Sankhyā : The Indian Journal of Statistics
Special issue in memory of D. Basu
2002, Volume 64, Series A, Pt. 3, pp 884-893

SOME EXTENSIONS OF THE SKOROHOD REPRESENTATION THEOREM

By JAYARAM SETHURAMAN
Florida State University, USA

SUMMARY. Let $\{\mu_n\}_{n=0}^{\infty}$ be a sequence of probability measures on a measurable space $(\mathcal{X}, \mathcal{A})$. The famous Skorohod representation theorem Skorohod (1956) says that when μ_n converges to μ_0 weakly, we can find random variables $\{X_n, n = 0, 1, \ldots\}$ distributed marginally as $\{\mu_n, n = 0, 1, \ldots\}$, which converge almost surely to X_0. If the hypothesis is strengthened to $||\mu_n - \mu_0|| \to 0$, where $|| \cdot ||$ is the variation norm for measures, one can get the stronger conclusion that the above random variables X_n can be chosen so that $P(X_n = X_0) \to 1$. This result is a corollary of available results in the literature; see e.g. Dobrushin (1970). In this paper we show that if we assume a still stronger condition, namely that the pdf's f_n of μ_n w.r.t a measure ν, satisfy $\liminf_n f_n(x) \geq f_0(x)$ a.e $[\nu]$ where f_0 is the pdf of μ_0, then the random variables above can be chosen so that $P(X_n = X_0 \text{ ultimately}) = 1$. We also show that these results are tight by showing that the conditions are necessary and sufficient. We conclude the paper with illustrative examples and show that $\limsup_n f_n(x) = f_0(x)$ may not imply $\mu_n \to \mu_0$ in any sense.

1. Introduction

Let $\{\mu_n\}_{n=0}^{\infty}$ be a sequence of probability measures on a measurable space $(\mathcal{X}, \mathcal{A})$. When $\mathcal{X}$ is a separable complete metric space and $\mathcal{A}$ is the collection of Borel sets in $\mathcal{X}$, the famous Skorohod representation theorem (for a proof see Skorohod, 1956) states that

$$\mu_n \to \mu_0 \text{ weakly}, \tag{1.1}$$

if and only if there exist $\mathcal{X}$-valued random variables $\{X_n\}$ on some probability space $(\Omega, \mathcal{S}, P)$ such that

$$PX_n^{-1} = \mu_n, n = 0, 1, \ldots, \text{ and } P(X_n(\omega) \to X_0(\omega)) = 1. \tag{1.2}$$

Paper received January 2001; revised May 2002.
AMS (2000) *subject classification.* 60B10, 60F15.
Keywords and phrases. Skorohod representation theorem, Vasershtein metric, Scheffé's theorem.

An elegant alternate proof of this result can be found in Blackwell and Dubins (1983). Throughout this paper, unless otherwise stated, the convergence is as $n \to \infty$. The convergence of $X_n(\omega)$ to $X_0(\omega)$ in (1.2) is in the metric topology of $\mathcal{X}$. Dudley (1968) and Wichura (1970) have relaxed the separability condition on $\mathcal{X}$ and extended the Skorohod representation theorem.

In the rest of this paper, we will dispense with the assumption that $\mathcal{X}$ is a separable complete metric space. Instead, from now on, we will assume that $(\mathcal{X}, \mathcal{A})$ is an arbitrary measurable space possessing a countably generated sub σ-field of $\mathcal{A}$ containing all singletons. It is known that this is equivalent to saying that the diagonal subset of $\mathcal{X} \times \mathcal{X}$ is in the product σ-field $\mathcal{A} \times \mathcal{A}$.

We now ask the question how does the conclusion (1.2) get strengthened if we strengthen the weak convergence assumption (1.1) to

$$\sup_{A \in \mathcal{A}} |\mu_n(A) - \mu_0(A)| \to 0, \tag{1.3}$$

i.e. to convergence in variation norm? The answer to this question is already implicit in the work of many authors including that of Dobrushin (1970), but we will repeat it here for completeness in Theorem 2.1 of Section 2. Given a sequence of probability measures $\{\mu_n\}_{n=0}^{\infty}$, there is no loss of generality in assuming that they possess pdf's $\{f_n(x)\}_{n=0}^{\infty}$ with respect to some dominating measure ν. In Section 3 we will strengthen the weak convergence assumption (1.1) and the convergence in variation norm assumption (1.3) to the stronger condition on the pdf's

$$\liminf_n f_n(x) \geq f(x) \text{ a. e. } x \ [\nu]. \tag{1.4}$$

and ask how we can strengthen the conclusion (1.2).

This result is stated and proved as Theorem 3.1 in Section 3. Theorems 2.1 and 3.1 are tight in the sense that they are necessary and sufficient.

A small variation of the proof of Scheffé's theorem shows that (1.4) implies (1.3). However the condition

$$\limsup_n f_n(x) = f(x) \text{ a. e. } x \ [\nu] \tag{1.5}$$

does not imply (1.3). In Section 4 we give an example to illustrate what can happen if condition (1.5) holds but condition (1.4) does not hold; in such instances condition (1.3) may not hold and μ_n may not converge to μ_0 in any sense.

2. When μ_n Converges in Variation Norm

We can state the following theorem.

THEOREM 2.1 *The probability measures μ_n converge to μ_0 in variation norm, i.e. condition (1.3) holds if and only if there exist random variables $\{X_n\}_{n=0}^{\infty}$ such that*

$$PX_n^{-1} = \mu_n, n = 0, 1, \dots \text{ and } P(\{\omega : X_n(\omega) \neq X_0(\omega)\}) \to 0. \tag{2.1}$$

PROOF. The proof is immediate from the following lemma.

LEMMA 2.1 *Let μ_n, μ_0 be two probability measures on $(\mathcal{X}, \mathcal{A})$. Let*

$$||\mu_n - \mu_o|| \stackrel{def}{=} \sup_{A \in \mathcal{A}} |\mu_n(A) - \mu_0(A)|$$

be the variation norm of $\mu_n - \mu_0$. There exist random variables X_n, X_0 on some probability space $(\Omega, \mathcal{B}, P)$ such that

$$PX_n^{-1} = \mu_n, PX_0^{-1} = \mu_0 \tag{2.2}$$

and

$$P(X_n \neq X_0) = ||\mu_n - \mu_0||. \tag{2.3}$$

PROOF. The Vasershtein metric between μ_n and μ_0 is defined to be

$$\inf P(X_n \neq X_0)$$

where the infimum is taken over all random variables X_n, X_0 such that $PX_n^{-1} = \mu_n, PX_0^{-1} = \mu_0$, i. e. (2.2) holds. The content of this lemma is that Vasershtein metric is the same as the variation metric.

Let X_n, X_0 be random variables on some probability space such that (2.2) holds. Then for any subset $A \in \mathcal{A}$,

$$\begin{aligned} P(X_n \neq X_0) &\geq P(X_n \in A, X_0 \in A^c) = 1 - P(X_n \in A^c \textit{ or } X_0 \in A) \\ &\geq 1 - P(X_n \in A^c) - P(X_0 \in A) = \mu_n(A) - \mu_0(A) \end{aligned}$$

and thus

$$P(X_n \neq X_0) \geq \sup_{A \in \mathcal{A}} |\mu_n(A) - \mu_0(A)| = ||\mu_n - \mu_0||. \tag{2.4}$$

We will now construct a particular case of such a pair of random variables (X_n, X_0) such that the opposite inequality holds. This will complete the proof of the lemma.

Such constructions are well known, for instance see Dobrushin (1970). However, for the sake of completeness and for the proof of the Theorem 3.1 in the next section, we will give a construction that treats X_n and X_0 in an asymmetric fashion.

There is no loss of generality in assuming that μ_n, μ_0 have pdf's f_n, f_0 respectively with respect to some measure ν on $(\mathcal{X}, \mathcal{A})$. It is well known that

$$||\mu_n - \mu_0|| = \frac{1}{2}\int |f_n(x) - f_0(x)| d\nu(x) = \int (f_n(x) - f_0(x))^+ d\nu(x). \quad (2.5)$$

Let $(\Omega, \mathcal{B}, P)$ be a sufficiently rich probability space supporting three independent random variables, X_0, V_n, U such that U is uniform $[0,1]$ and the $\mathcal{X}$-valued random variables X_0, V_n have pdf's $f_0(x)$ and $\frac{(f_n(x)-f_0(x))^+}{||\mu_n-\mu_0||}$, respectively w.r.t ν. Define

$$X_n = \begin{cases} X_0 & \text{if } \begin{cases} f_0(X_0) \le f_n(X_0) \text{ or} \\ f_0(X_0) > f_n(X_0) \text{ and } U \le \frac{f_n(X_0)}{f_0(X_0)} \end{cases} \\ V_n & \text{if } f_0(X_0) > f_n(X_0) \text{ and } U > \frac{f_n(X_0)}{f_0(X_0)}. \end{cases} \quad (2.6)$$

It is easy to see check that

$$\begin{aligned} P(X_n \in A|X_0) &= I(X_0 \in A, f_0(X_0) \le f_n(X_0)) \\ &\quad + I(X_0 \in A, f(X_0) > f_n(X_0))\frac{f_n(X_0)}{f_0(X_0)} \\ &\quad + I(V_n \in A, f_0(X_0) > f_n(X_0))\frac{f_0(X_0) - f_n(X_0)}{f_0(X_0)} \end{aligned}$$

where $I(\cdot)$ stands for the usual indicator function. Hence

$$\begin{aligned} P(X_n \in A) &= \int_{A\cap(f_0\le f_n)} f_0 d\nu + \int_{A\cap(f_0>f_n)} f_n d\nu \\ &\quad + \frac{\int_A (f_n - f_0)^+ d\nu}{||\mu_n - \mu_0||}\int (f_0 - f_n)^+ d\nu \\ &= \int_A [\min(f_0, f_n) + (f_n - f_0)^+] d\nu = \int_A f_n d\nu = \mu_n(A) \end{aligned}$$

for all $A \in \mathcal{A}$.

Also,

$$\{X_n \neq X_0\} \subseteq \{f_0(X_0) > f_n(X_n)\} \cap \left\{U > \frac{f_n(X_0)}{f_0(X_0)}\right\}$$

and thus

$$P(X_n \neq X_0) \leq P\left(f_0(X_0) > f_n(X_n)\} \cap \left\{U > \frac{f_n(X_0)}{f_0(X_0)}\right\}\right) = ||\mu_n - \mu_0||.$$

Together with (2.4) this shows that

$$P(X_n \neq X_0) = ||\mu_n - \mu_0||.$$

This completes the proof of Lemma 2.1 and Theorem 2.1. □

Remark 1 There are many ways to construct random variables X_n, X_0 satisfying (2.2) and (2.3). The usual construction, see for instance Dobrushin (1970), treats X_n and X_0 in a symmetric fashion. The importance of this asymmetry will become apparent in the next section.

The paper of Sethuraman and Tiwari (1982) contains an interesting application of Theorem 2.1 establishing the weak convergence of a particular sequence of random probability measures.

3. When μ_n Converges in a Sense Stronger than in Variation Norm

Let μ_n be a sequence of probability measures as in the previous section. As before we will assume that μ_n has pdf $f_n(x)$ w.r.t some measure ν on $(\mathcal{X}, \mathcal{A})$, $n = 0, 1, \ldots$. The result, popularly known as Scheffé's theorem, states that

$$\lim_n f_n(x) = f_0(x) \text{ a.e. } [\nu] \tag{3.1}$$

(or just in ν-measure) is sufficient for μ_n to converge to μ_0 in variation norm. A small variation of the proof of Scheffé's theorem shows that

$$\liminf_n f_n(x) \geq f_0(x) \text{ a.e. } [\nu]$$

is also enough for μ_n to converge to μ_0 in variation norm, whereas

$$\limsup_n f_n(x) = f_0(x) \text{ a.e. } [\nu]$$

is also sufficient for the same conclusion only if $\sup_{n\geq N} f_n(x)$ is integrable for some N. This section explores how Theorem 2.1 can be extended under a condition of convergence on pdf's.

THEOREM 3.1 *The condition*

$$\liminf f_n(x) \geq f_0(x) \text{ a.e. } [\nu] \tag{3.2}$$

is necessary and sufficient for the existence of random variables $\{X_n\}_{n=0}^{\infty}$ *such that*

$$\begin{array}{l} PX_n^{-1} = \mu_n, n = 0, 1, \dots \text{ and} \\ P(\{\omega : X_n(\omega) = X_0(\omega) \text{ ultimately}\}) = 1. \end{array} \tag{3.3}$$

PROOF. We say that $X_n(\omega) = X_0(\omega)$ ultimately, if $X_n(\omega) = X_0(\omega)$ for $n \geq N(\omega)$ where $P(N(\omega) < \infty) = 1$. Suppose that condition (3.2) holds. Using the procedure outlined in Lemma 2.1, we can find a sufficiently rich probability space $(\Omega, \mathcal{B}, P)$ supporting independent random variables $X_0, V_n, n = 1, 2, \dots$ on $(\mathcal{X}, \mathcal{A})$ with pdf's $f_0(x), \frac{(f_n(x)-f_0(x))^+}{||\mu_n-\mu_0||}, n = 1, 2, \dots,$ respectively and a real valued random variable U which is uniform on $[0, 1]$ and independent of the previous random variables. As before define

$$X_n = \begin{cases} X_0 & \text{if} \begin{cases} f_0(X_0) \leq f_n(X_0) \text{ or} \\ f_0(X_0) > f_n(X_0) \text{ and } U \leq \frac{f_n(X_0)}{f_0(X_0)} \end{cases} \\ V_n & \text{if } f_0(X_0) > f_n(X_0) \text{ and } U > \frac{f_n(X_0)}{f_0(X_0)} \end{cases} \tag{3.4}$$

for $n = 1, 2, \dots$. It is clear that $PX_n^{-1} = \mu_n, n = 0, 1, \dots$. Let $h_N(x) = \inf_{n\geq N} f_n(x)$. Condition (3.2) implies that the sequence of functions $(f_0(x) - h_N(x))^+$ are bounded above by $f_0(x)$ and converge to 0 a.e. $[\nu]$. Thus $\int (f_0 - h_N)^+ d\nu \to 0$ as $N \to \infty$. We note that

$$\begin{aligned} & P(X_n \neq X_0 \text{ for some } n \geq N) \\ \leq \;& P\left(\left\{U > \frac{f_n(X_0)}{f_0(X_0)} \text{ and } f_0(X_0) > f_n(X_0)\right\} \text{ for some } n \geq N\right) \\ \leq \;& P\left(\left\{U > \frac{h_N(X_0)}{f_0(X_0)} \text{ and } f_0(X_0) > h_N(X_0)\right\}\right) \\ = \;& \int (f_0 - h_N)^+ d\nu \;\to\; 0 \end{aligned}$$

as $N \to \infty$. This establishes (3.3).

To prove the converse, we will now assume that there are random variables $\{X_n\}$ satisfying condition (3.3). The second part of this condition can also be expressed as

$$I(X_n = X_0) \to \ a.e.\ 1.$$

This implies that

$$g_n(X_0) \stackrel{def}{=} E(I(X_n = X_0)|X_0) \to 1 \ a.e.$$

which is the same as

$$g_n(x) \to 1 \ a.e.\ [\mu_0].$$

Now, for any $A \in \mathcal{A}$,

$$\begin{aligned}
\int_A f_n(x) d\nu(x) &= P(X_n \in A) \ \geq \ P(X_n \in A, X_n = X_0) \\
&= P(X_0 \in A, X_n = X_0) \\
&= E(I(X_0 \in A) E(X_n = X_0 | X_0)) \ = \ \int_A f_0(x) g_n(x) d\nu(x).
\end{aligned}$$

Hence

$$f_n \geq f_0 g_n \ a.e.\ [\nu].$$

Since $g_n(x) \to 1$ $a.e.$ $[\mu_0]$, we obtain

$$\liminf f_n \geq f_0 \ a.e.\ [\mu_0],$$

which also holds a.e. $[\nu]$, since $\liminf f_n(x) \geq f_0(x)$ trivially holds when $f_0(x) = 0$. This completes the proof of Theorem 3.1. □

Remark 2 Note that condition (3.2) is equivalent to

$$\liminf f_n = f_0 \text{ a.e. } [\nu].$$

This follows from the facts $\liminf f_n = \lim_N h_N \geq f_0$, $\int f_0 d\nu = 1$ which lead to $1 \leq \int h_N d\nu \leq \int f_N d\nu = 1$. It is well known that the stronger condition (3.2) implies (1.3). This can also be seen from the implications

$$(3.2) \Leftrightarrow (3.3) \Rightarrow (2.1) \Leftrightarrow (1.3).$$

Goldstein (1979) has considered coupling of random variables, which is similar in spirit to Theorem 3.1 but quite different at the same time. For two probability measures μ, ν on $(\mathcal{X}^\infty, \mathcal{A}^\infty)$ where $(\mathcal{X}, \mathcal{A})$ is a standard Borel space, Goldstein showed that there exists a probability measure P on $(\mathcal{X}^\infty \times \mathcal{X}^\infty, \mathcal{A}^\infty \times \mathcal{A}^\infty)$ such that the co-ordinate wise projection random variables (X, Y) satisfy $PX^{-1} = \mu, PY^{-1} = \nu$ and $P(X_n = Y_n \text{ ultimately}) = 1$ if and only if $\mu = \nu$ on the tail σ-field.

4. Some Examples

In the three examples given in this section, we consider sequences of probability measures $\mu_{n,r}$ with pdf's $f_{n,r}(x)$ with respect to the Lebesgue measure μ_0 on $[0,1]$, where the index (n,r) ranges over $n = 1,2,\dots,r = 1,\dots,2^n - 1$. The double indices (n,r) are ordered co-ordinate wise and limits are taken as $n \to \infty$. The pdf of μ_0 with respect to itself will be denoted by $f_0(x) \equiv 1$.

EXAMPLE 1 Let

$$f_{n,r}(x) = \begin{cases} a_n(2^n - 1) + 1 & \text{if } \frac{r-1}{2^n} \le x < \frac{r}{2^n} \\ 1 - a_n & \text{otherwise} \end{cases} \tag{4.1}$$

where $0 < a_n < 1, a_n \to 0$ and $2^n a_n \to \infty$. In this example

$$\liminf f_{n,r}(x) = f_0(x)$$

and

$$\limsup f_{n,r}(x) = \infty$$

for all x. From Theorem 3.1 there exist random variables $X_{n,r}$ with distribution $\mu_{n,r}$ such that

$$P(X_{n,r} = X_0 \text{ ultimately}) = 1.$$

Notice that (1.3) holds even though $f_{n,r}(x)$ does not converge to $f_0(x)$.

It is easy to construct examples where

$$\liminf_n f_n(x) < f_0(x) \text{ and } \limsup_n f_n(x) = f_0(x)$$

for all x. From Theorem 3.1 it is clear that there will not exist random variables satisfying (3.3). We illustrate this in Example 2 where (1.3), or equivalently (2.1), continues to hold and Example 3 where (1.3) does not hold.

EXAMPLE 2 Let

$$f_{n,r} = \begin{cases} 0 & \text{if } \frac{r-1}{2^n} \le x < \frac{(r)}{2^n} \\ 1 + a_n & \text{otherwise,} \end{cases}$$

where

$$a_n = \frac{1}{2^n - 1},$$

be the pdf's of $\mu_{n,r}$ with respect to μ_0. It is clear that

$$\liminf_{n,r} f_{n,r}(x) = 0 < 1 = f_0(x)$$

and

$$\limsup_{n,r} f_{n,r}(x) = 1 = f_0(x)$$

for all $x \neq 0$. Suppose that there are random variables X_0 and $X_{n,r}$ on some probability space such that X_0 has distribution μ_0 and $X_{n,r}$ has distribution $\mu_{n,r}$ for $n = 1, 2, \ldots, r = 1, \ldots, 2^n - 1$. Then

$$\{X_{n,r} \neq X_0\} \supset \left\{ \frac{r-1}{2^n} \leq X_0 \leq \frac{r}{2^n} \right\}.$$

It is a standard example in text books that the indicator functions $I(\frac{r-1}{2^n} \leq X_0 \leq \frac{r}{2^n})$ converge in measure to zero but not converge almost surely, as (n, r) goes to ∞. This means that

$$P(X_{n,r} = X_0 \text{ ultimately}) = 0$$

and thus (3.3) will not hold in this example. It is interesting to note that condition (1.3), or the equivalent condition (2.1), does hold in this example.

Example 3 We will now give an example where $\limsup f_{n,r}(x) = f_0(x)$ for all $x < 1$ but $||\mu_{n,r} - \mu_0|| \not\to 0$. Let $I_{n,r} = [\frac{r-1}{2^n} < x < \frac{r}{2^n}]$. Define

$$f_{n,r}(x) = \begin{cases} f_0(x) & \text{if } x \in I_{n,r} \\ 0 & \text{if } 0 \leq x \leq 1 - \frac{1}{2^n} \text{ and } x \notin I_{n,r} \\ 2^n - 1 & \text{if } x > 1 - \frac{1}{2^n}. \end{cases} \tag{4.2}$$

It is easy to check that $f_{n,r}(x)$ is a pdf of some probability measure $\mu_{n,r}$,

$$\liminf f_{n,r}(x) = 0 \text{ for all } x < 1,$$

and

$$\limsup f_{n,r}(x) = f_0(x) \text{ for all } x < 1.$$

However, we notice that

$$\sup_{(n,r) \geq (N,R)} f_{n,r}(x) = \begin{cases} 2^m - 1 & \text{if } 1 - \frac{1}{2^m} < x \leq 1 - \frac{1}{2^{m+1}} \text{ and } m \geq N \\ 1 & \text{if } 0 \leq x \leq 1 - \frac{1}{2^N} \end{cases}$$

This means that $\sup_{(n,r) \geq (N,R)} f_{n,r}(x)$ is not μ_0 integrable, $\mu_{n,r}(A) \to 0$ for $A \in [0, 1)$, $\mu_{n,r} \to \delta_1$ weakly, where δ_1 is the degenerate distribution at 1, and that $||\mu_{n,r} - \mu_0|| \to 1$. Notice that we can easily modify this example for arbitrary pdf $f_0(x)$, and still have $\limsup f_{n,r}(x) = f_0(x)$ and $||\mu_{n,r} - \mu_0|| \to 1$.

References

BLACKWELL, D. and DUBINS, L.E. (1983). An extension of Skorohod's almost sure representation theorem, *Proc. Amer. Math. Soc.*, **89**, 691-692.

DOBRUSHIN, R.L. (1970). Prescribing a system of random variables by conditional distributions, *Theory Probab. Appl.*, **15**, 458-486.

DUDLEY, R.M. (1968). Distance of probability measures and random variables, *Ann. Math. Statist.*, **39**, 1563-1572.

GOLDSTEIN, S.(1979). Maximal coupling, *Z. Wahrsch. Verw. Gebiete*, **46**, 193-204.

SETHURAMAN, J. and TIWARI, R.C. (1982). Convergence of Dirichlet processes and the interpretation of their parameter, *Statistical Decision Theory and Related Topics III Vol. 2.*, S.S. Gupta and J.O. Berger, eds., Academic Press, New York, 305-315.

SKOROHOD A.V.(1956). Limit theorems for stochastic processes, *Theory Probab. Appl.*, **1**, 261-290.

VASERSHTEIN L.N. (1969). Markov processes over denumerable products of spaces describing large system of automata, *Problemy Peradači Informacii*, **5**, 64-72; translated in *Problems of Information Transmission*, **5**, 47-52.

WICHURA M.J. (1970). On the construction of almost uniformly convergent random variables with given weakly convergent image laws, *Ann. Math. Statist.*, **41**, 284-291.

JAYARAM SETHURAMAN
DEPARTMENT OF STATISTICS
FLORIDA STATE UNVERSITY
TALLAHASSEE, FL 32309-4330, USA
E-mail: sethu@stat.fsu.edu