Journal of Multivariate Analysis **67**, 154–168 (1998)
Article No. MV981758

Variational Inequalities for Arbitrary Multivariate Distributions*

N. Papadatos

University of Cyprus, 1678 Nicosia, Cyprus

and

V. Papathanasiou

University of Athens, 157 84 Athens, Greece

Received August 1, 1996; revised September 19, 1997

Upper bounds for the total variation distance between two arbitrary multivariate distributions are obtained in terms of the corresponding **w**-functions. The results extend some previous inequalities satisfied by the normal distribution. Some examples are also given. © 1998 Academic Press

AMS 1991 subject classifications: primary 60E15, 60E05; secondary 60F15.

Key words and phrases: Total variation distance, multivariate distributions, **w**-functions.

1. INTRODUCTION

In a recent paper, Cacoullos, Papathanasiou, and Utev (1994) obtained some upper bounds for the total variation distance between an arbitrary continuous d.f. and the standard normal one. Utev (1989) had also obtained upper bounds for the total variation distance using integrodifferential inequalities. More recently, Papadatos and Papathanasiou (1995a, 1995b) extended the above results in two directions; they gave analogous upper bounds for arbitrary (continuous and discrete) d.f.'s (a) via the corresponding w-functions and (b) in terms of a Fisher-type information (see also Mayer-Wolf (1990) for a similar approach to the normal). Some applications of the above results were also given.

* This work was partially supported by the Ministry of Industry, Energy and Technology of Greece under Grant 1369.

0047-259X/98 $25.00

Papathanasiou (1996) extended the results of Cacoullos, Papathanasiou, and Utev (1994) to the multivariate case. Specifically, he showed under general conditions imposed on the d.f. F that

$$\sup_A |F(A)-\Phi(A)| \equiv \rho(F, \Phi) \leqslant 2 \sum_{i=1}^{p} \mathrm{E}\, |w^i(\mathbf{X})-1|, \tag{1.1}$$

where the supremum is taken over the Borel sets of R^p, Φ is the d.f. of the standard normal variate with independent components, $\mathbf{X}=(X_1, ..., X_p)'$ is an arbitrary standardized p-dimensional random vector with d.f. F and density f, and $\mathbf{w}(\mathbf{X})=(w^1(\mathbf{X}), ..., w^p(\mathbf{X}))'$ is the $\mathbf{w}$-function associated with $\mathbf{X}$ (see (3.1), below). Furthermore, as a by-product of (1.1), the multivariate CLT was obtained and, moreover, the rate of convergence in the CLT was investigated by Cacoullos, Papadatos, and Papathanasiou (1997); for other results on the rate of convergence see also Gotze (1991) and Sweeting (1977). Papathanasiou (1996) obtained also some upper bounds for the total variation distance between an arbitrary multivariate d.f. F (with density f) and the multivariate standard normal one with density φ, in terms of the (relative) Fisher information matrix of f with respect to φ (cf. Barron, 1986).

In the present paper the above results are extended to the general case of an arbitrary density g instead of φ; it is proved that, under general conditions, the total variation distance between two arbitrary continuous d.f.'s F and G is bounded by some expressions of the form (1.1).

When G is discrete and has independent components, the main result is presented in Theorem 2.1, while the continuous case (part of which has been discussed in Papathanasiou, 1996) is studied in Section 3 (Theorems 3.1, 3.2, and Corollary 3.1). Finally, some particular illustrative examples are presented in Section 4.

It should be noted that upper bounds for the total variation distance between multivariate d.f.'s in terms of a generalized Fisher-type information matrix are also valid for a quite large class of multivariate d.f.'s and they will be the object of a future paper.

2. TOTAL VARIATION DISTANCE FOR DISCRETE RANDOM VECTORS

In order to obtain the upper bounds for the total variation distance between two arbitrary discrete random vectors, we first give the definition of the $\mathbf{w}$-function in this case.

Consider a random vector $\mathbf{X}=(X_1, ..., X_p)'$ with d.f. F and probability function $f(\mathbf{x})$ supported by a "convex" set $C^p \subset \{0, 1, ...\}^p$ such that

$(0, 0, ..., 0)' \in C^p$ (in the sense that if $\mathbf{x} = (x_1, ..., x_p)' \in C^p$ then $\{0, ..., x_1\} \times \cdots \times \{0, ..., x_p\} \subset C^p$). Assume that the mean vector $\boldsymbol{\mu}$ and the dispersion matrix Σ of $\mathbf{X}$ are well-defined ($\Sigma > 0$) and consider the linear functions $q^i(\mathbf{x}) = \sum_{j=1}^{p} \sigma_{ij}^* x_j$ where $(\sigma_{ij}^*) = \Sigma^{-1}$. Them the $\mathbf{w}$-function of $\mathbf{X}$ is defined for every $\mathbf{x}$ in C^p by $\mathbf{w}(\mathbf{x}) = (w^1(\mathbf{x}), ..., w^p(\mathbf{x}))'$ with

$$w^i(\mathbf{x})\, f(\mathbf{x}) = \sum_{k=0}^{x_i} (\mu^i - q^i(\mathbf{u}_i, k, \mathbf{v}_i))\, f(\mathbf{u}_i, k, \mathbf{v}_i), \tag{2.1}$$

where $\mu^i = \mathrm{E}[q^i(\mathbf{X})]$, $\mathbf{u}_i = (x_1, ..., x_{i-1})'$, and $\mathbf{v}_i = (x_{i+1}, ..., x_p)'$ for $i = 1, ..., p$ (see Cacoullos and Papathanasiou, 1992).

In this section, we derive upper bounds for the total variation distance between F and G, only when G is a discrete d.f. with independent components; similar bounds can also be found for the general case, but we will not investigate them because of their complexity and since the most interesting situations suggest convergence to a limiting d.f. with independent components (like the multivariate Poisson; see Corollary 2.1 below and Examples 1 and 2 of Section 4).

Cacoullos and Papathanasiou (1992) established the identity

$$\mathrm{Cov}[q^i(\mathbf{X}), \psi(\mathbf{X})] = \mathrm{E}[w^i(\mathbf{X})\, \Delta_i \psi(\mathbf{X})], \tag{2.2}$$

where ψ is an arbitrary function defined on C^p such that the two sides of (2.2) exist, and Δ_i denotes the ith partial difference operator

$$\Delta_i \psi(\mathbf{x}) = \psi(x_1, ..., x_i + 1, ..., x_p) - \psi(\mathbf{x}).$$

Consider now another d.f., $G = G_1 \cdot \cdots \cdot G_p$, with probability function $g = g_1 \cdot \cdots \cdot g_p$ supported on a set of the form $I^p = \{0, ..., b_1\} \times \cdots \times \{0, ..., b_p\}$, where $0 < b_i \leqslant \infty$. It is furthermore assumed that the mean m_i and the variance s_i^2 of g_i exist. Then, according to the definition (2.1), the $\mathbf{w}$-function of g exists and it is given by $\mathbf{w}_g(\mathbf{x}) = (w_g^1(x_1), ..., w_g^p(x_p))'$, where

$$s_i^2 w_g^i(x_i)\, g_i(x_i) = \sum_{k=0}^{x_i} (m_i - k)\, g_i(k), \qquad \text{for} \quad x_i \in \{0, ..., b_i\}.$$

Assume that h is a given bounded function defined on I^p and consider the special function (cf. Papathanasiou, 1996; Papadatos and Papathanasiou, 1995a)

$$\psi^i(\mathbf{x}) = \frac{1}{g_i(x_i - 1)\, w_g^i(x_i - 1)} \sum_{k=0}^{x_i - 1} (\mathrm{E}_g[h \mid \mathbf{v}_i] - \mathrm{E}_g[h \mid k, \mathbf{v}_i])\, g_i(k), \tag{2.3}$$

defined for all $\mathbf{x} \in I^p$ such that $x_i \geqslant 1$ (for simplicity, set $\psi^i(\mathbf{x}) \equiv 0$ for all other $\mathbf{x}$), where $\mathbf{v}_i = (x_{i+1}, ..., x_p)'$ as before. The following lemma gives sufficient conditions for ψ^i and $\Delta_j \psi^i$ to be uniformly bounded and will be used in the sequel.

LEMMA 2.1. *If $b_i < +\infty$ or*

$$b_i = +\infty \quad \textit{and} \quad \limsup_{x_i \to +\infty} \frac{1 - G_i(x_i)}{w_g^i(x_i)\, g_i(x_i)} < \infty,$$

then, there exist finite constants c_i such that

$$|\psi^i(\mathbf{x})| \leqslant (b - \alpha)\, c_i \tag{2.4}$$

and

$$\Delta_j \psi^i(\mathbf{x}) \equiv 0 \textit{ for } j < i \qquad \textit{while} \qquad |\Delta_j \psi^i(\mathbf{x})| \leqslant 2(b - \alpha)\, c_i \textit{ for } j \geqslant i \tag{2.5}$$

for all $\mathbf{x}$ in I^p and all bounded h with $\alpha \leqslant h(\mathbf{x}) \leqslant b$.

Proof. Observe that

$$\sum_{k=0}^{x_i - 1} (\mathrm{E}_g[h \mid \mathbf{v}_i] - \mathrm{E}_g[h \mid k, \mathbf{v}_i])\, g_i(k)$$
$$= \sum_{k=x_i}^{b_i} (\mathrm{E}_g[h \mid k, \mathbf{v}_i] - \mathrm{E}_g[h \mid \mathbf{v}_i])\, g_i(k)$$

and thus

$$|\psi^i(\mathbf{x})| \leqslant (b - \alpha) \frac{\min\{G_i(x_i - 1), 1 - G_i(x_i - 1)\}}{w_g^i(x_i - 1)\, g_i(x_i - 1)}. \tag{2.6}$$

The rest of the proof for (2.4) follows by using exactly the same arguments as those of Lemma 3.2 in Papadatos and Papathanasiou (1995a), while (2.5) is a simple consequence of (2.4) and the fact that for $j < i$, ψ^i is independent of x_j.

In the case of a multivariate Poisson d.f. where

$$g(\mathbf{x}) = \prod_{i=1}^{p} g_i(x_i) = \prod_{i=1}^{p} e^{-\lambda_i} \frac{\lambda_i^{x_i}}{x_i!}, \qquad \lambda_i > 0, \quad x_i = 0, 1, ..., \quad i = 1, ..., p, \tag{2.7}$$

we have $w_g^i(x_i) \equiv 1$ and the conditions of Lemma 2.1 are obviously satisfied. However, we can go further if we apply the results of Barbour, Holst, and Janson (1992) for a special choice of h as an indicator function.

LEMMA 2.2. *If g is a multivariate Poisson as in* (2.7) *and* $h(\mathbf{x}) = I_A(\mathbf{x})$, *where A is an arbitrary subset of* $\{0, 1, ...\}^p$, *then*

$$|\psi^i(\mathbf{x})| \leqslant \min\{\lambda_i, \sqrt{\lambda_i}\} \tag{2.8}$$

and

$$|\Delta_i \psi^i(\mathbf{x})| \leqslant 1 - e^{-\lambda_i} \qquad \text{while for } j > i, \qquad |\Delta_j \psi^i(\mathbf{x})| \leqslant 2 \min\{\lambda_i, \sqrt{\lambda_i}\}. \tag{2.9}$$

Proof. Fix $\mathbf{x} \in A$ and consider the set

$$A_i \equiv A_i(x_1, ..., x_{i-1}, x_{i+1}, ..., x_p) = \{k : (\mathbf{u}_i, k, \mathbf{v}_i) \in A\}.$$

It is not hard to verify that

$$\psi^i(\mathbf{x}) = -\lambda_i \, \mathrm{E}_{\lambda_1, ..., \lambda_{i-1}}[g_{\lambda_i, A_i}(x_i)], \tag{2.10}$$

where $g_{\lambda, A}(\cdot)$ is given by (1.10) in Barbour, Holst, and Janson (1992, p. 6) and the expectation in (2.10) is taken with respect to the first $i-1$ components of the multivariate Poisson defined in (2.7). Similarly,

$$\Delta_i \psi^i(\mathbf{x}) = -\lambda_i \, \mathrm{E}_{\lambda_1, ..., \lambda_{i-1}}[\Delta_i g_{\lambda_i, A_i}(x_i)]$$

and the estimates (2.8) and (2.9) follow from Lemma 1.1.1 in Barbour, Holst, and Janson (1992, p. 7).

The main result of this section is given in the following

THEOREM 2.1. *Let G be as in Lemma* 2.1 *and consider an arbitrary d.f. F with probability function f satisfying the above conditions. Suppose also that* $C^p \subset I^p$ *and that f has mean vector* $\boldsymbol{\mu} = (\mu_1, ..., \mu_p)'$ *and positive definite dispersion matrix* $\Sigma = (\sigma_{ij})$ *with diagonal elements* $\sigma_{ii} = \sigma_i^2$. *Then, there exist constants c_i (depending only on G_i) such that*

$$\rho(F, G) \leqslant \sum_{i=1}^{p} 2 \frac{c_i}{s_i^2} \mathrm{E}_f |s_i^2 w_g^i(X_i) - \sigma_i^2 w^i(\mathbf{X})| + \sum_{i=1}^{p} c_i \frac{|\mu_i - m_i|}{s_i^2}$$

$$+ \sum\sum_{1 \leqslant i < j \leqslant p} 2 \frac{c_i}{s_i^2} |\sigma_{ij}| \, \mathrm{E}_f |w^j(\mathbf{X})|, \tag{2.11}$$

where $\mathbf{w}$ *is given by* (2.1) *and the total variation distance in this case is given by*

$$\rho(F, G) = \sup_{A} |F(A) - G(A)|,$$

where the supremum is taken over the sets $A \subset \{0, 1, ...\}^p$.

Proof. Taking forward differences with respect to x_i in (2.3) we have

$$\mathrm{E}_g[h \mid \mathbf{v}_i] - \mathrm{E}_g[h \mid \mathbf{v}_{i-1}] = w_g^i(x_i)\, \Delta_i \psi^i(\mathbf{x}) - \frac{x_i - m_i}{s_i^2}\, \psi^i(\mathbf{x})$$

and thus, adding for all i,

$$\mathrm{E}_g[h] - h(\mathbf{x}) = \sum_{i=1}^{p} \left[w_g^i(x_i)\, \Delta_i \psi^i(\mathbf{x}) - \frac{x_i - m_i}{s_i^2}\, \psi^i(\mathbf{x}) \right].$$

Taking expectations with respect to f we get

$$\begin{aligned}
\mathrm{E}_g[h] - \mathrm{E}_f[h] &= \sum_{i=1}^{p} \left[\mathrm{E}_f[w_g^i(X_i)\, \Delta_i \psi^i(\mathbf{X})] - \frac{1}{s_i^2}\, \mathrm{E}_f[(X_i - m_i)\, \psi^i(\mathbf{X})] \right] \\
&= \sum_{i=1}^{p} \mathrm{E}_f[w_g^i(X_i)\, \Delta_i \psi^i(\mathbf{X})] + \sum_{i=1}^{p} \frac{m_i - \mu_i}{s_i^2}\, \mathrm{E}_f[\psi^i(\mathbf{X})] \\
&\quad - \sum_{i=1}^{p} \frac{1}{s_i^2}\, \mathrm{E}_f[(X_i - \mu_i)\, \psi^i(\mathbf{X})].
\end{aligned}$$

Applying identity (2.2) for $\psi = \psi^i$ (a bounded function) in combination with the identity

$$X_i - \mu_i = \sum_{j=1}^{p} \sigma_{ij}(q^j(\mathbf{X}) - \mu^j)$$

in the last summand, we conclude that

$$\begin{aligned}
&\mathrm{E}_f[(X_i - \mu_i)\, \psi^i(\mathbf{X})] \\
&\qquad = \sigma_i^2\, \mathrm{E}_f[w^i(\mathbf{X})\, \Delta_i \psi^i(\mathbf{X})] + \sum_{j \neq i} \sigma_{ij}\, \mathrm{E}_f[w^j(\mathbf{X})\, \Delta_j \psi^i(\mathbf{X})]
\end{aligned}$$

so that

$$
\begin{aligned}
&\mathrm{E}_g[h]-\mathrm{E}_f[h]\\
&\quad=\sum_{i=1}^{p}\frac{1}{s_i^2}\,\mathrm{E}_f[\Delta_i\psi^i(\mathbf{X})(s_i^2 w_g^i(X_i)-\sigma_i^2 w^i(\mathbf{X}))]\\
&\qquad+\sum_{i=1}^{p}\frac{m_i-\mu_i}{s_i^2}\,\mathrm{E}_f[\psi^i(\mathbf{X})]-\sum_{i=1}^{p}\frac{1}{s_i^2}\sum_{j\neq i}\sigma_{ij}\,\mathrm{E}_f[w^j(\mathbf{X})\,\Delta_j\psi^i(\mathbf{X})].
\end{aligned}
$$

Now, (2.11) follows from Lemma 2.1 by choosing $h(\mathbf{x})=I_A(\mathbf{x})$, where A is an arbitrary subset of $\{0, 1, ...\}^p$.

By using Lemma 2.2 for the multivariate Poisson (2.7) we have the following bound.

Corollary 2.1. *If $G=\mathrm{P}_{\boldsymbol{\lambda}}$ is the multivariate Poisson* (2.7), *then for any arbitrary d.f. F,*

$$
\begin{aligned}
\rho(F,\mathrm{P}_{\boldsymbol{\lambda}})\leqslant&\sum_{i=1}^{p}\frac{1-e^{-\lambda_i}}{\lambda_i}\,\mathrm{E}_f\,|\lambda_i-\sigma_i^2 w^i(\mathbf{X})|+\sum_{i=1}^{p}\min\{1,\lambda_i^{-1/2}\}\;|\mu_i-\lambda_i|\\
&+2\sum\sum_{1\leqslant i<j\leqslant p}\min\{1,\lambda_i^{-1/2}\}\;|\sigma_{ij}|\;\mathrm{E}_f\,|w^j(\mathbf{X})|. \qquad (2.12)
\end{aligned}
$$

It should be noted that the general case of a discrete random vector $\mathbf{X}$ with correlated components has been treated here, since a linear transform of $\mathbf{X}$ to another random vector $\mathbf{Z}$ with uncorrelated components would lead (in general) to a non-integer valued $\mathbf{Z}$. This is the reason for using formulas involving the covariances in (2.11) and (2.12), in contrast to the continuous analogue (see Theorem 3.2 and Corollary 3.1, below).

3. CONTINUOUS CASE

Suppose that the p-dimensional random vector $\mathbf{X}=(X_1, ..., X_p)'$ has a density f supported by a convex set C^p (in the sense that $f(\mathbf{x}_1)>0$ and $f(\mathbf{x}_2)>0$ implies $f(\lambda\mathbf{x}_1+(1-\lambda)\,\mathbf{x}_2)>0$ for all $\lambda\in[0,1]$) and dispersion matrix $\Sigma>0$. The $\mathbf{w}$-function of $\mathbf{X}$, $\mathbf{w}(\mathbf{x})=(w^1(\mathbf{x}), ..., w^p(\mathbf{x}))'$, is defined, as in the discrete case, for every $\mathbf{x}$ in the support of f by the relations (see Cacoullos and Papathanasiou, 1992)

$$
w^i(\mathbf{x})\,f(\mathbf{x})=\int_{-\infty}^{x_i}(\mu^i-q^i(\mathbf{u}_i,t,\mathbf{v}_i))\,f(\mathbf{u}_i,t,\mathbf{v}_i)\,dt,\qquad i=1,...,p, \qquad (3.1)
$$

where $q^i(\mathbf{x})$, μ^i, $\mathbf{u}_i$ and $\mathbf{v}_i$ are as in the discrete case (see (2.1)).

As in the discrete case, Cacoullos and Papathanasiou (1992) established the identity

$$\mathrm{Cov}[q^i(\mathbf{X}), \psi(\mathbf{X})] = \mathrm{E}[w^i(\mathbf{X})\, \psi_i(\mathbf{X})] \tag{3.2}$$

satisfied by any function ψ defined on C^p with $\nabla\psi = (\psi_1, ..., \psi_p)'$, provided that both sides of (3.2) exist and that $w^i(\mathbf{x})\, f(\mathbf{x}) \to 0$ monotonically as $\mathbf{x}$ approaches any boundary point of C^p. It should be noted that Cacoullos and Papathanasiou proved (3.2) when the support of f is a p-rectangle, though their assertions continue to hold also for the case of a general convex support.

An interesting application of (3.2) arises when $\mathbf{X}$ is normal. In this case, $w^i(\mathbf{x}) \equiv 1$ for all i and $\mathbf{x}$ and thus (3.2) provides a tool for obtaining lower variance bounds (and the corresponding characterizations) in terms of the partial derivatives of ψ (cf. Chernoff (1981) and Chen (1982) for upper bounds in the case of independence; see also Hudson (1978) and Chou (1988) for analogous identities involving the multivariate exponential family).

Now let G be another arbitrary d.f. with density g supported by a convex set E^p such that $C^p \subset E^p$. We further assume that g has mean vector $\mathbf{m} = (m_1, ..., m_p)'$ and dispersion matrix $S = (s_{ij}) > 0$. For any bounded function h and all $\mathbf{v}_{i-1} \in E^{p-i+1}$ (the projection of E^p to the $p-i+1$ last components), we define the functions (cf. Papathanasiou, 1996; Papadatos and Papathanasiou, 1995a)

$$\psi^{(i)}(\mathbf{v}_{i-1}) = \frac{1}{g(\mathbf{v}_{i-1})\, w_g^{(i)}(\mathbf{v}_{i-1})} \times \int_{-\infty}^{x_i} (\mathrm{E}_g[h \mid \mathbf{v}_i] - \mathrm{E}_g[h \mid t, \mathbf{v}_i])\, g(t, \mathbf{v}_i)\, dt, \tag{3.3}$$

where $g(x_i, \mathbf{v}_i) = g(\mathbf{v}_{i-1})$ is the marginal density of the last $p-i+1$ components of g and

$$g(x_i, \mathbf{v}_i)\, w_g^{(i)}(x_i, \mathbf{v}_i) = \int_{-\infty}^{x_i} (m^{(i)} - q^{(i)}(t, \mathbf{v}_i))\, g(t, \mathbf{v}_i)\, dt, \tag{3.4}$$

where $q^{(i)}(x_i, \mathbf{v}_i) = \sum_{j=i}^{p} s^*_{ij} x_j$, $S^{-1} = (s^*_{ij})$, $m^{(i)} = \mathrm{E}_g[q^{(i)}]$. Then, one can easily establish the following

Lemma 3.1. *If*

$$\sup_{\mathbf{v}_{i-1}} \frac{\min\{G(x_i \mid \mathbf{v}_i), 1 - G(x_i \mid \mathbf{v}_i)\}}{g(x_i \mid \mathbf{v}_i)\, |w_g^{(i)}(\mathbf{v}_{i-1})|} < \infty \tag{3.5}$$

and

$$\sup_{\mathbf{v}_{i-1}} \frac{|q^{(i)}(\mathbf{v}_{i-1}) - m^{(i)}| \min\{G(x_i \mid \mathbf{v}_i), 1 - G(x_i \mid \mathbf{v}_i)\}}{g(x_i \mid \mathbf{v}_i) \, |w_g^{(i)}(\mathbf{v}_{i-1})|} < \infty, \tag{3.6}$$

where $g(x_i \mid \mathbf{v}_i)$ *is the conditional density of the* i*th component of* g *given the* $p-i$ *last components,* $G(x_i \mid \mathbf{v}_i)$ *is the corresponding conditional d.f. and the supremum in* (3.5) *and* (3.6) *is taken over* $\mathbf{v}_{i-1} \in E^{p-i+1}$, *then, there exist finite constants* c_i *and* c_i' *such that, for all* $\mathbf{v}_{i-1} \in E^{p-i+1}$ *and all bounded* h *with* $\alpha \leqslant h(\mathbf{x}) \leqslant b$,

$$|\psi^{(i)}(\mathbf{v}_{i-1})| \leqslant (b-\alpha)\, c_i' \qquad \text{and} \qquad |w_g^{(i)}(\mathbf{v}_{i-1}) \, \psi_i^{(i)}(\mathbf{v}_{i-1})| \leqslant (b-\alpha)\, c_i, \tag{3.7}$$

where $\psi_i^{(i)} = \partial \psi^{(i)} / \partial x_i$.

Proof. Observe that

$$\int_{-\infty}^{+\infty} (\mathrm{E}_g[h \mid \mathbf{v}_i] - \mathrm{E}_g[h \mid t, \mathbf{v}_i]) \, g(t, \mathbf{v}_i) \, dt = 0$$

and thus,

$$|\psi^{(i)}| \leqslant (b-\alpha) \frac{\min\{G(x_i \mid \mathbf{v}_i), 1 - G(x_i \mid \mathbf{v}_i)\}}{g(x_i \mid \mathbf{v}_i) \, |w_g^{(i)}(\mathbf{v}_{i-1})|}.$$

On the other hand, by (3.3),

$$\begin{aligned} &w_g^{(i)}(\mathbf{v}_{i-1}) \, \psi_i^{(i)}(\mathbf{v}_{i-1}) \\ &\quad = (q^{(i)}(\mathbf{v}_{i-1}) - m^{(i)}) \, \psi^{(i)}(\mathbf{v}_{i-1}) + \mathrm{E}_g[h \mid \mathbf{v}_i] - \mathrm{E}_g[h \mid \mathbf{v}_{i-1}]. \end{aligned} \tag{3.8}$$

Therefore,

$$|w_g^{(i)} \psi_i^{(i)}| \leqslant (b-\alpha) \frac{|q^{(i)} - m^{(i)}| \min\{G(x_i \mid \mathbf{v}_i), 1 - G(x_i \mid \mathbf{v}_i)\}}{|w_g^{(i)}| \, g(x_i \mid \mathbf{v}_i)} + (b-\alpha)$$

and (3.7) follows from (3.5) and (3.6).

We can now state the following

THEOREM 3.1. *Under the preceding conditions and, furthermore,* $\boldsymbol{\mu} = \mathbf{m}$ *and* $\Sigma = S$, *there exist constants* c_i *(depending only on* G*) such that*

$$\rho(F, G) \leqslant \sum_{i=1}^{p} c_i \, \mathrm{E}_f \left| \frac{w^{(i)}(X_i, ..., X_p)}{w_g^{(i)}(X_i, ..., X_p)} - 1 \right|, \tag{3.9}$$

where $w_g^{(i)}(x_i, ..., x_p)$ *is given by* (3.4) *and* $w^{(i)}(x_i, ..., x_p)$ *is defined by*

$$f(x_i, ..., x_p)\, w^{(i)}(x_i, ..., x_p) = \int_{-\infty}^{x_i} (\mu^{(i)} - q^{(i)}(t, \mathbf{v}_i))\, f(t, \mathbf{v}_i)\, dt,$$

where $q^{(i)}(x_i, \mathbf{v}_i) = \sum_{j=i}^{p} s_{ij}^* x_j = \sum_{j=i}^{p} \sigma_{ij}^* x_j$, $S^{-1} = (s_{ij}^*) = (\sigma_{ij}^*) = \Sigma^{-1}$, *and* $m^{(i)} = \mathrm{E}_g[q^{(i)}] = \mathrm{E}_f[q^{(i)}] = \mu^{(i)}$ (*i.e.,* $w^{(i)}$ *is the first component of the* **w**-*function associated with the marginal density* $f(x_i, ..., x_p)$; *cf.* (3.1)).

Proof. Adding (3.8) for all i and taking expectations with respect to f we get (here $\mathbf{V}_i = (X_{i+1}, ..., X_p)'$)

$$\begin{aligned}
&\mathrm{E}_g[h] - \mathrm{E}_f[h] \\
&\quad = \sum_{i=1}^{p} [\mathrm{E}_f[w_g^{(i)}(X_i, \mathbf{V}_i)\, \psi_i^{(i)}(X_i, \mathbf{V}_i)] - \mathrm{Cov}_f[q^{(i)}(X_i, \mathbf{V}_i), \psi^{(i)}(X_i, \mathbf{V}_i)]] \\
&\quad = \sum_{i=1}^{p} \mathrm{E}_f\left[w_g^{(i)}(X_i, \mathbf{V}_i)\, \psi_i^{(i)}(X_i, \mathbf{V}_i) \left(1 - \frac{w^{(i)}(X_i, \mathbf{V}_i)}{w_g^{(i)}(X_i, \mathbf{V})}\right)\right],
\end{aligned}$$

by (3.2) for $\psi = \psi^{(i)}$ applied to the marginal density $f(x_i, ..., x_p)$. Taking $h(\mathbf{x}) = I_A(\mathbf{x})$, where A is an arbitrary Borel set of R^p, and applying Lemma 3.1 we conclude (3.9).

Similarly, one can easily show the following

THEOREM 3.2. *If* G *is as in Lemma* 3.1 *and the matrices* S *and* Σ *are diagonal with elements* $s_1^2, ..., s_p^2$ *and* $\sigma_{11}, ..., \sigma_{pp}$, *respectively, then there exist constants* c_i *and* c_i' (*depending only on* G) *such that*

$$\rho(F, G) \leqslant \sum_{i=1}^{p} c_i\, \mathrm{E}_f \left| \frac{\sigma_{ii} w^i(\mathbf{X})}{s_i^2 w_g^{(i)}(X_i, ..., X_p)} - 1 \right| + \sum_{i=1}^{p} c_i' \frac{|\mu_i - m_i|}{s_i^2}, \tag{3.10}$$

where $w_g^{(i)}(x_i, ..., x_p)$ *is given by* (3.4) *and* $w^i(\mathbf{x})$ *is given by* (3.1).

The proof of Theorem 3.2 is similar to that of Theorem 2.1 in Papathanasiou (1996) and is omitted. A special case is considered in the following

COROLLARY 3.1. *If the d.f.* G *is as in Theorem* 3.2 *and moreover it has independent components* (*i.e*, $G = G_1 \cdot \cdots \cdot G_p$ *with density* $g = g_1 \cdot \cdots \cdot g_p$), *then there exist constants* c_i *and* c_i' (*depending only on* G_i) *such that*

$$\rho(F, G) \leqslant \sum_{i=1}^{p} c_i\, \mathrm{E}_f \left| \frac{\sigma_{ii} w^i(\mathbf{X})}{s_i^2 w_g^i(X_i)} - 1 \right| + \sum_{i=1}^{p} c_i' \frac{|\mu_i - m_i|}{s_i^2}, \tag{3.11}$$

where $\mathbf{w}$ *is given by* (3.1) *and* $w_g^i(x_i)$ *is the* (*univariate*) *w-function associated with the density* g_i.

4. EXAMPLES

Here we present some illustrative examples, applying the above results.

EXAMPLE 1. Let $\mathbf{X}_n$ be the multinomial random vector $\mathrm{MN}(n; \theta_1, ..., \theta_p)$, where the parameters $\theta_i = \theta_i(n) = \lambda_i/n$ ($\lambda_i > 0$ fixed) for all $i = 1, ..., p$. The $\mathbf{w}$-function of $\mathbf{X}_n$ is given by (see Cacoullos and Papathanasiou, 1992)

$$w^i(\mathbf{x}) = \frac{n - x_1 - \cdots - x_p}{n(1 - \theta_1 - \cdots - \theta_p)} = \frac{n - x}{n - \lambda}, \qquad i = 1, ..., p,$$

where $x = x_1 + \cdots + x_p$ and $\lambda = \lambda_1 + \cdots + \lambda_p$. The support of $\mathbf{X}_n$ is obviously the "convex" set $C^p = \{x_1 \geqslant 0, ..., x_p \geqslant 0 : x_1 + \cdots + x_p \leqslant n\}$ and, of course, $\mu_i = \lambda_i$, $\sigma_i^2 = \lambda_i(1 - \lambda_i/n)$ and $\sigma_{ij} = -\lambda_i \lambda_j/n$. Since $X = X_1 + \cdots + X_p$ has a Binomial d.f. $\mathrm{Bi}(n; \lambda/n)$ and $w^i(\mathbf{x}) \geqslant 0$ for all $\mathbf{x}$ in C^p, we have

$$\mathrm{E}\,|w^i(\mathbf{X})| = \mathrm{E}[w^i(\mathbf{X})] = 1$$

and Corollary 2.1 leads to the estimate

$$\begin{aligned}\rho(F_n, \mathrm{P}_{\boldsymbol{\lambda}}) &\leqslant \sum_{i=1}^{p} (1 - e^{-\lambda_i})\, \mathrm{E}\left|1 - \left(1 - \frac{\lambda_i}{n}\right) \frac{n - X}{n - \lambda}\right| \\ &\quad + \frac{2}{n} \sum\sum_{1 \leqslant i < j \leqslant p} \min\{1, \lambda_i^{-1/2}\}\, \lambda_i \lambda_j \\ &\leqslant \frac{2p\lambda}{n - \lambda} + \frac{2\lambda^2}{n} = O(n^{-1}) \qquad \text{as} \quad n \to \infty,\end{aligned}$$

and the well-known convergence of multinomial to the multivariate Poisson follows, with a rate of convergence of order at least n^{-1}.

EXAMPLE 2. Let $\mathbf{X}_n$ be the negative multinomial random vector $\mathrm{NM}(n; \theta_1, ..., \theta_p)$; that is, the probability function of $\mathbf{X}_n$ is given by

$$f_n(\mathbf{x}) = \binom{n + x_1 + \cdots + x_p - 1}{x_1, ..., x_p} \theta_1^{x_1} \cdot \cdots \cdot \theta_p^{x_p} \cdot \theta^n, \qquad \mathbf{x} \in \{0, 1, ...\}^p,$$

where $\theta = 1 - \theta_1 - \cdots - \theta_p$ is the failure probability. Assume that the parameters $\theta_i = \theta_i(n) = \lambda_i/(n + \lambda)$ ($\lambda_i > 0$ fixed) for all $i = 1, ..., p$, where

$\lambda = \lambda_1 + \cdots + \lambda_p$ as in the previous example. The $\mathbf{w}$-function of $\mathbf{X}_n$ is given in this case by

$$w^i(\mathbf{x}) = \frac{1}{n}(1 - \theta_1 - \cdots - \theta_p)(n + x_1 + \cdots + x_p) = \frac{n + x}{n + \lambda}, \qquad i = 1, ..., p,$$

where $x = x_1 + \cdots + x_p$ (note that $(1 + \theta_1 + \cdots + \theta_p)^{-1}$ in the corresponding result of Cacoullos and Papathanasiou (1992, p. 179) has been replaced by the corrected one, $1 - \theta_1 - \cdots - \theta_p$). Since $\mu_i = \lambda_i$, $\sigma_i^2 = \lambda_i(1 + \lambda_i/n)$, and $\sigma_{ij} = \lambda_i \lambda_j / n$, we get from Corollary 2.1 (note that $w^i(\mathbf{x}) \geqslant 0$ for all $\mathbf{x}$ in $\{0, 1, ...\}^p$ and $\mathrm{E}\,|w^i(\mathbf{X})| = \mathrm{E}[w^i(\mathbf{X})] = 1$ as in the previous example)

$$\begin{aligned}\rho(F_n, \mathrm{P}_{\boldsymbol{\lambda}}) &\leqslant \sum_{i=1}^{p} (1 - e^{-\lambda_i})\, \mathrm{E}\left|1 - \left(1 + \frac{\lambda_i}{n}\right) \frac{n + X}{n + \lambda}\right| \\ &\quad + \frac{2}{n} \sum\sum_{1 \leqslant i < j \leqslant p} \min\{1, \lambda_i^{-1/2}\}\, \lambda_i \lambda_j \\ &\leqslant \frac{(2p - 1)\, \lambda + \lambda^2/n}{n + \lambda} + \frac{2\lambda^2}{n} = O(n^{-1}),\end{aligned}$$

and thus, the negative multinomial d.f. converges to the multivariate Poisson with a rate of convergence of order at least n^{-1}.

EXAMPLE 3. Let $\mathbf{X}$ be an n-dimensional random vector uniformly distributed in the interior of a sphere with center $(0, ..., 0)'$ and radius $\sqrt{n+2}$. Since the r.v.'s $X_1, ..., X_n$ are exchangeable, it is clear that each pair (say $(X, Y)' = (X_i, X_j)'$) has the same density,

$$f_n(x, y) = \frac{n}{2\pi(n+2)^{n/2}} (n + 2 - x^2 - y^2)^{n/2 - 1}, \qquad x^2 + y^2 < n + 2,$$

with marginal densities

$$\begin{aligned} f_X(x) &= f_Y(x) \\ &= \frac{1}{B(1/2, (n+1)/2)(n+2)^{n/2}} (n + 2 - x^2)^{(n-1)/2}, \qquad x^2 < n + 2,\end{aligned}$$

where $B(\alpha, b) \equiv \Gamma(\alpha)\, \Gamma(b)/\Gamma(\alpha + b)$.

It is evident that X and Y are uncorrelated with mean 0 and variance 1. Furthermore, the $\mathbf{w}$-function is given by

$$w^1(x, y) = w^2(x, y) = \frac{1}{n}(n+2-x^2-y^2), \qquad x^2+y^2<n+2.$$

Applying (1.1) (or Corollary 3.1 with $G=\Phi$, the bivariate standard normal d.f.), we have

$$\begin{aligned}\rho(F_n, \Phi) &\leqslant \frac{4}{n}\,\mathrm{E}\,|2-X^2-Y^2| \leqslant \frac{8}{n}\,\mathrm{E}\,|1-X^2|\\ &\leqslant \frac{8}{n}\sqrt{\mathrm{E}[X^4]-1} = \frac{8}{n}\sqrt{\frac{2(n+1)}{n+4}} \approx \frac{8\sqrt{2}}{n}.\end{aligned}$$

In order to apply Theorem 3.1, we first calculate

$$w^{(1)}(x, y) = w^1(x, y) = \frac{1}{n}(n+2-x^2-y^2), \qquad x^2+y^2<n+2$$

and

$$w^{(2)}(y) = \frac{1}{n+1}(n+2-y^2), \qquad y^2<n+2.$$

Therefore, Theorem 3.1 gives

$$\begin{aligned}\rho(F_n, \Phi) &\leqslant \frac{2}{n}\,\mathrm{E}\,|2-X^2-Y^2| + \frac{2}{n+1}\,\mathrm{E}\,|1-Y^2|\\ &\leqslant \left(\frac{4}{n}+\frac{2}{n+1}\right)\mathrm{E}\,|1-X^2| \leqslant \left(\frac{4}{n}+\frac{2}{n+1}\right)\sqrt{\frac{2(n+1)}{n+4}} \approx \frac{6\sqrt{2}}{n},\end{aligned}$$

which is slightly better than the previous one.

EXAMPLE 4. Let $U_{(1)} < \cdots < U_{(n)}$ be the order statistics corresponding to a random sample of size n from the uniform (0, 1) r.v. Fix p and q with $0<p<q<1$ and suppose that $i=(n+1)\,p$, $j=(n+1)\,q$. It is well known that the random pair

$$(X_n, Y_n)' = \sqrt{n+2}\,(U_{(i)}-p, U_{(j)}-q)'$$

has, as $n \to \infty$, a limiting normal distribution with mean $(0, 0)'$ and dispersion matrix

$$\Sigma = \begin{pmatrix} p(1-p) & p(1-q) \\ p(1-q) & q(1-q) \end{pmatrix}.$$

If $\Phi_{p,q}$ denotes this normal d.f. and F_n is the d.f. of $(X_n, Y_n)'$, then we have after some algebra

$$w^{(1)}(x_n, y_n) = \frac{n+2}{p(q-p)(n+1)} \left(\frac{x_n}{\sqrt{n+2}} + p \right) \left(\frac{y_n - x_n}{n+2} + q - p \right),$$

$$w^{(2)}(y_n) = \frac{n+2}{q(1-p)(n+1)} \left(\frac{y_n}{n+1} + q \right) \left(1 - q - \frac{y_n}{\sqrt{n+2}} \right).$$

Since $\mathrm{E}[X_n] = \mathrm{E}[Y_n] = 0$, $\mathrm{D}[X_n, Y_n] = \Sigma$ for all n, one can apply Theorem 3.1 (it can be shown that the constants $c_i = 2$, as in the case of two independent normal r.v.'s) leading to the bound

$$\begin{aligned} \rho(F_n, \Phi_{p,q}) &\leqslant 2 \left\{ \mathrm{E} \left| \frac{(n+2)\, U_{(i)}(U_{(j)} - U_{(i)})}{p(q-p)(n+1)} - 1 \right| \right. \\ &\quad \left. + \mathrm{E} \left| \frac{(n+2)\, U_{(j)}(1 - U_{(j)})}{q(1-p)(n+1)} - 1 \right| \right\} \\ &\leqslant \frac{2(n+2)}{n+1} \left\{ \frac{1}{p(q-p)} \sqrt{\mathrm{Var}[U_{(i)}(U_{(j)} - U_{(i)})]} \right. \\ &\quad \left. + \frac{1}{q(1-q)} \sqrt{\mathrm{Var}[U_{(j)}(1 - U_{(j)})]} \right\} \\ &\approx \frac{2}{\sqrt{n}} \left\{ \sqrt{\frac{q}{p(q-p)} - 4} + \sqrt{\frac{1}{q(1-q)} - 4} \right\}, \end{aligned}$$

which gives an estimate of order $n^{-1/2}$ for the rate of convergence, by only using the moments of uniform order statistics.

REFERENCES

A. D. Barbour, L. Holst, and S. Janson, "Poisson Approximation," Oxford Univ. Press, New York, 1992.

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

T. Cacoullos, N. Papadatos, and V. Papathanasiou, Variance inequalities for covariance kernels and applications to central limit theorems, *Theory Probab. Appl.* **42** (1997), 195–201.

T. Cacoullos and V. Papathanasiou, Lower variance bounds and a new proof of the central limit theorem, *J. Multivariate Anal.* **43** (1992), 173–184.

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.

L. Chen, An inequality for the multivariate normal distribution, *J. Multivariate Anal.* **12** (1982), 306–315.

H. Chernoff, A note on an inequality involving the normal distribution, *Ann. Probab.* **9** (1981), 533–535.

J. P. Chou, An identity for multidimensional continuous exponential families and its applications, *J. Multivariate Anal.* **24** (1988), 129–142.

F. Gotze, On the rate of convergence in the multivariate CLT, *Ann. Probab.* **19** (1991), 724–739.

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

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

N. Papadatos and V. Papathanasiou, Distance in variation between two arbitrary distributions via the associated *w*-functions, *Theory Probab. Appl.* **40** (1995a), 685–694.

N. Papadatos and V. Papathanasiou, Distance in variation and a Fisher-type information, *Math. Methods Statist.* **4** (1995b), 230–237.

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

V. Papathanasiou, Multivariate variational inequalities and the central limit theorem, *J. Multivariate Anal.* **58** (1996), 189–196.

T. J. Sweeting, Speed of convergence for the multidimensional central limit theorem, *Ann. Probab.* **5** (1977), 28–41.

S. Utev, Probability problems connected with a certain integrodifferential inequality, *Siberian Math. J.* **30** (1989), 490–493.