The Annals of Probability
1974, Vol. 2, No. 3, 480–485

CONDITIONAL DISTRIBUTIONS AND TIGHTNESS[1]

By Patrick Billingsley

The University of Chicago

Conditions for tightness and weak convergence of sequences of stochastic processes are given in terms of restrictions on the conditional probabilities of large increments and of large jumps.

1. Results. Let D be the space of functions on $[0, 1]$ with discontinuities of at most the first kind, with the Skorohod J_1 topology (see [2] for the theory of D and for the other weak-convergence concepts required here). For a random element X of D, let $J(X)$ be the maximum of the jumps $|X(t) - X(t-)|$, and let $J_{[s,t]}(X)$ be the maximum jump in the interval $[s, t]$. Let $M(X) = \sup_t |X(t)|$. In all that follows, $0 < \delta < 1$, $\varepsilon > 0$, $r > 0$, and A is a Borel set on the line.

Let $\alpha(A, \varepsilon, \delta)$ be a number such that

$$0 \leqq t_1 \leqq \cdots \leqq t_m \leqq s \leqq 1 , \qquad s - t_m \leqq \delta \tag{1}$$

implies that

$$P[|X(s) - X(t_m)| > \varepsilon \,||\, X(t_1), \cdots, X(t_m)] \leqq \alpha(A, \varepsilon, \delta) \tag{2}$$

holds with probability 1 on the set

$$[X(t_1) \in A, \cdots, X(t_m) \in A] . \tag{3}$$

Let $\gamma(A, r, \delta)$ be a number such that (1) implies that

$$P[J_{[t_m,s]}(X) \geqq r \,||\, X(t_1), \cdots, X(t_m)] \leqq \gamma(A, r, \delta) \tag{4}$$

holds with probability 1 on the set (3).

If the functions α and γ are small in an appropriate sense, then the random function X has small probability of oscillating violently; it is therefore possible to formulate tightness conditions in terms of these functions. Suppose that, for each X_n in a sequence of random elements of D, α_n and γ_n are related to X_n by the analogues of (2) and (4).

THEOREM 1. *The sequence $\{X_n\}$ is tight in D if the sequence $\{M(X_n)\}$ is tight on the line, if*

$$\lim_{\delta\to 0} \limsup_{n\to\infty} \alpha_n(A, \varepsilon, \delta) = 0 \tag{5}$$

for each positive ε and each bounded set A, and if

$$\lim_{r\to\infty} \limsup_{\delta\to 0} \limsup_{n\to\infty} \gamma_n(A, r, \delta) = 0 \tag{6}$$

for each bounded set A.

Received July 23, 1973; revised September 19, 1973.

[1] Supported by a Guggenheim Fellowship, an SRC grant, and NSF GP 32037X.

AMS 1970 *subject classifications.* Primary 60B10; Secondary 60F05.

Key words and phrases. Weak convergence, tightness.

Theorem 1 has been applied by Barbour to prove convergence of epidemic processes. It can also be used to give a simple derivation of the results of Section 4 of [4]. Note that (5) is not a necessary condition for tightness even if X_n is continuous and does not depend on n: If $X = x_k$ with probability 2^{-k}, where $x_k(t) = k^{-1} + kt$, then $P[X(\delta) - X(0) > \varepsilon \,||\, X(0) = k^{-1}] = 1$ for $k > \varepsilon/\delta$.

THEOREM 2. *If* (5) *and* (6) *hold, if the sequence* $\{J(X_n)\}$ *is tight on the line, and if the finite-dimensional distributions of* X_n *converge weakly to those of* X, *then* $X_n \to_{\mathscr{D}} X$.

These results become weaker but simpler if given in terms of $\alpha_n(\varepsilon, \delta) = \alpha_n(R^1, \varepsilon, \delta)$.

THEOREM 3. *The sequence* $\{X_n\}$ *is tight in* D *if*

$$\lim_{\delta\to 0} \limsup_{n\to\infty} \alpha_n(\varepsilon, \delta) = 0 \tag{7}$$

for each positive ε *and if for each* t *the sequence* $\{X_n(t)\}$ *is tight on the line.*

Theorem 3 is due to Grigelionis [3]. A variant may prove convenient:

THEOREM 4. *The sequence* $\{X_n\}$ *is tight in* D *if* (7) *holds and if the sequence* $\{X_n(0)\}$ *and* $\{J(X_n)\}$ *are tight on the line.*

THEOREM 5. *If* (7) *holds, and if the finite-dimensional distributions of the* X_n *converge weakly to those of* X, *then* $X_n \to_{\mathscr{D}} X$.

Theorem 5, together with Chebyshev's inequality, gives a simple proof of Donsker's theorem.

2. Three lemmas. It will be convenient to take the α in (2) to be minimal: $\alpha(A, \varepsilon, \delta)$ is the supremum over (1) of the essential supremum of the left side of (2) over the set (3). In the same way, let $\gamma(A, r, \delta)$ be the smallest number such that (1) implies that (4) holds with probability 1 on the set (3). Let $\beta(A, \varepsilon, \delta)$ be the smallest number for which (1) implies that

$$P[\sup_{t_m \leq u \leq s} |X(u) - X(t_m)| > \varepsilon \,||\, X(t_1), \cdots, X(t_m)] \leq \beta(A, \varepsilon, \delta) \tag{8}$$

holds with probability 1 on (3). Put $\beta(\varepsilon, \delta) = \beta(R^1, \varepsilon, \delta)$. Denote by A^ε the ε-neighborhood of A—the set of points x such that there exists in A a y with $|x - y| < \varepsilon$.

LEMMA 1. *We have*

$$\beta(A, \varepsilon, \delta) \leq 2\alpha(A^{\varepsilon+r}, \tfrac{1}{2}\varepsilon, \delta) + \gamma(A, r, \delta) \tag{9}$$

and

$$\beta(\varepsilon, \delta) \leq 2\alpha(\tfrac{1}{2}\varepsilon, \delta)\,. \tag{10}$$

PROOF. We shall use the fact that, if M_0 lies in a σ-field $\mathscr{F}$, then requiring $P[H \,||\, \mathscr{F}] \leq \alpha$ with probability 1 on M_0 is the same thing as requiring $P(M \cap H) \leq \alpha P(M)$ for all M with $M \in \mathscr{F}$ and $M \subset M_0$.

In proving (9) we may assume that $s = t_m + \delta$; write t in place of t_m. Suppose M is a subset of (3) and lies in the σ-field generated by $X(t_1), \cdots, X(t_m)$. Fix a positive integer k for the moment and for $1 \leqq j \leqq 2^k$ define

$$M_j = \left[\max_{0 \leqq i < j} \left|X\left(t + \frac{i}{2^k}\delta\right) - X(t)\right| < \varepsilon \leqq \left|X\left(t + \frac{j}{2^k}\delta\right) - X(t)\right|\right].$$

Notice that, on the set $M \cap M_j$, we have $X(t + i2^{-k}\delta) \in A^\varepsilon$ for $0 \leqq i \leqq j - 1$. Now

$$\begin{aligned} P\left(M \cap \left[\max_{i \leqq 2^k} \left|X\left(t + \frac{i}{2^k}\delta\right) - X(t)\right| > \varepsilon\right]\right) & \\ \leqq P(M \cap [|X(t + \delta) - X(t)| > \tfrac{1}{2}\varepsilon]) & \\ + \textstyle\sum_{j<2^k} P\Big(M \cap M_j \cap \left[X\left(t + \frac{j}{2^k}\delta\right) \in A^{\varepsilon + r}\right] & \\ \cap \left[\left|X(t + \delta) - X\left(t + \frac{j}{2^k}\delta\right)\right| > \tfrac{1}{2}\varepsilon\right]\Big) & \\ + \textstyle\sum_{j<2^k} P\left(M \cap M_j \cap \left[\left|X\left(t + \frac{j}{2^k}\delta\right) - X\left(t + \frac{j-1}{2^k}\delta\right)\right| \geqq r\right]\right). & \end{aligned}$$

By the defining property of α, this is at most

$$2\alpha(A^{\varepsilon + r}, \tfrac{1}{2}\varepsilon, \delta)P(M) + P(M \cap J_k),$$

where J_k is the set where $|X(t + j2^{-k}\delta) - X(t + (j-1)2^{-k}\delta)| \geqq r$ for some $j \leqq 2^k$. Letting k tend to infinity and using the defining property of γ gives

$$P(M \cap [\sup_{t \leqq u \leqq t+\delta} |X(u) - X(t)| > \varepsilon]) \leqq 2\alpha(A^{\varepsilon + r}, \tfrac{1}{2}\varepsilon, \delta)P(M) + \gamma(A, r, \delta)P(M),$$

from which (9) follows. Replacing A and $A^{\varepsilon+r}$ by R^1 in this argument and omitting the J_k gives a proof of (10) which is formally (9) with $A = R^1$ and $r = \infty$.

For the definition of w' see ([2] page 110).

LEMMA 2. *For each pair δ and δ_0* $(0 < \delta, \delta_0 < 1)$,

$$\begin{aligned} (11) \qquad & P[w_X'(\tfrac{1}{3}\delta) \geqq 2\varepsilon] \\ & \leqq 2P(\textstyle\bigcup_{t \leqq 1} [X(t) \notin A]) + \beta(A, \varepsilon, \delta) + \dfrac{2\beta(A, \varepsilon, \delta)}{\delta_0(1 - \beta(A, \varepsilon, \delta_0))} \end{aligned}$$

and

$$(12) \qquad P[w_X'(\tfrac{1}{3}\delta) \geqq 2\varepsilon] \leqq \beta(\varepsilon, \delta) + \frac{2\beta(\varepsilon, \delta)}{\delta_0(1 - \beta(\varepsilon, \delta_0))}.$$

PROOF. We prove (11), (12) being the case $A = R^1$. Fix ε, δ, and δ_0, and for a positive integer k, temporarily fixed, define random times $T_0, T_1, \cdots$ as follows. Take $T_0 = 0$; T_{l-1} having been defined, take T_l to be the smallest u of the form $i2^{-k}$ satisfying $T_{l-1} < u \leqq 1$ and $|X(u) - X(T_{l-1})| > \varepsilon$, with $T_l = 2$ if there is no such u. Note that $T_{l-1} \geqq 1$ implies $T_l = 2$. Put $D_l = T_l - T_{l-1}$; note that $\sum_l D_l \leqq 2$.

If $T_{l-1} = j2^{-k} < 1$, then (since $\delta < 1$) $D_l < \delta$ implies that $|X(u) - X(j2^{-k})| > \varepsilon$

for some u with $j2^{-k} < u < j2^{-k} + \delta$ and $u \leqq 1$. Therefore

$$P[D_l < \delta \,||\, X(i2^{-k}),\, i \leqq j] \leqq \beta(A, \varepsilon, \delta) \tag{13}$$

holds with probability 1 on the set

$$M_{lj} = [T_{l-1} = j2^{-k},\, X(i2^{-k}) \in A,\, i \leqq j]\,.$$

The same relation holds for δ_0, so that

$$E[D_l \,||\, X(i2^{-k}),\, i \leqq j] \geqq \delta_0(1 - \beta(A, \varepsilon, \delta_0)) \tag{14}$$

holds with probability 1 on M_{lj}.

We want to show that the event

$$T_l - T_{l-1} \geqq \delta \quad \text{if} \quad T_{l-1} < 1\,, \qquad i = 1, 2, \cdots \tag{15}$$

has high probability. Write $\eta = P(\bigcup_{t \leqq 1} [X(t) \notin A])$, $\beta = \beta(A, \varepsilon, \delta)$, and $\beta_0 = \beta(A, \varepsilon, \delta_0)$. Since (13) and (14) hold with probability 1 on M_{lj}, (15) holds except on a set of probability at most

$$\begin{aligned} &\eta + \textstyle\sum_{l \geqq 1} P[T_{l-1} < 1;\, D_l < \delta;\, X(t) \in A,\, t \leqq 1] \\ &\quad \leqq \eta + \textstyle\sum_{l \geqq 1} \sum_{j < 2^k} P(M_{lj} \cap [D_l < \delta]) \leqq \eta + \beta \sum_{l \geqq 1} \sum_{j < 2^k} P(M_{lj}) \\ &\quad \leqq \eta + \frac{\beta}{\delta_0(1 - \beta_0)} \textstyle\sum_{l \geqq 1} \sum_{j < 2^k} \int_{M_{lj}} D_l \, dP \\ &\quad \leqq \eta + \frac{\beta}{\delta_0(1 - \beta_0)} \textstyle\sum_{l \geqq 1} E[D_l] \leqq \eta + \frac{2\beta}{\delta_0(1 - \beta_0)}\,. \end{aligned}$$

If (15) holds, then there exist points s_i^k, $i = 0, 1, \cdots, N_k$ such that

$$0 = s_0^k < \cdots < s_{N_k}^k = 1\,, \qquad s_i^k - s_{i-1}^k \geqq \delta\,, \qquad i = 1, \cdots, N_k - 1$$

(the last inequality may fail for $i = N_k$) and such that $|X(t) - X(s)| \leqq 2\varepsilon$ if s and t have the form $i2^{-k}$ and lie in a common interval (s_{i-1}^k, s_i^k), $i = 1, \cdots, N_k$; such s_i^k exist except on a set of probability $\eta + 2\beta/\delta_0(1 - \beta_0)$. If such s_i^k exist for infinitely many k, then, since $N_k \leqq 1 + \delta^{-1}$, there is a subsequence of integers k along which the N_k have a common value N and along which s_i^k converges for each $i = 0, 1, \cdots, N$ to some s_i, in which case

$$0 = s_0 < \cdots < s_{N-1} \leqq s_N = 1\,, \qquad s_i - s_{i-1} \geqq \delta\,, \qquad i = 1, \cdots, N - 1\,. \tag{16}$$

Since $|X(t) - X(s)| \leqq 2\varepsilon$ if s and t are dyadic rationals in a common interval (s_{i-1}, s_i), it follows by right-continuity that

$$w_X[s_{i-1}, s_i) \leqq 2\varepsilon\,, \qquad i = 1, \cdots, N\,. \tag{17}$$

And outside a set of probability $\eta + 2\beta/\delta_0(1 - \beta_0)$ there exist s_i satisfying (16) and (17).

By the definitions of η and β,

$$\sup_{1-\delta \leqq u \leqq 1} |X(u) - X(1 - \delta)| \leqq \varepsilon \tag{18}$$

outside a set of probability $\eta + \beta$. Suppose (16), (17), and (18) all hold. If s_{N-1} exceeds $1 - \delta/2$, replace it by $1 - \delta/2$; the new s_i satisfy (17), and $s_i - s_{i-1} \geqq \delta/2$

for $i = 1, \cdots, N$. Thus $w'(\delta/3) \leqq 2\varepsilon$ outside a set of probability $2\eta + \beta + 2\beta/\delta_0(1 - \beta_0)$, which proves (11).

These two lemmas suffice for the proof of Theorem 1. The remaining theorems require a third lemma.

LEMMA 3. *If*

$$0 = s_0 < \cdots < s_N = 1, \quad s_i - s_{i-1} \leqq \delta, \quad i = 1, \cdots, N, \tag{19}$$

then

$$P(\bigcup_{t\leqq 1} [X(t) \notin A^\varepsilon]) \leqq P(\bigcup_{i\leqq N} [X(s_i) \notin A]) + \alpha(A^{\varepsilon+r}, \varepsilon, \delta) + P[J(X) \geqq r] \tag{20}$$

and

$$P(\bigcup_{t\leqq 1} [X(t) \notin A^\varepsilon]) \leqq P(\bigcup_{i\leqq N} [X(s_i) \notin A]) + \alpha(\varepsilon, \delta). \tag{21}$$

PROOF. Let $0 = u_1 < \cdots < u_k = 1$ be points that include the s_i among them, and for $j = 1, \cdots, k$ define

$$M_j = [X(u_l) \in A^\varepsilon, 0 \leqq l < j; X(u_j) \notin A^\varepsilon].$$

Let $\eta = P(\bigcup_{i\leqq N} [X(s_i) \notin A])$. If the inner sums below denote summation over those j for which $s_{i-1} < u_j < s_i$, then

$$\begin{aligned} P(\textstyle\bigcup_{j\leqq k} [X(u_j) \notin A^\varepsilon]) \\ &\leqq \eta + \textstyle\sum_{i=1}^N \sum P(M_j \cap [X(u_j) \in A^{\varepsilon+r}] \cap [|X(u_j) - X(s_i)| > \varepsilon]) \\ &\quad + \textstyle\sum_{i=1}^N \sum P(M_j \cap [|X(u_j) - X(u_{j-1})| \geqq r]). \end{aligned}$$

By the defining property of α, this is at most $\eta + \alpha(A^{\varepsilon+r}, \varepsilon, \delta) + P(M')$, where M' is the set where $|X(u_j) - X(u_{j-1})| \geqq r$ for some $j \leqq k$. Letting the u_j become dense in [0, 1] yields (20). Replacing $A^{\varepsilon+r}$ by R^1 in this argument and omitting M' gives a proof of (21), formally the case $r = \infty$.

3. Proof of Theorem 1. By Lemma 1, the hypotheses (5) and (6) together imply

$$\lim_{\delta\to 0} \limsup_{n\to\infty} \beta_n(A, \varepsilon, \delta) = 0 \tag{22}$$

for bounded A. Suppose positive ε and η are given. Since $\{M(X_n)\}$ is tight by hypothesis, there is a bounded set A such that $P(\bigcup_{t\leqq 1} [X_n(t) \notin A]) < \eta$ for all n. And by (22) there exists a δ_0 such that $\beta_n(A, \varepsilon, \delta_0) < \frac{1}{2}$ for large n. That δ_0 fixed, there exists a δ such that $\beta_n(A, \varepsilon, \delta) < \eta\delta_0 < \eta$ for large n. By Lemma 2, $P[w_X'(\delta/3) \geqq 2\varepsilon] < 7\eta$ for large n, which proves tightness (see Theorem 15.2 of [2]).

4. Proof of Theorem 2. Weak convergence will follow if we prove $\{X_n\}$ tight, and by Theorem 1 this will follow if we prove $\{M(X_n)\}$ tight. Suppose η is given. Since $\{J(X_n)\}$ is tight by hypothesis, there is an r such that $P[J(X_n) \geqq r] < \eta$ for all n. Choose a so that $P[\sup_t |X(t)| \geqq a] < \eta$, put $A = (-a, a)$, and then choose δ so that $\alpha_n(A^{1+r}, 1, \delta) < \eta$ for large n. Fix points s_i satisfying (19). Since the finite-dimensional distributions converge,

$$\limsup_{n\to\infty} P[\max_{i\leqq N} |X_n(s_i)| \geqq a] \leqq P[\max_{i\leqq N} |X(s_i)| \geqq a] < \eta,$$

so that $P(\bigcup_{i \leq N} [X_n(s_i) \notin A]) < \eta$ for large n. It follows by Lemma 3 that $P[\sup_t |X_n(t)| \geqq a + 1] < 3\eta$ for large n, which proves that $\{M(X_n)\}$ is tight.

5. Proof of Theorem 3. From (7) and Lemma 1 it follows that

$$\lim_{\delta \to 0} \limsup_{n \to \infty} \beta_n(\varepsilon, \delta) = 0 , \tag{23}$$

which, together with Lemma 2 (use (12)), implies

$$\lim_{\delta \to 0} \limsup_{n \to \infty} P[w'_{X_n}(\delta) \geqq \varepsilon] = 0 . \tag{24}$$

That $\{X_n\}$ is tight will now follow if we show that $\{M(X_n)\}$ is tight. Given a positive η, choose δ so that $\alpha_n(1, \delta) < \eta$ for large n. That δ fixed, choose s_i satisfying (19) and then, using the tightness of each sequence $\{X_n(t)\}$, choose a so that $P[\max_{i \leq N} |X_n(s_i)| \geqq a] < \eta$ for all n. From Lemma 3 it follows that $P[\sup_t |X_n(t)| \geqq a + 1] < 2\eta$ for large n, which proves $\{M(X_n)\}$ tight.

6. Proof of Theorem 4. By Theorem 3, it suffices to prove that the sequence $\{X_n(t)\}$ is tight for each t. As before, (23) holds and hence so does (24).

We first show that

$$\sup_t |x(t)| \leqq |x(0)| + \frac{1}{\delta} w_x'(\delta) + \frac{1}{\delta} J(x) \tag{25}$$

for x in the space D. Indeed, given ε choose points s_i such that $0 = s_0 < \cdots < s_k = 1$ and such that $s_i - s_{i-1} > \delta$ and $w_x[s_{i-1}, s_i) < w_x'(\delta) + \varepsilon$ for $i = 1, \cdots, k$. Since $k \leqq 1/\delta$ and $|x(t)| \leqq |x(0)| + k(w_x'(\delta) + \varepsilon) + kJ(x)$, letting ε tend to 0 gives (25).

From (25) it follows that

$$\begin{aligned} &P[\sup_t |X_n(t)| \geqq 3a] \\ &\quad \leqq P[|X_n(0)| \geqq a] + P[w'_{X_n}(\delta) \geqq \delta a] + P[J(X_n) \geqq \delta a] . \end{aligned} \tag{26}$$

For $\eta > 0$, there exists by (24) a δ such that $P[w'_{X_n}(\delta) \geqq 1] < \eta$ for large n. Choose v so large that $P[|X_n(0)| \geqq v] < \eta$ and $P[J(X_n) \geqq v] < \eta$ for all n. If a exceeds v, $1/\delta$, and v/δ, the left side of (26) is at most 3η. Thus $\{M(X_n)\}$ is tight.

7. Proof of Theorem 5. If the sequence $\{X_n(t)\}$ is weakly convergent, it is tight. Hence Theorem 5 is an immediate consequence of Theorem 3.

REFERENCES

[1] Barbour, Andrew (1973). Paper to appear.

[2] Billingsley, Patrick (1968). *Convergence of Probability Measures.* Wiley, New York.

[3] Grigelionis, B. (1973). On relative compactness of sets of probability measures in $D_{[0,\infty)}(\mathscr{X})$. *Litovsk. Mat. Sb.* to appear.

[4] Liggett, Thomas M. (1970). On convergent diffusions: the densities and the conditioned processes. *Indiana Univ. Math. J.* **20** 265–279.

Department of Statistics
The University of Chicago
1118 East 58th Street
Chicago, Illinois 60637