PRESERVATION OF WEAK CONVERGENCE UNDER MAPPINGS

By Flemming Topsoe

University of Copenhagen

Suppose we have a weakly convergent sequence of probability measures defined on some space S and that we carry the measures over to another space S' by means of a measurable mapping, or perhaps by means of a whole sequence of mappings. How far is weak convergence preserved?

Throughout what follows, S and S' denote two separable metric spaces. The letter h will always denote a measurable mapping from S into S' (Borel measurability), the letter g will denote a measurable mapping from S' into the reals $\dot{R}$, P will be used for a probability measure on S, and Q will be used for a probability measure on S'. Weak convergence of a sequence of probability measures, notationally indicated by the symbol $\rightarrow_w$, is defined in the usual way requiring convergence of the integrals for every real, bounded and continuous function.

If h is a P-continuity function (i.e. continuous a.e. P) and if $P_n \rightarrow_w P$ then weak convergence is preserved, i.e. $P_n h^{-1} \rightarrow_w Ph^{-1}$. This is almost trivial and one would guess that the P-continuity of h is also necessary for the preservation of weak convergence; indeed, this is so as demonstrated in [4].

A more complicated problem arises if, instead of one h, we have a whole sequence $\{h_n\}$ of mappings and ask whether $P_n h_n^{-1} \rightarrow_w Ph^{-1}$ holds for every sequence $\{P_n\}$ with $P_n \rightarrow_w P$. A powerful sufficient condition has been given by Rubin in an unpublished paper ([5]). Here we shall find necessary and sufficient conditions.

Since it is of no importance that the limit measure in the above formulation of the problem is generated from P via a mapping h, we shall replace it by a measure Q. To be precise, we are given a sequence of mappings $\{h_n\}_{n \geqq 1}$, a probability measure P, and a probability measure Q; and we search after conditions that $P_n h_n^{-1}$ converges weakly to Q whenever P_n converges weakly to P. When this holds, we shall say that *weak convergence is preserved*; from the context it should always be clear which mappings and measures we have in mind. Clearly, weak convergence is preserved iff

$$(1) \qquad \forall_{g \text{ bd. cont.}}, \quad \forall_{P_n \rightarrow_w P} \int g(h_n)\, dP_n \rightarrow \int g\, dQ$$

holds (bd. cont. = "bounded continuous").

For every fixed g we can solve the problem suggested by (1). If f is a function from S into $\dot{R}$ and if δ and ϵ are positive, we denote by $\partial_{\delta,\epsilon}(f)$ or $\partial_{\delta,\epsilon}f$, the *$\delta,\epsilon$-boundary* of f, the set of those points x in S for which the distance between $f(x')$ and $f(x'')$ exceeds ϵ for some pair of points x', x'' in the open sphere with center x and radius δ (see [4], [6]).

Received 1 May 1967.

THEOREM 1. *Given P, a sequence $\{f_n\}_{n\geq 1}$ of bounded, real-valued, and measurable functions defined on S, and a real number α (which, in many applications, is of the form $\int f\,dP$). Then a necessary and sufficient condition that $\int f_n\,dP_n \to \alpha$ for every sequence $\{P_n\}$ converging weakly to P is that*

(i) *the sequence $\{f_n\}$ be uniformly bounded,*

(ii) $\int f_n\,dP \to \alpha$, *and*

(iii) $\forall_{\epsilon>0} \lim_{\delta\to 0} \limsup_{n\to\infty} P(\partial_{\delta,\epsilon} f_n) = 0$

hold.

This result, interesting in its own right, is easily proved by adapting the ideas and methods presented in [4] to the present problem and there seems to be no point in running through a detailed proof.

Using the methods of [6] one finds that the condition (iii) above holds iff for every $\epsilon > 0$, for every sequence $\{\delta_k\}$ of positive numbers converging to 0, and for every subsequence $\{f_{n_k}\}$ we have

$$P\left(\bigcap_{k=1}^{\infty} \partial_{\delta_k,\epsilon}(f_{n_k})\right) = 0. \tag{2}$$

This condition, which can be easier to check than (iii), will not be used in the sequel.

Applying our new knowledge we find that (1) holds iff $Ph_n^{-1} \to_w Q$ and

$$\forall_{\epsilon>0} \forall_{g \text{ bd. cont.}} \lim_{\delta\to 0} \limsup_{n\to\infty} P(\partial_{\delta,\epsilon} g(h_n)) = 0 \tag{3}$$

hold. As one might guess from this, we find the following answer to our problem:

THEOREM 2. *A necessary and sufficient condition that weak convergence is preserved is that*

$$Ph_n^{-1} \to_w Q \tag{4}$$

and

$$\forall_{\epsilon>0} \lim_{\delta\to 0} \limsup_{n\to\infty} P(\partial_{\delta,\epsilon} h_n) = 0 \tag{5}$$

hold.

The δ,ϵ-boundaries occurring in (5) are defined in analogy with the definition for real-valued functions.

It is implicit in the theorem that the validity of the conditions is independent of the choice of metrics (as long as they generate the correct topologies). The condition (5) can, just as well as the condition (iii) of Theorem 1, be reformulated by requiring that sets of the form $\bigcap_{k=1}^{\infty} \partial_{\delta_k,\epsilon}(h_{n_k})$ have P-measure 0.

Note that (5) does not involve the measure Q.

PROOF OF THEOREM 2. The derivation of (3) tells us that we need only consider those bounded continuous functions g which are uniformly continuous. The sufficiency follows from this remark. Assume now that weak convergence is preserved. Then (3) and (4) hold. We must prove that (5) is satisfied. Clearly, this is so in case S' is the unit interval. Then the result follows in case S' is a countable product of unit intervals (with the usual metric for product spaces). Since any separable metric space is homeomorphic to a subset of the countable product space just considered, the result follows for the general S'. The theorem is now proved except for the fact that we only proved it with a specific distance function in S' (one generated by a homeomorphism to a space with a specific distance function). Assume then that d and d^* are two equivalent metrics in S' and assume that the condition of the theorem based on d holds; we must prove that the condition based on d^* holds. The reader should observe that the condition (5) is not independent of the metric (it is easy to construct a suitable example). However, we only claim that (5) taken together with (4) is independent of the metric. Denoting open spheres by $S(\cdot, \cdot)$ and using obvious notation, we define for every y in S' and for every positive ϵ a number $\rho_\epsilon(y)$ by

$$\rho_\epsilon(y) = \sup\{\rho : S(y, \rho) \subset S^*(y, \epsilon)\}.$$

We need the following easily established facts:

$$(6) \qquad \forall_{y \in S'} \forall_\epsilon \, \rho_\epsilon(y) > 0,$$

$$(7) \qquad \forall_{\epsilon,\eta} \, \mathrm{cl}\,\{y : \rho_\epsilon(y) \leqq \eta\} \subset \{y : \rho_{\frac{1}{2}\epsilon}(y) < 2\eta\},$$

$$(8) \qquad \forall_h \forall_{\epsilon,\eta,\delta} \, \partial^*_{\delta,2\epsilon}(h) \subset h^{-1}(\mathrm{cl}\,\{y : \rho_\epsilon(y) \leqq \eta\}) \cup \partial_{\delta,\eta}(h);$$

here, cl denotes closure.

Fix $\epsilon > 0$. Then, for any $\eta > 0$ we have

$$\begin{aligned}\lim_{\delta\to 0} \limsup_{n\to\infty} P(\partial^*_{\delta,2\epsilon} h_n) &\leqq \limsup_{n\to\infty} P h_n^{-1}(\mathrm{cl}\{y : \rho_\epsilon(y) \leqq \eta\}) \\ &\leqq Q(\mathrm{cl}\,\{y : \rho_\epsilon(y) \leqq \eta\}).\end{aligned}$$

Thus

$$\begin{aligned}\lim_{\delta\to 0} \limsup_{n\to\infty} P(\partial^*_{\delta,2\epsilon} h_n) &\leqq \lim_{\eta\to 0} Q(\mathrm{cl}\,\{y : \rho_\epsilon(y) \leqq \eta\}) \\ &= Q(\textstyle\bigcap_{\eta>0} \mathrm{cl}\,\{y : \rho_\epsilon(y) \leqq \eta\}) = Q(\varnothing) = 0. \quad \text{Q.E.D.}\end{aligned}$$

Let us turn to the case where Q is generated from P via a given mapping h, i.e. assume that $Q = Ph^{-1}$. In Theorem 2 we only have to change (4) to

$$(9) \qquad Ph_n^{-1} \to_w Ph^{-1},$$

that is we demand that h_n, considered as a random element, converges in distribution to h. We remind the reader of the following implications:

$$\begin{aligned}Ph_n^{-1} \to_w Ph^{-1} &\Leftrightarrow \forall_{g \text{ bd. cont.}} \int \{g(h_n) - g(h)\}\, dP \to 0 \\ &\Leftarrow \forall_{g \text{ bd. cont.}} \int |g(h_n) - g(h)|\, dP \to 0 \\ &\Leftrightarrow h_n \to h \text{ in measure} \\ &\Leftarrow h_n \to h \quad \text{a.e.}\end{aligned}$$

It is of course easy to construct examples of essentially different mappings having the same distribution, thus the second implication arrow can not be reversed.

Rubin considers the set

$$(10) \qquad E = \{x \in S : h_n(x_n) \to h(x), \forall\, x_n \to x\}$$

and finds that if $PE = 1$ then weak convergence is preserved. To get a feeling for the content of this condition note that $E = S$ amounts to the same thing as the continuity of h and the uniform convergence of h_n to h on every compact subset of S. For the purpose of bringing Rubin's condition in a form closer to our condition, observe that a point x in S belongs to E iff $h_n x \to hx$ and $\lim_{\delta\to 0} \limsup_{n\to\infty} \omega_{h_n}(S(x, \delta)) = 0$ hold; in the last expression we look at the oscillation of h_n in the sphere $S(x, \delta)$. Denoting by C the set of convergence:

$$C = \{x \varepsilon S : h_n x \to hx\},$$

we find that

$$E = C \cap \bigcap_{\epsilon>0} \bigcup_{\delta>0} \bigcup_n \bigcap_{m\geq n} (S \backslash \partial_{\delta,\epsilon}(h_n)). \tag{11}$$

The second set occurring in (11) is clearly measurable and, since S' is separable, the set C is also measurable (consider the mappings $x \to (h_n x, hx) \to d(h_n x, hx)$), thus E is measurable[1]).

From (11) we deduce that Rubins condition is equivalent to the two requirements:

$$h_n \to h \quad \text{a.e.} \tag{12}$$

and

$$\forall_{\epsilon>0} \lim_{\delta\to 0} \lim_{n\to\infty} P(\bigcup_{m\geq n} \partial_{\delta,\epsilon}(h_n)) = 0. \tag{13}$$

If we wish, the limit in (13) can be written in the form $\lim_\delta P(\limsup_n \partial_{\delta,\epsilon} h_n)$.

Clearly, (12) implies (9) and (13) implies (5) so that we have now proved Rubin's theorem.

Rubin's condition is sufficient but not necessary. In fact neither (12) nor (13) are necessary. As far as (12) is concerned this follows from previous remarks. To see that (13) is not necessary consider the example $S = S' = [0, 1]$, P = Lebesgue measure, $h \equiv 0$, and $h_n x = 0$ except for $x = r_n$ when $h_n x = 1$; here $\{r_n\}$ is a sequence of points dense in $[0, 1]$. Clearly, weak convergence is preserved but (13) fails. Also note that (12) holds in this example (we even have convergence everywhere). Qualitatively speaking, weak convergence is preserved if the mappings are well behaved except for a limited number of peaks (say that each h_n has one peak) even if the locations of all the peaks are well spread out over S. This result can not be proved by Rubin's method.

We shall now indicate some situations to which the necessity of our condition for preservation of weak convergence applies. Consider the random variables $\{h_n\}$ associated with the central limit theorem and assume that these have been realized on the space $\dot{R}^\infty$. That is, our basic probability space is $\dot{R}^\infty$ with some (Borel-) probability measure P and for $x = (x_1, x_2, \cdots)$ in $\dot{R}^\infty$ $h_n(x)$ denotes the normalized sum $A_n^{-1}(x_1 + x_2 + \cdots + x_n) + B_n$. For suitable normalizations

[1] Rubin's result holds even if we drop the separability assumption; however, there is some difficulty in proving the rather unessential claim that E is measurable (is it ?).

(the A's and B's) and suitable P, Ph_n^{-1} converges weakly to a normal distribution. If we replace P by a sequence $\{P_n\}$ converging weakly to P, we can no longer be sure that we have weak convergence to the normal distribution, that is weak convergence is not preserved. This follows from Theorem 2. In fact, it turns out that, no matter what P is and no matter how we choose the A's and B's, the condition (5) always breaks down. The same negative result is found if one considers the random variables associated with the arch-sine law or those associated with the limit theorem for the maximal among the n first sums.

Lastly, some bibliographical remarks, most of them communicated to me by Professor T. W. Anderson. Rubin's theorem was established in the unpublished paper [5]. I believe that this paper contains a proof of precisely the version we have called "Rubin's theorem." A special case of the result proved in [5] was proved by T. W. Anderson ([1]). Another special case was employed by T. W. Anderson and H. Rubin in [2]. A proof, different from ours, of Rubin's theorem is soon to be published in a monograph by P. Billingsley ([3]). Further remarks can be found in [1].

REFERENCES

[1] ANDERSON, T. W. (1963). Asymptotic theory for principal component analysis. *Ann. Math. Statist.* **34** 122–148.

[2] ANDERSON T. W., and RUBIN, H. (1950). The saymptotic properties of estimates of the parameters of a single equation in a complete system of stochastic equations. *Ann. Math. Statist.* **21** 570–582.

[3] BILLINGSLEY, P. (1967). Weak convergence of probability measures. To appear.

[4] BILLINGSLEY, P. and TOPSOE, F. (1967). Uniformity in weak convergence. *Z. Wahrscheinlichkeitstheorie verw. Geb.* **7** 1–16.

[5] RUBIN, H. Topological properties of measures on topological spaces. Unpublished.

[6] TOPSOE, F. (1967). On the connection between P-continuity and P-uniformity in weak convergence. *Teor. Verojatnost. i Primenen.* **12** 281–290.