ANNALES DE L'I. H. P., SECTION B

S. T. RACHEV
J. E. YUKICH

Smoothing metrics for measures on groups

Annales de l'I. H. P., section B, tome 25, nº 4 (1989), p. 429-441
<http://www.numdam.org/item?id=AIHPB_1989__25_4_429_0>

Ann. Inst. Henri Poincaré,
Vol. 25, n° 4, 1989, p. 429-441. *Probabilités et Statistiques*

Smoothing metrics for measures on groups

by

S. T. RACHEV
University of California at Santa Barbara,
Santa Barbara, CA 93106

and

J. E. YUKICH (*)
Lehigh University, Bethlehem, PA 18015

ABSTRACT. — We investigate and study smoothing metrics d_g on the space $\mathcal{M}^+(G)$ of bounded positive measures on a complete, separable locally compact metrizable group G. Via the use of a tauberian theorem we characterize smoothing metrics metrizing the topology of weak convergence in $\mathcal{M}^+(G)$. Properties of $d_g(P_n, P)$ are investigated, where P_n, $n \geqq 1$, denote the empirical measures for P; the limiting distribution of $n^{1/2} d_g(P_n, P)$ is determined in certain instances.

Key words : Weak convergence, probability metrics, tauberian theorems, convolutions.

RÉSUMÉ. — Nous étudions les métriques régularisantes d_g sur l'espace $\mathcal{M}^+(G)$ des mesures positives bornées sur un groupe localement compact, séparable métrisable G. En utilisant un théorème taubérien nous caractérisons les métriques régularisantes qui définissent la topologie de la convergence faible sur $\mathcal{M}^+(G)$. Si P_n, $n \geqq 1$ désignent les mesures empiriques pour une mesure P nous étudions les propriétés de $d_g(P_n, P)$; la loi limite de $n^{1/2} d_g(P_n, P)$ est déterminée dans certains exemples.

Classification A.M.S. : Primary: 60 B 15, 60 B 10, 40 E 05; Secondary: 47 D 05.
(*) Research supported in part by a grant from the Reidler Foundation.

1. INTRODUCTION AND NOTATION

The study of the metrization of the space of probability measures over a given measure space has received wide attention and by now there is a wide variety of metrics available for study and use. See Dudley's monograph [Du2] as well as Zolotarev's survey article [Zo]. Some of the more interesting metrics are quite difficult to calculate and estimate, thereby limiting their use. On the other hand, metrics of simpler structure generally seem to have limited theoretical and practical importance. In this article we develop the study of smoothing metrics, denoted by d_g (also called convolution metrics [Yu1]). Despite their weak structure, smoothing metrics possess some noteworthy properties. It is shown that they are natural choices for the metrization of weak convergence of probability measures on groups and that they are of practical and theoretical interest in the study of empirical measures.

First, we extend the results of [Yu1], where it is shown that smoothing metrics d_g metrize weak star convergence of probability measures. The extended results are conveniently expressed in the language of tauberian theory and it seems that this aspect is of independent interest, especially from the functional analytic point of view.

Using the fact that the correspondence between positive measures and convolution operators is bicontinuous [Si], it is proved that smoothing metrics metrize the topology of weak star convergence of bounded positive measures on locally compact abelian metrizable groups. In fact, the class of smoothing metrics possessing this property is characterized via a tauberian condition. This represents an extension of earlier work on the subject. See section two.

Properties of $d_g(P_n, P)$, where P_n, $n \geqq 1$, denote the empirical measures for P, are considered. The weak structure of d_g is exploited to deduce several interesting and potentially useful probabilistic estimates for $d_g(P_n, P)$. It is shown that if P is the uniform law on the circle then the limiting distribution of $n^{1/2} d_g(P_n, P)$ can be explicitly determined. Actually, in certain instances $d_g(P_n, P)$ may be computed exactly, making its use in statistical problems attractive. See section three.

Notation

Throughout G denotes a separable locally compact abelian (LCA) metrizable group. Every such group is second countable (*i.e.*, has a countable basis), complete and has the Lindelöf property.

Let Γ denote the dual group of G; by the Pontryagin duality theorem Γ is also second countable [HR], p. 381. Let $\mathcal{M}(G)$ denote the space of all finite signed Baire measures on G, $\mathcal{M}^+(G)$ the subset of non-negative measures and $\mathcal{M}_1^+(G)$ the subset of probability measures. Let m denote Haar measure on G.

Let C(G) be the space of bounded continuous functions on G with values in $\mathbb{R}$ and $C_0(G)$ the subspace of functions vanishing at infinity; equip both spaces with their canonical sup norms. For all $p \geqq 1$ let $\mathcal{L}^p(G)$ be the space of real valued functions f on G such that $|f|^p$ is integrable with respect to m.

A sequence $(\mu_n)_{n \in \mathbb{N}^+}$ in $\mathcal{M}(G)$ is said to converge weak star to $\mu \in \mathcal{M}(G)$ iff

$$\lim_{n \to \infty} \int f \, d\mu_n = \int f \, d\mu \quad \text{for all } f \in C(G);$$

in this case write $\mu_n \xrightarrow{w} \mu$. The corresponding topology on $\mathcal{M}(G)$ is called the weak star topology (weak topology by probabilists).

We point out that smoothing metrics have been previously discussed by Kerstan and Matthes [KM], Szasz [Sz] and the authors ([RY], [Yu1], [RI]). Actually, Kolmogorov considered $\mathcal{L}^1$ versions of smoothing metrics as early as 1953 [Ko]. As in [Ko], let g_σ denote the density for the normal distribution with mean 0 and variance σ^2. For probability measures P and Q on $\mathbb{R}$ define

$$\Delta_\sigma(P, Q) := \int_{-\infty}^{\infty} dx \left| \int_{-\infty}^{\infty} g_\sigma(x-y)(dP - dQ)(y) \right|.$$

Kolmogorov observed that $P_n \xrightarrow{w} Q$ iff for all $\sigma > 0$ $\Delta_\sigma(P_n, Q) \to 0$ as $n \to \infty$ [Ko]. Although Δ_σ is not used in the sequel, it may be noted that it bears close resemblance to our smoothing metrics; see (2.1).

2. SMOOTHING METRICS VIA A TAUBERIAN THEOREM

For each $g \in \mathcal{L}^1(G)$ let $Z(g) := \{\gamma \in \Gamma : \hat{g}(\gamma) = 0\}$, where $\hat{g}$ denotes the Fourier transform of g. A famous theorem of Wiener states that the translates of $g \in \mathcal{L}^1(G)$ span $\mathcal{L}^1(G)$ if and only if $Z(g) = \varnothing$; *see* e.g. [Ru1], Theorem 7.2.5 (*d*). Taking this tauberian theorem as our starting point, we first note that a similar, easier result holds in the space $C_0(G)$. See Theorem 2.5 below. Using this tauberian theorem we deduce theorems for weak convergence of measures in $\mathcal{M}^+(G)$.

A salient feature of the theory involves the use of *metric kernels*, defined as those $g \in C_0(G) \cap \mathscr{L}^1(G)$ such that $Z^0(g) = \varnothing$, *i.e.*, $Z(g)$ has empty interior. The motivation and justification for the term metric kernel, apparently used here for the first time, will become clear in the sequel. For the moment, note that metric kernels always exist for second countable LCA groups:

LEMMA 2.1. — *There exist metric kernels on* G.

Proof. — This appears in [Yu1]. By the assumed second countability, it is actually possible to choose g so that $Z(g) = \varnothing$.

Q.E.D.

Each metric kernel g induces a smoothing metric d_g on $\mathscr{M}(G)$ defined by [Yu1]

$$d_g(\mu, \nu) := \sup \left| \int g(xy^{-1})(d\mu - d\nu)(y) \right|, \qquad \mu, \nu \in \mathscr{M}(G). \tag{2.1}$$

Using the condition $Z^0(g) = \varnothing$, it is easily seen that d_g defines a norm on $\mathscr{M}(G)$ and is thus a metric on $\mathscr{M}(G)$. Notice that $g * (\mu - \nu)$ may be regarded as the averages of $\mu - \nu$ and in this sense d_g is a natural measure of distance.

Alternatively, d_g may be viewed as the difference of convolution operators. More precisely, given $\mu \in \mathscr{M}^+(G)$, define the associated convolution operator $T_\mu : C_0(G) \to C_0(G)$ by

$$T_\mu f(x) := \int f(xy^{-1})\, d\mu(y), \qquad \forall f \in C_0(G).$$

T_μ is a contraction operator and $d_g(\mu, \nu)$ is just the action of $T_\mu - T_\nu$ on g, evaluated in the strong operator topology on $C_0(G)$.

It is known [Si] that sequential weak star convergence of μ_n to μ can be viewed as the convergence of T_{μ_n} to T_μ in the strong operator topology on $C_0(G)$; when G is the real line this equivalence is a classic result of Feller [Fe], p. 257. Among other things, the following theorem shows that it is enough to restrict attention to the strong convergence of $T_{\mu_n}(g)$ to $T_\mu(g)$ for any metric kernel g.

The following theorem also shows that the smoothing metric d_g metrizes the topology of weak star convergence in $\mathscr{M}^+(G)$; this is possible since G is Lindelöf and hence $\mathscr{M}^+(G)$ is metrizable (Varadarajan [Va], IV, p. 49, Teorema 13, p. 62; Dudley [Du1], §4). For the sake of simplicity, attention is restricted to sequential convergence; nonetheless, the results remain true in the setting of generalized sequences, or nets.

THEOREM 2.2. — *Let g be a metric kernel on* G, μ_n, $n \in \mathbb{N}^+$, *a sequence in* $\mathcal{M}^+(G)$, ν *in* $\mathcal{M}^+(G)$ *and* $\mu_n(G) \to \nu(G)$. *The following are equivalent:*

(1) $\mu_n \overset{w}{\to} \nu$;

(2) $d_g(\mu_n, \nu) \to 0$ *as* $n \to \infty$;

(3) $\forall f \in C_0(G)$, $\sup_x \left| \int f(xy^{-1})(d\mu_n - d\nu)(y) \right| \to 0$;

(4) *if* $T_n f(x) := \int f(xy^{-1})\, d\mu_n(y)$ $\forall f \in C_0(G)$ *and* $\forall x \in G$, *then the sequence* T_n, $n \in \mathbb{N}^+$, *converges in the strong operator topology on* $C_0(G)$ *to an operator* T_0 *such that* $\|T_0\| = \|\nu\|$.

The next theorem characterizes the d_g-metrizability of weak star convergence in $\mathcal{M}^+(G)$ in terms of the metric kernels, thereby justifying the definition and use of the term. This result represents a substantial generalization of Theorem 2.2 of [Yu1].

THEOREM 2.3. — *Let* $g \in C_0(G) \cap \mathcal{L}^1(G)$. *The following are equivalent statements for the pseudo-metric* d_g:

(1) d_g *metrizes the topology of weak star convergence in* $\mathcal{M}^+(G)$;

(2) *g is a metric kernel, and*

(3) $d_g(\mu, \nu) \neq 0$ *if* $\mu \neq \nu$.

The above theorems fail to hold if $\mathcal{M}^+(G)$ is replaced by $\mathcal{M}(G)$; on the other hand, condition (1) of Theorem 2.2 implies any of the others in this more general setting (*cf.* Theorem 7 of [Du1]). That the converse is false is an immediate consequence of Theorem 11 of [Du1].

The proofs of the above theorems center around the use of an apparently new [Ru2] tauberian theorem, one which shows that metric kernels are precisely those functions spanning $C_0(G)$ (see Theorem 2.5 below). In this way we obtain the equivalence of (2) and (3) of Theorem 2.2 and in this sense d_g is a natural choice for a metric on $\mathcal{M}^+(G)$.

Finally, before proving Theorems 2.2 and 2.3, let us mention an easy but interesting consequence of Theorem 2.2: the equivalence of (2) and (4) shows that a convolution semigroup Q_t, $t > 0$, on $C_0(G)$ is continuous (*i.e.*, $Q_t \to I$ as $t \to 0^+$ in the strong operator topology on $C_0(G)$, where I is the identity operator; *cf.* Berg and Forst [BF], p. 48) if and only if

$$(Q_t - I)(g) \to 0 \quad \text{in } (C_0(G), \|\cdot\|),$$

where g is any metric kernel.

Proofs of Theorem 2.2 *and* 2.3. — As for Theorem 2.2, one could hope to prove the implication (2) $\Rightarrow$ (1) by showing that (2) implies the pointwise convergence of Fourier transforms $\hat{\mu}_n(t) \to \hat{\nu}(t)$; this argument works well on $\mathbb{R}^k$ [Yu1] where it is possible to show that the μ_n, $n \geqq 1$, are uniformly tight. Instead, we adopt a different approach, one which shows that metric kernels span $C_0(G)$. Three auxiliary results are needed; the first is technical.

LEMMA 2.4 ([Ru1], Theorem 1.6.4). – *If* E *is a non-empty open set in* Γ, *there exists* $f \in \mathscr{L}^1(G)$ *such that* $\hat{f}=0$ *outside* E.

The next theorem does not seem to appear in the literature and may be thought of as a companion to Wiener's famous tauberian theorem, described earlier.

THEOREM 2.5. – *If* $g \in C_0(G) \cap \mathscr{L}^1(G)$, *then the translates of* g span $C_0(G)$ *if and only if* $Z^0(g)=\varnothing$.

Proof. – First, suppose $Z^0(g)=\varnothing$. Let $\varphi \in C_0^*(G)$. By the well-known isometric isomorphism between $C_0^*(G)$ and $\mathscr{M}(G)$, there exists $\nu \in \mathscr{M}(G)$ such that

$$\varphi(f)=\int f(y)\,\nu(dy), \qquad \forall f \in C_0(G).$$

Define $\mu(dy)$ by $\nu(d(y^{-1}))$. If $\int g_x(y)\,\nu(dy)=0$ for every translate g_x of g then $(g * \mu)(x)=0$, where we note that since $g \in \mathscr{L}^1$, $g * \mu$ is well defined. The uniqueness theorem for Fourier-Stieltjes transforms implies $\hat{g}\hat{\mu}=0$. Since $\hat{g} \neq 0$ except possibly on a set with empty interior, it follows by unicity and continuity of Fourier-Stieltjes transforms that $\varphi=0$.

Thus, by the Hahn-Banach theorem the translates of g span $C_0(G)$.

Conversely, suppose $Z(g)$ contains an open set E. Then by Lemma 2.4, there exists $f \in \mathscr{L}^1(G)$ such that $\hat{f}=0$ outside E.

Now $\hat{g}\hat{f}=0$ and since g and f are both in $\mathscr{L}^1(G)$ this implies that $\int g_x f=0$ for all $x \in G$, showing that f is orthogonal to the span of the translates of g.

Q.E.D.

Using Theorem 2.5 and a simple approximation argument, it is easy to deduce the next result:

LEMMA 2.6. – *Let* g *be a fixed metric kernel and* $f \in C_0(G)$. *For all* $\varepsilon > 0$ *and all* $\mu, \nu \in \mathscr{M}(G)$ *with finite total variation norm* $\|\mu-\nu\|$ *there exists a finite constant* $A:=A(\varepsilon, f, \|\mu-\nu\|)$ *such that*

$$\sup_x \left| \int f(xy^{-1})(d\mu-d\nu)(y) \right| \leqq \varepsilon + A\, d_g(\mu, \nu).$$

Proof of Theorem 2.2. — The equivalence of (2) and (3) follows from Lemma 2.6. The equivalence of (1), (3) and (4) follows from a generalization of Proposition 4.2 of Siebert [Si] to $\mathscr{M}^+(G)$.

Proof of Theorem 2.3. — We only need to demonstrate (3) $\Rightarrow$ (2). If (2) is not valid, *i.e.*, $Z^0(g) \neq \varnothing$ then there is a $\varphi \in C_0^*(G)$ such that $\varphi(g_x)=0$ for all x. The signed measure $\mu \in \mathcal{M}(G)$ associated with φ may be written as $\mu=\mu^+-\mu^-$ where $\mu^+ \in \mathcal{M}^+(G)$ and $\mu^- \in \mathcal{M}(G)$. Now $d_g(\mu^+, \mu^-)=0$, thus $\mu^+=\mu^-$, $\varphi=0$ and the conclusion follows from the Hahn-Banach theorem and Theorem 2.5.

Q.E.D.

Theorems 2.2 and 2.3 do not extend to non-metrizable G. Indeed, when G is non-metrizable there are no metric kernels, that is, there is no g simultaneously satisfying $g \in \mathcal{L}^1(G)$, $g \in C_0(G)$ and $Z^0(g)=\varnothing$. If there were such a g, then since the proof of the equivalence of (2) and (3) of Theorem 2.2 doesn't directly use the metrizability of G, $\mathcal{M}^+(G)$ would be metrizable. However, weak star convergence in $\mathcal{M}^+(G)$ is not metrizable, since the set of unit masses is homeomorphic to G (Varadarajan [Va], Teorema 13). We have thus proved the following result, which adds to Wiener's tauberian theorem:

COROLLARY 2.5. — *If* G *is a non-metrizable* LCA *group then there are no metric kernels,* i.e., *there is no function* $g \in \mathcal{L}^1(G) \cap C_0(G)$ *with* $Z^0(g)=\varnothing$. *Thus, if the translates of g* span $\mathcal{L}^1(G)$ *then* $g \notin C_0(G)$.

As a final remark, it is of interest to note that even for non-metric kernels g, d_g occupies a role in the theory of probability metrics. Classical metrics may be cast into the d_g form, g a kernel which is not necessarily in $\mathcal{L}^1 \cap C_0$. For example, for all μ, $\nu \in \mathcal{M}_1^+([a, b])$, the uniform metric $\| F_\mu(x)-F_\nu(x) \|$ between distribution functions F_μ and F_ν may be represented as $(b-a)\, d_u$, where u denotes the uniform density on $[a, b]$.

3. PROPERTIES OF d_g; EMPIRICAL MEASURES

It is evident that d_g has a relatively weak structure and one would expect its uniform structure to be strictly weaker than, for example, that of the Prokhorov or dual-bounded Lipschitz metric, denoted by ρ and β respectively. (See Dudley's monograph [Du2] for the definitions of ρ and β and a detailed discussion of probability metrics.) Indeed, simple examples show that d_g and β, do not, in general, induce the same uniformity. Clearly, if g is a Lipschitz probability density on $\mathbb{R}$ then for all P, $Q \in \mathcal{M}_1^+(\mathbb{R})$, $d_g(P, Q) \leqq C\,\beta(P, Q)$, where C is a constant depending only on g, but an inequality in the other direction will not hold in general. For example, let g have support on [0, 1] and for all $n \geqq 1$ let P_n and Q_n be probability measures on $\mathbb{R}$ defined by

$$P_n(2j)=Q_n(2j+1)=n^{-1}, \qquad j=1, \ldots, n.$$

Then $d_g(\mathrm{P}_n, \mathrm{Q}_n) = O(n^{-1})$ and $\beta(\mathrm{P}_n, \mathrm{Q}_n) \geqq 1/2$, showing that the uniform structure of β on $\mathbb{R}$ is strictly stronger than that of d_g. D'Aristotile, Diaconis and Freedman [DDF] attempt to classify the uniformities compatible with the weak toplogy on the set of probability measures on a separable metric space, starting with the uniformity induced by ρ. As we have shown in the case of separable LCA groups, the classification of uniformities may even start with that of d_g, g a metric kernel. When $\mathrm{G} = \mathbb{R}$, however, the uniformity of d_g is not the weakest one inducing the weak topology (see [Se] for characterizations of the weakest metrics). To see that the d_g-uniformity is not stronger than that generated by the Lévy metric ρ_{L}, take P_n to be uniform on $[0, n]$, Q_n uniform on $[n, 2n]$ and note that $d_g(\mathrm{P}_n, \mathrm{Q}_n) = O(n^{-1})$, whereas $\rho_{\mathrm{L}}(\mathrm{P}_n, \mathrm{Q}_n) \cong 1$.

(*a*) Limiting behavior of $n^{1/2} d_g(\mathrm{P}_n, \mathrm{P})$, $\mathrm{P} \in \mathcal{M}_1^+(\mathrm{T})$

Throughout take G to be the circle group T and P_n the empirical measures for the probability measure $\mathrm{P} \in \mathcal{M}_1^+(\mathrm{T})$, *i.e.*

$$\mathrm{P}_n := n^{-1}(\delta_{\mathrm{X}_1} + \ldots + \delta_{\mathrm{X}_n}),$$

where X_i, $i = 1, \ldots, n$ denotes an i.i.d. sequence of random variables with law P and δ_x the unit mass at x. The requirement that g be a metric kernel implies $\hat{g}(k) \neq 0 \ \forall k \in \mathbb{Z}$. Additionally, suppose that g is of bounded variation and has the Fourier series expansion

$$g(x) = \sum_{n=1}^{\infty} a_n \cos nx, \qquad 0 < x \leqq 2\pi, \tag{3.1}$$

where $a_n \neq 0$. Since $g \in \mathscr{L}^2(\mathrm{T})$, Parseval's theorem implies $\sum a_n^2 < \infty$.

Since g is of bounded variation, the restriction to T of the translates of the 2π-periodic extension of g to $\mathbb{R}$ is a Vapnik-Chervonenkis class of functions (see [Du3], [Po] for definitions). A straightforward application of Dudley's CLT result [Du3] for the function-indexed empirical process (see also Giné and Zinn [GZ1], [GZ2], Pollard [Po]) gives for all $\mathrm{P} \in \mathcal{M}_1^+(\mathrm{T})$

$$n^{1/2} d_g(\mathrm{P}_n, \mathrm{P}) \xrightarrow{w} \sup_{x \in (0, 2\pi]} |\mathrm{G}_{\mathrm{P}}(g_x)|, \tag{3.2}$$

where G_{P} denotes a Gaussian process indexed by the translates of g; G_{P} has mean zero and covariance [Du3]:

$$\mathrm{EG}_{\mathrm{P}}(g_x g_y) = \int_0^{2\pi} g_x g_y \, d\mathrm{P} - \int_0^{2\pi} g_x \, d\mathrm{P} \int_0^{2\pi} g_y \, d\mathrm{P}.$$

Therefore, for any $\mathrm{P} \in \mathcal{M}_1^+(\mathrm{T})$, $n^{1/2} d_g(\mathrm{P}_n, \mathrm{P})$ may be approximated in probability by the supremum of the Gaussian process G_{P}. A bounded and compact law of the iterated logarithm for $d_g(\mathrm{P}_n, \mathrm{P})$ follows (Theorem 1.3, [DP]).

It turns out that if P denotes the uniform probability measure on $[0, 2\pi]$ then the covariance structure of the limiting Gaussian process $\mathrm{G_P}$ process may be given explicitly. In fact the $\mathrm{G_P}$ process is stationary.

THEOREM 3.1. — *Let g be a metric kernel of bounded variation of the form* (3.1), i. e.,

$$g(x)=\sum_{n=1}^{\infty} a_n \cos nx,$$

and let P *denote the uniform probability measure on* T. *Then the limiting Gaussian process* $\mathrm{G_P}$ *in* (3.2) *is stationary and has covariance structure*

$$\mathrm{EG_P}(g_u g_v)=\frac{1}{2}\sum_{n=1}^{\infty} a_n^2 \cos n(u-v). \tag{3.3}$$

Proof. — It is easily seen that

$$\begin{aligned}\mathrm{EG_P}(g(u-y)g(v-y)) &= \frac{1}{2\pi}\int_0^{2\pi} g(u-y)g(v-y)\,dy \\ &= \frac{1}{2\pi}\int_0^{2\pi} \sum_{n=1}^{\infty} a_n \cos n(u-y) \sum_{n=1}^{\infty} a_n \cos n(v-y)\,dy \\ &= \frac{1}{2\pi}\sum_{n=1}^{\infty} a_n^2 \int_0^{2\pi} \cos n(u-y)\cos n(v-y)\,dy \quad (3.4)\end{aligned}$$

since $\int_0^{2\pi} \cos n(u-y)\cos m(v-y)\,dy=0,\ m\neq n$. Expansion of $\cos n(u-y)$ and $\cos n(v-y)$ shows that (3.4) reduces to the finite sum $\frac{1}{2}\sum_{n=1}^{\infty} a_n^2 \cos n(u-v)$.

Q.E.D.

(*b*) The calculation and approximation of $d_g(\mathrm{P}_n, \mathrm{P})$, $\mathrm{P}\in\mathcal{M}_1^+(\mathbb{R})$

For the remainder of this section, take G to be the additive group $\mathbb{R}$.

Since d_g is expressed as a supremum over translates, a relatively small class of functions, the estimation of d_g is relatively easy. That we may estimate d_g makes its use in statistical problems attractive, especially since d_g possesses useful properties (e. g. CLT properties-see paragraph above) not enjoyed by many other metrics. In fact, in certain instances $d_g(\mathrm{P}_n, \mathrm{P})$ may be easily computed, where P_n are the usual empirical measures for

$P \in \mathcal{M}_1^+(\mathbb{R})$: in such cases, the supremum over all translates can be replaced by the maximum over a *finite* number N of translates, where $N = O(n)$.

The remainder of this section discusses ways to either probabilistically estimate or exactly calculate the value of $d_g(P_n, P)$. As for the first, exponential bounds for $n^{1/2} d_g(P_n, P)$ may be obtained through the use of general bounds available for the empirical process indexed by translates of a fixed function. Such bounds hold uniformly in $n \in \mathbb{N}^+$; see Alexander [Al] for sharp bounds with large constants and Yukich [Yu2] for less refined bounds, but with smaller constants. Additionally, for any law P on $\mathbb{R}^d$, $d \geqq 1$, a. s. rates of convergence of $n^{1/2} d_g(P_n, P)$ to the supremum of a Gaussian process may be found. This is achieved by almost surely approximating $n^{1/2} d_g(P_n, P)$ by an empirical Brownian bridge related to P; by doing so one obtains a. s. rates of convergence of the form $O(n^{-\alpha})$, $\alpha > 0$ (see e. g. Theorem 6.3 of Massart [Ma]).

The calculation of $d_g(P_n, P)$ is facilitated by both its weak structure and the latitude in the choice of g. As an example, if P is the uniform measure on [0, 1] and g is the bilateral exponential $\frac{1}{2}\exp\{-|x|\}$, then $d_g(P_n, P)$ can be explicitly calculated. Indeed, with probability one

$$\begin{aligned} 2n\, d_g(P_n, P) &= \sup_{x \in \mathbb{R}} \left| \sum_{i=1}^{n} e^{-|x-X_i|} - n \int_0^1 e^{-|x-y|}\, dy \right| \\ &= \sup_{0 \leqq x \leqq 1} \left| \sum_{i=1}^{n} e^{-|x-X_i|} - n \int_0^1 e^{-|x-y|}\, dy \right| \\ &= \sup_{0 \leqq x \leqq 1} \left| \sum_{i=1}^{n} e^{-|x-X_i|} + n(e^{-x} + e^{x-1} - 2) \right|. \end{aligned} \tag{3.5}$$

Defining new random variables Y_i, $1 \leqq i \leqq 3n$, by

$$Y_i = \begin{cases} X_i, & 1 \leqq i \leqq n \\ 0, & n+1 \leqq i \leqq 2n, \\ 1, & 2n+1 \leqq i \leqq 3n \end{cases}$$

yields with probability one

$$2n\, d_g(P_n, P) = \sup_{0 \leqq x \leqq 1} \left| \sum_{i=1}^{3n} e^{-|x-Y_i|} - 2n \right|;$$

moreover, the sup is attained at an extremal point of $\tau_n(x) := \sum_{i=1}^{3n} e^{-|x-Y_i|}$. Letting $Y_{(i)}$, $i \geqq 1$, denote the usual order statistics, notice that $\tau_n(x)$ is piecewise concave up over the intervals $(0, Y_{(1)})$, $(Y_{(1)}, Y_{(2)})$, . . ., $(Y_{(3n-1)}, Y_{(3n)})$, and clearly $\sup \tau_n(x)$ occurs at either 0,1 or one of the Y_i, $1 \leqq i \leqq n$. Elementary calculations show that $\inf \tau_n(x)$

occurs when x assumes one of the $3n$ values $\{x_1, \ldots, x_{3n}\}$, where

$$x_k := \frac{1}{2}\log\left(\sum_{i=1}^{k} e^{Y(i)} / \sum_{k+1}^{3n} e^{-Y(i)}\right), \qquad 1 \leqq k \leqq 3n.$$

Returning to (3.5) we thus obtain

$$d_g(P_n, P) = \max_{s_j} \frac{1}{2}\left|\frac{1}{n}\sum_{i=1}^{n} (e^{-|s_j - X_i|}) + (e^{-s_j} + e^{s_j - 1} - 2)\right|,$$

where s_j ranges over the set of $4n+2$ points $\{0, 1, X_1, \ldots, X_n, x_1, \ldots, x_{3n}\}$.

Using the same method, it is clear that if P is uniform on $[a, b]$ then $d_g(P_n, P)$ reduces to a maximum over $O(n)$ points. It is not known whether similar methods can be used to calculate $d_g(P_n, P)$ for non-uniform P:

QUESTION. — Given a law P on $\mathbb{R}$, is there a metric kernel g for which one may explicitly calculate the d_g distance between P and the empirical measure P_n?

More generally, can the d_g distance between arbitrary probability measures P and Q be interpreted (or calculated) in terms of a distance between random variables X and Y having respective marginals P and Q? For example if d_g is replaced by the Prokhorov metric, the answer to the question is in the affirmative as shown by Strassen's famous theorem [St]. These observations lead to the following question.

QUESTION. — Is the d_g distance between elements P and Q of $\mathcal{M}_1^+(G)$ equal to a "probabilistic distance" between random variables X and Y, where X and Y have marginal distributions P and Q, respectively?

In the particular case of d_u defined at the end of section two, the second question has the following solution:

$$d_u(\mu, \nu) = \inf D(X, Y), \qquad \mu, \nu \in \mathcal{M}_1^+([a, b]),$$

where the infimum is taken over the set of all joint distributions of X and Y with fixed marginal distributions μ and ν, and where

$$D(X, Y) := (b-a)^{-1} \sup_{x \in \mathbb{R}} [\Pr(X \leqq x < Y) + \Pr(Y \leqq x < Y)];$$

see [Ra].

ACKNOWLEDGEMENTS

The authors gratefully thank Professors B. Eisenberg and R. Dudley for their helpful comments and remarks during the course of this work.

REFERENCES

[Al] K. S. ALEXANDER, Probability Inequalities for Empirical Processes and a Law of the Iterated Logarithm, *Ann. Prob.*, Vol. **12**, 1984, pp. 1041-1067; *Correction*, Vol. **15**, 1987, pp. 428-430.

[BF] C. BERG and G. FORST, *Potential Theory on Locally Compact Abelian Groups*, Springer-Verlag, 1975.

[DDF] A. D'ARISTOTILE, P. DIACONIS and D. FREEDMAN, *On Merging of Probabilities*, Technical Report No. 301, 1988, Dept of Statistics, Stanford, California.

[Du1] R. M. DUDLEY, Convergence of Baire Measures, *Studia Mathematica*, Vol. **XXVII**, 1966, pp. 251-268.

[Du2] R. M. DUDLEY, Probabilities and Metrics, *Aarhus Universitet Lecture Notes Series*, No. **45**, 1976.

[Du3] R. M. DUDLEY, Universal Donsker Classes and Metric Entropy, *Ann. Prob.*, Vol. **15**, 1987, pp. 1306-1326.

[DP] R. M. DUDLEY and W. PHILIPP, Invariance Principles for Sums of Banach Space Valued Random Elements and Empirical Processes, *Z. Wahr. v. Geb.*, Vol. **62**, 1983, pp. 509-552.

[Fe] W. FELLER, *An Introduction to Probability Theory and its Applications*, Vol. **2**, 1971, second edition, Wiley and Sons.

[GZ1] E. GINÉ and J. ZINN, Some Limit Theorems for Empirical Processes, *Ann. Prob.*, Vol. **12**, 1984, pp. 929-989.

[GZ2] E. GINÉ and J. ZINN, Lecture on the Central Limit Theorem for Empirical Processes, *Lecture Notes in Mathematics*, No. **1221**, 1985, Springer, pp. 50-113.

[HR] E. HEWITT and K. A. ROSS, *Abstract Harmonic Analysis*, Vol. **I**, 1979, Springer-Verlag.

[KM] J. KERSTAN and K. MATTHES, Gleichverteilungseigenschaften von Faltungen von Verteilungsgesetzen auf lokalkompakten abelschen Gruppen, I., *Math. Nachr.*, Vol. **37**, 1968, pp. 267-312.

[Ko] A. N. KOLMOGOROV, Some Recent Work in the Field of Limit Theorems in Probability Theory, *Moscow Vestnik*, Vol. **10**, 1953, pp. 29-38.

[Ma] P. MASSART, Rates of Convergence in the Central Limit Theorem for Empirical Processes, *Lecture Notes Math.*, No. **1193**, 1986, pp. 73-109.

[Po] D. POLLARD, *Convergence of Stochastic Processes*, Springer-Verlag, 1984.

[Ra] S. T. RACHEV, Minimal Metrics in a Space of Real-Valued Random Variables, *Soviet Math. Dokl.*, Vol. **23/2**, 1981, pp. 425-428.

[RI] S. T. RACHEV, and Ts. G. IGNATOV, Ideal Quadratic Metrics, Stability Problems of Stochastic Models, *Seminar Proceedings*, V.N.I.I.S.I., Moscow), pp. 119-128 (Russian); *J. Soviet Math*, Vol. **35**, No. 2, 1986, pp. 2386-2394 (English).

[RY] S. T. RACHEV and J. E. YUKICH, Rates for the CLT via New Ideal Metrics, *Ann. Prob.*, Vol. **17**, 1989, pp. 775-788.

[Ru1] W. RUDIN, *Fourier Analysis on Groups*, Interscience Publishers, 1962.

[Ru2] W. RUDIN, Personal communication, August 1986.

[Se] V. V. SENATOV, On Some Properties of Metrics on the Set of Distribution Functions, *Math. U.S.S.R. Sbornik*, Vol. **31**, No. 3, 1977, pp. 379-387.

[Si] E. SIEBERT, Convergence and Convolutions of Probability Measures on a Topological Group, *Ann. Prob.*, Vol. **4**, 1976, pp. 433-443.

[St] V. STRASSEN, The Existence of Probability Measures with Given Marginals, *Ann. Math. Statist.*, Vol. **36**, 1965, pp. 423-439.

[Sz] D. SZASZ, On a Metrization of the Vague Convergence, *Stud. Sci. Math. Hungar.*, Vol. **9**, 1974, pp. 219-222.

[Va] V. S. VARADARAJAN, Measures on Topological Spaces (in Russian), *Mathematicheskii Sbornik,* Vol. **55**, (97), 1961, pp. 35-100.

[Yu1] J. E. YUKICH, The Convolution Metric d_g, *Math. Proc. Camb. Phil. Soc.,* Vol. **98**, 1985, pp. 533-540.

[Yu2] J. E. YUKICH, Uniform Exponential Bounds for the Normalized Empirical Process, *Studia Mathematica,* Vol. **84**, No. 2, 1986, pp. 71-78.

[Zo] V. M. ZOLOTAREV, Probability Metrics, *Theory Prob. Appl.,* Vol. **28**, (2), 1986, pp. 278-302.

(*Manuscript received April* 11*th,* 1988)
(*accepted April* 24*th,* 1989.)