STOCHASTICA Vol. IX Nº 1, 1985

NEW METRICS FOR WEAK CONVERGENCE OF DISTRIBUTION FUNCTIONS

Michael D. Taylor

ABSTRACT

Sibley and Sempi have constructed metrics on the space of probability distribution functions with the property that weak convergence of a sequence is equivalent to metric convergence. Sibley's work is a modification of Lévy's metric, but Sempi's construction is of a different sort. Here we construct a family of metrics having the same convergence properties as Sibley's and Sempi's but which does not appear to be related to theirs in any simple way. Some instances are brought out in which the metrics have probabilistic interpretations.

1. Introduction.

Lévy introduced a metric on Δ, the set of probability distribution functions, with the property that weak convergence of a sequence is equivalent to metric convergence so long as one works in the subspace Δ_0, the set of distribution functions generated by random variables for which $P[|X| = +\infty] = 0$; see [2, p. 228]. Sibley in [7] produced a modification of Lévy's metric with the pleasant property that weak convergence and metric convergence are equivalent on <u>all</u> of Δ. A slight modification of Sibley's metric is discussed in [3]. Sempi in [4] has exhibited yet a third way of constructing a metric with the desired convergence properties on Δ. In [5] he investigates the connection between the weak convergence of r-dimensional distribution functions and the product topology on $\Delta \times \Delta \times \ldots \Delta$ induced by his and Sibley's metrics on Δ. In [6] he extends his metric to one for weak convergence of multiple distribution functions.

This paper exhibits a family of metrics on Δ which appear to be unrelated to Sibley's but which still enjoy the property that weak and metric convergence are equivalent on Δ. The basic idea is rather simple-minded:

Think of distribution functions as graphs in the infinite strip $(-\infty,\infty) \times [0,1]$. We shrink this strip, horizontally, onto the square $(0,1) \times [0,1]$ and then use some standard metrics to measure "distance" between images of distribution functions. We will also see some consequences of the equivalence of metric and weak convergence and some probabilistic interpretations of some of the metrics under certain conditions.

A metric for spaces of real-valued randon variables which is suggested by this work is described in [8].

I am indebted to the referee for his helpful comments.

2. The Main Result.

By Δ, the set of distribution functions, we mean the set of nondecreasing, left-continuous functions from R into [0,1]. The set of distribution functions for which $P[|X| = +\infty] = 0$, Δ_0, consists of the members of Δ satisfying

$$\lim_{x \to \infty} F(x) = 1 \quad \text{and}$$

$$\lim_{x \to -\infty} F(x) = 0.$$

If F, $F_1, F_2, F_3, \ldots$ belong to Δ, we say the sequence $\{F_n\}$ converges **weakly** to F and write $F_n \xrightarrow{w} F$ provided $\{F_n\}$ converges pointwise to F at every point of continuity of F.

Our first requirement is a lemma about nondecreasing functions from (0,1) to [0,1]. Making an obvious modification of terminology, we say $f_n \xrightarrow{w} f$ on (0,1) if $\{f_n\}$ converges pointwise to f at every point of continuity of f in the interval (0,1). (Note: all integration is with respect to Lebesgue measure).

Lemma 1. Suppose that $0 < p < \infty$ and that $f, f_1, f_2, f_3, \ldots$ are nondecreasing functions from (0,1) into [0,1]. Then $f_n \xrightarrow{w} f$ if and only if $\int_0^1 |f_n - f| \to 0$.

Proof. If $f_n \xrightarrow{w} f$ on (0,1), then $\int_0^1 |f_n - f| \to 0$ by the Lebesgue convergence theorem.

Now assume that $\int_0^1 |f_n - f| \to 0$ and that there is a point of continuity, x_0, of f for which $f_n(x_0) \not\longrightarrow f(x_0)$. Then we must be able to find $\varepsilon > 0$ and $\delta > 0$ satisfying

$$|f_n(x_0) - f(x_0)| \geqslant \varepsilon \text{ for an infinite number of n's,}$$

$$|f(x) - f(x_0)| < \varepsilon/2 \quad \text{for all } x$$

$$\text{satisfying } |x - x_0| < \delta \text{ , and}$$

$$[x_0 - \delta, x_0 + \delta] \subset (0,1).$$

Case 1. Suppose we can find a subsequence $\{f_{n_k}\}$ of $\{f_n\}$ satisfying both

$$|f_{n_k}(x_0) - f(x_0)| \geqslant \varepsilon \quad \text{and}$$

$$f_{n_k}(x_0) > f(x_0)$$

for all k. Since the functions involved are nondecreasing, we must have $f_{n_k}(x) - f(x) \geqslant \varepsilon/2$ for all x satisfying $x_0 \leqslant x \leqslant x_0+\delta$. Thus

$$\int_0^1 |f_{n_k} - f|^p \geqslant \int_{x_0}^{x_0+\delta} (f_{n_k} - f)^p$$

$$\geqslant \delta(\varepsilon/2)^p.$$

Hence $\int_0^1 |f_n - f|^p \not\to 0$, a contradiction.

Case 2. We now suppose there is a subsequence $\{f_{n_k}\}$ of $\{f_n\}$ satisfying

$$|f_{n_k}(x_0) - f(x_0)| \geqslant \varepsilon \text{ and } f_{n_k}(x_0) < f(x_0)$$

for all k. The rest of the proof is a simple modification of case 1.

Since case 1 or case 2 must occur, this establishes the lemma.

<u>Theorem 1.</u> Let h be a homeomorphism of the open interval (0,1) onto $\mathbb{R}$ and let p be a positive real number. For F and G members of Δ define

$$d_{h,p}(F,G) = \begin{cases} \int_0^1 |F\circ h - G\circ h|^p & \text{if } 0 < p < 1 \text{ and} \\ (\int_0^1 |F\circ h - G\circ h|^p)^{1/p} & \text{if } 1 \leqslant p < \infty. \end{cases}$$

Then $d_{h,p}$ is a metric on Δ, and for any F and sequence $\{F_n\}$ in Δ it follows that $F_n \xrightarrow{w} F$ if and only if $d_{h,p}(F_n,F) \to 0$.

Proof. To see that $d_{h,p}$ is a metric, note first that for F and G in Δ, (1) $F = G$ implies $\int_0^1 |F\circ h - G\circ h|^p = 0$, and (2) $F \neq G$

implies by the left-continuity of F and G that $\int_0^1 |F\circ h - G\circ h|^p > 0$. So F = G if and only if $d_{h,p}(F,G) = 0$. The triangle inequality follows from the fact that we have adapted to our uses well-known metric definitions from analysis; see, for example, [2].

To finish the proof, let F be a member of Δ and $\{F_n\}$ be a sequence in Δ and note that the following statements are equivalent:

$$F_n \xrightarrow{w} F.$$

$$F_n \circ h \xrightarrow{w} F\circ h \quad \text{on} \quad (0,1).$$

$$\int_0^1 |F_n\circ h - F\circ h|^p \longrightarrow 0.$$

Naturally the use of (0,1) in this result is quite arbitrary. Any finite, open interval would do just as well.

3. Some Consequences and Interpretations.

It is known that Δ is sequentially compact with respect to weak convergence [2], therefore

Theorem II. If h is any homeomorphism of (0,1) onto R and p any positive real number, then $(\Delta, d_{h,p})$ is a compact metric space.

There is a nice way to rewrite the metrics $d_{h,p}$. Let G be any distribution function which is a homeomorphism of R onto (0,1). We can take $h = G^{-1}$ and appeal to the change of variables formula to write

$$d_{h,p}(F_1,F_2) = \int_R |F_1 - F_2|^p \, dG \text{ for } 0 < p < 1 \text{ and}$$

$$\left(\int_R |F_1 - F_2|^p \, dG\right)^{1/p} \text{ for } 1 \leq p < \infty.$$

This in turn leads to probabilistic interpretations of the metrics in some simple situations. To see this, it is convenient to have the following lemma:

Lemma 2. If X and Y are independent, real-valued random variables on a probability space Ω and F_X and F_Y are the distribution functions of X and Y respectively, then for any (extended-) real numbers u and v, $u < v$, we have

$$\int_u^v F_X dF_Y = P[X < Y \text{ and } u \leqslant Y < v]$$

where P is the probability measure on Ω.

Proof. Let $\{x_n\}$ be a dense sequence with no repetitions in the interval $[u,v]$ such that $x_1 = u$ and $x_2 = v$. Define

$$\{t_{k1}, t_{k2}, t_{k3}, \ldots, t_{kk}\} = \{x_1, x_2, x_3, \ldots, x_k\}$$

where $t_{k1} < t_{k2} < \ldots < t_{kk}$ and $k = 1,2,3,\ldots$

Define the function ϕ_k over $[u,v)$ by

$$\phi_k(t) = F_X(t_{kj}) \text{ if } t_{kj} \leqslant t < t_{k,j+1}.$$

Clearly $\phi_1 < \phi_2 \leqslant \ldots \leqslant F_X$ over $[u,v)$. Choose t from $[u,v)$. For every k there is a unique t_{km} satisfying $t_{km} \leqslant t < t_{k,m+1}$ and $\phi_k(t) = F_X(t_{km})$. Because of the denseness of $\{x_n\}$, we see that $t_{km} \to t$ from the left as $k \to \infty$, and hence, by virtue of the left-continuity of F_X, we have $F_X(t_{km}) \to F_X(t)$. So $\phi_k \to F_X$ pointwise as $k \to \infty$, and by the monotone convergence theorem

$$\int_u^v \phi_k dF_Y \to \int_u^v F_X dF_Y.$$

Note that

$$\int_u^v \phi_k dF_Y = \Sigma_{j=1}^{k-1} F_X(t_{kj}) \cdot [F_Y(t_{k,j+1}) - F_Y(t_{kj})]$$

$$= \Sigma_{j=1}^{k-1} [P\ X < t_{kj}] \cdot P[t_{kj} \leqslant Y < t_{k,j+1}]$$

$$= P[X < t_{kj} \leqslant Y < t_{k,j+1} \text{ for some } j].$$

Letting $k \to \infty$, we obtain

$$\int_u^v F_X dF_Y = P[X < Y \text{ and } u \leqslant Y < v]$$

which completes the proof.

A probabilistic interpretation of $d_{h,1}$ under certain very special conditions is an immediate consequence of the next lemma.

<u>Lemma 3.</u> Let X, Y, and Z be real-valued random variables on a probability space Ω such that X and Z are mutually independent, Y and Z are mutually independent, and $X \leqslant Y$. Then

$$\int_R |F_X - F_Y| dF_Z = P[X < Z \leqslant Y]$$

where P is the probability measure on Ω.

Proof. Let

$$A = \{\omega \epsilon \Omega: X(\omega) < Z(\omega)\},$$

$$B = \{\omega \epsilon \Omega: Y(\omega) < Z(\omega)\}, \quad \text{and}$$

$$C = \{\omega \epsilon \Omega: X(\omega) < Z(\omega) \leqslant Y(\omega)\}.$$

Note that A is the disjoint union of B and C so that $P(A) = P(B) + P(C)$.

For any real number t, we have $F_X(t) = P[X<t] \geqslant P[Y<t] = F_Y(t)$ since $X \leqslant Y$.

Therefore

$$\int_R |F_X - F_Y| dF_Z = \int_R F_X dF_Z - \int_R F_Y dF_Z$$

$$= P(A) - P(B) \text{ (by Lemma 2)}$$

$$= P(C)$$

$$= P[X < Z \leqslant Y].$$

Theorem III. If, in the last lemma, F_Z is a homeomorphism of R onto (0,1), then taking $h = F_Z^{-1}$ we have

$$d_{h,1}(F_X, F_Y) = P[X < Z \leqslant Y].$$

A second probabilistic interpretation is found by considering maps X and Y from a probability space Ω into a metric space (M,d) such that $d(X,Y)$ is a random variable. Let P be the probability measure on Ω. We ought to have $P[d(X,Y)<t] = 0$ if $t \leqslant 0$ and $P[d(X,Y)<t]>0$ for some $t>0$. It also seems reasonable to expect $P[d(X,Y)<t] = 1$ for all $t>0$ precisely in the case when $X=Y$. Let us define

$$\varepsilon_0(t) = \begin{cases} 0 & \text{if } t \leqslant 0 \text{ and} \\ 1 & \text{if } t > 0. \end{cases}$$

Then one measure of how "close" X and Y are to one another ought to be the "distance" between the distribution functions $P[d(X,Y)<t]$ and $\varepsilon_0(t)$.

Theorem IV. Let X and Y be maps from a probability space Ω into a metric space (M,d) such that $d(X,Y)$ is a random variable on Ω, and let Z be a real-valued random variable on Ω such that $d(X,Y)$ and Z are independent and F_Z, the distribution function of Z, is a homeomorphism of R onto (0,1). If F is the distribution function of d(X,Y) (i.e., F(t) = P[d(X,Y) < t]), then taking $h = F_Z^{-1}$ we have

$$d_{h,1}(\varepsilon_0, F) = P[0 \leqslant Z < d(X,Y)].$$

Proof. Note that P[d(X,Y) < t] = 0 if t ⩽ 0, therefore $F \leqslant \varepsilon_0$. Then

$$\begin{aligned} d_{h,1}(\varepsilon_0, F) &= \int_R |\varepsilon_0 - F| dF_Z \\ &= \int_R \varepsilon_0 dF_Z - \int_R F dF_Z \\ &= \int_0^\infty dF_Z - (F.F_Z|_{-\infty}^{\infty} - \int_R F_Z dF) \\ &= 1 - F_Z(0) - 1 + P[Z < d(X,Y)] \\ &= P[Z < d(X,Y)] - P[Z < 0] \\ &= P[0 \leqslant Z < d(X,Y)]. \end{aligned}$$

As a final result we show that weak convergence of distribution functions is equivalent to weak convergence of their quasi-inverses. Another proof of this can be found in the lemma to Theorem 7 of [1].

If F, a distribution function, is both continuous and strictly increasing, then by its quasi-inverse we shall mean just its inverse. If on the other hand F is not always continuous or not always increasing, this means it well have gaps in the graph or horizontal line segments or both. See Figure 1. To get the quasi-inverse, F^, we fill in the missing vertical line segments and erase the horizontal segments and

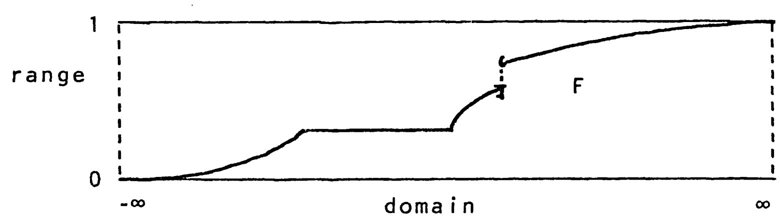

Figure 1

think of (0,1) (or [0,1]) and (-∞,∞) (or [-∞,∞]) as the range of the new function. See Figure 2.

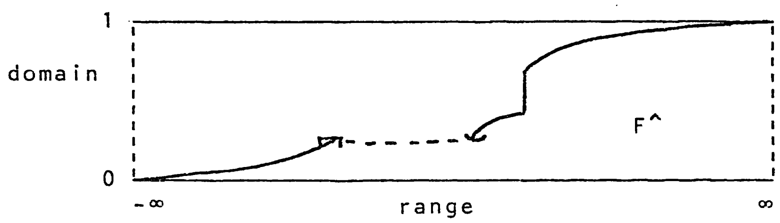

Figure 2

Here is an analytical definition:

Definition. If $F: [a,b] \to [c,d]$ is nondecreasing (where the intervals can be finite or infinite), then $F^{\wedge}: [c,d] \to [a,b]$ is defined by

$$F^{\wedge}(x) = \sup\{t \in [a,b] : F(t) < x\}.$$

We note several things about this definition. Firstly, if there is no t satisfying $F(t) < x$, we take $F^{\wedge}(x)$ to be a. Secondly, when dealing with distribution functions, members of Δ, we automatically think of their domains as being extended to $[-\infty,\infty]$; we might, for example, set $F(-\infty) = 0$ and $F(\infty) = \lim_{t\to\infty} F(t)$, but it will turn out to be irrelevant which extension we use so long as

F is nondecreasing. Thirdly, under this definition $F^\wedge$ turns out to be nondecreasing and left-continuous. Fourthly, this particular definition is only a special case of a more general definition of quasi-inverse to be found in [3].

Lemma 4. Let $F:[a,b] \to [c,d]$ be a nondecreasing function and $h:[c,d] \to [a,b]$ an onto, orientation-preserving homeomorphism. Then $(F \circ h)^\wedge = h^{-1} \circ (F^\wedge)$.

Proof. Choose x_0 a member of $[c,d]$. Let

$$A = \{t \in [c,d] : F(h(t)) < x_0\} \quad \text{and}$$

$$B = \{w \in [a,b] : F(w) < x_0\}.$$

Since the functions involved are nondecreasing, the sets A and B must be intervals or empty. Note that $B = h(A)$. Note also that $(F \circ h)^\wedge(x_0) = \sup A$. If we call this number y_0, then it must be either the right-hand endpoint of A in the case that A is an interval or c in case A is empty. Thus $h(y_0)$ is either the right-hand endpoint of the interval B or a in case B is empty. It follows that $F^\wedge(x_0) = \sup B = h(y_0)$. Hence, $h^{-1}(F^\wedge(x_0)) = y_0 = (F \circ h)^\wedge(x_0)$.

Theorem V. Let F be a member of Δ and $\{F_n\}$ a sequence in Δ. Then $F_n \xrightarrow{w} F$ if and only if $F_n^\wedge \xrightarrow{w} F^\wedge$ on $(0,1)$.

Proof. Let $h:[0,1] \to [-\infty,\infty]$ be an onto, orientation-preserving homeomorphism. Notice that for every n,

$$\begin{aligned}\int_0^1 |F_n \circ h - F \circ h| &= \text{the area between the graphs of } F_n \circ h \text{ and } F \circ h \\ &= \text{the area between the graphs of } (F_n \circ h)^\wedge \text{ and } (F \circ h)^\wedge \\ &= \int_0^1 |(F_n \circ h)^\wedge - (F \circ h)^\wedge| \\ &= \int_0^1 |h^{-1} \circ (F_n^\wedge) - h^{-1} \circ (F^\wedge)|.\end{aligned}$$

Then the following statements are equivalent:

$$F_n \xrightarrow{w} F.$$

$$\int_0^1 |F_n \circ h - F \circ h| \to 0.$$

$$\int_0^1 |h^{-1} \circ (F_n^\wedge) - h^{-1} \circ (F^\wedge)| \to 0.$$

$$h^{-1} \circ (F_n^\wedge) \xrightarrow{w} h^{-1} \circ (F^\wedge) \text{ on } (0,1).$$

Since h is a homeomorphism, the points of continuity of $h^{-1} \circ (F^\wedge)$ in (0,1) are precisely those of $F^\wedge$ in (0,1). Hence $F_n \xrightarrow{w} F$ if and only if $F_n^\wedge \xrightarrow{w} F^\wedge$ on (0,1).

References

[1] J. B. BROWN, Stochastic metrics, Z. Wahrsch. Verw. Gebiete, 24 (1972), 49-62.

[2] M. LOÈVE, Probability Theory I, 4th ed., Springer, New York, 1977.

[3] B. SCHWEIZER and A. SKLAR, Probabilistic Metric Spaces, Elsevier-North-Holland, New York, 1983.

[4] C. SEMPI, On the space of distribution functions, Riv. Mat. Univ. Parma (4)8 (1982), 243-250.

[5] C. SEMPI, Product topologies on the space of distribution functions, Riv. Mat. Univ. Parma (4)9 (1983), 425-432.

[6] C. SEMPI, A metric for weak convergence of multiple distribution functions, to appear in Note di Matematica.

[7] D. SIBLEY, A metric for weak convergence of distribution functions, Rocky Mountain J. Math., 1, (1971), 427-430.

[8] M. D. TAYLOR, Separation metrics for real-valued random variables, Internat. J. Math. Math. Sci. 7 (1984), 407-408.

Manuscript received January 22, 1985, and in final form May 7, 1985.

Department of Mathematics.
University of Central Florida.
Orlando, Florida 32816.