Volume 4 (1995), No.2, pp. 230–237 Allerton Press, Inc.
MATHEMATICAL METHODS OF STATISTICS

DISTANCE IN VARIATION AND A FISHER-TYPE INFORMATION

N. Papadatos[1] and V. Papathanasiou

Department of Mathematics, University of Athens
Panepistemiopolis, Athens 157 84, Greece

Upper bounds for the distance in total variation, in terms of Fisher-type information, are obtained for two arbitrary probability measures. Applications to extreme-value theory provide estimates of the rate of convergence. Another application simplifies Barron's result [1] concerning showing convergence to the normal law by using entropy.

Key words: distance in variation, Fisher-type information, extreme-value theory.

1. Introduction

Recently Cacoullos, Papathanasiou and Utev [4] established upper bounds for the distance in variation between an arbitrary probability measure and standard normal one via some integro-differential functionals including information. Moreover, another proof of the Central Limit Theorem was obtained. Also, in the discrete case Papathanasiou and Utev [10] extended these results to approximate an arbitrary discrete probability measure by a Poisson distribution.

In the present paper, by using a Fisher-type information we extend these results to obtain upper bounds for the distance in total variation between two arbitrary probability measures, both in the discrete and in the continuous case. This is achieved by using an extension of an identity due to Hudson [6] (see (2.2)) and by choosing an appropriate function $h(x)$, which plays the role of the Bolthausen function [2] (see (2.3)).

In Section 4 we give some applications to the Central Limit Theorem and to extreme-value theory.

2. Distance in Variation and Fisher-type Information

Let X and Y be two arbitrary random variables (r.v.'s) with absolutely continuous distributions and let F and G respectively be their distribution functions (d.f.'s). Furthermore, suppose that the corresponding densities $f(x)$ and $g(x)$ are differentiable in their interval supports (a, b), (a', b') such that

$$-\infty \le a' \le a < b \le b' \le +\infty$$

[1]The first author was partially supported by Greek state scolarship foundation.

and

$$f(a) = \lim_{x\to a} f(x) < \infty, \qquad f(b) = \lim_{x\to b} f(x) < \infty.$$

The distance in total variation between F and G is defined by

$$\rho(F,G) \equiv \sup_A |F(A) - G(A)|, \tag{2.1}$$

where the supremum is taken over the class of Borel sets A, and $F(A) = P[X \in A]$, $G(A) = P[Y \in A]$. In this section we obtain upper bounds for $\rho(F,G)$ in terms of a Fisher-type information. For this purpose, we use the following lemmas.

Lemma 2.1. *If h is absolutely continuous and bounded then*

$$\mathbf{E}[h'(X)] = h(b)f(b) - h(a)f(a) - \mathbf{E}\left[h(X)\frac{f'(X)}{f(X)}\right] \tag{2.2}$$

provided $\mathbf{E}|h'(X)| < \infty$.

Proof. We conclude (2.2) integrating by parts (cf. [11]). □

Define the function $h(x,A)$ by the formula (cf. [2])

$$h(x,A) = \frac{1}{g(x)} \int_{a'}^{x} (I_A(u) - G(A))g(u)\,du, \tag{2.3}$$

where $I_A(x)$ is the indicator function of A and $x \in (a', b')$.

Lemma 2.2. *If*

$$\limsup_{x\to a'} \frac{G(x)}{g(x)} < \infty, \qquad \limsup_{x\to b'} \frac{1-G(x)}{g(x)} < \infty,$$

then there exists a constant $c < \infty$ *such that*

$$\sup_A \sup_{a'<x<b'} |h(x,A)| \le c. \tag{2.4}$$

This constant can be taken as $c = \sup_{a'<x<b'} \left[\min\{G(x), 1-G(x)\}/g(x)\right]$.

Proof. By definition,

$$h(x,A) = \frac{1}{g(x)} \int_{a'}^{x} (I_A(u) - G(A))g(u)\,du = -\frac{1}{g(x)} \int_{x}^{b'} (I_A(u) - G(A))g(u)\,du.$$

Hence

$$|h(x,A)| \le \frac{1}{g(x)} \min\{G(x), 1-G(x)\}, \qquad a' < x < b'. \tag{2.5}$$

Obviously, the function $\min\{G(x), 1-G(x)\}$ is continuous and is equal to $G(x)$ if $x < \delta$ and to $1-G(x)$ if $x > \delta$, where δ is the median of X. Under the assumptions of Lemmas 2.1, 2.2, the continuous function $\min\{G(x), 1-G(x)\}/g(x)$ is bounded at both endpoints a', b' and hence it is bounded for each x in the interval and the proof of Lemma 2.2 is complete. ⊡

Note that this constant $c = c(G)$ depends on G only. For example, if $G = \Phi$ (the standard normal distribution) then $c = \sup_x [\min\{\Phi(x), \Phi(-x)\}/\varphi(x)] = \Phi(0)/\varphi(0) = \sqrt{\pi/2}$.

Now we give the following

Theorem 2.1. *Under the assumptions of Lemma 2.2, there is a constant c, which depends on G only, such that*

$$\rho(F,G) \le c\,\mathbf{E}\left|\frac{f'(X)}{f(X)} - \frac{g'(X)}{g(X)}\right| + \sup_A \left|h(b,A)f(b) - h(a,A)f(a)\right|. \tag{2.6}$$

Proof. Differentiating (2.3) and writing $h(x)$ instead of $h(x,A)$ we obtain

$$h'(x) = -\frac{g'(x)}{g(x)}h(x) + I_A(x) - G(A). \tag{2.7}$$

Taking expectations in (2.7) with respect to f we obtain

$$F(A) - G(A) = \mathbf{E}\left[h(X)\frac{g'(X)}{g(X)}\right] + \mathbf{E}[h'(X)].$$

In virtue of Lemma 2.1,

$$F(A) - G(A) = \mathbf{E}\left[-h(X)\left(\log\frac{f(X)}{g(X)}\right)'\right] + h(b)f(b) - h(a)f(a),$$

and thus

$$|F(A) - G(A)| \le \sup_x |h(x)|\mathbf{E}\left|\left(\log\frac{f(X)}{g(X)}\right)'\right| + |h(b)f(b) - h(a)f(a)|.$$

Taking the supremum over A completes the proof of the Theorem, with c as in Lemma 2.2. ⊡

Corollary 2.1. *Under any of the following conditions*

(i) $f(a) = f(b) = 0$,
(ii) $f(a) = 0$, $b = b'$ *and* $\liminf_{x\to b} g(x) > 0$,
(iii) $a = a'$, $f(b) = 0$ *and* $\liminf_{x\to a} g(x) > 0$,
(iv) $a = a'$, $\liminf_{x\to a} g(x) > 0$, $b = b'$ *and* $\liminf_{x\to b} g(x) > 0$,

one has $\left|h(b,A)f(b) - h(a,A)f(a)\right| = 0$ *and*

$$\rho(F,G) \le c\,\mathbf{E}\left|\frac{f'(X)}{f(X)} - \frac{g'(X)}{g(X)}\right|. \tag{2.8}$$

Remark 2.1. Using the Cauchy–Schwarz inequality in (2.8) we obtain a weaker bound. Namely,

$$\rho(F,G) \le c\sqrt{\mathbf{E}\left(\frac{f'(X)}{f(X)} - \frac{g'(X)}{g(X)}\right)^2}, \tag{2.9}$$

and thus a Fisher-type information gives an estimate of the distance $\rho(F,G)$.

3. Discrete Case

In this section we extend the above results to the discrete case.

Let X and Y be discrete r.v.'s with probability functions $p(x)$, $q(y)$ and d.f.'s $P(x)$, $Q(y)$, respectively. Suppose $p(x) > 0$ for $x = 0,1,\dots,b$ and $q(y) > 0$ for $y = 0,1,\dots,b'$, where $0 < b \le b' \le +\infty$. As in the continuous case, we are interested in establishing upper bounds for the distance in variation between P and Q.

Lemma 3.1. *For any bounded function* $h(x)$

$$\mathbf{E}[\Delta h(X)] = -\mathbf{E}\left[\frac{\Delta p(X-1)}{p(X)}h(X)\right] + p(b)h(b+1). \tag{3.1}$$

Proof. Cf. [11]. □

We define the function $h(x,A)$ by

$$h(x,A) = \frac{1}{q(x-1)}\sum_{k=0}^{x-1}(I_A(k) - Q(A))q(k), \qquad x = 1,2,\dots,b', \tag{3.2}$$

with $h(x,A) = 0$ if $x \ne 1,2,\dots,b'$ ($Q(A) = P[Y \in A]$, $P(A) = P[X \in A]$ as in the continuous case).

Lemma 3.2. *Suppose that one of the following statements holds:*

(i) $b' < \infty$,

(i) $b' = +\infty$ *and* $\limsup_{x\to+\infty} \dfrac{1-Q(x)}{q(x)} < +\infty$.

Then there exists a constant c, *which depends only on* Q, *such that*

$$\sup_A \sup_x |h(x,A)| \le c. \tag{3.3}$$

Proof. Observe that

$$|h(x,A)| \le \frac{\min\{Q(x-1), 1-Q(x-1)\}}{q(x-1)}. \tag{3.4}$$

Relation (3.3) is obvious for $b' < \infty$, with

$$c = \max_{x=0,1,\ldots,b'-1} \left[\min\{Q(x), 1-Q(x)\}/q(x)\right].$$

If $b' = +\infty$ then (3.4) and (ii) imply (3.3) with

$$c = \sup_{x \in \mathbb{N}} \left[\min\{Q(x), 1-Q(x)\}/q(x)\right].$$

The next theorem is a discrete analogue of Theorem 2.1 concerning Fisher-type information.

Theorem 3.1. *Under the assumptions of Lemma 3.2, there is a constant c which depends on Q only, such that*

$$\rho(P,Q) \le c\left\{\mathbf{E}\left|\frac{\Delta p(X-1)}{p(X)} - \frac{\Delta q(X-1)}{q(X)}\right| + I(b < b')\,p(b)\right\}. \tag{3.5}$$

Proof. Taking forward differences in (3.2) shows that

$$\Delta h(x,A) = -\frac{\Delta q(X-1)}{q(X)} h(x,A) + I_A(x) - Q(A). \tag{3.6}$$

We take expected values in (3.6) and, by virtue of Lemma 3.1, we obtain

$$P(A) - Q(A) = \mathbf{E}\left[h(X,A)\Big(\frac{\Delta q(X-1)}{q(X)} - \frac{\Delta p(X-1)}{p(X)}\Big)\right] + h(b+1,A)\,p(b). \tag{3.7}$$

Since $h(b+1,A)\,p(b) = 0$ for $b = b'$,

$$h(b+1,A)\,p(b) = I(b < b')\,h(b+1,A)\,p(b).$$

By applying Lemma 3.2 to (3.7) the proof of Theorem 3.1 is completed. □

If $b = b'$, then one can obtain the discrete analogue of Corollary 2.2.

4. Applications

First we give applications of the above results to the Central Limit Theorem. We take $G = \Phi$, the standard normal d.f. Then, the conditions of Lemma 2.2 are satisfied and the constant $c = \sqrt{\pi/2}$, as follows from (2.5).

Corollary 4.1. *Under condition* (i) *of Corollary* 2.1

$$\rho(F,\Phi) \le \sqrt{\pi/2}\sqrt{J(X)}, \tag{4.1}$$

where

$$J(X) = \mathbf{E}\left(\frac{f'}{f} - \frac{\varphi'}{\varphi}\right)^2$$

is the standardized Fisher information according to Barron [1].

If the r.v. X has mean 0 and variance 1, then $I(X) = J(X) + 1$, where

$$I(X) = \mathbf{E}\left[\frac{f'(X)}{f(X)}\right]^2$$

is the Fisher information and

$$J(X) \geq 0 \tag{4.2}$$

with equality only if $f = \varphi$.

Barron [1] extended arguments from [3] to show that $J(S'_n) \to 0$. Here $S'_n = \sqrt{t}S_n + \sqrt{1-t}Z$, $0 \leq t < 1$, where S_n is the standardized sum of independent identically distributed r.v.'s and Z is an independent standard normal r.v. In addition he gave an identity which connects $J(X)$ with the relative entropy

$$D(X) = \int f(x) \log \frac{f(x)}{\varphi(x)}\, dx.$$

Then using the inequality $\left(\int |f-g|\right)^2 \leq 2D$ from [5], he established a strengthened version of Central Limit Theorem.

By using (4.1), the convergence $J(S'_n) \to 0$ leads directly to the convergence in the sense of the total variation. Hence (4.1) gives both the stability of characterization (4.2) with respect to the convergence in total variation and the rate of convergence (cf. Mayer-Wolf [7] for a different approach).

For another application concerning the convergence in extreme-value theory, suppose that $X_1, \ldots, X_n$ are i.i.d. r.v.'s with common d.f. F and

$$M_1 = \min(X_1, \ldots, X_n).$$

Commonly there exist normalizing constants $a_n > 0$ and $b_n \in \mathbb{R}$ and a nondegenerate d.f. G such that

$$\lim_{n\to\infty} P\left\{\frac{M_1 - b_n}{a_n} \leq x\right\} = G(x).$$

When this happens, F is said to be in the min domain of attraction of G, and G must be one of the following three Gnedenko extreme-value types:

$$\begin{aligned} L_{1,\gamma}(x) &= 1 - \exp[-(-x)^{-\gamma}], && x \leq 0, \quad \gamma > 0, \\ L_{2,\gamma}(x) &= 1 - \exp[-x^{\gamma}], && x > 0, \quad \gamma > 0, \\ L_{3,0}(x) &= 1 - \exp(-e^{x}), && x \in \mathbb{R}. \end{aligned}$$

Similarly, for $M_n = \max(X_1, \ldots, X_n)$, G must be one of the three extreme-value types:

$$\begin{aligned} H_{1,\gamma}(x) &= \exp[-x^{-\gamma}], && x > 0, \quad \gamma > 0, \\ H_{2,\gamma}(x) &= \exp[-(-x)^{\gamma}], && x \leq 0, \quad \gamma > 0, \\ H_{3,0}(x) &= \exp(-e^{-x}), && x \in \mathbb{R}. \end{aligned}$$

Recently in [8], [9] uniform rates of convergence in extreme-value theory were established. We give here bounds for the distance when $G = L_{2,1}(x) = 1 - e^{-x}$ (exponential) or $G = L_{3,0}(x)$.

Corollary 4.2. *Let $X_1, X_2, \ldots, X_n$ be i.i.d. r.v.'s with common d.f. F and density $f(x) = F'(x)$, $x \in (a, b)$, $a > -\infty$ as in Theorem 2.1. Let $a_n > 0$, b_n be sequences such that $a \cdot a_n + b_n \equiv 0$, $n = 1, 2, \ldots$. Then*

$$\rho(F_n, L_{2,1}) \le \mathbf{E}\left|1 + \frac{f_n'(X)}{f_n(X)}\right|, \qquad n = 2, 3, \ldots, \tag{4.3}$$

where $F_n(x) = P[a_n M_1 + b_n \le x]$.

Corollary 4.3. *Let $X_1, \ldots, X_n$ be i.i.d. r.v.'s with common d.f. F and density $f = F'$, $f(a) = 0$. Then for arbitrary sequences $a_n > 0$, b_n we have*

$$\rho(F_n, L_{3,0}) \le \frac{1}{\log 2}\mathbf{E}\left|\frac{f_n'(X)}{f_n(X)} - 1 + e^X\right|, \qquad n = 2, 3, \ldots, \tag{4.4}$$

where F_n is as in Corollary 4.2.

In these two cases the conditions of Lemma 2.2 are satisfied. In Corollary 4.2, it is obvious that $h(0) = 0$ and $f_n(a_n b + b_n) = 0$ for $n \ge 2$. In Corollary 4.3 the statement (i) of Corollary 2.1 is satisfied.

We give two examples when F is uniform on $(0, 1)$ or $F(x) = e^x$, $x \le 0$. Then (4.3) and (4.4) yield

$$\rho(F_n, L_{2,1}) \le 1/\sqrt{n(n-2)}, \qquad n \ge 3,$$
$$\rho(\widetilde{F}_n, L_{3,0}) \le 3/(n-1)\log 2, \qquad n \ge 2,$$

for $F_n = P[nM_1 \le x]$ and $\widetilde{F}_n = P[M_1 + \log n \le x]$. In both cases the rate of convergence is at least $O(1/n)$.

Of course similar results are obtained for the distance between F and $H_{2,1}$ or $H_{3,0}$ by using the fact that $\rho(F, G) = \rho(1 - F(-x), 1 - G(-x))$.

We give here an application in the discrete case. Let Q be a geometric distribution with parameter p. Theorem 3.1 gives the following

Corollary 4.4. *For an arbitrary d.f. P, as in Theorem 3.1,*

$$\rho(P, Q) \le \frac{1-p}{p}\left\{p(b) + \sum_{k=1}^{b} p(k)\left|\frac{p}{1-p} + \frac{\Delta p(k-1)}{p(k)}\right|\right\}.$$

Two examples for this corollary follow.

Let $X_1, X_2, \ldots, X_n$ be i.i.d. r.v.'s with common probability function $p(x) = 1/(v+1)$, $x = 0, 1, \ldots, v$. Suppose that $n/v \to \lambda > 0$ (constant) as $v \to \infty$

(equivalently, $n = n(v) = \lambda v + o(v)$). Let also $P_v(x) = P[M_1 \le x]$, where $M_1 = \min\{X_1, \dots, X_n\}$. Then Corollary 4.4 with $p = 1 - e^{-\lambda}$ shows that

$$\rho(P_v, Q) \to 0, \qquad \text{as} \quad v \to \infty.$$

Let $X_1, \dots, X_n$ be independent geometric r.v.'s with parameter $p = \lambda/n$. Then

$$\rho(P_n, Q) \le \frac{1}{e^\lambda - 1}\left|1 - e^\lambda\left(1 - \frac{\lambda}{n}\right)^n\right| \to 0$$

(where Q is as in the previous example and $P_n(x) = P[M_1 \le x]$). The rate of convergence here is at least $O(1/n)$.

References

[1] A. R. Barron, *Entropy and the Central Limit Theorem*, Ann. Probab., **14** (1986), 336–342.

[2] E. Bolthausen, *An estimate of the remainder in a combinatorial Central Limit Theorem*, Z. Wahrsch. verw. Gebiete, **66** (1984), 379–386.

[3] L. D. Brown, *A proof of the Central Limit Theorem motivated by the Cramér–Rao inequality*. In: *Statistics and Probability*, Essays in Honour of C. R. Rao, G. Kallianpur, P. R. Krishnaiah, and J. K. Ghosh, eds., North-Holland, Amsterdam, 1982, pp. 140–148.

[4] T. Cacoullos, V. Papathanasiou, and S. Utev, *Variational inequalities with examples and an application to the Central Limit Theorem*, Ann. Probab., **22** (1994), 1607–1618.

[5] I. Csiszar, *Information-type measures of difference of probability distributions and indirect observations*, Studia Sci. Math. Hungar., **2** (1967), 299–318.

[6] H. M. Hudson, *A natural identity for exponential families with applications in multiparameter estimation*, Ann. Statist., **6** (1978), 473–484.

[7] E. Mayer-Wolf, *The Cramér–Rao functional and limiting laws*, Ann. Probab., **18** (1990), 840–850.

[8] E. Omey and S. T. Rachev, *On the rate of convergence in extreme value theory*, Theory Probab. Appl., **33** (1988), 560–566.

[9] E. Omey, *Rates of convergence for densities in extreme value theory*, Ann. Probab., **16** (1988), 479–486.

[10] V. Papathanasiou and S. Utev, *Integro-differential inequalities and the Poisson approximation*, Siberian Advances in Math. (1995), to appear.

[11] V. Papathanasiou, *Some characteristic properties of the Fisher information matrix via Cacoullos-type inequalities*, J. Multivariate Anal., **44** (1993), 256–265.

[*Received October* 1993]