Ann. Inst. Statist. Math.
Vol. 41, No. 3, 451-460 (1989)

LIMITING PROPERTIES OF SOME MEASURES OF INFORMATION

K. ZOGRAFOS, K. FERENTINOS AND T. PAPAIOANNOU

Probability-Statistics and Operational Research Unit, Department of Mathematics, University of Ioannina, Greece

(Received September 28, 1987; revised July 20, 1988)

Abstract. In this paper we investigate the limiting behaviour of the measures of information due to Csiszár, Rényi and Fisher. Conditions for convergence of measures of information and for convergence of Radon-Nikodym derivatives are obtained. Our results extend the results of Kullback (1959, *Information Theory and Statistics*, Wiley, New York) and Kirmani (1971, *Ann. Inst. Statist. Math.*, **23**, 157-162).

Key words and phrases: Convergence of information measures, Csiszár information, ϕ-divergence, Rényi information, Fisher information.

1. Introduction

The study of limiting properties of sequences of distributions is of fundamental importance in probability and mathematical statistics. In statistical information theory, the limiting behaviour of measures of information has been investigated by Kullback (1959) and Kirmani (1971). These authors have studied only the measures of information due to Kullback-Leibler and Matusita, respectively. Related are the papers by Linnik (1959), Rényi (1961), Brown (1982) and Barron (1986) and Chapter 8 of a recent book by Liese and Vajda (1987). No results are available concerning Fisher's fundamental measure of statistical information or the other measures of information proposed by Rényi (1961) and Csiszár (1963).

The aim of the present paper is to investigate the limiting properties of the latter measures of information. The main result states that if a sequence of generalized probability density functions (gpdf's) converges uniformly to a gpdf, then the corresponding measures due to Csiszár (ϕ-divergence) and Rényi (information gain of order a) converge to their minimum value; conversely, if the measures converge to their minimum value, then the gpdf's converge in distribution. Thus we obtain a new criterion of convergence in distribution based on the Csiszár and Rényi measures of information which can be used to prove limit theorems in probability and statistics.

2. Main results

Consider the measurable transformations $T_N(x)$, $N = 1, 2, \ldots$, of the probability spaces $(\mathscr{X}, \mathscr{A}, \mu_i)$ onto the probability spaces $(\mathscr{Y}, \mathscr{F}, \upsilon_i^{(N)})$, where $\upsilon_i^{(N)}(G) = \mu_i(T_N^{-1}(G))$, $T_N^{-1}(G) = \{x \in \mathscr{X}: T_N(x) \in G\}$, $G \in \mathscr{F}$ and $i = 1$, 2. Let also $T(x)$ be another measurable transformation of $(\mathscr{X}, \mathscr{A}, \mu_i)$ onto $(\mathscr{Y}, \mathscr{F}, \upsilon_i)$. $T_N(x)$ and $T(x)$ are statistics and N may be the sample size. Suppose that the probability measures $\upsilon_i^{(N)}$, υ_i, $i = 1, 2$ are dominated by a σ-finite measure λ and let $g_i^{(N)} = d\upsilon_i^{(N)}/d\lambda$, $g_i = d\upsilon_i/d\lambda$, $i = 1, 2$ be the corresponding Radon-Nikodym derivatives (or gpdf's).

The Csiszár and Rényi (of order a) measures of information in υ_2 about υ_1 are defined, respectively, as

$$I^C(\upsilon_1, \upsilon_2) = \int_{\mathscr{Y}} g_2 \phi(g_1/g_2) d\lambda ,$$

$$I^R(\upsilon_1, \upsilon_2) = \frac{1}{\alpha - 1} \log \int_{\mathscr{Y}} g_1^{\alpha} g_2^{1-\alpha} d\lambda, \qquad \alpha > 0 \quad (\alpha \neq 1) ,$$

where ϕ is a real valued convex function on $[0, \infty)$ (cf. Rényi (1961) and Csiszár (1963)). The measures $I^C(\mu_1, \mu_2)$, $I^R(\mu_1, \mu_2)$, $I^C(\upsilon_1^{(N)}, \upsilon_2^{(N)})$ and $I^R(\upsilon_1^{(N)}, \upsilon_2^{(N)})$ are defined analogously.

2.1 *Csiszár's measure*

THEOREM 2.1. *If the $T_N(x)$, $N = 1, 2, \ldots$ are such that*

$$g_i^{(N)} \to g_i, \qquad i = 1, 2$$

in the mean, then

$$I^C(\mu_1, \mu_2) \geq \limsup_{N \to \infty} I^C(\upsilon_1^{(N)}, \upsilon_2^{(N)}) \geq \liminf_{N \to \infty} I^C(\upsilon_1^{(N)}, \upsilon_2^{(N)}) \geq I^C(\upsilon_1, \upsilon_2) .$$

PROOF. (a) Let $e = \{G_1, G_2, \ldots, G_k\}$ be a finite partition of $\mathscr{Y}$ into pairwise disjoint sets such that $G_i \in \mathscr{F}$ and $UG_i = \mathscr{Y}$, $i = 1, 2, \ldots, k$, then (Perez (1968) and Vajda (1972))

$$I^C(\upsilon_1^{(N)}, \upsilon_2^{(N)}) = \sup_{e \in H} \sum_{i=1}^{k} \upsilon_2^{(N)}(G_i) \phi\left(\frac{\upsilon_1^{(N)}(G_i)}{\upsilon_2^{(N)}(G_i)} \right),$$

where H is the class of all finite partitions of $\mathscr{Y}$ as indicated above.

Therefore

$$I^C(\upsilon_1^{(N)}, \upsilon_2^{(N)}) \geq \sum_{i=1}^{k} \upsilon_2^{(N)}(G_i)\phi\left(\frac{\upsilon_1^{(N)}(G_i)}{\upsilon_2^{(N)}(G_i)}\right).$$

Accordingly,

$$\liminf_{N\to\infty} I^C(\upsilon_1^{(N)}, \upsilon_2^{(N)}) \geq \liminf_{N\to\infty} \sum_{i=1}^{k} \upsilon_2^{(N)}(G_i)\phi\left(\frac{\upsilon_1^{(N)}(G_i)}{\upsilon_2^{(N)}(G_i)}\right). \tag{2.1}$$

Since $g_i^{(N)} \to g_i$ in the mean is equivalent to

$$\upsilon_i^{(N)}(G) \to \upsilon_i(G), \tag{2.2}$$

uniformly in $G \in \mathscr{F}$, $i = 1, 2$ (cf. Loéve (1963), p. 140, Problem 16 and Burrill (1972), p. 176, Theorem 9-4A) and since ϕ is continuous on $[0, \infty)$, relation (2.1) gives

$$\liminf_{N\to\infty} I^C(\upsilon_1^{(N)}, \upsilon_2^{(N)}) \geq \sum_{i=1}^{k} \upsilon_2(G_i)\phi\left(\frac{\upsilon_1(G_i)}{\upsilon_2(G_i)}\right). \tag{2.3}$$

Hence

$$\liminf_{N\to\infty} I^C(\upsilon_1^{(N)}, \upsilon_2^{(N)}) \geq I^C(\upsilon_1, \upsilon_2), \tag{2.4}$$

because the right-hand side of (2.4) is the supremum of the right-hand expressions of (2.3) over all such partitions of $\mathscr{Y}$. But Csiszár's measure of information satisfies the maximal information property (cf. Csiszár (1963)), that is,

$$I^C(\mu_1, \mu_2) \geq I^C(\upsilon_1^{(N)}, \upsilon_2^{(N)}).$$

Thus we have

$$I^C(\mu_1, \mu_2) \geq \limsup_{N\to\infty} I^C(\upsilon_1^{(N)}, \upsilon_2^{(N)}). \tag{2.5}$$

Inequalities (2.4) and (2.5) establish the theorem.

The proof of the theorem did not require the "uniform" part in the condition $\upsilon_i^{(N)}(G) \to \upsilon_i(G)$ uniformly in $G \in \mathscr{F}$, $i = 1, 2$. The theorem was stated with the stronger and equivalent condition $g_i^{(N)} \to g_i$, $i = 1, 2$ in the mean since this condition is easier to verify than the absolutely essential one.

If the limit of the sequence of information measures $I^C(\upsilon_1^{(N)}, \upsilon_2^{(N)})$ exists, then by Theorem 2.1, $I^C(\mu_1, \mu_2)$ and $I^C(\upsilon_1, \upsilon_2)$ are upper and lower bounds of this sequence.

COROLLARY 2.1. *If in addition to the conditions of Theorem* 2.1 $T(x)$ *is sufficient w.r.t.* $\{\mu_1, \mu_2\}$, *then*

$$\lim_{N \to \infty} I^C(\upsilon_1^{(N)}, \upsilon_2^{(N)}) = I^C(\upsilon_1, \upsilon_2) .$$

THEOREM 2.2. *Let*

$$\lim_{N \to \infty} \frac{g_i^{(N)}(y)}{g_i(y)} = 1 \quad [\lambda] ,$$

uniformly, $i = 1, 2$. *If*

$$\lim_{N \to \infty} \phi\left(\frac{g_1^{(N)}(y)}{g_2^{(N)}(y)} \right) = \phi\left(\frac{g_1(y)}{g_2(y)} \right) \quad [\lambda] ,$$

uniformly and $\int_{\mathscr{Y}} g_2 |\phi(g_1/g_2)| d\lambda$ *is finite, then*

$$\lim_{N \to \infty} I^C(\upsilon_1^{(N)}, \upsilon_2^{(N)}) = I^C(\upsilon_1, \upsilon_2) .$$

PROOF. We have

$$\begin{aligned}
&|I^C(\upsilon_1, \upsilon_2) - I^C(\upsilon_1^{(N)}, \upsilon_2^{(N)})| \\
&\quad = \left| \int_{\mathscr{Y}} g_2 \phi(g_1/g_2) d\lambda - \int_{\mathscr{Y}} g_2^{(N)} \phi(g_1^{(N)}/g_2^{(N)}) d\lambda \right| \\
&\quad \leq \int_{\mathscr{Y}} |g_2 \phi(g_1/g_2) - g_2^{(N)} \phi(g_1/g_2)| d\lambda \\
&\qquad + \int_{\mathscr{Y}} |g_2^{(N)} \phi(g_1/g_2) - g_2^{(N)} \phi(g_1^{(N)}/g_2^{(N)})| d\lambda \\
&\quad \leq \int_{\mathscr{Y}} |1 - (g_2^{(N)}/g_2)| |\phi(g_1/g_2)| g_2 d\lambda \\
&\qquad + \int_{\mathscr{Y}} |\phi(g_1/g_2) - \phi(g_1^{(N)}/g_2^{(N)})| (g_2^{(N)}/g_2) g_2 d\lambda .
\end{aligned}$$

By the conditions of the theorem, for sufficiently large N and $\varepsilon > 0$, we also have

$$\left| 1 - \frac{g_2^{(N)}}{g_2} \right| < \varepsilon, \quad \left| \phi\left(\frac{g_1}{g_2} \right) - \phi\left(\frac{g_1^{(N)}}{g_2^{(N)}} \right) \right| < \varepsilon \quad \text{and} \quad \frac{g_2^{(N)}}{g_2} < 1 + \varepsilon .$$

Therefore

$$|I^C(v_1^{(N)}, v_2^{(N)}) - I^C(v_1, v_2)| \leq \varepsilon \int_{\mathscr{Y}} g_2 |\phi(g_1/g_2)| d\lambda + \varepsilon(1+\varepsilon) ,$$

which establishes the theorem.

The results of Theorems 2.1 and 2.2 translate into simple inequalities and limiting forms when the second argument (distribution) of the measure of information remains fixed while the first argument $g_1^{(N)} \to g_1$ in the mean.

Next we consider the case where the sequence $\{g_1^{(N)}\}$, $N = 1, 2, \ldots$ of gpdf's in the first argument of the measure of information converges uniformly to the gpdf of the second argument. The result almost holds both ways, i.e., if the sequence of gpdf's converges uniformly to a gpdf, then the corresponding measures of information converge to their minimum value; conversely, if the measures converge to their minimum value, then the gpdf's converge in distribution. Thus we obtain a new criterion of convergence in distribution which avoids the Fourier transforms and is based on the Csiszár measure of information.

THEOREM 2.3. *If*

$$\lim_{N \to \infty} \frac{g_1^{(N)}(y)}{g_1(y)} = 1 \qquad [\lambda] ,$$

uniformly, then

$$\lim_{N \to \infty} I^C(v_1^{(N)}, v_1) = \phi(1) ,$$

if ϕ is differentiable and $\phi'(1)$ is nonnegative and finite.

PROOF. A Taylor series expansion of $\phi(g_1^{(N)}(y)/g_1(y))$ around the point 1 yields

$$\phi\left(\frac{g_1^{(N)}(y)}{g_1(y)}\right) = \phi(1) + \left(\frac{g_1^{(N)}(y)}{g_1(y)} - 1\right)\phi'(u^{(N)}(y)) ,$$

where $u^{(N)}(y)$ lies between $g_1^{(N)}(y)/g_1(y)$ and 1. Then

$$I^C(v_1^{(N)}, v_1) - \phi(1) = \int_{\mathscr{Y}} \left(\frac{g_1^{(N)}(y)}{g_1(y)} - 1\right) g_1(y)\phi'(u^{(N)}(y)) d\lambda .$$

By using convergence of $g_1^{(N)}(y)$ to $g_1(y)$ for sufficiently large N and $\varepsilon > 0$, we have

$$\left| \frac{g_1^{(N)}(y)}{g_1(y)} - 1 \right| < \varepsilon .$$

Since ϕ is convex, $\phi'' \geq 0$ and hence ϕ' is an increasing function. Thus $\phi'(u^{(N)}(y))$ is either less than or equal to $\phi'(1)$ or less than or equal to $\phi'(g_1^{(N)}(y)/g_1(y)) \leq \phi'(1+\varepsilon)$ for sufficiently large N, for all $y[\lambda]$. In addition, $\phi(1)$ is the minimum value of $I^C(v_1^{(N)}, v_1)$, for all N. Thus

$$0 \leq I^C(v_1^{(N)}, v_1) - \phi(1) \leq \varepsilon\phi'(1+\varepsilon) ,$$

for sufficiently large N and therefore $\lim_{N\to\infty} I^C(v_1^{(N)}, v_1) = \phi(1)$.

For the next theorem we shall assume that the probability spaces are Euclidean.

THEOREM 2.4. *If*

$$\lim_{N\to\infty} I^C(v_1^{(N)}, v_1) = \phi(1) ,$$

then $v_1^{(N)}(G) \to v_1(G)$ *for* $G \in \mathscr{F}$ *or* $g_1^{(N)}(y) \to g_1(y)$ *in distribution.*

PROOF. Consider again a partitioning of $\mathscr{Y}$ into pairwise disjoint sets such that $G_i \in \mathscr{F}$ and $UG_i = \mathscr{Y}$, $i = 1, 2, \ldots, k$. Then

$$I^C(v_1^{(N)}, v_1) \geq \sum_{i=1}^{k} v_1(G_i)\phi\left(\frac{v_1^{(N)}(G_i)}{v_1(G_i)}\right) \geq \phi(1) .$$

By the condition of the theorem we have

$$\phi(1) \geq \lim_{N\to\infty} \sum_{i=1}^{k} v_1(G_i)\phi\left(\frac{v_1^{(N)}(G_i)}{v_1(G_i)}\right) \geq \phi(1)$$

or

$$\lim_{N\to\infty} \sum_{i=1}^{k} v_1(G_i)\phi\left(\frac{v_1^{(N)}(G_i)}{v_1(G_i)}\right) = \phi(1) .$$

Our purpose is to prove that

(2.6) $$\lim_{N\to\infty} \upsilon_1^{(N)}(G) = \upsilon_1(G) \quad \text{in} \quad G \in \mathscr{F}.$$

Suppose (2.6) is not true. Then there exists a subsequence $N_1 < N_2 < \cdots < N_s < \cdots$ of the integers and a probability measure υ on $(\mathscr{Y}, \mathscr{F})$ such that

(2.7) $$\lim_{s\to\infty} \upsilon_1^{(N_s)}(G) = \upsilon(G) \quad \text{for} \quad G \in \mathscr{F} \quad \text{and} \quad \upsilon_1(G) \neq \upsilon(G).$$

The continuity of ϕ gives

$$\lim_{s\to\infty} \sum_{i=1}^{k} \upsilon_1(G_i)\phi\left(\frac{\upsilon_1^{(N_s)}(G_i)}{\upsilon_1(G_i)}\right) = \sum_{i=1}^{k} \upsilon_1(G_i)\phi\left(\frac{\upsilon(G_i)}{\upsilon_1(G_i)}\right).$$

But $\left\{\sum_{i=1}^{k} \upsilon_1(G_i)\phi[\upsilon_1^{(N_s)}(G_i)/\upsilon_1(G_i)]\right\}$ is a subsequence of $\left\{\sum_{i=1}^{k} \upsilon_1(G_i)\cdot \phi[\upsilon_1^{(N)}(G_i)/\upsilon_1(G_i)]\right\}$ which converges to $\phi(1)$. Thus

$$\sum_{i=1}^{k} \upsilon_1(G_i)\phi\left(\frac{\upsilon(G_i)}{\upsilon_1(G_i)}\right) = \phi(1).$$

This is possible only if $\upsilon_1(G_i) = \upsilon(G_i)$ (cf. Csiszár (1963) or Perez (1968)) which contradicts (2.7). Thus Theorem 2.4 is proved. Convergence in distribution follows from Theorem 29.1 of Billingsley ((1979), p. 392) or Barron ((1986), p. 339).

The above theorem presents, under somewhat different conditions, a particular case of Proposition 8.6 of Liese and Vajda (1987). It is useful in establishing limit laws. We have used it to prove the ergodic theorem for homogeneous Markov chains (cf. Zografos (1987)).

2.2 *Rényi's measure*

Rényi's measure of information can be obtained from Csiszár's measure by taking $\phi(u) = \operatorname{sign}(\alpha - 1)u^\alpha$, $u \geq 0$, $\alpha > 0$ $(\alpha \neq 1)$. In fact, for this choice of ϕ we have $I^R = (\alpha - 1)^{-1} \log |I^C|$. Another choice of ϕ is $\phi(u) = (\alpha - 1)^{-1} \cdot (u^\alpha - \alpha u + \alpha - 1)$ with $I^R = (\alpha - 1)^{-1} \log [(\alpha - 1)I^C + 1]$. In both cases, I^R is a continuous and increasing function of I^C. In view of this relationship, the following results are immediate consequences of the corresponding results of Subsection 2.1.

THEOREM 2.5. *If $g_i^{(N)} \to g_i$, $i = 1, 2$ in the mean, then*

$$I^R(\mu_1, \mu_2) \geq \limsup_{N\to\infty} I^R(\upsilon_1^{(N)}, \upsilon_2^{(N)}) \geq \liminf_{N\to\infty} I^R(\upsilon_1^{(N)}, \upsilon_2^{(N)}) \geq I^R(\upsilon_1, \upsilon_2).$$

COROLLARY 2.2. *If in addition to the conditions of Theorem* 2.5 $T(x)$ *is sufficient w.r.t.* $\{\mu_1, \mu_2\}$, *then*

$$\lim_{N \to \infty} I^R(\upsilon_1^{(N)}, \upsilon_2^{(N)}) = I^R(\upsilon_1, \upsilon_2), \qquad \alpha > 0 \quad (\alpha \neq 1) .$$

THEOREM 2.6. *Under the conditions of Theorem* 2.2, *we have*

$$\lim_{N \to \infty} I^R(\upsilon_1^{(N)}, \upsilon_2^{(N)}) = I^R(\upsilon_1, \upsilon_2), \qquad \alpha > 0 \quad (\alpha \neq 1) .$$

THEOREM 2.7. *Under the conditions of Theorem* 2.3, *we have*

$$\lim_{N \to \infty} I^R(\upsilon_1^{(N)}, \upsilon_1) = 0, \qquad \alpha > 0 \quad (\alpha \neq 1) .$$

THEOREM 2.8. *If*

$$\lim_{N \to \infty} I^R(\upsilon_1^{(N)}, \upsilon_1) = 0, \qquad \alpha > 0 \quad (\alpha \neq 1) ,$$

then $\upsilon_1^{(N)}(G) \to \upsilon_1(G)$ *in* $G \in \mathscr{F}$ *or* $g_1^{(N)}(y) \to g_1(y)$ *in distribution.*

The discrete version of this theorem has been proved by Rényi ((1970), p. 597).

2.3 *Fisher's measure*

As is natural to expect, the limiting behaviour of Fisher's parametric measure of information is analogous. This may be seen on using Theorem 2.6 and the relationship $I^R(\theta) = (\alpha/2)I^F(\theta)$ between Rényi's and Fisher's measures of information (cf. Ferentinos and Papaioannou (1981), Theorem 4.1). Certain smoothing conditions must be imposed for the limits to go through.

Let

$$I_N^F(\theta) = \int_{\mathscr{Y}} \left[\frac{\partial}{\partial\theta} \log g^{(N)}(y, \theta) \right]^2 g^{(N)}(y, \theta) d\lambda$$

and

$$I^F(\theta) = \int_{\mathscr{Y}} \left[\frac{\partial}{\partial\theta} \log g(y, \theta) \right]^2 g(y, \theta) d\lambda$$

be Fisher's measures of information based on the gpdf's $g^{(N)}(y, \theta)$, $N = 1, 2, \ldots$, and $g(y, \theta)$, respectively. Let also

$$I_N^R(\theta) = \liminf_{\Delta\theta \to 0} \frac{1}{(\Delta\theta)^2} I^R(g^{(N)}(y, \theta), g^{(N)}(y, \theta + \Delta\theta))$$

and

$$I^R(\theta) = \liminf_{\Delta\theta \to 0} \frac{1}{(\Delta\theta)^2} I^R(g(y, \theta), g(y, \theta + \Delta\theta))$$

be Rényi's corresponding parametric measures. The parameter space Θ is an open subset of the real line.

THEOREM 2.9. *Suppose that the regularity conditions of Fisher's information measure are satisfied. Suppose moreover that for all gpdf's involved,* $\int_{\mathscr{Y}} |(\partial^2/\partial\theta^2)g(y,\theta)|\,d\lambda < \infty$ *for all* $\theta \in \Theta$ *and the third partial derivative of* $g(y,\theta)$ *with respect to* θ *exist for all* $\theta \in \Theta$ *and all* $y[\lambda]$. *If*

i) $\displaystyle\lim_{N\to\infty} \frac{g^{(N)}(y,\theta)}{g(y,\theta)} = 1$ *and*

$$\lim_{N\to\infty} \left(\frac{g^{(N)}(y,\theta)}{g^{(N)}(y, \theta + \Delta\theta)} \right)^{\alpha} = \left(\frac{g(y,\theta)}{g(y, \theta + \Delta\theta)} \right)^{\alpha} \quad \textit{and} \quad \alpha > 0 \quad (\alpha \neq 1),$$

uniformly for $y[\lambda]$ *and* $\theta, \theta + \Delta\theta \in \Theta$.

ii) $I^R(g^{(N)}(y,\theta), g^{(N)}(y, \theta + \Delta\theta))$ *is continuous with respect to* $\Delta\theta$ *for* $\theta \in \Theta$ *and* $N = 1, 2, \ldots$ *then*

$$\lim_{N\to\infty} I_N^F(\theta) = I^F(\theta) \qquad \textit{for} \quad \theta \in \Theta .$$

PROOF. By Theorem 4.1 of Ferentinos and Papaioannou (1981)

$$I_N^F(\theta) = \frac{2}{\alpha} I_N^R(\theta) = \frac{2}{\alpha} \liminf_{\Delta\theta \to 0} \frac{1}{(\Delta\theta)^2} I^R(g^{(N)}(y, \theta), g^{(N)}(y, \theta + \Delta\theta))$$

and

$$I^F(\theta) = \frac{2}{\alpha} I^R(\theta) = \frac{2}{\alpha} \liminf_{\Delta\theta \to 0} \frac{1}{(\Delta\theta)^2} I^R(g(y, \theta), g(y, \theta + \Delta\theta)) .$$

By Theorem 2.6, taking the limit as $N \to \infty$ and for $\Delta\theta$ taking values outside a neighborhood of zero, we have

$$\lim_{N\to\infty} I_N^F(\theta) = \frac{2}{\alpha} \liminf_{\Delta\theta\to 0} \frac{1}{(\Delta\theta)^2} \left\{ \lim_{N\to\infty} I^R(g(y,\theta), g^{(N)}(y,\theta+\Delta\theta)) \right\}$$
$$= \frac{2}{\alpha} \liminf_{\Delta\theta\to 0} \frac{1}{(\Delta\theta)^2} I^R(g(y,\theta), g(y,\theta+\Delta\theta))$$
$$= I^F(\theta) \quad \text{for} \quad \theta\in\Theta .$$

The converse of this theorem remains an open question.

Acknowledgements

The authors wish to thank the referees for their helpful suggestions which improved the presentation of the paper. In particular, they are thankful to the referee who pointed out the Liese and Vajda reference.

REFERENCES

Barron, E. R. (1986). Entropy and the central limit theorem, *Ann. Probab.*, **14**, 336–342.

Billingsley, P. (1979). *Probability and Measure*, Wiley, New York.

Brown, L. D. (1982). A proof of the central limit theorem motivated by the Cramér-Rao inequality, *Statistics and Probability: Essays in Honor of C. R. Rao*, (eds. G. Kallianpur, P. R. Krishnaiah and J. K. Ghosh), North-Holland, Amsterdam.

Burrill, C. W. (1972). *Measure, Integration and Probability*, McGraw-Hill, New York.

Csiszár, I. (1963). Eine informationstheoretische ungleichung und ihre anvendung auf den beweis der ergodizitat von Markoffschen Ketten, *A Mayar Tudomanyos Academia Mathematikai Kutato Intezelent Kozlemezyri*, **8**, 85–108.

Ferentinos, K. and Papaioannou, T. (1981). New parametric measures of information, *Inform. and Control*, **51**, 193–208.

Kirmani, S.N.U.A. (1971). Some limiting properties of Matusita's measure of distance, *Ann. Inst. Statist. Math.*, **23**, 157–162.

Kullback, S. (1959). *Information Theory and Statistics*, Wiley, New York.

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

Linnik, Yu. V. (1959). An information-theoretic proof of the central limit theorem with Lindeberg condition, *Theory Probab. Appl.*, **4**, 288–299.

Loéve, M. (1963). *Probability Theory*, 3rd ed., D. Van Nostrad Company Inc., New Jersey.

Perez, A. (1968). Risk estimates in terms of generalized *f*-entropies, *Proc. Colloquium on Information Theory, Debreen* (1967), *J. Bolyai Math. Society*, Budapest 1968, 299–315.

Rényi, A. (1961). On measures of entropy and information, *Proc. Fourth Berkeley Symp. on Math. Statist. Prob.*, Vol. 1, 547–561, University of California Press.

Rényi, A. (1970). *Probability Theory*, North-Holland, Amsterdam.

Vajda, I. (1972). On the *f*-divergence and singularity of probability measures, *Period. Math. Hungar.*, **2**, 223–234.

Zografos, K. (1987). Contributions to Statistical Information Theory, Ph.D. Thesis, Probability-Statistics and Operational Research Unit., Department of Mathematics, University of Ioannina (in Greek).