BULLETIN OF THE
GREEK MATHEMATICAL SOCIETY
VOLUME 28, 1987A

A COMPARISON BETWEEN THE LIPSCHITZ AND PROHOROV METRICS

BY
George A. Anastassiou

Introduction. We consider a Polish space S with a metric d bounded by 1. Let $\mathscr{P}(S)$ be the space of all the probability measures on S defined on the corresponding Borel σ-algebra, metrized by the $d_{BL}^{(1)}$ (Lipschitz) metric, where

$$d_{BL}^{(1)}(F, G) := \sup_{\psi}\left\{\left|\int\psi\, dF - \int\psi\, dG\right|\right\} : F, G \in \mathscr{P}(S)$$

with

$$\psi \in \text{Lip } 1 := \{\psi : |\psi(x) - \psi(y)| \leq M_\psi d(x, y)\}.$$

($d_{BL}^{(1)}$ defines the weak topology on $\mathscr{P}(S)$, see [2].) Similarly, for $0 < \alpha < 1$ we define

$$d_{BL}^{(\alpha)}(F, G) := \sup_{\psi}\left\{\left|\int\psi\, dF - \int\psi\, dG\right|\right\} : F, G \in \mathscr{P}(S),$$

where

$$\psi \in \text{Lip } \alpha := \{\psi : |\psi(x) - \psi(y)| \leq M_\psi d^\alpha(x, y)\}.$$

We observe, easily, that for $0 < \alpha < \beta \leq 1$ we have $\text{Lip } 1 \subset \text{Lip } \beta \subset \text{Lip } \alpha$ (from $d^\beta(x, y) < d^\alpha(x, y) < 1$, by $d(x, y) < 1$).

Next, if $K \subseteq S$ and $\epsilon \geq 0$ then by K^ϵ we denote the set of all points $y \in S : d(x, y) < \epsilon$ for some $x \in K$. The Prohorov distance between the probability measures F, G on S is defined by

$$d_P(F, G) := \inf\{\epsilon > 0 : F(K) \leq G(K^\epsilon) + \epsilon\},$$

where K runs in the set of all compact subsets of S, (see [1], [3], [7]). (d_P is also a metric and describes as well the weak topology on $\mathscr{P}(S)$).

Our main result has as follows:

Theorem 1. *For $F, G \in \mathscr{P}(S)$ we have*

$$d_{BL}^{(\alpha)}(F, G) \leq 2 d_P^\alpha(\Phi, \Gamma), \quad 0 < \alpha \leq 1 \tag{1.1}$$

where $d_P(F, G)$ depends upon d.

For the proof of the above theorem we need the following:

Lemma 2. *$d^\alpha(x, y)$, $(0 < \alpha \leq 1)$, is a metric on S.*

Proof. Obvious.

Lemma 3. *$d_{BL}^{(\alpha)}$ $(0 < \alpha \leq 1)$, is a metric on $\mathscr{P}(S)$.*

Proof. We easily see that

$$d_{BL}^{(\alpha)}(F, G) < d_{BL}^{(\alpha)}(F, H) + d_{BL}^{(\alpha)}(H, G),$$

for all $F, G, H \in \mathscr{P}(S)$.

Also, it is trivial that

$$d_{BL}^{(\alpha)}(F, G) = d_{BL}^{(\alpha)}(G, F).$$

If $F \neq G$ then $d_{BL}^{(1)}(F, G) > 0$, consequently $d_{BL}^{(\alpha)}(F, G) > 0$.

The only nontrivial part of the proof is to show: if $d_{BL}^{(\alpha)}(F, G) = 0 \Rightarrow F = G$. From the assumption we have

$$\int \psi \, dF = \int \psi \, dG, \quad \text{for all } \psi \in \text{Lip } \alpha$$

We consider the function

$$\psi(x) := (1 - c \cdot d^\alpha(x, A))_+$$

where $d^\alpha(x, A) := \inf\{d^\alpha(x, y) : y \in A\}$ and $c > 0$ a constant.

We prove that $|\psi(x) - \psi(y)| < c \cdot d^\alpha(x, y)$ and $I_A \leq \psi \leq I_{A^{(1/c)}}$. Here I_A is the indicator function on A and

$$A^{(1/c)} := \left\{ x \in S : \inf d(x, y) \leq \frac{1}{c}, \ \forall y \in A \right\}.$$

First case: $\psi(x) = 1 - c \cdot d^\alpha(x, A)$, $\psi(y) = 1 - c \cdot d^\alpha(y, A)$. Since

$$d^\alpha(x, y) \geq (d(x, A) - d(y, A))^\alpha \geq d^\alpha(x, A) - d^\alpha(y, A).$$

we get

$$|\psi(x) - \psi(y)| = c \cdot |d^\alpha(x, A) - d^\alpha(y, A)| \leq c \cdot d^\alpha(x, y).$$

Thus,

$$|\psi(x) - \psi(y)| \leq c \cdot d^\alpha(x, y)$$

and consequently

$$\frac{1}{c}\,\psi \in \mathrm{Lip}\,\alpha$$

Second case: $\psi(x) = 0, \quad \psi(y) = 0 \Rightarrow \frac{1}{c}\,\psi \in \mathrm{Lip}\,\alpha$

Third case: Let $\psi(y) = 0$ and $\psi(x) = 1 - c \cdot d^{\alpha}(x, A)$. From $\psi(y) = 0$ we have $1 - c \cdot d^{\alpha}(y, A) < 0$, thus $c \cdot d^{\alpha}(y, A) > 1$, which is used in the triangle property of d^{α}. We have,

$$\begin{aligned}\psi(x) - \psi(y) &= 1 - c \cdot d^{\alpha}(x, A) \le 1 + c[d^{\alpha}(x, y) - d^{\alpha}(y, A)] \\ &\le c \cdot d^{\alpha}(y, A) - c \cdot d^{\alpha}(y, A) + c \cdot d^{\alpha}(x, y) = c \cdot d^{\alpha}(x, y).\end{aligned}$$

That is

$$\frac{1}{c}\,\psi \in \mathrm{Lip}\,\alpha,$$

true in all cases.

Letting $c \to +\infty$ we get $I_A \le \psi \le I_A$, thus

$$\int I_A \cdot dF = \int I_A \cdot dG,$$

i.e., $F(A) = G(A)$ for all the closed subsets A of S. Finally $F = G$.

Lemma 4. *Let $0 < \alpha \le l$, and random variables X, Y with values on S having distributions F, G (respectively), such that the expectation*

$$E(d^{\alpha}(X, Y)) \le \epsilon^{\alpha}, \quad \epsilon > 0.$$

Then

$$d_{BL}^{(\alpha)}(F, G) \le \epsilon^{\alpha}.$$

Proof. We observe that

$$\begin{aligned}\left|\int\psi\, dF - \int\psi\, dG\right| &= |E(\psi(X)) - E(\psi(Y))| \\ &\le E(|\psi(X) - \psi(Y)|) \le E(d^{\alpha}(X, Y)) \le \epsilon^{\alpha}.\end{aligned}$$

Consequently,

$$d_{BL}^{(\alpha)}(F, G) \le \epsilon^{\alpha}.$$

Comment 5. *By a theorem of Strassen (1965), see [8], we have the following expression $d_P(F, G) = \inf(\epsilon > 0 : \exists\ \lambda$ probability measure on*

$S \times S$ with marginals F, G such that $\lambda(d(x, y) > \epsilon) \leq \epsilon$). Since $d(x, y) \leq 1$, $\forall x, y \in S$ we get

$$d_P(F, G) \leq 1 \quad \forall F, G \in \mathscr{P}(S).$$

Proof of Theorem 1. Let $0 < \epsilon \leq 1$. For any probability measure P on $S \times S$ we have

$$\int d^\alpha(x, y) \cdot dP \leq \epsilon^\alpha \cdot P\{d^\alpha(x, y) \leq \epsilon^\alpha\} + P\{d^\alpha(x, y) > \epsilon^\alpha\}$$
$$= \epsilon^\alpha + (1 - \epsilon^\alpha) \cdot P\{d^\alpha(x, y) > \epsilon^\alpha\} \leq 2 \cdot \epsilon^\alpha.$$

The above is true: since we assumed that if $d_P(F, G) \leq \epsilon$ than there exists a probability measure P on $S \times S$ such that

$$P\{\omega : d(X(\omega), Y(\omega)) > \epsilon\} \leq \epsilon < 1$$

(from Commet 5), equivalently)

$$P\{d^\alpha(X, Y) > \epsilon^\alpha\} \leq \epsilon < \epsilon^\alpha.$$

Now, from Lemma 4 we get

$$d_{BL}^{(\alpha)}(F, G) \leq 2 \cdot \epsilon^\alpha.$$

This completes the proof of the theorem.

Comment 6. *Inequality (1.1) is new for $0 < \alpha < 1$ and known for $\alpha = 1$, see [3]. This could be proved easier (all $0 < \alpha \leq 1$) with the help of the very strong theorem of Kantorovich-Rubinstein, see [4]. However the above give an independent and simple proof of (1.1).*

These metrics and their relationships play an active role in the contemporary statistics, as well as in other applied areas of Mathematics (see [3], [5], [6]).

REFERENCES

1. Billingsley, P. (1968). Convergence of Probability Measures. John Wiley and Sons, New York.

2. Dudley, R.M. (1976). Probabilities and Metrics. Lecture Notes No. 45, Aarhus Unibersitet.

3. Huber, P.J. (1981). Robust Statistics. John Wiley and Sons, New York.

4. Kantorovich L and Rubinstein G. (1958). On a space of completely additive functions (Russian). Vestnik Leningrad Univ., Ser. Mat. Astronom. 13, 52-59.

5. Kemperman J. H. B. (1982). On the role of duality in the theory of moments. Proceedings Intern. Symp. Semi-Infinite Programming and Applications. Springer-Verlag.

6. Kemperman J. H. B. (1983). On the optimal measure of a set when only the marginals are prescribed (to appear).

7. Prohorov, Yu. V. (1956). Convergence of random processes and limit theorems in probability theory. Theory Prob. Appl. 1, 157-214.

8. Strassen V. (1965). The existence of probability measures with given marginals. Annals Math. Stat. 36. 423-439.

9. Whitt W. (1974). Preservation of rates of convergence under mappings. Z. Wahrscheinlichkeitstheorie verw. Gebiete 29, 39-44.

(Received by the editors, June 10, 1986)
Department of Mathematical Seienes, Memphis State University, U.S.A.