Z. Wahrscheinlichkeitstheorie verw. Gebiete 29, 39 – 44 (1974)


# Preservation of Rates of Convergence under Mappings

Ward Whitt*

For appropriate metrics characterizing various modes of stochastic convergence, it is shown that rates of convergence are preserved by a large class of functions. For example, the extensions of a Lipschitz function on a separable metric space $S$ to the space of all probability measures on $S$ with the Prohorov metric and to the space of all $S$-valued random variables with the usual metric associated with convergence in probability inherit the Lipschitz property. Consequently, just as with the continuous mapping theorem associated with ordinary convergence, new rate of convergence theorems can sometimes be obtained from old ones by applying appropriate mappings.

## 1. Introduction and Summary

The purpose of this note is to show that rates of convergence in law and in probability are preserved under a large class of mappings. Consequently, just as with the continuous mapping theorems associated with the various modes of stochastic convergence, new rate of convergence theorems can often be obtained from old ones by applying appropriate mappings. Quite naturally, a stronger property than continuity is needed to preserve rates of convergence. It turns out that the stronger property is for the function to be w.p.1 Lipschitz or Hölder continuous of order $t$, cf. (2.4). Thus, corresponding to the continuous mapping theorems associated with various modes of stochastic convergence, we speak of Lipschitz mapping theorems associated with rates of stochastic convergence.

By their very nature, rates of convergence must be stated in terms of metrics, and some care must be given to the choice of metrics. In particular, we shall show that the Lipschitz mapping theorem for convergence in law *does not* hold for the Lévy and supremum metrics applied to c.d.f.'s whereas it *does* hold for the Prohorov and dual-bounded Lipschitz metrics, cf. Section 2. If final statements in terms of the Lévy or supremum metrics are desired, then they can be obtained from the Prohorov metric. This suggests that it would be desirable to express as many rate of convergence results as possible in terms of the Prohorov metric. That this is often possible and natural has been amply demonstrated by Dudley [4].

We refer the reader to Dudley [4] for relevant background. The various metrics are defined and related in Section 2. Lipschitz mapping theorems for some metrics and counterexamples for other metrics appear in Section 3.

* Partially supported by National Science Foundation Grant GK-27866 and a Yale University Junior Faculty Fellowship in the Social Sciences.

## 2. The Metrics

Let $S$ be a SMS (separable metric space) with metric $d$ and Borel $\sigma$-field $\mathscr{S}$ (generated by the open subsets of $S$). Let $\mathscr{V} \equiv \mathscr{V}(S, d)$ be the set of all $S$-valued random variables on a fixed underlying probability space. Let $\mathscr{P} \equiv \mathscr{P}(S, d)$ be the space of all probability measures on $(S, \mathscr{S})$. We shall use lower case Greek letters for metrics on $\mathscr{V}$ and $\mathscr{P}$ in contrast with the lower case Roman letters to be used for metrics on the underlying spaces, e.g., $d$ on $S$ and $\alpha$ on $\mathscr{V}$.

Let $\alpha \equiv \alpha(S, d)$ denote the usual metric on $\mathscr{V}(S, d)$ corresponding to convergence in probability, that is, for $X, Y \in \mathscr{V}$, let

$$\alpha(X, Y) = \inf\{\varepsilon \geqq 0 \colon P[d(X, Y) \geqq \varepsilon] \leqq \varepsilon\}. \tag{2.1}$$

We use the separability assumption to have (2.1) well defined, cf. [1], p. 225. Obviously $\alpha(X_1, Y_1) \leqq \alpha(X_2, Y_2)$ if $d(X_1, Y_1) \leqq d(X_2, Y_2)$ w.p.1 in $\mathscr{V}(S, d)$.

We now turn to the space of probability measures $\mathscr{P}(S, d)$. Recall that $P_t \to P$ for a sequence or net $\{P_t\}$ in $\mathscr{P}(S, d)$ in the topology of weak convergence if $P_t(f) \to P(f)$ as $t \to \infty$ for all bounded continuous real-valued functions $f$ on $S$, where $P(f) = \int_S f\,dP$, cf. p. 11 of [1] or p. 40 of [7]. For $A \in \mathscr{S}$ on $(S, d)$, let

$$A^\varepsilon = \{y \colon \exists\, x \in A,\, d(x, y) < \varepsilon\}.$$

Then the *Prohorov metric* $\rho \equiv \rho(S, d)$, which induces the topology of weak convergence on $\mathscr{P}$, is defined by

$$\begin{aligned} \rho(P, Q) &= \max\{\gamma(P, Q), \gamma(Q, P)\}, \\ \gamma(P, Q) &= \inf\{\varepsilon \geqq 0 \colon P(F) \leqq \varepsilon + Q(F^\varepsilon),\, F \text{ closed}\}, \end{aligned} \tag{2.2}$$

cf. Section 2 of [3]. We also define $\rho$ on $\mathscr{V}$ by interpreting $\rho(X, Y)$ as $\rho[\mathscr{L}(X), \mathscr{L}(Y)]$ where $\mathscr{L}(X)$ is the probability law (measure) on $(S, \mathscr{S})$ induced by $X$. Of course, $\rho$ is only a pseudometric on $\mathscr{V}$. It is significant that $\rho(X, Y) \leqq \alpha(X, Y)$, cf. Theorem 1 of [3]. For example, Theorems 4.1 and 4.2 of [1] are elementary consequences of this fact.

The *Lévy metric* $\lambda$ on $\mathscr{P}(R)$ is the Prohorov metric restricted to closed sets of the form $(-\infty, x]$, cf. [5], p. 33. If we use the metric $r_k$ on $R^k$, where

$$r_k[(x_1, \ldots, x_k), (y_1, \ldots, y_k)] = \max_{1 \leqq i \leqq k} |x_i - y_i|,$$

then a natural generalization of the Lévy metric $\lambda$ to $\mathscr{P}(R^k, r_k)$ can be defined by using the Prohorov metric $\rho$ on $\mathscr{P}(R^k, r_k)$ restricted to closed sets of the form $(-\infty, x_1] \times \cdots \times (-\infty, x_k]$. Obviously, $\lambda \leqq \rho$, but $\lambda$ also induces the weak convergence topology on $\mathscr{P}(R^k, r_k)$, cf. p. 18 of [1]. Also note that $\gamma(P, Q) = \gamma(Q, P)$ for $\rho$ but not for $\lambda$.

The sup-metric $\sigma \equiv \sigma(S, d)$ on $\mathscr{P}(S, d)$ is defined by

$$\sigma(P, Q) = \sup\{|P(A) - Q(A)|,\, A \in \mathscr{S}\}. \tag{2.3}$$

Obviously, in general $\sigma$ induces a stronger topology than weak convergence on $\mathscr{P}$. Other metrics can be obtained by restricting the class of sets over which we take the supremum. Let $\mu$ denote the restriction of $\sigma$ on $\mathscr{P}(R^k, r_k)$ to sets of the form

$(-\infty, x_1] \times \cdots \times (-\infty, x_k]$. Obviously, $\lambda \leqq \mu$, but it is well known that $\mu$ also induces weak convergence on $\mathscr{P}(R^k)$ at those $P$ with continuous c.d.f.'s. We introduce $\mu$ because many existing rate of convergence theorems are stated in terms of it.

Following Dudley [2, 3, 4], we also introduce the dual-bounded Lipschitz metric $\beta$. Recall that a function $f: (S_1, d_1) \to (S_2, d_2)$ is *Hölder continuous of order t* if $\|f\|_t < \infty$, where $\|f\|_t = \|f, f\|_t$ and

$$\|f, g\|_t = \sup_{x \neq y} \{d_2[f(x), g(y)]/d_1(x, y)^t\} < \infty. \tag{2.4}$$

The function $f$ is said to be *Lipschitz* if it is Hölder continuous of order 1. Let $BL(S)$ denote the Banach space of all bounded real-valued Lipschitz functions $f$ on $(S, d)$ with norm

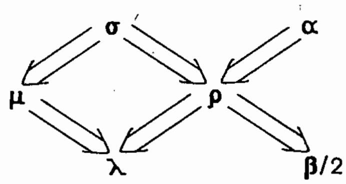
$$\|f\|_{BL} = \|f\|_1 + \|f\|_\infty, \tag{2.5}$$

where $\|f\|_\infty = \sup\{|f(x)|, x \in S\}$, cf. [2]. Then the dual-bounded Lipschitz metric $\beta \equiv \beta(S, d)$ is defined on $\mathscr{P}(S, d)$ by

$$\beta(P, Q) = \sup\{|P(f) - Q(f)| : f \in BL(S), \|f\|_{BL} \leqq 1\}. \tag{2.6}$$

Dudley has shown that $\beta$ is also a metric which induces the weak convergence topology on $\mathscr{P}(S, d)$ ([2], Theorems 6 and 8), that $\beta \leqq 2\rho$ ([3], Corollary 2), and that $\beta$ and $\rho$ define the same uniformity ([3], Remark after Corollary 3).

In conclusion, the inequality relationships for the metrics above, all of which can be regarded as pseudometrics on $\mathscr{V}$, are summarized in the diagram below:




## 3. Lipschitz Mapping Theorems and Generalizations

We say a family of functions $\mathscr{F} = \{f: S_1 \to S_2\}$ is *uniformly Hölder continuous of order t* if

$$\|\mathscr{F}\|_t = \sup\{\|f\|_t, f \in \mathscr{F}\} < \infty, \tag{3.1}$$

where $\|f\|_t$ is defined in (2.4). (The domain and range of $f$ could also depend on $f$.) We say that $\mathscr{F}$ is *uniformly Lipschitz* if $\|\mathscr{F}\|_1 < \infty$. We say that two families of functions $\mathscr{F} = \{f: S_1 \to S_2\}$ and $\mathscr{G} = \{g: S_1 \to S_2\}$ are *bi-uniformly Hölder continuous of order t* if

$$\|\mathscr{F}, \mathscr{G}\|_t = \sup\{\|f, g\|_t, f \in \mathscr{F} \text{ and } g \in \mathscr{G}\} < \infty, \tag{3.2}$$

where $\|f, g\|_t$ is defined in (2.4). We say that the pair $(\mathscr{F}, \mathscr{G})$ is bi-uniformly Lipschitz if $\|\mathscr{F}, \mathscr{G}\|_1 < \infty$. The special case of greatest interest to us arises when $\mathscr{F} = \{f_n\}$ and $\mathscr{G} = \{f\}$. Then

$$\|\mathscr{F}, \mathscr{G}\|_t = \sup_{x \neq y} \{d_2[f_n(x), f(y)]/d_1(x, y)^t\} \tag{3.3}$$

can be used to obtain rates of convergence of $f_n(x_n) \to f(x)$ from rates of convergence of $x_n \to x$. We shall only state results for $\|\mathscr{F}\|_t$. Corresponding results also hold for $\|\mathscr{F}, \mathscr{G}\|_t$ by essentially the same arguments. Let $x \vee y = \max\{x, y\}$.

**Theorem 3.1.** *If* $\|\mathscr{F}\|_t = c$, *then*

$$\alpha[f(X), f(Y)] \leqq c\,\alpha(X, Y)^t \vee \alpha(X, Y)$$

*for all* $f \in \mathscr{F}$, *where* $\alpha$ *is defined in* (2.1).

*Proof.* Since $\|\mathscr{F}\|_t = c$, $d_1(x, y) < b$ implies $d_2[f(x), f(y)] < c\,b^t$ for all $f \in \mathscr{F}$. Hence,

$$P\{d_2[f(X), f(Y)] \geqq c\,\varepsilon^t\} \leqq P\{d_1(X, Y) \geqq \varepsilon\} \leqq \varepsilon$$

for all $f \in \mathscr{F}$ and $\varepsilon \geqq \alpha(X, Y)$. ||

If we are concerned with the rate of convergence in $\mathscr{V}$ of a sequence $\{X_n\}$ to a specified limit $X$, then the Hölder continuity only needs to hold w.p.1. We say that $\|\mathscr{F}\|_t = c$ w.p.1 with respect to $X$ if (3.1) holds with the supremum in (2.4) being over all $y$ and some set of $x$ which has probability one with respect to $X$.

**Corollary 3.1.** *If* $\|\mathscr{F}\|_t = c$ w.p.1 *with respect to* $X$, *then*

$$\alpha[f(X_n), f(X)] \leqq c\,\alpha(X_n, X)^t \vee \alpha(X_n, X).$$

*Similar corollaries exist for the other theorems in this section.*

Each function $f: S_1 \to S_2$ has extensions $\hat{f}: \mathscr{V}(S_1) \to \mathscr{V}(S_2)$ and $\hat{f}: \mathscr{P}(S_1) \to \mathscr{P}(S_2)$, where $\hat{f}(X) = f(X)$ and $f(P) = Pf^{-1}$. Theorem 3.1 implies that $\hat{\mathscr{F}} = \{\hat{f}\}$ inherits the Lipschitz property from $\mathscr{F} = \{f\}$, that is,

**Corollary 3.2.** *If* $\|\mathscr{F}\|_1 = c$, *then* $\|\hat{\mathscr{F}}\|_1 = 1 \vee c$ *for the extension* $\hat{\mathscr{F}}$ *of* $\mathscr{F}$ *to* $(\mathscr{V}, \alpha)$.

**Lemma 3.1.** *If* $\|\mathscr{F}\|_t = c$ *and* $\delta = [\varepsilon/c]^{(1/t)}$, *then*

$$f^{-1}(A)^\delta \subseteq f^{-1}(A^\varepsilon)$$

*for any* $A \in \mathscr{S}$ *and* $f \in \mathscr{F}$.

*Proof.* If $x \in f^{-1}(A)^\delta$, then $d(x, y) < [\varepsilon/c]^{(1/t)}$ for some $y \in f^{-1}(A)$. Since $\|\mathscr{F}\|_t = c$, $d[f(x), f(y)] < \varepsilon$ for this $x$ and $y$. But this means that $x \in f^{-1}(A^\varepsilon)$. ||

**Theorem 3.2.** *If* $\|\mathscr{F}\|_t = c$, *then*

$$\rho(Pf^{-1}, Qf^{-1}) \leqq c\,\rho(P, Q)^t \vee \rho(P, Q)$$

*for all* $f \in \mathscr{F}$, *where* $\rho$ *is defined in* (2.2).

*Proof.* For each $f \in \mathscr{F}$,

$$\begin{aligned}
\rho(Pf^{-1}, Qf^{-1}) &= \inf\{\varepsilon \geqq 0\colon\ Pf^{-1}(F) \leqq \varepsilon + Qf^{-1}(F^\varepsilon),\ F \text{ closed}\} \\
&= \inf\{\varepsilon \geqq 0\colon\ P(f^{-1}(F)) \leqq \varepsilon + Q(f^{-1}(F^\varepsilon)),\ F \text{ closed}\} \\
&\leqq \inf\{\varepsilon \geqq 0\colon\ P(f^{-1}(F)) \leqq \varepsilon + Q(f^{-1}(F)^\delta),\ F \text{ closed}\} \\
&\leqq \inf\{\varepsilon \geqq 0\colon\ P(H) \leqq \varepsilon + Q(H^\delta),\ H \text{ closed}\} \\
&\leqq c\,\rho(P, Q)^t \vee \rho(P, Q),
\end{aligned}$$

where $\delta$ and the first inequality come from Lemma 3.1. ||

**Corollary 3.3.** *If* $\|\mathscr{F}\|_1 = c$, *then* $\|\hat{\mathscr{F}}\|_1 = 1 \vee c$ *for the extension* $\hat{\mathscr{F}} = \{\hat{f}\}$ *of* $\mathscr{F} = \{f\}$ *to* $(\mathscr{P}, \rho)$.

For the dual-bounded Lipschitz metric $\beta$, we only have results for Lipschitz mappings.

**Lemma 3.2.** *If* $\|\mathscr{F}\|_1 = c$, *then*

$$\|g \circ f\|_{BL} \leqq (1 \vee c)\, \|g\|_{BL}.$$

*Proof.* From (2.5), we have

$$\begin{aligned}\|g \circ f\|_{BL} &= \|g \circ f\|_1 + \|g \circ f\|_\infty = \|g\|_1\, \|f\|_1 + \|g\|_\infty \\ &= c\, \|g\|_1 + \|g\|_\infty \leqq (1 \vee c)\, \|g\|_{BL}. \quad \|\end{aligned}$$

**Theorem 3.3.** *If* $\|\mathscr{F}\|_1 = c$, *then*

$$\beta(Pf^{-1}, Qf^{-1}) \leqq (1 \vee c)\, \beta(P, Q)$$

*for all* $f \in \mathscr{F}$, *where* $\beta$ *is defined in* (2.6).

*Proof.* From (2.5), for any $f \in \mathscr{F}$,

$$\begin{aligned}\beta(Pf^{-1}, Qf^{-1}) &= \sup\{|Pf^{-1}(g) - Qf^{-1}(g)|:\ g \in BL(S_2),\ \|g\|_{BL} \leqq 1\} \\ &= \sup\{|P(g \circ f) - Q(g \circ f)|:\ g \in BL(S_2),\ \|g\|_{BL} \leqq 1\} \\ &\leqq \sup\{|P(g \circ f) - Q(g \circ f)|:\ g \circ f \in BL(S_1),\ \|g \circ f\|_{BL} \leqq 1 \vee c\} \\ &\leqq \sup\{|P(h) - Q(h)|:\ h \in BL(S_1),\ \|h\|_{BL} \leqq 1 \vee c\} \\ &= (1 \vee c)\, \beta(P, Q),\end{aligned}$$

with Lemma 3.2 being applied to obtain the first inequality. ||

In the way of positive results, we conclude with the following trivial result for $\sigma$ in (2.3).

**Theorem 3.4.** *For any* $\mathscr{F}$,

$$\sigma(Pf^{-1}, Qf^{-1}) \leqq \sigma(P, Q)$$

*for all* $f \in \mathscr{F}$.

Unfortunately, the results just obtained for $\alpha$, $\rho$, $\beta$, and $\sigma$ do not apply to the Lévy metric $\lambda$ and the supremum c.d.f. metric $\mu$. Dudley's example on p. 1572 of [4] for showing that the uniformity of Lévy's metric is strictly weaker than the uniformity of the dual-bounded Lipschitz metric provides

*Counterexample 3.1.* Let $P_n, Q_n \in \mathscr{P}(R)$ be defined by

$$P_n(2j) = Q_n(2j+1) = 1/n, \qquad 1 \leqq j \leqq n. \tag{3.4}$$

Let $f\colon R \to R$ be defined by

$$f(x) = \sin(\pi x/2), \qquad -\infty < x < \infty. \tag{3.5}$$

Obviously $\|f\|_1 = 1$, but $\lambda(P_n, Q_n) \leqq \mu(P_n, Q_n) = 1/n$, while $P_n f^{-1}(\{0\}) = 1$ and $Q_n f^{-1}(\{-1, 1\}) = 1$, so that $\mu(P_n f^{-1}, Q_n f^{-1}) \geqq \lambda(P_n f^{-1}, Q_n f^{-1}) = 1/2$ for all $n$.

As a consequence, we see that the metrics $\alpha$, $\rho$, $\beta$, and $\sigma$ are much better than $\lambda$ and $\mu$ for the purpose of generating additional rate of convergence results. Since $\lambda \leqq \rho$, final statements in terms of $\lambda$ can always be obtained immediately from statements in terms of $\rho$.

Many rates of convergence for convergence in law are actually stated for real-valued random variables in terms of the c.f.d. metric $\mu$. This is the case with Theorems 1 and 5 of Rosenkrantz [7] for example. Since $\mu$ is not dominated by $\lambda$, $\rho$, or $\alpha$, no general corollary to results for $\lambda$, $\rho$, or $\alpha$ can be obtained for $\mu$ without additional conditions, but additional conditions are provided by the easily verified

**Theorem 3.5.** *Let $F$ be the c.d.f. corresponding to an arbitrary $P \in \mathscr{P}(R^k)$. If $\|F\|_1 = c$, then*

$$\mu(P, Q) \leqq (1+c)\,\lambda(P, Q).$$

Obviously $\|F\|_1 = c$ whenever $F$ has a bounded density. Note that the two conditions in Theorem 5 of Rosenkrantz [7] are just the conditions appearing in Corollary 3.3 and Theorem 3.5 here. In other words, the conditions in this section have been used before, but we indicate what each condition yields in a more general setting.

## References

1. Billingsley, P.: Convergence of Probability Measures. New York: Wiley 1968
2. Dudley, R.M.: Convergence of Baire measures. Studia Math. **27**, 251–268 (1966)
3. Dudley, R.M.: Distances of probability measures and random variables. Ann. Math. Statist. **39**, 1563–1572 (1968)
4. Dudley, R.M.: Speeds of metric probability convergence. Z. Wahrscheinlichkeitstheorie verw. Gebiete **22**, 323–332 (1972)
5. Gnedenko, B.V., Kolmogorov, A.W.: Limit Distributions for Sums of Independent Random Variables. Second Edition of English translation. Reading, Mass.: Addison-Wesley 1968
6. Parthasarathy, K.R.: Probability Measures on Metric Spaces. New York: Academic Press 1967
7. Rosenkrantz, W.: On rates of convergence for the invariance principle. Trans. Amer. Math. Soc. **129**, 542–552 (1967)

Ward Whitt
Department of Administrative Sciences
Yale University
2 Hillhouse Avenue
New Haven, Connecticut 06520
USA

*(Received October 30, 1972; in revised form October 10, 1973)*