Volume 7, No. 2 (1998), pp. 230–244 Allerton Press, Inc.

MATHEMATICAL METHODS OF STATISTICS

TOTAL VARIATION DISTANCE AND GENERALIZED COVARIANCE KERNELS

N. Papadatos and V. Papathanasiou

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

Variational inequalities between two arbitrary probability measures are obtained in terms of a generalization on the w-function, called the z-function or Generalized Covariance Kernel. Some illustrative examples are given and, furthermore, the connection with a generalisation of the Stein–Chen approach is discussed.

Key words: Generalized Covariance Kernels, Stein–Chen approach, total variation distance, covariance identity.

AMS 1991 Subject Classification: Primary 62E17; secondary 62E10.

1. Introduction and Summary

Cacoullos, Papathanasiou, and Utev (henceforth, CPU) [7] established upper bounds for the total variation distance between an arbitrary distribution and the standard normal, in terms of a Fisher-type information and the w-function. Papathanasiou and Utev [12] extended these results in the discrete case to approximate an arbitrary discrete distribution by the Poisson distribution.

Papadatos and Papathanasiou [10, 11] gave upper bounds for the distance in variation between two arbitrary distributions F, G in terms of a Fisher-type information (see also Mayer-Wolf [9]) and the corresponding w-functions, respectively.

The above results are based on the covariance identity

$$\mathrm{Cov}[X, \ell(X)] = \sigma^2 \mathbf{E}[w(X)\ell'(X)], \tag{1.1}$$

where σ^2 is the variance of the continuous (with interval support) random variable X and w is the w-function associated with the density of X (see Cacoullos and Papathanasiou [4], CPU [6, 7], and references therein).

Recently, Cacoullos and Papathanasiou [5] established the general covariance identity (2.1), satisfied by any differentiable function ℓ (with derivative ℓ').

In the present paper, by exploiting properties of the general covariance identity, upper bounds are obtained for the total variation distance between two arbitrary distributions. The main feature of these bounds is that they involve the associated z-functions (or *Generalized Covariance Kernels*) corresponding to a given function h (see (2.2)).

In Section 2, as a by-product of the main result (Theorem 2.1), a necessary and sufficient condition for the L^1-convergence of the densities is also given (Theorem 2.3).

In Section 3, the main result is given by Theorem 3.1. Some illustrative examples are given in Section 4, while the connection of the present approach with the Stein–Chen method (in a general setting) is discussed in Section 5.

2. Variational Inequalities in Terms of a Specific Choice of h

Let X be a continuous random variable with distribution function F, density f and support a finite or infinite interval (a,b). Suppose that the limits of f at the boundary points a, b exist (and are finite) and denote them by $f(a)$, $f(b)$, respectively. Consider also a function h defined on (a,b). Then we have the identity (see Cacoullos and Papathanasiou [5])

$$\mathrm{Cov}[h(X),\ell(X)] = \sigma^2 \mathrm{E}[z(X)\ell'(X)], \tag{2.1}$$

where ℓ is an arbitrary differentiable function and the function $z(x) = z_f(x;h)$ is defined for every x in (a,b) by

$$z(x)f(x) = \int_a^x \big(\mathrm{E}[h(X)] - h(t)\big) f(t)\,dt \tag{2.2}$$

(it is, of course, assumed that the expectations in (2.1) and (2.2) are well-defined in the sense that $\mathrm{E}|h(X)| < \infty$ and $\mathrm{E}\big|\big[h(X) - \mathrm{E}[h(X)]\big]\ell(X)\big| < \infty$).

As it was pointed out by Cacoullos and Papathanasiou [5], the z-function is a generalization of the w-function (sometimes called *Covariance Kernel*) since, by (1.1), $\mathrm{Var}[X]w(x) \equiv z_f(x;x)$ (observe that (1.1), applied to the standard normal distribution, reduces to the classical *Stein identity*, see Stein [14, 15]; see also Hudson [8]). Thus, z may be called the *Generalized Covariance Kernel* (GCK) associated with the density f with respect to the function h.

The following lemma presents some basic properties of the GCK.

Lemma 2.1. (i) $z_f(x; ch+\mu) = cz_f(x;h)$ *for any constants* μ, c.

(ii) $z(x) \equiv c \neq 0$ *in* (a,b) *if and only if* f *is differentiable in* (a,b), $f(a) = f(b) = 0$ *and* $h(x) = \mu - c\big(f'(x)/f(x)\big)$ *for some constant* c.

(iii) *If* h *is strictly increasing (decreasing) in* (a,b), *then* z *is positive (negative) in* (a,b).

Proof. (i) is obvious.

(ii) Suppose that $z(x) \equiv c$. Then, from (2.2) we have

$$f(x) = \frac{1}{c}\int_a^x \big(\mathrm{E}[h(X)] - h(t)\big) f(t)\,dt \tag{2.3}$$

and therefore f is differentiable with $cf'(x) = (\mu - h(x))f(x)$, where $\mu = \mathbb{E}[h(X)]$; thus $h(x) = \mu - c(f'(x)/f(x))$. Since

$$\mu = \mathbb{E}[h(X)] = \mu - c\mathbb{E}[f'(X)/f(X)] = \mu - c(f(b) - f(a)),$$

we must have $f(a) = f(b)$. Substituting $h(x) = \mu - c(f'(x)/f(x))$ in (2.3) we take

$$f(x) = \frac{1}{c}\int_a^x \left(\mu - \left(\mu - c\frac{f'(t)}{f(t)}\right)\right) f(t)\,dt = f(x) - f(a),$$

which shows that $f(a) = 0$. Since the converse is obvious, the proof of (ii) is complete.

(iii) If h is strictly increasing, then observe that

$$z(x)\frac{f(x)}{1-F(x)} = \frac{1}{1-F(x)}\int_x^b (h(t) - \mathbb{E}[h(X)])f(t)\,dt, \tag{2.4}$$

or, equivalently,

$$z(x)m(x) = \mathbb{E}[h(X) \mid X > x] - \mathbb{E}[h(X)], \tag{2.5}$$

where m is the hazard rate. This means that z is positive. The case of decreasing h is similar and the proof of (iii) is complete. □

Consider now another r.v. Y, with distribution function G, differentiable density g (with continuous derivative g') and support an interval (a', b') such that $-\infty \le a' \le a < b \le b' \le +\infty$. Suppose also that the limits of g at the boundary points a', b' exist (and are finite) and denote them by $g(a')$, $g(b')$, respectively. Consider also the special function (see also Papadatos and Papathanasiou [10]) $\ell(x) = \ell_A(x; g)$ defined for all x in (a', b') by

$$\ell(x) = \frac{1}{g(x)}\int_{a'}^x (I_A(t) - G(A))g(t)\,dt, \tag{2.6}$$

where I_A is the indicator function of a (Borel) set A and $G(A) = \mathbb{P}(Y \in A)$.

The following lemma is useful for the main variational inequality of this section.

Lemma 2.2. *(i) Set* $\|\ell\| = \sup_x \sup_A |\ell_A(x; g)|$. *If*

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

then there exists a constant c_2 *(depending on* G *only) such that*

$$\|\ell\| = c_2 < \infty.$$

(ii) Set $\|\ell'\| = \sup_x \sup_A |\ell'_A(x; g)|$. *If*

$$\limsup_{x\to a'_+}\frac{|g'(x)|G(x)}{g^2(x)} < \infty \qquad \text{and} \qquad \limsup_{x\to b'_-}\frac{|g'(x)|(1-G(x))}{g^2(x)} < \infty,$$

then there exists a constant c_1 (depending on G only) such that

$$\|\ell\| = c_1 < \infty.$$

Proof. (i) See Papadatos and Papathanasiou [10], Lemma 2.2.
(ii) Differentiating (2.6) we have

$$\ell'_A(x;g) = \ell'(x) = -\frac{g'(x)}{g(x)}\ell(x) + I_A(x) - G(A), \tag{2.7}$$

and thus,

$$|\ell'(x)| \le \frac{g'(x)}{g(x)}|\ell(x)| + 1 \le 1 + \frac{|g'(x)|}{g^2(x)}\min\{G(x), 1 - G(x)\}. \tag{2.8}$$

Under the above assumptions, the continuous function on the RHS of (2.8) is bounded at both endpoints and hence it is bounded over (a', b') by some constant $c_1' < \infty$. Taking the supremum over the Borel sets A and x in (a', b'), we conclude the desired result (in fact, $c_1 \le 1 + c_1'$). □

We can now prove the main result of this section, stated as follows.

Theorem 2.1. *Let X be a random variable with density f and distribution F as above. If g satisfies the conditions of Lemma 2.2 and $\mathrm{E}\big[|g'(X)|/g(X)\big] < \infty$, then, there exist finite constants c_1 and c_2, depending only on G, such that*

$$\rho(F, G) \le c_1 \mathrm{E}|z(X) - 1| + c_2\big|\mathrm{E}[g'(X)/g(X)]\big|, \tag{2.9}$$

where $\rho(F,G) = \sup_A |F(A) - G(A)|$ is the total variation distance between F and G, $z(\cdot) \equiv z_f(\,\cdot\,; -g'/g)$ is the z-function associated with the density f when $h = -g'/g$ and the constants c_1 and c_2 can be taken as in Lemma 2.2.

Proof. Taking expectations in (2.7) with respect to f, we have

$$\begin{aligned} \mathrm{E}[\ell'(X)] &= \mathrm{E}\Big[-\frac{g'(X)}{g(X)}\ell(X)\Big] + F(A) - G(A) \\ &= \mathrm{Cov}\Big[-\frac{g'(X)}{g(X)}, \ell(X)\Big] + \mathrm{E}\Big[-\frac{g'(X)}{g(X)}\Big]\mathrm{E}[\ell(X)] + F(A) - G(A). \end{aligned} \tag{2.10}$$

In virtue of (2.1), relation (2.10) becomes

$$F(A) - G(A) = \mathrm{E}\big[\ell'(X)(1 - z(X))\big] - \mathrm{E}[-g'(X)/g(X)]\mathrm{E}[\ell(X)], \tag{2.11}$$

where

$$z(x) \equiv z_f(x; -g'/g) = \frac{1}{f(x)}\int_a^x \Big(\mathrm{E}\Big[-\frac{g'(X)}{g(X)}\Big] + \frac{g'(t)}{g(t)}\Big) f(t)\,dt.$$

From (2.11) we have

$$|F(A) - G(A)| \le \mathrm{E}|\ell'(X)(1 - z(X))| + \left|\mathrm{E}\left[\frac{g'(X)}{g(X)}\right]\right| |\mathrm{E}[\ell(X)]|$$
$$\le \|\ell'\|\mathrm{E}|1 - z(X)| + \|\ell\| \left|\mathrm{E}\left[\frac{g'(X)}{g(X)}\right]\right|,$$

which, taking the supremum over the Borel sets A and in virtue of Lemma 2.2, completes the proof. □

For $g = \varphi$ (the standard normal density) we have as an immediate consequence the following corollary (see CPU [7], Theorem 1.1).

Corollary 2.1. *If* $\mathrm{E}[X] = 0$, $\mathrm{Var}[X] = 1$, *then, under the same assumptions imposed on* F, *we have*

$$\rho(F, \Phi) \le 2\mathrm{E}|w(X) - 1|, \tag{2.12}$$

where w *is the* w*-function associated with* X.

Applying the Cauchy–Schwarz inequality to (2.9), we obtain a weaker bound

$$\rho(F, G) \le c_1\sqrt{\mathrm{E}[z(X) - 1]^2} + c_2|\mathrm{E}[g'(X)/g(X)]|, \tag{2.13}$$

whereas (2.11) similarly yields the following

Corollary 2.2. *If* $\sup_A \sqrt{\mathrm{E}[\ell'_A(X; g)]^2} = c_3 < \infty$ (*obviously,* $c_3 \le c_1$), *then*

$$\rho(F, G) \le c_3\sqrt{\mathrm{E}[z(X) - 1]^2} + c_2|\mathrm{E}[g'(X)/g(X)]|, \tag{2.14}$$

where c_2 *is as in Lemma* 2.2 *and* c_3 *depends on* F *and* G.

We give now the following characterization which constitutes an extension of that in CPU [6], Characterization 4.

Theorem 2.2. *If, in addition to the conditions of Theorem* 2.1,

$$\mathrm{E}[g'(X)/g(X)] = 0 \quad \textit{and} \quad g(a') = g(b') = 0$$

then

$$\mathrm{E}[z(X) - 1]^2 \ge 0, \tag{2.15}$$

where the equality holds if and only if $f(x) = g(x)$ *for almost all* x.

Proof. Obviously, $\mathrm{E}[z(X) - 1]^2 \ge 0$. If $\mathrm{E}[z(X) - 1]^2 = 0$, then from (2.13) (or (2.14)) we have $\rho(F, G) = 0$ and thus $f = g$ for almost all x. Suppose now that $f = g$ for almost all x. Then, $z(x) = z_f(x; -g'/g) = z_g(x; -g'/g)$ for almost all x. Since, from Lemma 2.1, $z_g(x; -g'/g) \equiv 1$, the desired result follows from the absolute continuity of F. □

Remark 2.1. It should be noted that the condition $\mathrm{E}[g'(X)/g(X)] = 0$ is necessary for the validity of the above theorem. Indeed, if we take $g(x) = 6x(1-x)$ and $f(x) = (3-e)^{-1}x(1-x)e^x$ for $x \in (0,1)$, then $f \neq g$ while $z_f(x; -g'/g) \equiv 1$ for $x \in (0,1)$. However, $\mathrm{E}[g'(X)/g(X)] = -1 \neq 0$ in this case.

It is well known that convergence in total variation is equivalent to the L^1 convergence of the densities. In fact,

$$\rho(F_n, G) = \frac{1}{2} \int |f_n - g|.$$

The following theorem gives necessary and sufficient conditions for the L^1 convergence.

Theorem 2.3. *Let Y have density g as in Lemma 2.2, let, furthermore, $g(a') = g(b') = 0$, and assume that (the score function of Y has finite second moment):*

$$\mathrm{E}[g'(Y)/g(Y)]^2 < \infty.$$

Consider a sequence of absolutely continuous r.v.'s $X_1, X_2, \ldots, X_n, \ldots$ with densities $f_1, f_2, f_3, \ldots, f_n, \ldots$ each with support an interval $(a_n, b_n) \subset (a', b')$. Suppose further that

(i) $\mathrm{E}[g'(X_n)/g(X_n)] = -\varepsilon_n \to 0$ *as* $n \to \infty$;

(ii) $\mathrm{E}|z_n(X_n)| \leq \omega_n \to 1$ *as* $n \to \infty$, *where* $z_n(x) \equiv z_{f_n}(x; -g'/g)$;

(iii) $\mathrm{E}[g'(X_n)/g(X_n)]^2 \leq C < \infty$, $n = 1, 2, \ldots$

Then

$$\int |f_n - g| \to 0 \quad \textit{if and only if} \quad z_n(X_n) \to 1 \quad \textit{in probability.} \tag{2.16}$$

Proof. For $\varepsilon < 1$ arbitrarily small, set

$$A_{n,\varepsilon} = \{x \in (a_n, b_n) : |z_n(x) - 1| \leq \varepsilon\}, \qquad A^c_{n,\varepsilon} = (a_n, b_n) \setminus A_{n,\varepsilon}.$$

Suppose first that $z_n(X_n) \to 1$ in probability. Then,

$$\begin{aligned}\int_{A^c_{n,\varepsilon}} |z_n(x)| f_n(x)\,dx &= \mathrm{E}|z_n(X_n)| - \int_{A_{n,\varepsilon}} |z_n(x)| f_n(x)\,dx \\ &\leq \omega_n - (1-\varepsilon) \int_{A_{n,\varepsilon}} f_n(x)\,dx.\end{aligned}$$

Thus,

$$\begin{aligned}\mathrm{E}|z_n(X_n) - 1| &= \int_{A^c_{n,\varepsilon}} |z_n(x) - 1| f_n(x)\,dx + \int_{A_{n,\varepsilon}} |z_n(x) - 1| f_n(x)\,dx \\ &\leq \int_{A^c_{n,\varepsilon}} f_n(x)\,dx + \omega_n - (1-2\varepsilon) \int_{A_{n,\varepsilon}} f_n(x)\,dx - 2\varepsilon,\end{aligned}$$

since $\int_{A^c_{n,\varepsilon}} f_n(x)\,dx \to 0$, $\int_{A_{n,\varepsilon}} f_n(x)\,dx \to 1$ by the assumption that $z_n(X_n) \to 1$ in probability. From the arbitrariness of ε, it follows that

$$\mathbf{E}|z_n(X_n) - 1| \to 0 \qquad \text{as} \quad n \to \infty,$$

and from Theorem 2.1,

$$\frac{1}{2}\int |f_n - g| = \rho(F_n, G) \le c_1 \mathbf{E}|z_n(X_n) - 1| + c_2|\varepsilon_n| \to 0,$$

as is to be proved.

Conversely, let $\int |f_n - g| \to 0$. It suffices to prove that each subsequence $\{z_{n_k}(X_{n_k})\}$ contains a further subsequence $\{z_{n_{k(i)}}(X_{n_{k(i)}})\}$ such that $z_{n_{k(i)}}(X_{n_{k(i)}}) \to 1$ in probability, as $i \to \infty$ (cf. Billingsley [2], Theorem 20.5).

Choose $n_{k(i)}$ such that $\int |f_{n'} - g| < 2^{-i}$ (we write n' instead of $n_{k(i)}$). Then, it follows from the first Borel–Cantelli lemma that $f_{n'}(x) \to g(x)$ for almost all x; that is, for the set $B = \{x \in (a', b') : f_{n'}(x) \to g(x)\}$ we have $\lambda\big((a', b') \setminus B\big) = 0$, where λ is the Lebesgue measure on the Borel subsets of (a', b'). Fix $x \in B$. Since $(a_{n'}, b_{n'}) \to (a', b')$, $x \in (a_{n'}, b_{n'})$, for all large enough i, and

$$z_{n'}(x) - 1 = \varepsilon_{n'}\frac{F_{n'}(x)}{f_{n'}(x)} + \frac{1}{f_{n'}(x)}\int_{a_{n'}}^{x} \frac{g'(t)}{g(t)}(f_{n'}(t) - g(t))\,dt + \frac{g(x) - f_{n'}(x) - g(a_{n'})}{f_{n'}(x)}.$$

Hence,

$$|z_{n'}(x) - 1| \le |\varepsilon_{n'}|\frac{F_{n'}(x)}{f_{n'}(x)} + \frac{1}{f_{n'}(x)}\int_{a_{n'}}^{x} \frac{|g'(t)|}{g(t)}|f_{n'}(t) - g(t)|\,dt + \frac{|g(x) - f_{n'}(x)|}{f_{n'}(x)} + \frac{g(a_{n'})}{f_{n'}(x)} \to 0,$$

since $g(a_{n'}) \to g(a') = 0$, $f_{n'}(x) \to g(x) > 0$ (because $x \in B$), $\varepsilon_{n'} \to 0$ (by the assumption), and

$$\left(\int_{a_{n'}}^{x} \frac{|g'(t)|}{g(t)}|f_{n'}(t) - g(t)|\,dt\right)^2 \le \int_{a_{n'}}^{x} \left[\frac{g'(t)}{g(t)}\right]^2 |f_{n'}(t) - g(t)|\,dt \int |f_{n'} - g|$$
$$\le \left(\mathbf{E}[g'(X_{n'})/g(X_{n'})]^2 + \mathbf{E}[g'(Y)/g(Y)]^2\right)\int |f_{n'} - g| \to 0.$$

Therefore,

$$\text{(2.17)} \quad \mathbf{P}[|z_{n'}(X_{n'}) - 1| > \varepsilon] = \int_{A^c_{n',\varepsilon}} f_{n'}(x)\,dx = \int_{A^c_{n',\varepsilon}} |f_{n'}(x) - g(x) + g(x)|\,dx \le \int |f_{n'} - g| + G(A^c_{n',\varepsilon}).$$

Observe that

$$\limsup_i A^c_{n',\varepsilon} = \{x \in (a',b') : |z_{n'}(x) - 1| > \varepsilon \text{ i.o.}\} \subset (a',b') \setminus B,$$

and from Fatou's Lemma,

$$\limsup_i G(A^c_{n',\varepsilon}) \le G(\limsup_i A^c_{n',\varepsilon}) \le G\big((a',b') \setminus B\big) = 0,$$

because $\lambda\big((a',b') \setminus B\big) = 0$ and G is absolutely continuous with respect to Lebesgue measure λ. This, combined with (2.17), completes the proof. □

The above theorem gives the stability of the characterization presented by Theorem 2.2. Furthermore, for $g = \varphi$, it reduces to an extension of the result in CPU [7], Theorem 1.2, stated as follows.

Corollary 2.3. *Let $X_1, X_2, \ldots, X_n, \ldots$ be absolutely continuous random variables with densities $f_1, f_2, \ldots, f_n, \ldots$ each with an interval support. Suppose that* $\mathrm{E}[X_n] = \mu_n \to 0$ *and* $\mathrm{Var}[X_n] = \sigma_n^2 \to 1$ *as $n \to \infty$. Then,*

$$\int |f_n - \varphi| \to 0 \qquad \text{if and only if} \quad w_n(X_n) \to 1 \quad \text{in probability,}$$

where w_n is the w-function associated with X_n.

Proof. Since $-\varphi'(x)/\varphi(x) = x$, the conditions of Theorem 2.3 are satisfied. Furthermore, $z_n(x) = \sigma_n^2 w_n(x) > 0$ and since $\mathrm{E}[w_n(X_n)] = 1$, the result follows from (2.16). □

3. Variational Inequalities via the Associated z-Functions

In this section, we extend the results obtained by Papadatos and Papathanasiou [11]. To this end, define for x in (a',b') the special function $\psi(x) = \psi_A(x; h, g)$ by the relation

$$\psi(x) = \frac{1}{z_g(x;h)g(x)} \int_{a'}^{x} (I_A(t) - G(A))g(t)\,dt, \tag{3.1}$$

with h a given strictly increasing function, A a Borel set, and the function $z_g(x;h)$ given by

$$z_g(x;h) = \frac{1}{g(x)} \int_{a'}^{x} (\mathrm{E}_g[h] - h(t))g(t)\,dt \tag{3.2}$$

(i.e., z is the z-function associated with g with respect to h). Here, as usual, E_g denotes the expectation with respect to g.

The following lemma is required for the main result.

Lemma 3.1. *Suppose that h is differentiable in (a',b') and $\mathrm{Var}_g[h] < \infty$. Set*

$$\|\psi\| = \sup_x \sup_A |\psi_A(x;h,g)|, \qquad \|z_g\psi'\| = \sup_x \sup_A |z_g(x;h)\psi'_A(x;h,g)|.$$

If

(i) *either* $h(a') > -\infty$, *or*

$$h(a') = -\infty \quad \text{and} \quad \lim_{x \to a'_+} \frac{h'(x)G(x)}{-h(x)g(x)} \quad \text{exists (not necessarily finite)}$$

and

(ii) *either* $h(b') < +\infty$, *or*

$$h(b') = +\infty \quad \text{and} \quad \lim_{x \to b'_-} \frac{h'(x)(1-G(x))}{h(x)g(x)} \quad \text{exists (not necessarily finite)}$$

then there exist finite constants c_1, c_2 *such that*

$$\|\psi\| = c_2, \qquad \|z_g \psi'\| = c_1. \tag{3.3}$$

Proof. Suppose that $h(a') = -\infty$, $h(b') < +\infty$ (the other cases are similar). The existence of $\lim_{x \to a'_+} (h'G)/(-hg) = s \in [0, +\infty]$ guarantees that $s \in [0, 1]$, since the non-negative (for x in a neighborhood of a') function

$$s(x) = \frac{-h(x)G(x)}{g(x)z_g(x;h)} \to 1 - s \geq 0 \qquad \text{as} \quad x \to a'_+$$

by Hospital's rule (note that, from Chebyshev's inequality, $\mathrm{Var}_g[h] < \infty$ implies that $\lim_{x \to a'_+} h(x)G(x) = 0$).

It is easy to see that for all A and x in (a', b'),

$$|\psi_A(x; h, g)| \leq \frac{\min\{G(x), 1 - G(x)\}}{g(x)z_g(x;h)} \tag{3.4}$$

and since the continuous function on the RHS of (3.4) has finite limits at both end-points, it is bounded by a constant $c_2 < \infty$.

Taking derivatives in (3.1) we have

$$\psi'(x) = \frac{-\big(g(x)z_g(x;h)\big)'}{g(x)z_g(x;h)}\psi(x) + \frac{I_A(x) - G(A)}{z_g(x;h)} \tag{3.5}$$

or

$$z_g(x;h)\psi'(x) = -\big(\mathrm{E}_g[h] - h(x)\big)\psi(x) + I_A(x) - G(A). \tag{3.6}$$

Thus, from (3.4),

$$|z_g(x;h)\psi'_A(x;h,g)| \leq 1 + |\mathrm{E}_g[h] - h(x)| \frac{\min\{G(x), 1-G(x)\}}{g(x)z_g(x;h)} \to 2 - s \tag{3.7}$$

as $x \to a'_+$ (by the above argument). Similarly, the limit is finite as $x \to b'_-$ and thus, the continuous function on the RHS of (3.7) is bounded by a finite constant c_1 and the proof is complete. □

The main result of this section is stated in the following

Theorem 3.1. *Let X and Y be two arbitrary continuous random variables with distribution functions F, G, densities f, g, and interval supports (a,b), (a',b'), respectively, such that $-\infty \le a' \le a < b \le b' \le +\infty$. Suppose that g satisfies the conditions of Lemma 3.1 and, furthermore, the (strictly increasing) function h has a finite first moment with respect to f. Then*

$$\rho(F,G) \le c_1 \mathbf{E}_f\left|\frac{z_f(X;h)}{z_g(X;h)} - 1\right| + c_2\left|\mathbf{E}_f[h] - \mathbf{E}_g[h]\right|, \tag{3.8}$$

where the constants c_1 and c_2 depend on G only.

Proof. Taking expectations in (3.6) with respect to the density f, we obtain

$$\mathbf{E}_f[z_g\psi'] - \mathbf{E}_f[(h - \mathbf{E}_g[h])\psi] = F(A) - G(A).$$

Hence,

$$\mathbf{E}_f[z_g\psi'] - \mathrm{Cov}_f[h,\psi] - \mathbf{E}_f[\psi](\mathbf{E}_f[h] - \mathbf{E}_g[h]) = F(A) - G(A) \tag{3.9}$$

and, in virtue of identity (2.1),

$$F(A) - G(A) = \mathbf{E}_f\left[(z_g(X;h) - z_f(X;h))\psi'(X)\right] + \mathbf{E}_f[\psi](\mathbf{E}_g[h] - \mathbf{E}_f[h]), \tag{3.10}$$

where $z_f(x;h)$ is the z-function associated with f with respect to h (see (2.2)).

Thus,

$$|F(A) - G(A)| \le \mathbf{E}_f\left|[(z_f/z_g) - 1]\psi' z_g\right| + \left|\mathbf{E}_f[\psi](\mathbf{E}_g[h] - \mathbf{E}_f[h])\right|,$$

which, taking into account (3.3), completes the proof. □

For $h(x) = x$, $z_g(x;x) = w_Y(x)\,\mathrm{Var}[Y]$ and $z_f(x;x) = w_X(x)\,\mathrm{Var}[X]$ and relation (3.8) gives

$$\rho(F,G) \le c_1 \mathbf{E}\left|\frac{\mathrm{Var}[X]w_X(X)}{\mathrm{Var}[Y]w_Y(X)} - 1\right| + c_2\left|\mathbf{E}[X] - \mathbf{E}[Y]\right|, \tag{3.11}$$

which is the main result obtained by Papadatos and Papathanasiou [11] (w_X and w_Y are the w-functions associated with the densities f and g, respectively).

If $\mathbf{E}_f[h] = \mathbf{E}_g[h]$, relation (3.10) becomes

$$F(A) - G(A) = \mathbf{E}_f\left[\psi'(X)(z_g(X;h) - z_f(X;h))\right].$$

Hence,

$$\rho^2(F,G) = \sup_A \left(\mathbf{E}_f\left[\psi'_A(X;h,g)(z_f(X;h) - z_g(X;h))\right]\right)^2. \tag{3.12}$$

Therefore, if $\sup_A \mathbf{E}_f[\psi'_A(X;h,g)]^2 = c_3^2 < \infty$, then (3.12), by the Cauchy–Schwarz inequality, gives the following

Corollary 3.1. *Under the above assumptions, we have*

$$\rho(F,G) \le c_3\sqrt{\mathrm{E}_f\left[z_f(X;h) - z_g(X;h)\right]^2}, \tag{3.13}$$

where c_3 depends on F and G.

The assumption that $\sup_A \mathrm{E}_f\left[z_g(X;h)\psi'_A(X;h,g)\right]^2 = c_4^2 < \infty$ (note that $c_4 \le c_1$) gives the next

Corollary 3.2. *Under the above assumptions we have*

$$\rho(F,G) \le c_4\sqrt{\mathrm{E}_f\left(\frac{z_f(X;h)}{z_g(X;h)} - 1\right)^2}, \tag{3.14}$$

where c_4 depends on F and G.

However, in Corollaries 3.1 and 3.2, the constants depend on F and G. Therefore, the bounds are inappropriate in order to apply them to sequences of random variables. So, we give the following

Theorem 3.2. *Let f and g be as in Theorem 3.1. Furthermore, suppose that $g(a') = g(b') = 0$ and g is unimodal, so that $g'(x) \ge 0$ in (a', x_0) and $g'(x) \le 0$ in (x_0, b'), where x_0 is a mode of g. If h satisfies the conditions:*
(i) $\mathrm{E}_g[h] = h(x_0)$ *and*
(ii) *for all x in (a', b'), $|h(x) - h(x_0)| \ge \delta|g'(x)|/g(x)$ for some $\delta > 0$,*
then

$$\rho(F,G) \le \frac{c_1}{\delta}\mathrm{E}_f|z_f(X;h) - z_g(X;h)| + c_2|\mathrm{E}_f[h] - \mathrm{E}_g[h]|, \tag{3.15}$$

where c_1 and c_2 are as in Theorem 4.1 (they depend only on G).

Proof. First we show that $z_g(x;h) \ge \delta$. For x in (a', b'), consider the function

$$q(x) = \int_{a'}^{x} \left(\mathrm{E}_g[h] - h(t)\right)g(t)\,dt - \delta g(x).$$

Obviously, $q(a') = q(b') = 0$ and $q'(x) = \left(h(x_0) - h(x)\right)g(x) - \delta g'(x)$, which is non-negative for $x < x_0$ and non-positive for $x > x_0$. Hence, $q(x) \ge 0$, which is equivalent to $z_g(x;h) \ge \delta$. Since

$$\sup_x \sup_A |\psi'_A(x;h,g)| \le \frac{\sup_x \sup_A |\psi'_A(x;h,g)z_g(x;h)|}{\inf_x z_g(x;h)} \le \frac{c_1}{\delta},$$

the desired result follows from (3.10). $\square$

For example, if $g = \varphi$ (with mode 0), the result applies to any increasing h satisfying $\mathrm{E}_g[h] = h(0)$ and $|h(x) - h(0)| \ge \delta|x|$, while, for $\alpha > 2$ and $g(x) = x^{\alpha-1}e^{-x}/\Gamma(\alpha)$, $x > 0$, one can take h satisfying $h(\alpha-1) = \mathrm{E}_g[h]$ and $|h(x) - h(\alpha-1)| \ge \delta|1 - (\alpha-1)/x|$.

4. Illustrative Examples

In this section we illustrate the previous results with some examples.

Example 4.1. Let $F_n(x) = (1 - e^{-x}/n)^n$, $x > -\log(n)$, be the distribution of $X_n = \max\{Z_1, Z_2, \ldots, Z_n\} - \log(n)$, where Z_j are i.i.d. standard exponential random variables and $G(x) = \exp(-e^{-x})$, $-\infty < x < \infty$, is the limiting Gumbel distribution. In this case, the conditions of Lemma 2.2 are satisfied and one finds $c_1 = 2$, $c_2 = 1/\log(2)$. Since $g'(x)/g(x) = e^{-x} - 1$, $\mathrm{E}[|g'(X_n)|/g(X_n)] \le 1 + \mathrm{E}[e^{-X_n}] = 1 + n/(n+1) < \infty$ and thus Theorem 2.1 applies here. Therefore, $z_n(x) \equiv z_{f_n}(x; 1 - e^{-x}) = (n - e^{-x})/(n+1)$ for $x > -\log(n)$ (observe that z_n is positive, which also follows from the fact that $h(x) = -g'(x)/g(x)$ is increasing; see Lemma 2.1). Thus, from Theorem 2.1 we have (cf. Papadatos and Papathanasiou [10], Corollary 4.3)

$$\rho_n \equiv \rho(F_n, G) \le c_1 \mathrm{E}|z_n(X_n) - 1| + c_2 |\mathrm{E}[g'(X_n)/g(X_n)]|$$
$$= 2\mathrm{E}\Big[1 - \frac{n - e^{-X_n}}{n+1}\Big] + \frac{1}{\log(2)} |\mathrm{E}[e^{-X_n}] - 1| = \frac{1}{n+1}\Big(2 + \frac{2n}{n+1} + \frac{1}{\log(2)}\Big).$$

On the other hand,

$$\rho_n \ge G(0) - F_n(0) = e^{-1} - \Big(1 - \frac{1}{n}\Big)^n \ge \frac{e^{-1}}{2n},$$

which shows that the estimate (2.9) is of the correct order $1/n$.

Example 4.2. Let $U_{1:n} < U_{2:n} < \cdots < U_{n:n}$ be the ordered sample corresponding to a random sample of size n from the uniform $(0,1)$ distribution, and for $n \ge i > 1$ set $X_n = nU_{i:n}$ (i fixed). The density of X_n is given by

$$f_n(x) = [n^n B(i, n+1-i)]^{-1} x^{i-1} (1 - x)^{n-i}, \qquad 0 < x < n,$$

which, as $n \to \infty$, tends to the Erlang density

$$g(x) = [\Gamma(i)]^{-1} x^{i-1} e^{-x}, \qquad x > 0.$$

Let F_n and G be the corresponding d.f.'s. It is easily verified that the conditions of Lemma 2.2 are satisfied and $\mathrm{E}[-g'(X_n)/g(X_n)] = 0$. Thus, we simply get by Theorem 2.1 (in this case $z_n(x) = 1 - x/n$, $0 < x < n$)

$$\rho(F_n, G) \le c_1 \mathrm{E}[U_{i:n}] = ic_1/(n+1) = O(1/n).$$

Observe that $z_n(X_n)$ has the same d.f. as $1 - U_{i:n}$, which, of course, tends in probability to 1, as required by Theorem 2.3.

Example 4.3. Consider again X_n, F_n, and G as in the previous example, and set $h(x) = x^2$. In this case we have

$$\mathrm{E}_g(h) = i(i+1), \qquad z_g(x; h) = x(x + i + 1)$$

and

$$\mathrm{E}_{f_n}(h) = \frac{n^2}{(n+1)(n+2)} i(i+1), \qquad z_{f_n}(x;h) = \frac{x(n-x)[(n+1)x + n(i+1)]}{(n+1)(n+2)}.$$

Since the conditions of Lemma 3.1 are satisfied, Theorem 3.1 is applicable (for all fixed $i \geq 1$) and gives

$$\rho(F_n, G) \leq \frac{1}{(n+1)(n+2)} \Big\{ c_1 \mathrm{E}\Big(\frac{(n+1)X_n^2 + (ni + 3n + 2)X_n + (3n+2)(i+1)}{X_n + i + 1} \Big) + c_2(3n+2)i(i+1) \Big\}$$

$$\leq \frac{1}{(n+1)(n+2)} \Big[c_1 \mathrm{E}((n+1)X_n + 3n + 2) + c_2 i(i+1)(3n+2) \Big] = O(1/n),$$

while, if we take $h(x) = x$, we get from (3.11) (see also Papadatos and Papathanasiou [11], Examples 5.1 and 5.4)

$$\rho(F_n, G) \leq \frac{1}{n+1} \Big(\frac{n(i+1)+1}{n+1} c_1 + i c_2 \Big) = O(1/n).$$

5. Connections of the Covariance Identity With a Generalization of the Stein–Chen Approach

Consider the differential equation

$$z_g(x;h)\ell'(x) - \big(h(x) - \mathrm{E}_g[h]\big)\ell(x) = q(x) - \mathrm{E}_g[q], \tag{5.1}$$

where q belongs to a suitable family of functions $\mathcal{N}$, h is a given function and $z_g(x;h)$ is given by (3.2). Then, under appropriate conditions, the unique solution of (5.1) is

$$\ell(x) = \ell_q(x;h,g) = \frac{1}{z_g(x;h)g(x)} \int_{a'}^{x} (q(t) - \mathrm{E}_g[q]) g(t)\, dt. \tag{5.2}$$

For $g = \varphi$ and $h(x) = x$, we have $z_g(x;h) = w_\varphi(x) \equiv 1$. Thus (5.1) leads to the differential equation

$$\ell'(x) - x\ell(x) = q(x) - \mathrm{E}_\varphi[q], \tag{5.3}$$

that is, the basic equation of the Stein–Chen method for the normal approximation (see Stein [12]). If, for example, S_n is the standardized sum of n i.i.d. continuous random variables with mean zero and variance one and $\mathcal{N} = \{q \colon q(x) = e^{itx} = \cos tx + i \sin tx\}$, then, taking expectations in (5.3) with respect to the density f_n of S_n, we obtain

$$\mathrm{E}[\ell'(S_n)] - \mathrm{E}[S_n \ell(S_n)] = \varphi_{S_n}(t) - \varphi_Y(t), \tag{5.4}$$

and, from the covariance identity,

$$\mathrm{E}\big[(w_{S_n}(S_n)-1)\ell'(S_n)\big] = \varphi_Y(t)-\varphi_{S_n}(t), \tag{5.5}$$

where w_{S_n} is the w-function associated with f_n and $\varphi_Y(t)$, $\varphi_{S_n}(t)$ are the characteristic functions of the standard normal r.v. Y and the standardized sum S_n, respectively.

Since ℓ', the derivative of

$$\ell(x) = \frac{1}{\varphi(x)}\int_{-\infty}^{x}(e^{itu}-\varphi_Y(t))\varphi(u)\,du,$$

is absolutely bounded (over $\mathcal{N}$ and x) by a constant c (cf. Bolthausen [3] and CPU [7]), (5.5) gives

$$\sup_t|\varphi_{S_n}(t)-\varphi_Y(t)| \le c\mathrm{E}|w_{S_n}(S_n)-1| \le c\sqrt{\mathrm{Var}[w_{S_n}(S_n)]}. \tag{5.6}$$

Since $\mathrm{Var}[w_{S_n}(S_n)] \to 0$ under general conditions (in fact, when $\mathrm{Var}[w_{X_1}(X_1)] < \infty$; see CPU [6]) we have proved the following central limit theorem.

Theorem 5.1. *Under the above conditions, the sequence $\varphi_{S_n}(t)$ of the characteristic functions of the standardized sums tends to the characteristic function of the standard normal.*

Observe that the same arguments, applied to $\mathcal{N} = \{q : |q(x)| \le C < \infty\}$, lead to the total variation convergence of f_n to φ (see Billingsley [2]).

For g an arbitrary density and $h = -g'/g$, (5.1) reduces to the equation

$$\ell'(x) - (-g'(x)/g(x))\ell(x) = q(x) - \mathrm{E}_g[q], \tag{5.7}$$

provided that $\mathrm{E}_g[-g'/g] = 0$ and $g(a') = g(b') = 0$ (cf. (2.7)), while, if h is an arbitrary increasing function, (5.1) provides an extension of (3.6) (note that for $\mathcal{N} = \{q : q(x) = I_A(x),\ A \text{ Borel}\}$, (5.7) and (5.1) is the basis of the results proved in Section 2 and 3, respectively).

Taking expectations in (5.1) with respect to the density f and assuming that $\mathrm{E}_f[h] = \mathrm{E}_g[h]$, we have (using identity (2.1))

$$\mathrm{E}_f[q] - \mathrm{E}_g[q] = \mathrm{E}_f\big[(z_g - z_f)\ell'\big]. \tag{5.8}$$

Under the assumption that $\sup_q \mathrm{E}_f[\ell']^2 \le c^2 < \infty$ (where the supremum is taken over $q \in \mathcal{N}$), we have the following stronger result stated as

Theorem 5.2. *Under the above conditions we have*

$$\sup_q\big|\mathrm{E}_f[q]-\mathrm{E}_g[q]\big| \le c\sqrt{\mathrm{E}_f\big[z_f(X;h)-z_g(X;h)\big]^2}. \tag{5.9}$$

Furthermore, the quantity $\mathrm{E}_f\big[(z_g - z_f)\ell'\big]$ in (5.8), in virtue of the covariance identity, can be written as

$$\mathrm{Cov}_f[h^*, \ell] = \mathrm{E}_f[z^*\ell'], \tag{5.10}$$

where $z^* = z_g - z_f$ and h^* is given by the relation

$$z^*(x)f(x) = \int_a^x \left(\mathrm{E}_f[h^*] - h^*(t)\right) f(t)\, dt. \tag{5.11}$$

Therefore, we have the following

Theorem 5.3. *Under the above conditions,*

$$\sup_q \left|\mathrm{E}_f[q] - \mathrm{E}_g[q]\right| \le c^* \sqrt{\mathrm{Var}_f[h^*]}, \tag{5.12}$$

where $c^* = \sup_q \sqrt{\mathrm{E}_f[\ell - \mathrm{E}_f(\ell)]^2}$, *provided that* $c^* < \infty$.

From the above discussion, we see that there are many possibilities of choosing h. It would be of some interest to find the function h which gives the best approximation (in some sense) of an arbitrary g by f.

References

[1] P. Billingsley, *Convergence of Probability Measures*, Wiley, New York, 1968.
[2] P. Billingsley, *Probability and Measure*, 2nd ed., Wiley, New York, 1986.
[3] E. Bolthausen, *An estimate of the remainder in a combinatorial central limit theorem*, Z. Wahrsch. verw. Gebiete, 66 (1984), 379–386.
[4] T. Cacoullos and V. Papathanasiou, *Characterizations of distributions by variance bounds*, Statist. Probab. Lett., 7 (1989), 351–356.
[5] T. Cacoullos and V. Papathanasiou, *A generalization of covariance identity and related characterizations*, Math. Methods Statist., 4 (1995), 106–113.
[6] T. Cacoullos, V. Papathanasiou, and S. Utev, *Another characterization of the normal law and a proof of the central limit theorem connected with it*, Theory Probab. Appl., 37 (1993), 581–588.
[7] 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.
[8] H. M. Hudson, *A natural identity for exponential families with applications in multiparameter estimation*, Ann. Statist., 6 (1978), 473–484.
[9] E. Mayer-Wolf, *The Cramer-Rao functional and limiting laws*, Ann. Probab., 18 (1990), 840–850.
[10] N. Papadatos and V. Papathanasiou, *Distance in variation and a Fisher-type information*, Math. Methods Statist., 4 (1995), 230–237.
[11] N. Papadatos and V. Papathanasiou, *Distance in variation between two arbitrary distributions via the associated w-functions*, Theory Probab. Appl., 40 (1995), 685–694.
[12] V. Papathanasiou and S. Utev, *Integro-differential inequalities and the Poisson approximation*, Siberian Advances Math., 5 (1995), 120–132.
[13] C. M. Stein, *A bound for the error in the normal approximation to the distribution of a sum of dependent random variables*, In: *Proc. Sixth Berkeley Symp. Math. Statist. Probab.*, Vol. 2, Univ. California Press, Berkeley, 1972, pp. 583–602.
[14] C. M. Stein, *Estimation of the mean of a multivariate normal distribution*, Stanford Univ. Technical Report No. 48, 1973.
[15] C. M. Stein, *Estimation of the mean of a multivariate normal distribution*, Ann. Statist., 9 (1981), 1135–1151.

[*Received November 1996*]