coordinates of $\tilde{\mathbf{y}}(0) = (G^{-1}(j/m) : 1 \leq j \leq m-1)$ used in the calculations described below. Finally,

$$\frac{\dot{f}(x)}{f(x)} = -\frac{2x}{1+x^2}$$

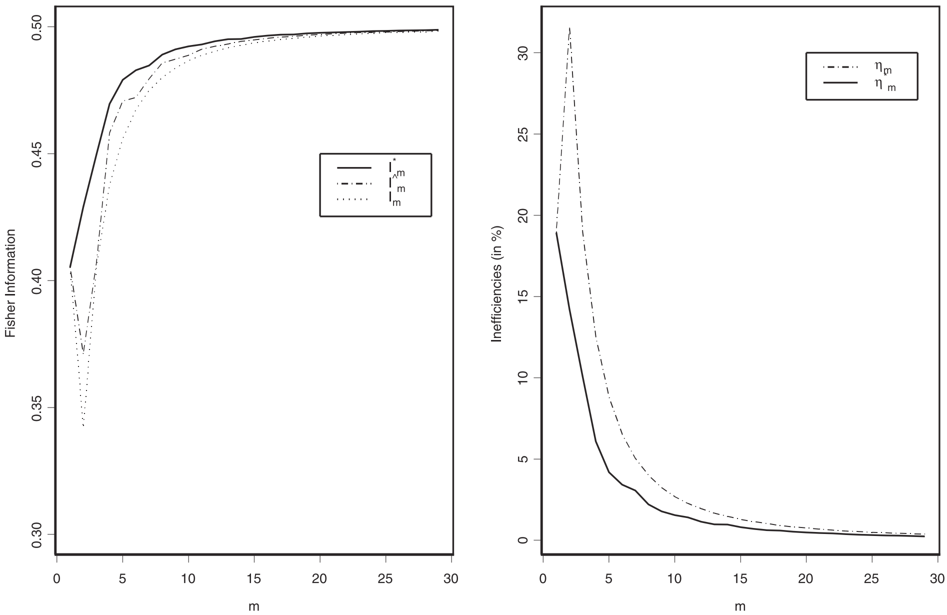

Fig. 2. Fisher information measures and relative asymptotic inefficiencies for the uniform, asymptotic and optimal partitions, evaluated for $m = 1, \ldots, 30$ in the Cauchy location model.

so that for $\alpha = \pi(\lambda - 1/2)$

$$\frac{\mathrm{d}}{\mathrm{d}\lambda} f(F^{-1}(\lambda)) = -\frac{2\,\mathrm{tg}\,\alpha}{1 + \mathrm{tg}^2\,\alpha} = -2\sin\alpha\cos\alpha = \sin[\pi(2\lambda - 1)]$$

which is not monotone on $(0, 1)$. Therefore $\varphi(\lambda)$ is neither convex nor concave on $(0, 1)$. This excludes similar application of Theorem 3 as above, and also leads to the strange non-monotone behaviour of the Fisher information I_m and $\tilde{I}_m$ for $m = 2, 3$ and 4, visible in Figure 2 and Table 3 below. This behavior was first observed without explanation on p. 141 in Pötzelberger and Felsenstein (1993). Even stranger appears to be that there are two different solutions of the system (56) for $m = 4$. Since the application of the algorithm based on Theorem 3 and outlined in the previous example is impossible, we had no choice but to apply the same gradient method for evaluation of the optimal partitions $\boldsymbol{y}_*(0)$ as in that example, with the above specified initial estimates $\tilde{\boldsymbol{y}}(0)$ of $\boldsymbol{y}_*(0)$. The same characteristics as considered in the previous example are for the Cauchy family presented in Table 3.

As it can be seen in Figure 2, the Fisher informations $I_m, \tilde{I}_m$ and I_m^* differ for the three quantizations when m is small. The differences are reasonably small for $m \geq 10$ and they are negligible when $m \geq 20$.

Example 4 (Double exponential model). We observe an even stranger behaviour of the informations I_m and I_m^* than in the previous example in the double exponential location family where

Table 3. The same characteristics as in Table 2 for the Cauchy model. For $m = 4$ the table presents all 3 coordinates of $\boldsymbol{y}(0)$, $\tilde{\boldsymbol{y}}(0)$ and $\boldsymbol{y}_*(0)$ because those of $\boldsymbol{y}_*(0)$ are in this case not symmetric about 0. In fact, there are two optimal partitions $\boldsymbol{y}_*(0)$ – the second is obtained by multiplying the coordinates of the first one by -1

m	$\boldsymbol{y}(0)$	I_m	η_m	$\tilde{\boldsymbol{y}}(0)$	$\tilde{I}_m$	$\tilde{\eta}_m$	$\boldsymbol{y}_*(0)$	I_m^*	η_m^*
2	0.000	0.405	18.94	0.000	0.405	18.94	0.000	0.405	18.94
3	0.577	0.342	31.60	0.491	0.371	25.82	0.179	0.429	14.22
4	−1.000 0.000 1.000	0.405	18.94	−1.188 0.000 1.888	0.407	18.64	−1.157 0.239 3.285	0.450	10.08
5	0.325 1.376	0.438	12.48	0.234 1.851	0.458	8.36	0.214 3.334	0.470	6.08
10	0.000 0.325 0.727 1.376 3.0777	0.484	3.24	0.000 0.234 0.706 1.851 4.306	0.487	2.56	0.000 0.199 0.444 2.250 5.022	0.491	1.78
15	0.105 0.105 0.325 0.577 0.900 1.376 2.246 4.705	0.493	1.46	0.075 0.075 0.234 0.491 0.996 1.851 3.137 6.593	0.494	1.16	0.075 0.075 0.265 0.497 1.803 2.682 4.272 8.818	0.495	0.96
20	0.000 0.158 0.325 0.510 0.727 1.000 1.376 1.963 3.078 6.314	0.496	0.82	0.000 0.113 0.234 0.396 0.706 1.188 1.851 2.733 4.306 8.851	0.497	0.67	0.000 0.116 0.240 0.385 0.578 1.658 2.317 3.297 5.147 10.521	0.497	0.53

$$f(x) = \frac{1}{2}e^{-|x|} \quad \text{for } x \in R \quad \text{and} \quad F(x) = \begin{cases} 1 - \frac{1}{2}e^{-x} & \text{for } x \geq 0 \\ \frac{1}{2}e^{x} & \text{for } x < 0. \end{cases}$$

Here $I(0) = 1$ and

$$F^{-1}(\lambda) = \begin{cases} \ln\dfrac{1}{2(1-\lambda)} & \text{for } \lambda \in [1/2, 1) \\ \ln(2\lambda) & \text{for } \lambda \in (0, 1/2). \end{cases}$$

Therefore

$$\varphi(\lambda) = \begin{cases} 1-\lambda & \text{for } \lambda \in [1/2, 1) \\ \lambda & \text{for } \lambda \in (0, 1/2). \end{cases}$$

We see that $\varphi(\lambda)$ is in this case concave on $(0,1)$, but not strictly concave as required by Theorem 3. Hence Theorem 3 is again not applicable in the same manner as in Examples 1 and 2. As observed already on p. 135 of Pötzelberger and Felsenstein (1993), there are infinitely many optimal vectors $\boldsymbol{y}_*(0)$ in this case – one of them is that with all coordinates equal to 0. Therefore $I_m^* = I(0) = 1$ for all $m > 1$. Moreover, $I_m = I(0)$ if m is even and $I_m = 1/m$ if m is odd.

However, these strange facts have a sound statistical explanation. Namely, for the partition of R into the intervals $(-\infty, 0]$, $(0, \infty)$ all minimum ϕ-disparity estimators of location $\theta_{2,n}^{\phi,\boldsymbol{y}}$ reduce to the sample median which is known to be efficient in the original unreduced doubly exponential model. Similarly for the empirical quantization of R by the sample median $\boldsymbol{y}_n$, all empirical ϕ-disparity estimators $\tilde{\theta}_{2,n}^{\phi}$ reduce to $\boldsymbol{y}_n$ (to see this, put $m = 2$ in (31) and apply the definition of $p_1(\theta|\boldsymbol{y}_n)$). Thus the conclusions of the theory do not contradict the common sense.

5 Discussion

Asymptotic theory of point estimation is usually focused on asymptotically efficient estimators. In Sections 2 and 4 it has been shown that this optimality property is usually lost when data is quantized to m cells. However, if m increases to infinity and the partition is selected in a convenient way, the optimality is achievable by the estimators using quantized data. This fact follows from the possibility to achieve a monotone convergence of the quantized Fisher information to its counterpart in the continuous models. The larger Fisher's amount of information is, the more efficient the estimators are. Therefore, the best partition should be selected by maximizing this information quantity.

Asymptotic theory of testing hypotheses usually concerns with maximization of asymptotic power in contiguous alternatives. As shown in Sections 3 and 4, asymptotic power of the tests based on normalized ϕ-disparities of the discrete empirical and theoretical cell probability vectors can be considered as an increasing function of Fisher's amount of information. Selection of the best partition is thus here the same problem as in the statistical estimation.

Examples in Section 4 clarify in a quantitative way the above recommendation about selection of partitions. To be more precise, we see from the right-hand columns of Tables 2 and 3, and from the explicit formulas in Examples 1 and 4, that the relative asymptotic inefficiencies achievable in the (empirically) quantized continuous models are practically negligible already for $m \geq 10$. The rate with which the inefficiency vanishes seems to be quadratic in the sense of (25). The only exception was the doubly exponential model, with the lack of regularity. There $O(1/m^2)$ is replaced by $1/m$ for m odd and by 0 for m even. From the middle columns of Tables 2 and 3 we see that the method of Theorem 4 provides practically acceptable approximations to the optimal partitions already for $m \geq 10$. From the left-hand columns one can see that the loss of efficiency due to the use of simplified universal quantization by the fractiles of equidistant orders is relatively small, especially for $m \geq 10$. Thus we can conclude that the computationally feasible and robust discrete methods of statistical inference are applicable in the continuous models without a noticeable loss of efficiency.

Acknowledgements. We thank the referee for his comments.

References

Basu A, Sarkar S (1994a) Minimum disparity estimation in the errors-invariables model. Statistics and Probability Letters 20:69–73

Basu A, Sarkar S (1994b) The trade-off between robustness and efficiency and the effect of model smoothing. Journal of Statistics, Computation and Simulation 50:173–185

Birch MW (1964) A new proof of the Pearson–Fisher theorem. Annals of Mathematical Statistics 35:817–824

Brent RP (1973) Algorithms for minimization without derivatives. Prentice Hall

Brown BW, Lovato J, Russell K (1994) Library of routines for cumulative distribution functions, inverses, and other parameters. http://www.stat.unipg.it/stat/dcdflib/

Cheng SW (1975) A unified approach to choosing optimum quantiles for the ABLEs. Journal of the American Statistical Association 70:155–159

Csiszár I (1963) Eine informationstheoretische ungleichung und ihre anwendung auf den beweis der ergodizität on markoffschen ketten. Publications of the Mathematical Institute of the Hungarian Academy of Sciences 8, Ser. A, 84–108

Csiszár I (1967) Information-type measured of difference of probability distributions and indirect observations. Studia Scientiarium Mathematica Hungarica 2:299–318

Ghurye SG, Johnson BR (1981) Discrete approximations to the information integral. The Canadian Journal of Statistics 9:27–37

Inglot T, Jurlewicz T, Ledwina T (1991) Asymptotics for multinomial goodness of fit tests for simple hypotheses. Theory of Probability and Applications 35:771–777

Liese F, Vajda I (1987) Convex statistical distances. Leipzig, Teubner

Lindsay BG (1994) Efficiency versus robustness: The case of minimum Hellinger distance and other methods. Annals of Statistics 22:1081–1114

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

Menéndez ML, Morales D, Pardo L, Vajda I (2001a) Minimum disparity estimators for discrete and continuous models. Applications of Mathematics (in print)

Menéndez ML, Morales D, Pardo L, Vajda I (2001b) Approximations to powers of ϕ-disparity goodness of fit tests. Communications in Statistics–Theory and Methods 30:105–134

Millar PW (1984) A general approach to the optimality of minimum distance estimators. Transactions of the American Mathematical Society 286:377–418

Morales D, Pardo L, Vajda I (1995) Asymptotic divergence of estimates of discrete distributions. Journal of Statistical Planning and Inference 48:347–369

Park Ch, Basu A, Basu S (1995) Robust minimum distance inference based on combined distances. Communications in Statistics – Simulation and Computation 24:653–673

Pötzelberger K, Felsenstein K (1993) On the Fisher information of discretized data. Journal of Statistics, Computation and Simulation 46:125–144

Powell MJD (1964) An efficient method for finding the minimum of a function of several variables without calculating derivatives. Computation Journal 7:155–162

Rao CR (1961) Asymptotic efficiency and limiting information. Proceedings of the Fourth Berkeley Symposium on Mathematical Statistics and Probability, Vol. 1:531–545. Berkeley, CA, Berkeley University Press

Read RC, Cressie NAC (1988) Goodness-of-fit statistics for discrete multivariate data. New York, Springer Verlag

Serfling RJ (1980) Approximation theorems of mathematical statistics. Wiley, New York

Tsaridis Ch, Zografos K, 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 (1973) χ^{α}-divergence and generalized Fisher information. Trans. of the Sixth Prague Conference on Information Theory, Statistical Decision Functions and Random Processes. Prague, Academia, pp. 223–234

Vajda I (1989) Theory of statistical inference and information. Boston, Kluwer Academic Publishers

Vajda I (1995) Conditions equivalent to consistency of approximate MLE's for stochastic processes. Stochastic Processes and their Applications 56:35–56

Vajda I (2001) On the convergence of information contained in quantized observations. IEEE Transactions on Information Theory (in print)

Vajda I, Janžura M (1997) On asymptotically optimal estimates for general observations. Stochastic Processes and their Applications 72:27–45

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