Stochastic Processes and their Applications 2 (1974) 211–241. © North-Holland Publishing Company

WEAK CONVERGENCE IN APPLIED PROBABILITY *

Donald L. IGLEHART

Department of Operations Research, Stanford University, Stanford, Calif., 94305, U.S.A.

Received 19 December 1973

Abstract. Weak convergence of probability measures on function spaces has been an active area of research in recent years. While the theory has a somewhat abstract base, it is extremely useful in a wide variety of problems and we believe has much to offer to applied probability. Our aim in this survey paper is to discuss those aspects of the theory which are relevant to work in applied probability. After an introduction to the foundations of weak convergence, we shall discuss partial sum, point, Markov and extremal processes. These processes form the building blocks for many of the important models of applied probability.

applied probability	Markov processes
birth and death processes	partial sum processes
extremal processes	point processes
functional central limit theorems	renewal processes
invariance principle	weak convergence

1. Introduction and summary

The idea of weak convergence of probability measures on function spaces has been present in the literature for at least twenty years. However, in the last decade the number of papers in this area has increased dramatically. A major stimulus for this work has been the excellent book on the subject by Billingsley [5]. For an up-dated version of parts of [5] see [6]. Our aim in this paper is to discuss those aspects of weak convergence which are relevant for applied probability. The paper is organized primarily by the type of stochastic process involved. We shall discuss partial sum, point, Markov and extremal processes. These processes form

* This work was partially supported by National Science Foundation grant GP-31392X and Office of Naval Research contract N00014-67-A-0112-0031. Presented by invitation at the Second Conference on Stochastic Processes and Applications, Katholieke Universiteit Leuven, August 21–25, 1972.

the building blocks for many of the important models of applied probability. As with any paper of this sort, the selection of topics is greatly influenced by the author's interests. Some effort has been made to include those topics of greatest interest and use to people working in applied probability.

Classical probability theory deals in large part with limit theorems for partial sums of independent, identically distributed (i.i.d.) random variables (r.v.'s). In particular, the central limit theorem (c.l.t.) in the case where the summands have mean zero and finite variance states that the distribution functions (d.f.'s) or probability measures (p.m.'s) of the partial sums, properly normalized, converge weakly to the d.f. or p.m. of the standard normal r.v. The p.m.'s in this case are defined on the Borel sets of the real line. Once this result is established, one would like to be able to use the d.f. of the limit r.v. as an approximation for the d.f. of the partial sums for "large" values of the parameter n, say.

Weak convergence theory for p.m.'s on the Borel sets of a function space is directed at obtaining comparable results for sequences of stochastic processes or random functions (r.f.'s) whose sample paths lie in the function space. In other words, we would like to show that a sequence of stochastic processes converges to a limit process in an appropriate sense. Again the limit process, which is generally easy to compute with, is suggested as an approximation for the processes in the sequence, which are usually untractable for computations. One of the chief advantages of dealing with the entire process, rather than its value at a fixed time point, is that weak convergence results can often be obtained immediately for functionals of the process which are of greater interest than the original process. The standard method for showing weak convergence of an appropriately normed sequence of processes is to first show convergence of the finite-dimensional distributions (f.d.d.'s) and then to demonstrate a certain compactness property that prevents probability from "escaping to infinity". In showing convergence of the f.d.d.'s one must resort to classical methods such as those of characteristic functions, differential equations, or moments. To show the compactness property one must convert conditions for compactness of sets in the function space to easily verified probabilistic conditions. It is the latter task which creates the greatest technical difficulties for the theory. The notation and terminology used in this paper follows, for the most part, that of [5].

The organization of the paper is as follows. Section 2 contains a discussion of the foundations of weak convergence theory and is taken primarily from [5]. In Section 3 we treat convergence of processes formed from partial sums of r.v.'s or r.f.'s. Sections 4, 5 and 6 are devoted to weak convergence of r.f.'s formed from point processes, Markov processes and extremal processes, respectively. In Section 7 we discuss functionals of Brownian motion and the Poisson processes. Finally, in Section 8 a few concluding remarks are made dealing with future work in this area.

2. Foundations

Our discussion in this section follows [5] very closely and is meant to summarize those basic results which have proved most useful in applied probability. Weak convergence is concerned with p.m.'s on the Borel sets $\mathcal{S}$ of a metric space (S, m). A sequence of p.m.'s $\{P_n : n \geq 1\}$ is said to *converge weakly* to a p.m. P if

$$\lim_{n\to\infty} \int_S f \, \mathrm{d}P_n = \int_S f \, \mathrm{d}P \tag{2.1}$$

for all bounded, continuous, real-valued functions f on S. We write $P_n \Rightarrow P$ for this type of convergence. Equivalent conditions for weak convergence may be found in [5, Theorem 2.1]. Observe that weak convergence only involves S and its Borel sets $\mathcal{S}$, so that two different metrics inducing equivalent topologies lead to the same weak convergence.

A *random element* (r.e.) X is a measurable mapping from some probability space $(\Omega, \mathcal{F}, \mathbf{P})$ into S. The distribution of X is the p.m. $P = \mathbf{P}X^{-1}$ on $(S, \mathcal{S})$. We shall say that a sequence of random elements $\{X_n\}$ *converges weakly* to a r.e. X, and write $X_n \Rightarrow X$, if the distribution P_n of X_n converges weakly to the distribution P of X. We note that the r.e.'s X_n and X need not be defined on the same probability space $(\Omega, \mathcal{F}, \mathbf{P})$, but must only induce p.m.'s P_n and P on the same metric space (S, m).

Now assume S is separable. A sequence of r.e.'s $\{X_n\}$ *converges in probability* to X if X_n and X are defined on a common probability space and for all $\epsilon > 0$,

$$\mathbf{P}[m(X_n, X) \geq \epsilon] \to 0.$$

When X is a constant element (nonrandom), convergence in probability is equivalent to weak convergence. In such cases we shall write $m(X_n, X) \Rightarrow 0$ or $X_n \Rightarrow X$. If X_n and Y_n have a common domain, we also write $m(X_n, Y_n) \Rightarrow 0$ when for all $\epsilon > 0$,

$$\mathbf{P}[m(X_n, Y_n) > \epsilon] \to 0.$$

The notion of tightness, introduced by Prohorov [45], plays a key role in the weak convergence of p.m.'s. A family of p.m.'s Π on the space $(S, \mathcal{S})$ is said to be *tight* if for every $\epsilon > 0$ there exists a compact set K_ϵ such that $P(K_\epsilon) > 1 - \epsilon$ for all $P \in \Pi$. A family Π is said to be *relatively compact* if every sequence of elements of Π contains a convergent subsequence (the limit need not belong to Π). We shall say that a family of r.e.'s $\{X_n\}$ is tight if the corresponding distributions are tight. Intuitively, if a family of r.e.'s is tight, this prevents probability from "escaping to infinity" as one runs through the family. The main result which makes these concepts useful is the following theorem of Prohorov [45]; see also [5, §6].

Theorem 2.1. (a) *If Π is tight, then it is relatively compact.*
(b) *If S is a complete separable metric space and Π is relatively compact, then it is tight.*

Often one has occasion to deal with metric spaces S which are product space. Suppose S_i, $i = 1, ..., k$, are separable metric spaces with corresponding Borel sets $\mathcal{S}_i$. Let $S = S_1 \times ... \times S_k$ be endowed with the product topology and Borel sets $\mathcal{S}$. Since S is separable, $\mathcal{S} = \mathcal{S}_1 \times ... \times \mathcal{S}_k$, the product Borel field. For a given p.m. P on $\mathcal{S}$ we define the *marginal measures* P^i, $i = 1, ..., k$, by

$$P^i(A) = P(S_1 \times ... \times S_{i-1} \times A \times S_{i+1} \times S_k) \quad \text{for } A \in \mathcal{S}_i.$$

For a family Π of p.m.'s on S the notion of tightness can be expressed in terms of the tighiness of the families Π^i, $i = 1, ..., k$), of marginal measures. This result, which is elementary but very useful, was stated for $k = 2$ as [5, Problem 6, p. 41]. For a proof see [24].

Lemma 2.2. *Let Π be a family of probability measures on $(S, \mathcal{S})$ and let Π^i, $i = 1, ..., k$, be the corresponding families of marginal measures on $(S_i, \mathcal{S}_i)$. Then Π is tight if and only if each Π^i is tight.*

Next we state three extremely useful theorems for obtaining weak convergence results in applications. The first has come to be known as the "converging together theorem". For it we assume that X_n and Y_n are defined on a common domain and take values in a separable metric space (S, m). This result can be found in [5, Theorem 4.1].

Theorem 2.3. *If $X_n \Rightarrow X$ and $m(X_n, Y_n) \Rightarrow 0$, then $Y_n \Rightarrow X$.*

Now suppose h is a measurable mapping of S into S', a second metric space with Borel sets $\mathcal{S}'$. Each p.m. P on $(S, \mathcal{S})$ induces on $(S', \mathcal{S}')$ a unique p.m. $Ph^{-1}(A) = P(h^{-1} A)$ for $A \in \mathcal{S}'$. Let D_h be the set of discontinuities of h. The next result, known as the continuous-mapping theorem, is an analogue of the Mann–Wald theorem for Euclidean spaces; cf. [5, Theorem 5.1].

Theorem 2.4. *If $X_n \Rightarrow X$ and $\mathbf{P}[X \in D_h] = 0$, then $h \circ X_n \Rightarrow h \circ X$.*

In practice we use this result as follows. First we show $X_n \Rightarrow X$, often by just quoting known results. Then we find an appropriate mapping h which gives us the r.e.'s we are really interested in, $h \circ X_n$, and finally apply Theorem 2.4.

There is an alternative method for obtaining weak convergence of functionals. It is based on a result due to Skorohod [52] in the case of complete separable metric spaces. Dudley [14] removed the requirement for completeness and Wichura [67] further removed the separability requirement.

Theorem 2.5. *Let (S, m) be a metric space on which $X_n \Rightarrow X_0$. Then there exists a probability space $(\Omega^*, \mathcal{F}^*, \mathbf{P}^*)$ and random elements $X_n^* : \Omega^* \to S$ such that $P_{X_n} = P_{X_n^*}$ and $m(X_n^*, X_0^*) \to 0$ a.e.*

While we will not use this result, the approach is sometimes very useful and one that should be available to people applying weak-convergence methodology. For a very readable account of this method and its applications see [46].

Two function spaces have received the greatest attention in the weak convergence literature: the space of continuous functions on [0,1] and the space of functions on [0,1] having only jump discontinuities. These spaces are the natural ones for the sample paths of most processes which arise in applied probability. Sometimes, however, it is more natural to have these functions defined on $[0, r]$, $r > 0$, or even on $[0, \infty)$, particularly when dealing with first-passage problems.

Let $C[0,1]$ ($\equiv C$) denote the space of all continuous, real-valued functions on [0,1] with the metric of uniform convergence,

$$\rho(x, y) = \sup\{|x(t) - y(t)| : 0 \leq t \leq 1\},$$

and Borel sets $\mathcal{C}$. Thus if we think of the elements $x, y \in C$ as representing the trajectories of two particles moving on the real line $\mathbf{R}$, then the trajectories are within a distance ϵ of each other if at each time $t \in [0,1]$ the displacements are within ϵ. With this metric, C is a complete separable metric space (c.s.m.s.). For points $t_1, ..., t_k \in [0,1]$, let $\pi_{t_1 ... t_k}$ be the mapping that carries $x \in C$ into $(x(t_1, ..., x(t_k)) \in \mathbf{R}^k$. The mappings $\pi_{t_1 ... t_k}$ are clearly continuous. If $\{P_n : n \geq 1\}$ and P are p.m.'s on $(C, \mathcal{C})$ corresponding to random functions $\{X_n : n \geq 1\}$ and X, we say that the finite-dimensional distributions (f.d.d.'s) of X_n converge weakly to those of X provided $P_n \, \pi^{-1}_{t_1 ... t_k} \Rightarrow P\pi^{-1}_{t_1 ..., t_k}$ for all choices of $t_1, ..., t_k \in [0,1]$ and all $k \geq 1$. Weak convergence of X_n to X does not follow from the weak convergence of the f.d.d.'s; however, with the added assumption of relative compactness it does. The specification of the f.d.d.'s of a p.m. on (C, ρ) does determine it uniquely. Since (C, ρ) is a c.s.m.s., relative compactness is equivalent to tightness by Theorem 2.1. The key result which provides the principal method for proving weak convergence is the following (cf. [5, Theorem 8.1]):

Theorem 2.6. *Let $\{X_n\}$ and X be random functions taking values in $(C, \mathcal{C})$. Then the following are necessary and sufficient conditions for $X_n \Rightarrow X$:*

(a) *the f.d.d.'s of X_n converge weakly to those of X;*

(b) *the family $\{X_n\}$ is tight.*

As mentioned in Section 1, (a) must be shown by classical methods. Thus the primary task at this point is to give conditions for the tightness of a sequence $\{X_n\}$. This amounts to converting the Arzelà–Ascoli criterion for compactness of subsets of C to a probabilistic condition. Convenient sufficient conditions for tightness are given in [5, Theorems 8.3, 12.3].

The theory for $C[0,j], j > 0$, is exactly like that of $C[0,1]$ except that we replace 1 by j in the obvious places. For $C[0, \infty)$ the situation is a little more complicated. Generally, when one has weak convergence on $C[0,1]$, one has it on $C[0, \infty)$. For further discussion and results consult [58, 65].

We turn now to the space $D[0,1] \equiv D$, the space of all real-valued, right-continuous functions on $[0,1]$ having left limits. See [5, ch. 3] for a complete discussion of this space. This is the space which contains the sample paths of most processes of interest in applied probability. In order to describe the metric for D we let Λ denote the class of strictly increasing, continuous maps of [0,1] onto itself. For $\lambda \in \Lambda$, $\lambda(0) = 0$ and $\lambda(1) = 1$. Think of λ as being a new time scale. For $\lambda \in \Lambda$ let

$$\|\lambda\| = \sup_{s \neq t} \left| \log \frac{\lambda(t) - \lambda(s)}{t - s} \right|,$$

and define $d(x, y)$ as follows:

$$d(x, y) = \inf\{\epsilon > 0 \colon \|\lambda\| \leqslant \epsilon \text{ and } \rho(x, y \circ \lambda) \leqslant \epsilon \text{ for some } \lambda \in \Lambda\}.$$

The condition $\|\lambda\| \leqslant \epsilon$ means, of course, that the slopes of the chords of λ lie in the interval $[e^{-\epsilon}, e^{\epsilon}]$. When ϵ is small, these slopes are essentially one. The condition $\rho(x, y \circ \lambda) \leqslant \epsilon$ means that after the function y is "jiggled" left and/or right according to the time transformation λ, the ordinates of x and the transformed y are within ϵ of each other. The function d is a metric for D which renders it a c.s.m.s. This metric, introduced by Billingsley [5, §14], generates the Skorohod topology which relativized to C coincides with the uniform topology. A sequence of elements $\{x_n\}$ in D converges to x in the Skorohod topology if and only if there exist functions $\lambda_n \in \Lambda$ such that $\rho(x_n \circ \lambda_n, x) \to 0$ and $\rho(\lambda_n, e) \to 0$, where $e(t) \equiv t$. Let the Borel sets (the topological Borel field) of D be denoted by $\mathcal{D}$.

Next we would like to produce an analogue of Theorem 2.6. For $t_1, ..., t_k \in [0,1]$ the projections $\pi_{t_1, ..., t_k} \colon (D, \mathcal{D}) \to (R^k, \mathcal{R}^k)$ are measurable; however, they are not necessarily continuous. The latter fact complicates the analogue of 2.6(a). To this end let X be a random function in $(D, \mathcal{D})$ with distribution P. Let T_X consist of those $t \in [0,1]$ for which π_t is continuous a.e. If $d(x_n, x) \to 0$, then $x_n(0) \to x(0)$ and $x_n(1) \to x(1)$, so that π_0 and π_1 are continuous everywhere. For $0 < t < 1$, $t \in T_X$ if and only if

$$\mathbf{P}[X(t) \neq X(t-)] = 0,$$

since it turns out that π_t is continuous at x if and only if x is continuous at t. Since $(D, \mathcal{D})$ is a c.s.m.s., tightness and relative compactness are equivalent as for $(C, \mathcal{C})$. The analogue of Theorem 2.6 is:

Theorem 2.7. *Let $\{X_n\}$ and X be random functions taking values in $(D, \mathcal{D})$. Then the following are necessary and sufficient conditions for $X_n \Rightarrow X$:*

(a) $\pi_{t_1 \dots t_k} \circ X_n \Rightarrow \pi_{t_1 \dots t_k} \circ X$ *whenever* $t_1, ..., t_k \in T_X$;
(b) *the family* $\{X_n\}$ *is tight.*

Sufficient conditions for tightness in D can be found in [5, Theorems 15.2–15.4, 15.6].

We continue our discussion of C and D by stating two small results which have proved to be useful in applications. The first is a result of Liggett and Rosén, a proof of which can be found in [64, p. 46].

Lemma 2.8. *Let $\{X_n\}$ be a sequence of random functions in (D, d), $\{Y_n\}$ a sequence of random functions in (C, ρ), and X a random function in (C, ρ). If $d(X_n, Y_n) \Rightarrow 0$, then $X_n \Rightarrow X$ in (D, d) if and only if $Y_n \Rightarrow X$ in (C, ρ).*

For most practical applications this result means that we can work in C or D, whichever is more convenient. The second result can be found in [27, Lemma 3.2].

Lemma 2.9. *Let $\{X_n\}$ be a sequence of random functions in (D, d) and X a random function such that $\mathbf{P}[X \in C] = 1$. If $X_n \Rightarrow X$, then X_n is C-tight: for all positive ϵ and η, there exists a δ, $0 < \delta < 1$, and an integer n_0 such that*

$$\mathbf{P}[w_{X_n}(\delta) \geqslant \epsilon] \leqslant \eta$$

for $n \geqslant n_0$.

To conclude our discussion of D, we mention a special application of Theorem 2.4 known as the random-time change, which is extremely useful in practice; see [5, §17] for this material. Let D_0 be the subset of D consisting of elements φ which are nondecreasing and satisfy $0 \leqslant \varphi(t) \leqslant 1$ for $t \in [0,1]$. Think of φ as a time-change transformation of [0,1]. Give D_0 the relative Skorohod topology, and let $\mathcal{D}_0$ denote its Borel sets.

Theorem 2.10. *Let $\{X_n\}$ (resp. $\{\Phi_n\}$) be random functions in D (resp. D_0) with X_n and Φ_n defined on the same probability triple. If $X_n \Rightarrow X$ and $\Phi_n \Rightarrow \Phi$, where $\mathbf{P}[X \in C] = 1$ and $\mathbf{P}[\Phi = \varphi] = 1$, with $\varphi \in C$, then*

$$X_n \circ \Phi_n \Rightarrow X \circ \varphi.$$

Note that no independence assumption needs to be made with respect to X_n and Φ_n.

We conclude our discussion of D by mentioning the extensions to $D[0,j]$, $j > 0$, and $D[0, \infty)$. Again $D[0,j]$ can be treated exactly like $D[0,1]$. For $D[0, \infty)$ at least two topologies have been suggested; see

Stone [58] and Whitt [66]. Convergence in $D[0, \infty)$ with Stone's topology is equivalent to convergence in $D[0, r_n]$ for all n, where $\{r_n, n > 1\}$ is some sequence of positive numbers with $r_n \to \infty$ as $n \to \infty$. Hence, if we had weak convergence on $D[0,1]$, we would also expect to have it on $D[0, \infty)$ with Stone's topology. Whitt's topology is stronger and somewhat more complicated than Stone's and will not be discussed here.

3. Partial sum processes

In this section we shall discuss weak convergence results for processes formed by adding r.e's. The limit process for most of these results is Brownian motion whose distribution is Wiener measure.

Wiener measure W on $(C, \mathcal{C})$ is a p.m. which satisfies the following conditions:

(a) $W[x(0) = 0] = 1$,

(b) $W[x(t) \leqslant x] = (2\pi t)^{-1/2} \int_{-\infty}^{x} \exp[-u^2/2t]\, du$,

(c) under W, the process $\{x(t)\colon 0 \leqslant t \leqslant 1\}$ has independent increments.

The p.m. W is known to exist, of course; cf. [5, Theorem 9.1]. We can extend W to $(D, \mathcal{D})$ by giving it the value $W(A \cap C)$ for $A \in \mathcal{D}$, since $A \cap C \in \mathcal{C}$. Still the support of W on $(D, \mathcal{D})$ is C. We use the symbol W for Wiener measure (or the corresponding r.f., called Brownian motion) on $(C, \mathcal{C})$ or $(D, \mathcal{D})$, and count on the context to make the choice clear.

Most of the processes studied in this section are formed as follows. Let $\{\xi_i, i \geqslant 1\}$ be a sequence of r.v.'s defined on $(\Omega, \mathcal{F}, \mathrm{P})$ with $S_k = \xi_1 + \ldots + \xi_k$, $k \geqslant 1$, and $S_0 = 0$. Next form the r.f.'s $X_n \in C$ and $Y_n \in D$ as follows: for $0 \leqslant t \leqslant 1$,

$$X_n(t) = (\sigma\sqrt{n})^{-1} S_{[nt]} + (nt - [nt])(\sigma\sqrt{n})^{-1} X_{[nt]+1}, \qquad (3.1)$$

$$Y_n(t) = (\sigma\sqrt{n})^{-1} S_{[nt]},$$

where σ is a finite, positive constant. The r.f. X_n will be piecewise linear and Y_n piecewise constant, and they will agree at points t of the form k/n, $k = 0, 1, \ldots, n$.

3.1. Independent, identically distributed random variables

The first f.c.l.t. and the prototype for many succeeding results is due to Donsker [13]. It is the functional analogue of the classical Lindeberg–

Lévy c.l.t. for sums of independent identically distributed (i.i.d.) r.v.'s. Donsker's result as further refined is the following:

Theorem 3.1. *If the random variables* $\{\xi_i, i \geq 1\}$ *are independent and identically distributed, with mean* 0 *and variance* σ^2, $0 \leq \sigma^2 < \infty$, *then*

(a) $X_n \Rightarrow W$ *in* $(C, \mathcal{C})$;

(b) $Y_n \Rightarrow W$ *in* $(D, \mathcal{D})$.

A number of methods are available for proving (a) and (b); see [5, §§10, 16]. Convergence of the f.d.d.'s can be deduced immediately from the c.l.t. The easiest proof of (b) is completed by showing that

$$\mathrm{E}\{|Y_n(t)-Y_n(t_1)|^2\,|Y_n(t_2)-Y_n(t)|^2\} \leq 4(t_2-t_1)^2 \tag{3.3}$$

for $t_1 \leq t \leq t_2$, and then applying [5, Theorem 15.6]. To show (3.3), note that by the independence of the X_i's the left-hand side of (3.3) is

$$\sigma^{-4}\, n^{-2}\mathrm{E}\{|S_{[nt]} - S_{[nt_1]}|^2\}\;\mathrm{E}\{|S_{[nt_2]} - S_{[nt]}|^2\} =$$

$$= n^{-2}\,([nt]-[nt_1])\,([nt_2]-[nt]) \leq n^{-2}\,([nt_2]-[nt_1])^2. \tag{3.4}$$

In general, we have

$$n^{-1}\,([nt_2]-[nt_1]) \leq t_2-t_1+n^{-1},$$

which is less than or equal to $2(t_2-t_1)$ provided $t_2-t_1 \geq n^{-1}$. Thus (3.3) is satisfied if $t_2-t_1 \geq n^{-1}$. For the other case, $t_2-t_1 < n^{-1}$, the left-hand side of (3.4) is zero since either $[nt] = [nt_1]$ or $[nt] = [nt_2]$ or both. Thus we have shown that (3.3) holds in all cases, which yields [5, (15.21)] with $F(t) = 2t$ and $\alpha = 1$.

Once we have 3.1(b), 3.1(a) can be shown easily by applying Lemma 2.8.

Note that

$$d(X_n, Y_n) \leq \rho(X_n, Y_n) \leq n^{-1/2} \max\{|\xi_i|: 1 \leq i \leq n\} \tag{3.5}$$

Since ξ_i has a finite second moment, it is well known that the right-hand side of (3.5) converges weakly to zero. Thus $d(X_n, Y_n) \Rightarrow 0$, and since C supports W, Lemma 2.8 yields Theorem 3.1(a).

A vector version of Theorem 3.1 is easily obtained by using Lemma 2.2. In this case let $\{\boldsymbol{\xi}_i, i \geq 1\}$ be a sequence of i.i.d. k-dimensional random vectors with mean vector $\mathbf{0}$, finite positive definite covariance matrix $\boldsymbol{\Sigma}$, and partial sums $\{S_k, k \geq 0\}$. Let $(C^k, \mathcal{C}^k)$ and $(D^k, \mathcal{D}^k)$ be the product spaces formed from k copies of $(C, \mathcal{C})$ and $(D, \mathcal{D})$, respectively, with the product topology. The appropriate r.f.'s in $(C^k, \mathcal{C}^k)$ and $(D^k, \mathcal{D}^k)$ are

$$X_n(t) = n^{-1/2}\Sigma^{-1/2} S_{[nt]} + (nt - [nt])\, n^{-1/2}\Sigma^{-1/2}\xi_{[nt]+1},$$

$$Y_n(t) = n^{-1/2}\Sigma^{-1/2} S_{[nt]},$$

where $\Sigma^{-1/2}$ is the square root of Σ^{-1}, i.e., $\Sigma^{-1} = (\Sigma^{-1/2})'\,\Sigma^{-1/2}$. Then corresponding to Theorem 3.1 we have

$$X_n \Rightarrow W, \qquad Y_n \Rightarrow W, \tag{3.6}$$

where W is the r.f. with values in C^k or D^k having k-dimensional Wiener measure for its distribution. A complete proof of (3.6) can be found in [24].

A second generalization of Theorem 3.1 would be to relax the assumption that $\mathrm{E}\{\xi_i^2\} < \infty$. For the ordinary c.l.t., one then enters the area of weak converge to stable laws and the corresponding characterizations of domains of attraction. Functional analogues of these results are also available, where of course the limit processes are the stable processes. Since these problems do not generally arise in applied probability, we shall not discuss the results here. The basic reference for this material, however, is [52].

A third generalization of Theorem 3.1 would be to consider triangular arrays of r.v.'s

$$\{\xi_{ik_n},\ i = 1, 2, ..., k,\ n = 1, 2, ...\}\,.$$

Prohorov [45, Theorem 3.1] has shown the counterpart of (3.1) for this case under a natural Lindeberg condition. We mention in passing that many of the weak-convergence results that follow have triangular array analogues that will not be discussed.

3.2. Stationary, φ-mixing random variables

Another generalization of the case of i.i.d. summands is to drop the assumption of independence. Independence can be replaced by strict stationarity and a type of asymptotic independence. A complete theory has been developed in [5, ch. 4]. Let $\{\xi_i, i \geq 1\}$ be a strictly stationary sequence defined on $(\Omega, \mathcal{F}, \mathbf{P})$: all f.d.d.'s are invariant under shift. For two indices $i < j$, let $\mathcal{M}_i^j = \mathcal{B}(\xi_i, ..., \xi_j)$, the B.F. generated by $\xi_i, ..., \xi_j$. Let $\varphi(n) \searrow 0$ as $n \to \infty$. Then the sequence $\{\xi_i, i \geq 1\}$ is said to be *φ-mixing* if for each $k, n \geq 1$, $E_1 \in \mathcal{M}_1^k$ and $E_2 \in \mathcal{M}_{k+n}$,

$$|P(E_1 \cap E_2) - P(E_1)\,P(E_2)| \leq \varphi(n).$$

We mention two interesting examples of stationary φ-mixing sequences. The first is m-dependent sequences; i.e., ξ_i and ξ_j are independent whenever $|i - j| > m$ and the sequence is strictly stationary. The second involves a stationary finite-state Markov chain (M.c.), $\{\eta_i, i \geq 1\}$. For a real-valued function f on the state space, set $\xi_i = f(\eta_i)$. Then $\{\xi_i, i \geq 1\}$ is stationary φ-mixing with φ which convergest to 0 geometrically because of the geometric ergodicity in $\{\eta_i\}$. The same claim can be made when the state space is infinite, but $\{\eta_i\}$ is aperiodic, satisfies Doeblin's condition, and has one ergodic class.

The basic result for stationary, φ-mixing sequences $\{\xi_i, i \geq 1\}$ is due to Billingsley and can be found in [5, §20].

Theorem 3.2. *Let $\{\xi_i, i \geq 1\}$ be stationary, φ-mixing with ξ_1 having mean* 0 *and finite variance. If $\Sigma_{n=1}^{\infty} \varphi(n)^{1/2} < \infty$, then*

(a) *the series $\sigma^2 = \mathrm{E}\{\xi_1^2\} + 2\,\Sigma_{k=1}^{\infty} \mathrm{E}\{\xi_1 \xi_{k+1}\}$ converges absolutely;*

(b) *if $\sigma^2 > 0$, then $Y_n \Rightarrow W$, where Y_n is defined by* (3.2).

3.3. Random sums

In many problems in applied probability we are confronted with the following problem. On a single probability space $(\Omega, \mathcal{F}, \mathrm{P})$ are defined a renewal process $\{N(t), t \geq 0\}$ with rate $(0 < \lambda < \infty)$ and a sequence of r.v.'s $\{\xi_i, i \geq 1\}$ which satisfy either the conditions of Theorem 3.1 or Theorem 3.2. Next form the random sum process $\{S_{N(t)}, t \geq 0\}$ and the associated r.f.

$$Z_n(t) = (\sigma \sqrt{n})^{-1}\, S_{N(nt)}.$$

With Y_n defined as in (3.2), we have $Y_n \Rightarrow W$ by Theorem 3.1 or Theorem 3.2. By selecting an appropriate time scale we may assume that $\lambda < 1$ without loss of generality. Let

$$\Phi_n(t) = n^{-1} N(nt) \wedge 1, \qquad \Phi(t) = \lambda t, \qquad 0 \leq t \leq 1.$$

Clearly, Φ_n is a r.f. in D_0 and Φ is a deterministic r.f. in $C \cap D_0$. Then, using the strong law for renewal processes, it is not hard to show that $\rho(\Phi_n, \Phi) \to 0$ a.e. as $n \to \infty$. Since $\mathrm{P}[W \in C] = 1$, we can apply Theorem 2.10 to conclude that $Y_n \circ \Phi_n \Rightarrow W \circ \Phi$. But the probability that $\Phi_n(t) < 1$ for all $t \in [0,1]$ converges to one as $n \to \infty$, and we have $Z_n \Rightarrow W \circ \Phi$ or $\lambda^{-1/2} Z_n \Rightarrow W$. This argument can be found in [5, §10]. Summarizing we have:

Theorem 3.3. *Let $\{N(t), t \geq 0\}$ be a renewal process with rate λ and $\{\xi_i, i \geq 1\}$ a sequence of r.v.'s, satisfying the condition of either Theorem 3.1 or Theorem 3.2, both defined on a common probability space. If $Y_n \Rightarrow W$, then $\lambda^{-1/2} Z_n \Rightarrow W$.*

An alternative formulation of the r.f.'s in Donsker's theorem, random partial sums and renewal processes (to be discussed in Section 4) can be found in [25].

3.4. Functions of positive recurrent Markov chains

One important application of Theorem 3.3 is to partial sums of a function of a positive recurrent, irreducible, aperiodic discrete time Markov chain (M.c.). The basic references for this material are [11] and [20]. Suppose $\{x_i, i \geq 0\}$ is such a M.c. with state space $I = \{0, 1, ...\}$ and stationary distribution $\{\pi_i : i \in I\}$. Let f be a mapping of I into $\mathbf{R} = (-\infty, +\infty)$, and form the stochastic process

$$y_i = f(x_i), \quad i \geq 0.$$

Suppose for this discussion that $x_0 = 0$ a.e. Since the M.c. was assumed to be positive recurrent, it will return to the state 0 infinitely often, and the mean return times will be finite. Suppose the successive return times to state 0 are $\tau_0 = 0 < \tau_1 < \tau_2 < ...$. Using the strong Markov property enjoyed by this M.c., one can show that the three sequences of r.v.'s $\{\alpha_i, i \geq 1\}$, $\{\xi_i, i \geq 1\}$ and $\{\eta_i, i \geq 1\}$ are each i.i.d., where

$$\alpha_i = \tau_i - \tau_{i-1}, \qquad \xi_i = \sum_{j=\tau_{i-1}}^{\tau_i - 1} y_j, \qquad \eta_i = \sum_{j=\tau_{i-1}}^{\tau_i - 1} |y_j|.$$

Let $\{l(n) : n \geq 0\}$ be the discrete renewal process associated with the sequence $\{\tau_i : i \geq 0\}$. Then, if again $S_k = \xi_1 + ... + \xi_k$, we are interested in the process

$$Z_n(t) = n^{-1/2} S_{l(nt)} + n^{-1/2} \sum_{j=\tau_{l(nt)}+1}^{[nt]} y_j, \qquad 0 \leq t \leq 1. \tag{3.7}$$

Except for the second term on the right-hand side of (3.7) and the fact that $\mathrm{E}\{\xi_i\} = 0$, we would be able to apply Theorem 3.2. These obstacles are overcome in the following way. First we get rid of the sum in (3.7). Clearly,

$$\sup_{0 \leq t \leq 1} |Z_n(t) - n^{-1/2} S_{l(nt)}| \leq n^{-1/2} \max \{|\eta_i| : 1 \leq i \leq n+1\}. \tag{3.8}$$

The right-hand side of (3.8) converges weakly to 0, provided $E\{\eta_1^2\} < \infty$, which we shall assume is the case. Next let

$$\mu = \sum_{i \in I} f(i)\, \pi_i,$$

which we assume is absolutely convergent. Ignoring the sum in (3.7), we can write

$$Z_n(t) - \mu\, n^{1/2}\, t = n^{-1/2} \sum_{i=1}^{[nt]} (\xi_i - \mu\alpha_i) - \mu n^{-1/2} (nt - \tau_{l(nt)}). \quad (3.9)$$

From [11, Theorem 1.14.5], we know that $E\{\xi_i\} = E\{\alpha_i\} = \mu/\pi_0$, so $\xi_i - \mu\alpha_i$ has mean zero. Furthermore, the sequence $\{\xi_i - \mu\alpha_i : i \geq 1\}$ is i.i.d. The second term on the right-hand side of (3.9) can be disposed of as in (3.8), if $E\{\alpha_i^2\} < \infty$. Set

$$U_n(t) = Z_n(t) - \mu\, n^{1/2}\, t, \qquad \sigma^2 = E\{(\xi_1 - \mu\alpha_1)^2\}.$$

Combining these arguments and applying Theorem 3.3 yields:

Theorem 3.4. *If* $\Sigma_{i \in I} |f(i)|\, \pi_i < \infty$, $E\{\eta_1^2\} < \infty$, $E\{\alpha_1^2\} < \infty$ *and* $\sigma^2 > 0$, *then* $\pi_0^{-1/2}\, \sigma^{-1} U_n \Rightarrow W$.

The method used to obtain Theorem 3.4 can obviously be used to obtain comparable results for functionals of semi-Markov and cumulative processes. The ordinary c.l.t. for semi-Markov processes is discussed in [47] and for cumulative processes in [42].

4. Point processes

The previous section dealt with processes formed by adding r.e.'s. As one would expect, the limit processes obtained were Brownian motion or more general Gaussian processes. Now we turn to processes which are formed by counting points: point processes. For these processes not only are Gaussian limits permissible but also Poisson processes.

4.1. Renewal processes on the line

Let $\{u_n : n \geq 1\}$ be a sequence of nonnegative i.i.d. r.v.'s with mean λ^{-1}, $0 < \lambda < \infty$, and variance σ^2, $0 < \sigma^2 < \infty$. Form the partial sums $S_k = u_1 + \ldots + u_k$, $k \geq 1$, with $S_0 = 0$, and the associated renewal process

$\{N(t), t \geqslant 0\}$ defined by

$$N(t) = \begin{cases} n & \text{on } \{S_n \leqslant t < S_{n+1}\} , \\ +\infty & \text{on } \bigcap_{n=1}^{\infty} \{S_n \leqslant t\}. \end{cases}$$

Our goal now is to obtain a f.c.l.t. for this renewal process. The key to obtaining this result is the f.c.l.t. for random sums, Theorem 3.3, and the relation

$$S_{N(nt)} \leqslant nt < S_{N(nt)} + u_{N(nt)+1}. \tag{4.1}$$

Let

$$Z_n(t) = (\sigma\sqrt{n})^{-1} \sum_{i=1}^{N(nt)} (u_i - \lambda^{-1}),\; X_n(t) = (\sigma\sqrt{n})^{-1} \sum_{i=1}^{[nt]} (u_i - \lambda^{-1}),\; 0 \leqslant t \leqslant 1.$$

Then (4.1) yields

$$Z_n(t)\; (\sigma\sqrt{n})^{-1}\; (nt - \lambda^{-1} N(nt)) < Z_n(t) + (\sigma\sqrt{n})^{-1}\, u_{N(nt)+1}. \tag{4.2}$$

Since $\mathrm{E}\{u_1^2\} < \infty$, the term on the extreme right of (4.2) can be handled as in (3.8). But $Z_n \Rightarrow \lambda^{1/2} W$ from Theorem 3.3, since we assume the u_i's were i.i.d. Let

$$N_n(t) = (\sigma^2 \lambda^3 n)^{-1/2}\; [N(nt) - \lambda nt], \qquad 0 \leqslant t \leqslant 1.$$

Using the fact that $W \sim -W$ and the continuous mapping theorem we have:

Theorem 4.1. $N_n \Rightarrow W$.

Notice from (4.2) that Theorem 4.1 will follow whenever $Z_n \Rightarrow \lambda^{-1/2} W$, the i.i.d. assumption not being necessary. For example, the sequence $\{u_i, i \geqslant 1\}$ might be taken to be stationary, φ-mixing and Theorem 3.2 could be employed.

Thus we see that weak convergence of the sequence $\{u_i, i \geqslant 1\}$, properly normalized, implies weak convergence of the r.f.'s N_n formed from the renewal process $\{N(t), t \geqslant 0\}$. We would like to point out that the converse is also true; i.e., if $N_n \Rightarrow W$, then $X_n \Rightarrow W$; see [28] and [62]. This converse is not very useful when the $\{u_i, i \geqslant 1\}$ sequence is i.i.d., since we have $X_n \Rightarrow W$ directly. However, when the i.i.d. assumption is dropped, it is useful, and such situations occur in certain queueing prolems. This equivalence between partial sum processes and their associated point process is due to the fact that they are essentially inverses of each other.

4.2. Renewal processes in $\mathbf{R}^k$

We turn now to a f.c.l.t. for renewal processes in k dimensions. Let $\{u_n, n \geqslant 1\}$ be a sequence of random vectors in $\mathbf{R}^k$, $k \geqslant 2$. Set

$$S_0 = 0, \qquad S_k = u_1 + \ldots + u_k \quad \text{for } k \geqslant 1.$$

Suppose $h : \mathbf{R}^k \to [0, \infty)$ is a function with continuous first partial derivatives, $x \neq 0$ implies $h(x) > 0$, and h is homogeneous of degree 1; i.e., for all $x \in \mathbf{R}^k$ and $\lambda \geqslant 0$, $h(\lambda x) = \lambda h(x)$. Define the associated point process $\{M(t): t \geqslant 0\}$ by

$$M(t) = \begin{cases} n & \text{on } \{h(S_n) > t, \max\{h(S_k) : 1 \leqslant k \leqslant n-1\} \leqslant t\}, \\ +\infty & \text{on } \bigcap_{n=1}^{\infty} \{h(S_n) \leqslant t\}. \end{cases}$$

Note that in this case $M(t)$ is essentially the index of the first partial sum S_n, say, for which $h(S_n) > t$. The problem here is that there may be many values of n for which $h(S_n) \leqslant t < h(S_{n+1})$ since the sequence $\{h(S_n), n \geqslant 1\}$ need not be nondecreasing. We wish to show that a f.c.l.t. holds for the point process $\{M(t), t \geqslant 0\}$. The basic references for the ordinary c.l.t. are [4] and [18], and for the f.c.l.t. [32].

Let $\boldsymbol{\mu} \in \mathbf{R}^k$, $\boldsymbol{\mu} \neq 0$, and define the r.f.'s Y_n in D^k by

$$Y_n(t) = n^{-1/2} [S_{[nt]} - nt\boldsymbol{\mu}], \quad 0 \leqslant t \leqslant 1,$$

and M_n in D as

$$M_n(t) = n^{-1/2} [M(nt) - nt/h(\boldsymbol{\mu})], \quad 0 \leqslant t \leqslant 1.$$

Kennedy's [32] main result is

Theorem 4.2. *If* $Y_n \Rightarrow \xi$ *and* $\mathbf{P}[\xi \in C^k] = 1$, *then*

$$M_n \Rightarrow -h(\boldsymbol{\mu})^{-3/2} (\nabla h(\boldsymbol{\mu}) \cdot \xi), \tag{4.3}$$

where C^k *is the product of* k *copies of* C, $\nabla h = (\partial h/\partial x_1, \ldots, \partial h/\partial x_k)$, *and* $\cdot$ *is the scalar product in* $\mathbf{R}^k$.

The main idea of the proof is similar to that of Theorem 4.1, although a number of additional technical details must be overcome. Notice that when the u_k's are i.i.d. with mean μ and positive definite covariance matrix, ξ is Brownian motion with dependent coordinates by (3.6). Since ξ is symmetric in this case, the minus sign before the limit in (4.3) can be omitted.

4.3. *The superposition of point processes*

The Poisson process arises in applied probability with surprising frequency. This phenomenon led Palm [44] to consider the limit process obtained when one takes the superposition of a large number of independent sparse point processes. He showed that the intervals between points in the limit process were exponentially distributed. This suggested, of course, that the limit process should be Poisson. Khintchine [35] (1955 in Russian) proved a result of this type. Many others have since extended his results. For a comprehensive and up-to-date account of this work see [12]. Here we shall indicate the function-space approach to this problem. Our discussion follows that of Çinlar [12] and Straf [59, 60].

A *point process* on the line is a stochastic process $\{N(t), t \geqslant 0\}$ defined on a probability triple $(\Omega, \mathcal{F}, \mathbf{P})$ having almost all sample paths nondecreasing, right-continuous, nonnegative, finite, integer valued and $N(0) = 0$. A process with these properties increases by jumps only, and the times of these jumps are called the points of N. An alternative way of looking at point processes is as random counting measures. A *random counting measure* on the line is a stochastic process $\{N(A), A \in \mathcal{R}_+\}$ taking nonnegative integer values and satisfying $N(A \cup B) = N(A) + N(B)$ a.e. whenever $A \cap B = \emptyset$, where $\mathcal{R}_+$ denotes the Borel sets of $\mathbf{R}_+ = [0,\infty)$. Think of $N(A)$ as being the number of points in A with multiplicity of points permitted and so counted. Then for a point process N identify $N(t) - N(s)$ for $s < t$ with the random counting measure N evaluated at $(s, t]$:

$$N(t) - N(s) = N((s, t]).$$

This dual interpretation will be used repeatedly.

Let μ be a σ-finite measure on $(\mathbf{R}_+, \mathcal{R}_+)$ having no atoms. The random counting measure N_μ is said to be a Poisson random measure with mean measure μ provided $N_\mu(A_1), ..., N_\mu(A_m)$ are independent when

$A_1, ..., A_m \in \mathcal{R}_+$ are disjoint and

$$\mathbf{P}[N_\mu(A) = k] = \exp[-\mu(A)]\,\mu(A)^k/k!, \quad k = 0, 1, 2, ...,$$

whenever $\mu(A) < \infty$. The ordinary homogeneous Poisson process is associated with the Poisson random measure having measure a positive constant times Lebesgue measure. A nonhomogeneous Poisson process with parameter function $\mu(t)$ is associated with a Poisson random measure having mean measure

$$\mu(A) = \int_A \mu(\mathrm{d}x), \quad A \in \mathcal{R}_+.$$

Suppose now that we are given a triangular array of point processes

$$\begin{array}{l} N_{11}, ..., N_{1k_1} \\ \vdots \qquad \vdots \\ N_{n1}, ..., N_{nk_n}, \end{array}$$

where k_n is a monotone sequence of positive integers satisfying $k_n \to \infty$ as $n \to \infty$. We assume that all processes in the same row are defined on a common probability space and are mutually independent. The array of processes are said to be *infinitesimal* if

$$\lim_{n\to\infty} \sup_{1\leqslant i\leqslant k_n} \mathbf{P}[N_{ni}(B) \geqslant 1] = 0 \tag{4.4}$$

for any bounded interval $B \in \mathcal{R}_+$. If (4.4) holds for $B = [0, t]$, all $t > 0$, this is sufficient. Let N_n be the superposition of processes in the same row:

$$N_n = N_{n1} + ... + N_{nk_n}, \qquad n \geqslant 1.$$

Let $E \subset D$ consist of those elements of D that are nondecreasing, nonnegative, integer valued, and 0 at $t = 0$; i.e., $x \in E$ if it has the properties of a sample path of a point process. For $t \in [0,1]$, the r.f.'s N_n can be considered as elements of E. Let N_μ be the distribution of the r.f. associated with a Poisson process having mean measure μ. Since N_μ has no fixed discontinuities, $T_{N_\mu} = [0,1]$. The main result is given in the next theorem. Convergence of the f.d.d.'s is due to Franken [19] and Grigelionis [23], and the observation that this implies the full weak convergence in $(D, \mathcal{D})$ is due to Straf [59, §2; 60, p. 212].

Theorem 4.3. *Suppose the array* $\{N_{ni}\}$ *is infinitesimal and* μ *is a finite, nonatomic measure on the Borel sets of* $[0, r_m]$ *for* $m \geqslant 1$ *having a bounded density with respect to Lebesgue measure, where* $\{r_m\}$ *is the* *sequence entering in Stone's topology. Then*

$$N_n \Rightarrow N_\mu \text{ in } D[0,\infty) \text{ with Stone's topology} \tag{4.5}$$

if and only if

$$\lim_{n\to\infty} \sum_{i=1}^{k_n} \mathbf{P}[N_{ni}(B) = 1] = \mu(B) \tag{4.6}$$

$$\lim_{n\to\infty} \sum_{i=1}^{k_n} \mathbf{P}[N_{ni}(B) \geqslant 2] = 0 \tag{4.7}$$

for any finite interval $B \in \mathcal{R}_+$.

Proof. The fact that (4.6) and (4.7) are necessary and sufficient for the f.d.d.'s of N_n to converge weakly to those of N_μ is shown in [12] using the method of [19], and will not be repeated here. The condition that μ be finite with bounded density is not needed for this argument, but only for the tightness which follows. Recall that for weak convergence in $D[0,\infty)$ with Stone's topology, we need to show it on $D[0, r_n]$ with the Skorohod topology for a sequence $r_m \to \infty$. To show (4.5), then all that remains is to show that weak convergence of the f.d.d.'s implies tightness of $\{N_n, n \geqslant 1\}$.

To prove that $\{N_n\}$ is tight, we shall show that [5, (15.5), (15,6)] hold. We do it only for the case $r_m = 1$, the others being similar. For [5, (15.5)] we do note that

$$\sup_{0\leqslant t\leqslant 1} |N_n(t)| = N_n(1) \Rightarrow N(1),$$

and hence by Prohorov's Theorem 2.1 the sequence

$$\{\sup_{0\leqslant t\leqslant 1} |N_n(t)| : n \geqslant 1\}$$

is tight. For [5, (15.6)] we use Straf's idea. For $k > 1$, let $A_k \subset E$ be the set of functions in D which correspond to point processes having no points (jumps) in $[0, 1/k]$ or in $((k-1)/k, 1]$ and having no two points in the same or adjacent intervals of the form $(j/k, (j+1)/k]$, $j = 0, ..., k-1$. Formally,

$$\begin{aligned} A_k &= E \cap \{x : x(1/k) = 0, x(1) - x(1-1/k) = 0, \\ &\qquad x((j+2)/k) - x(j/k) \leqslant 1, j = 0, ..., k-2\}. \end{aligned} \tag{4.8}$$

Observe that for $x \in A_k$ the jump points of x are separated by a distance no less than $1/k$, and thus one of the allowable partitions of $[0,1]$ which enters in the definition of $w'_x(1/(k+1))$ is the one corresponding to the jumps. But if t_{i-1} and t_i are two such points, $w_x[t_{i-1}, t_i) = 0$. Hence, if $x \in A_k$, then $w'_x(1/(k+1)) = 0$. Thus

$$\mathbf{P}[N_n \in A_k] \leqslant \mathbf{P}[w'_{N_n}(1/(k+1)) = 0].$$

An equivalent form of [5, (15.6)] is that for every $\epsilon > 0$,

$$\lim_{\delta \searrow 0} \overline{\lim_{n\to\infty}}\, \mathbf{P}[w'_{X_n}(\delta) \geqslant \epsilon] = 0. \tag{4.9}$$

So in our present case, (4.9) will hold for N_n if

$$\lim_{k\to\infty} \overline{\lim_{n\to\infty}}\, \mathbf{P}[N_n \notin A_k] = 0.$$

From the definition (4.8) of A_k, we see that

$$\mathbf{P}[N_n \notin A_k] \leqslant \mathbf{P}[N_n(1/k) \geqslant 1] + \mathbf{P}[N_n(1) - N_n(1-1/k) \geqslant 1] + \sum_{j=0}^{k-2} \mathbf{P}[N_n((j+2)/k) - N_n(j/k) \geqslant 2].$$

Since we are assuming convergence of the f.d.d.'s, we have

$$\overline{\lim_{n\to\infty}}\, \mathbf{P}[N_n \notin A_k] \leqslant \mathbf{P}[N_\mu(1/k) \geqslant 1] + \mathbf{P}[N_\mu(1) - N_\mu(1-1/k) \geqslant 1] + \sum_{j=0}^{k-2} \mathbf{P}[N_\mu((j+2)/k) - N_\mu(j/k) \geqslant 2]. \tag{4.10}$$

These terms can be estimated as follows:

$$\mathbf{P}[N_\mu(1/k) \geqslant 1] \leqslant \mu([0, 1/k]),$$

$$\mathbf{P}[N_\mu(1) - N_\mu(1-1/k) \geqslant 1] \leqslant \mu((1-1/k, 1]),$$

$$\mathbf{P}[N_\mu((j+k)/k) - N_\mu(j/k) \geqslant 2] \leqslant \mu^2((j/k, (j+2)/k]).$$

Since μ is nonatomic,

$$\lim_{k\to\infty} \mathbf{P}[N_\mu(1/k) \geqslant 1] = 0, \qquad \lim_{k\to\infty} \mathbf{P}[N_\mu(1) - N_\mu(1-1/k) \geqslant 1] = 0.$$

The remaining term in (4.10) we estimate by

$$\sum_{j=0}^{k-2} \mu^2((j/k, (j+2)/k]) \leqslant \sup_{0 \leqslant j \leqslant k-2} \{\mu((j/k, (j+2)/k])\} \sum_{j=0}^{k-2} \mu((j/k, (j+2)/k]). \tag{4.11}$$

Next observe that

$$\lim_{k\to\infty} \sup_{0\leq j\leq k-2} \mu((j/k,\ (j+2)/k]) = 0$$

since μ has a bounded density, and that

$$\lim_{k\to\infty} \left\{ \sum_{j=0}^{k-2} \mu((j/k,\ (j+2)/k]) \right\} = 2\,\mu([0,1]) < \infty$$

since μ is finite. Hence letting $k \to \infty$ in (4.10) we have

$$\lim_{k\to\infty} \overline{\lim_{n\to\infty}}\ \mathbf{P}[N_n \notin A_k] = 0,$$

which completes the proof of [5, (15.6)] for N_n. Thus we have the theorem.

Theorem 4.3 can be generalized to m-dimensional point processes; see [12] for convergence of f.d.d.'s and [59, 60] for tightness. This enables one to obtain results about the superposition of compound point processes to compound Poisson processes; see [12, §4]. For a different treatment of this problem using characteristic functionals, see [30]. Another approach which begins by considering the weak convergence of partial sums of strong mixing sequences to the Poisson process can be found in [31]. Bingham [7] has also given a proof that for monotonic processes which are stochastically continuous, convergence of f.d.d.'s implies the full weak convergence.

4.4. The thinning of a point process

In the previous subsection we saw that superimposing a large number of independent uniformly spare point processes led to a Poisson process in the limit. Here we consider a single point process which shall be repeatedly thinned. To thin a point process, we randomly delete its point and then expand the time scale appropriately so as to leave its density of points essentially unchanged.

Let $\{N(t), t \geq 0\}$ be a point process and $\{Y_{ni},\ n,i \geq 1\}$ a doubly infinite array of Bernoulli r.v.'s both defined on a common probability triple $(\Omega, \mathcal{F}, \mathbf{P})$. Suppose the Y_{ni}'s are mutually independent and also independent of N. Assume that for all $n,i \geq 1$ that

$$q_n = \mathbf{P}[Y_{ni} = 1] = 1 - \mathbf{P}[Y_{ni} = 0].$$

Now define the new point processes $\{N_n;\ n \geq 0\}$ as r.f.'s in $D[0, \infty)$ as follows:

$$N_0(t) = N(t),$$

$$N_n(t) = \begin{cases} \Sigma_{i=1}^{N_{n-1}(t/q_n)} Y_{ni} & \text{on } \{N_{n-1}(t/q_n) \geqslant 1\}, \\ 0 & \text{otherwise.} \end{cases}$$

Our goal is to show that the sequence $\{N_n: n \geqslant 1\}$ converges weakly to a Poisson process under appropriate conditions on $\{q_n: n \geqslant 1\}$ and N.

Theorem 4.4. *Suppose that μ is a finite, nonatomic measure on the Borel sets of $[0, r_m]$ for $m \geqslant 1$ having a bounded density with respect to Lebesgue measure. If for all $t \geqslant 0$,*

$$N(ut)/u \Rightarrow \mu([0, t]) \quad \text{as } u \to \infty, \tag{4.12}$$

$$K_n = \prod_{i=1}^{n} q_i \to 0 \quad \text{as } n \to \infty, \tag{4.13}$$

then $N_n \Rightarrow N_\mu$.

Proof. See [31].

5. Markov processes

The problem of weak convergence of a sequence of Markov processes to a limiting diffusion process is of great interest in applied probability. A considerable literature exists on this problem, particularly for the special case in which the Markov processes are birth–death in discrete or continuous time. In our discussion here we shall only state some results for the simplest case of such convergence. This case covers, nevertheless, many of the models which arise in practice. For more general cases in which boundary behaviour of the processes are involved, the conditions become more extensive and complicated. The results discussed here are due to Stone [56].

Let $\{X_n, n \geqslant 1\}$ be the r.f.'s with values in $D[0, \infty)$ corresponding to a sequence of birth-and-death processes in continuous time with state space $E_n = \{\alpha_i^{(n)}: 0 \leqslant i < \infty\}$, where $\alpha_i^{(n)}$ is increasing in i and $\lim_{i\to\infty} \alpha_i^{(n)} = \alpha_\infty$. Let X be the r.f. corresponding to a diffusion process with state space $E = [a_1, a_2] \subset [-\infty, \infty]$. The specialization mentioned above is that we only consider diffusions for which $E = (-\infty, \infty)$ is a regular interval and both a_1 and a_2 are inaccessible. This means that for any two points $x, y \in (-\infty, \infty)$ there is positive probability of going from x to y in finite time, and that neither a_1 nor a_2 can be reached in finite time. Furthermore, we assume that the infinitesimal generator (i.g.) of the diffusion X is of the form

$$A = \tfrac{1}{2}\, \sigma^2(x) \frac{\mathrm{d}^2}{\mathrm{d}x^2} + m(x) \frac{\mathrm{d}}{\mathrm{d}x}, \qquad -\infty < x < \infty,$$

where $\sigma^2(x) > 0$, $\sigma^2(x)$ and $m(x)$ are continuous for $-\infty < x < \infty$, and the domain of A consists of all twice continuously differentiable functions on $(-\infty, \infty)$. As examples of such diffusions that arise frequently in practice, think of X as being Brownian motion ($\sigma^2(x) \equiv 1$, $m(x) \equiv 0$) or the Ornstein–Uhlenbeck process ($\sigma^2(x) \equiv 2$, $m(x) = -x$).

Now for the birth–death process X_n let

$$e_n(x) = \sup \{\alpha_i^{(n)} \in E_n : \alpha_i^{(n)} \leqslant x\}$$

for $x \in [\alpha_0^{(n)}, \alpha_\infty^{(n)}]$. Denote by $\lambda_i^{(n)}$ and $\mu_i^{(n)}$ the birth–death parameters associated with the state $\alpha_i^{(n)}$. Then set

$$m_n(x) = \mu_i^{(n)} (\alpha_{i-1}^{(n)} - \alpha_i^{(n)}) + \lambda_i^{(n)} (\alpha_{i+1}^{(n)} - \alpha_i^{(n)}),$$

$$\sigma_n^2(x) = \mu_i^{(n)} (\alpha_{i-1}^{(n)} - \alpha_i^{(n)})^2 + \lambda_i^{(n)} (\alpha_{i+1}^{(n)} - \alpha_i^{(n)})^2,$$

whenever $x \in [\alpha_0^{(n)}, \alpha_\infty^{(n)})$ and $e_n(x) = \alpha_i^{(n)}$. The function $m_n(x)$ ($\sigma_n^2(x)$) is known as the infinitesimal mean (variance) of the process X_n.

Stone's result can now be stated in the following theorem.

Theorem 5.1. *Let $X_n(0) = x_n$ and $X(0) = x$ a.e., with $x_n \to x$. Then the following two conditions are sufficient for $X_n \Rightarrow X$ as elements of $D[0,\infty)$:*

(a) *E_n becomes dense in $(-\infty, \infty)$ as $n \to \infty$;*

(b) *for every compact subinterval I of $(-\infty, \infty)$,*

$$\lim_{n\to\infty} m_n(x) = m(x), \qquad \lim_{n\to\infty} \sigma_n^2(x) = \sigma^2(x)$$

uniformly for $x \in I$.

As the limiting diffusion X has continuous paths, condition 5.1(a) is an obvious requirement. Condition 5.1(b) can be viewed as follows Each of the processes X_n and X has associated with it a transition function which in turn induces a semigroup of operators on an appropriate function space. These semigroups of operators are uniquely determined by their i.g.'s. The fact that the boundaries a_1 and a_2 of the diffusion X are inaccessible means that no boundary condition on the functions in the domain of A are required. Since both the X_n and X processes have continuous paths, their i.g.'s are local linear operators. Condition 5.1(a) is essentially a necessary and sufficient condition for the i.g.'s of the X_n's to converge to the i.g. of X. Hence the associated semigroups converge, which implies the convergence of the transition functions. Because of the Markov character of the processes, this implies convergence of the f.d.d.'s. Subsequent to the work of Stone, Liggett [39, §4] has shown that if the f.d.d.'s converge, then the processes are tight, and hence we have the full weak convergence. Liggett's proof also works for birth-and-death processes in higher dimensions. Hence a counterpart of 5.1(a) should be the appropriate condition in this case also.

To complete this section, we mention some of the other work related to this problem. Stone [57] has given an a.e. version of the work described above by appealing to local time arguments. Skorohod [54] and Trotter [61] discuss this problem from the semigroup point of view. Skorohod [55] and Gikhman and Skorohod [21] have treated the problem from the point of view of stochastic integrals and of the Skorohod representation. Borovkov [8–10] has dealt with the convergence of non-Markovian processes to diffusions under appropriate conditions on the infinitesimal moments. Another important method for problems in this area is due to Rosén [48, 49]; also see [5, ch. 4] and [43, ch. 8].

6. Extremal processes

Let $X_1, X_2, \ldots$ be a sequence of r.v.'s defined on a probability triple $(\Omega, \mathcal{F}, \mathbf{P})$, and define the r.v.'s

$$Y_n^1 = \max\{X_1, \ldots, X_n\}, \qquad Y_n^2 = \text{second largest among } \{X_1, \ldots, X_n\}.$$

The c.l.t. for Y_n^1 and the corresponding domains of attraction are classical problems treated by Gnedenko [22]; when the X_i's are i.i.d., Loynes [41] has dealt with the c.l.t. for Y_n^1 when the X_i's are a stationary φ-mixing sequence. In this section we shall discuss f.c.l.t.'s for (Y_n^1, Y_n^2) for both the i.i.d. and stationary φ-mixing cases, but only give a proof for the i.i.d. case.

6.1. Independent, identically distributed sequences

Assume that the sequence $X_1, X_2, \ldots$ is i.i.d. with d.f. F. Gnedenko [21] has determined that for constants $a_n > 0$ and b_n and non-degenerate d.f. G,

$$(Y_n^1 - a_n)/b_n \Rightarrow G \tag{6.1}$$

implies that G must be one of the following three d.f.'s (except for scale and location parameters):

$$\Phi_\alpha(x) = \begin{cases} 0, & x \leqslant 0, \\ \exp\{-x^{-\alpha}\}, & x > 0, \end{cases}$$

$$\Psi_\alpha(x) = \begin{cases} \exp\{-(-x)^\alpha\}, & x \leqslant 0, \\ 0, & x > 0, \end{cases}$$

or

$$\Lambda(x) = \exp\{-e^{-x}\}, \qquad -\infty < x < \infty.$$

Now form the r.f. M_n in $D[0,1]$ by setting

$$M_n(t) = \begin{cases} (Y_{[nt]}^1 - a_n)/b_n, & 1/n < t \leqslant 1, \\ (Y_1^1 - a_n)/b_n, & 0 \leqslant t \leqslant 1/n, \end{cases}$$

where the constants a_n and b_n are those that appear in (6.1) and give rise to a particular G. Our goal is to show that the sequence M_n converges weakly in the space $D[a,1]$, $a > 0$, to a limit process M. The M process has been called an extremal process by Dwass [16; 17], and is a r.f. on $D[a,1]$ characterized by its f.d.d.'s. For $a \leqslant t_1 < \ldots < t_k \leqslant 1$,

$$\mathrm{P}[M(t_1) \leqslant x_1, \ldots, M(t_k) \leqslant x_k] = G(x_1)^{t_1}\, G(x_2)^{t_2 - t_1} \ldots G(x_k)^{t_k - t_{k-1}}$$

whenever $x_1 \leqslant x_2 \leqslant \ldots \leqslant x_k$, and is equal to 0 otherwise. The following f.c.l.t. is due to Lamperti [36].

Theorem 6.1. *If $\{X_n, n \geqslant 1\}$ is a sequence of independent, identically distributed random variables and (6.1) holds, then in the space $D[a,1]$, $0 < a < 1$, with the Skorohod topology,*

$$M_n \Rightarrow M. \tag{6.2}$$

Proof. In order to show that the f.d.d.'s converge first take $t > 0$, x real and observe that

$$\mathrm{P}[M_n(t) \leqslant x] = [F^n(b_n x + a_n)]^{[nt]/n} \to G^t(x) \tag{6.3}$$

since we are assuming that (6.1) holds. Next take $a \leqslant t_1 < t_2 \leqslant 1$ and $x_1 < x_2$. Then by straightforward arguments we can write

$$\mathbf{P}[M_n(t_1) \leqslant x_1, M_n(t_2) \leqslant x_2] =$$

$$= \mathbf{P}[M_n(t_1) \leqslant x_1]\ \mathbf{P}[M_n(t_2) \leqslant x_2 \mid M_n(t_1) \leqslant x_1], \tag{6.4}$$

$$\mathbf{P}[M_n(t_2) \leqslant x_2 \mid M_n(t_1) \leqslant x_1] =$$

$$= \mathbf{P}[X_l \leqslant b_n x_2 + a_n,\ 1 \leqslant l \leqslant [nt_2] \mid X_l \leqslant b_n x_1 + a_n,\ 1 \leqslant l \leqslant [nt_1]]$$

$$= \mathbf{P}[X_l \leqslant b_n x_2 + a_n,\ [nt_1] + 1 \leqslant [nt_2]]. \tag{6.5}$$

Combining (6.3)–(6.5), we have

$$\mathbf{P}[M_n(t_1) \leqslant x_1, M_n(t_2) \leqslant x_2] \to G^{t_1}(x_1)\, G^{t_2 - t_1}(x_2).$$

The higher-dimensional distributions can be handled in a similar manner. For the proof of tightness of $\{M_n, n \geqslant 1\}$, see [63], where [5, Theorem 15.3] is used.

Now define the r.f. N_n as follows:

$$N_n(t) = \begin{cases} (Y^2_{[nt]} - a_n)/b_n, & 1/n < t \leqslant 1, \\ (Y^2_1 - a_n)/b_n, & 0 \leqslant t \leqslant 1/n. \end{cases}$$

Lamperti [36] has shown that under the assumption (6.1), the f.d.d.'s of (M_n, N_n) converge to those of (M, N), where

$$\mathbf{P}[M(t) \leqslant x, N(t) \leqslant y] = \begin{cases} G^t(x), & y \geqslant x, \\ G^t(y)\ [1 + t \log G(x)/G(y)], & y < x. \end{cases}$$

As usual, to verify tightness one needs only consider the marginal processes. The tightness of M_n is shown above, and the same proof works also for N_n. Thus we have $(M_n, N_n) \Rightarrow (M, N)$. The latter fact was pointed out by Welsch [63].

There are two further generalizations of this weak convergence result. Not only can we treat the first and second largest r.v.'s, but also the k^{th} largest and various vector-valued combinations of these. Secondly, the assumption that the X_i's be i.i.d. can be replaced by the weaker assumption of stationary φ-mixing. Both generalizations in the weak convergence setting are due to Welsch [63].

7. Functionals of limit processes

The most important feature of weak convergence of random functions is that a great variety of other limit theorems can be obtained almost immediately by employing the continuous mapping theorem (Theorem 2.4). Often in applied probability we first demonstrate weak convergence for a sequence of random functions for which such results are readily available. Then by introducing an appropriate functional on the original sequence of processes we obtain the desired weak convergence for the sequence of real interest. In order for this procedure to have real payoff, we must be able to calculate the distribution of these functionals of the limit processes. In this section we shall discuss this problem when the limit process is either Brownian motion or the Poisson process.

7.1. Functionals of Brownian motion

We begin by pointing out the following important fact. Suppose that $X_n \Rightarrow X$ as r.f.'s on $(D, \mathcal{D})$ and that $\mathbf{P}[X \in C] = 1$. Assume that $h : (D, \mathcal{D}) \to (S', \mathcal{S}')$ is a measurable map which is continuous on C with respect to the uniform topology, where $(S', \mathcal{S}')$ is a second complete separable metric space. Then because the uniform and Skorohod topologies agree on C and $\mathbf{P}[X \in C] = 1$, the continuous mapping theorem (Theorem 2.4) tells us that $h \circ X_n \Rightarrow h \circ X$. To complete the job, we should be able to specify the distribution of $h \circ X$. We shall list in Table 1 a number of specific instances when this can be done in the

Table 1
Functionals of Brownian motion

$h(x)$	$\mathbf{P}[h \circ W \leqslant y]$		References
$x(t)$	$\mathbf{P}[N(0,t) \leqslant y]$,	$y \in \mathbf{R}^1$	[5, p. 61]
$\lvert x(t)\rvert$	$2\mathbf{P}[N(0,t) \leqslant y] - 1$,	$y \geqslant 0$	[5, p. 61]
$x(t_2) - x(t_1)$, $0 \leqslant t_1 \leqslant t_2 \leqslant 1$	$\mathbf{P}[N(0, t_2 - t_1) \leqslant y]$,	$y \in \mathbf{R}^1$	[5, p. 61]
$\sup\{x(s): 0 \leqslant s \leqslant t\}$	$2\mathbf{P}[N(0,t) \leqslant y] - 1$,	$y \geqslant 0$	[5, p. 72]
$\inf\{x(s): 0 \leqslant s \leqslant t\}$	$2\mathbf{P}[N(0,t) \leqslant y]$,	$y \leqslant 0$	[5, p. 72]
$x(t) - \inf\{x(s): 0 \leqslant s \leqslant t\}$	$2\mathbf{P}[N(0,t) \leqslant y] - 1$,	$y \geqslant 0$	[29, p. 41]
$\inf\{t > 0: x(t) = a\}$, $a > 0$	$2\mathbf{P}[N(0,y) > a]$,	$y > 0$	[29, p. 25]
$t^{-1} m\{s: 0 \leqslant s \leqslant t, x(s) > 0\}$ ($m \equiv$ Lebesgue measure)	$(2/\pi) \arcsin \sqrt{y}$,	$0 \leqslant y \leqslant 1$	[29, p. 57]
$t^{-1} \sup\{s \in [0,t]: x(t) = 0\}$	$(2/\pi) \arcsin \sqrt{y}$,	$0 \leqslant y \leqslant 1$	[5, p. 28]
$\sup\{\lvert x(s)\rvert: 0 \leqslant s \leqslant t\}$	$\Sigma_{k=-\infty}^{\infty} (-1)^k \mathbf{P}[2(k-1)y < N(0,t) \leqslant (2k+1)y]$		[5, p. 79]

case where $X = W$, Brownian motion. The symbol $N(0, t)$ stands for a normal r.v. with mean 0 and variance t. The function $x(t) - \inf\{x(s): 0 \leqslant s \leqslant t\}$ makes the origin an impenetrable barrier and is of fundamental importance in queueing theory.

7.2. Functionals of the Poisson process

Let N_μ be a Poisson random measure or process with σ-finite mean measure μ having no atoms. We have seen in Section 4 that the Poisson process can arise as the limit process in a number of ways. To capitalize on weak convergence results we would like to be able to apply the continuous mapping theory as was done for Brownian motion.

The most interesting functional to consider is the mapping $h: D[0,\infty) \to (\mathbf{R}^\infty, \mathcal{R}^\infty)$ whose k^{th} coordinate is

$$h_k(x) = \begin{cases} +\infty, & \{t: x(t) = k\} = \emptyset, \\ \inf\{t: x(t) = k\}, & \text{otherwise,} \end{cases}$$

where $k \geqslant 1$. The mapping h is measurable, but not continuous with respect to the Skorohod topology on D. However, h is continuous on E; see Section 4 for definition. Since the distribution of N_μ concentrates on E, the continuous mapping theorem can be applied. Next we can look at $h_k(x) - h_{k-1}(x)$ and obtain weak convergence for the interarrival time between the $(k-1)^{\text{st}}$ and k^{th} arrivals.

8. Concluding remarks

We hope that this survey has given the reader a taste of the type of weak convergence results available. In some sense weak convergence theory is a way of life: whenever a c.l.t. is discovered, the first impulse is to extend it to a f.l.c.t. This extension is often pleasing from a purely esthetic point of view; however, it is also frequently very practical from an applied point of view.

There are three areas of work that have not yet been mentioned which we feel will be important for future work in applied probability. The first deals with a theory for processes with a multiple time parameter; cf. [3, 59, 60]. This work will have applications to models involving random measures in $\mathbf{R}^k$. It has already seen application to empirical d.f.'s for random vectors. The second area is the weak convergence of stochastic processes conditioned on a particular event. Most of this work has been carried out by Liggett [37–40]. For other examples of conditioned f.c.l.t's, see [1, 2, 26]. It seems very likely that this work will find applications in epidemic theory and other models which are essentially "one-shot affairs". The last area is that of rates of convergence for f.c.l.t.'s. This problem is, of course, a natural one to be addressed for any type of limit theorem. There are numerous results available among which we mention [50, 51, 33, 34, 15].

Acknowledgment

I am grateful to E. Çinlar, D.P. Kennedy and M. Straf for useful conversations on the material of Section 4 and for sharing their work [12, 33, 34, 60] prior to publication.

References

[1] B. Belkin, A limit theorem for conditioned recurrent random walk attracted to a stable law, Ann. Math. Statist. 41 (1970) 146–163.
[2] B. Belkin, An invariance principle for conditioned random walk attracted to a stable law, Z. Wahrscheinlichkeitstheorie Verw. Gebiete 21 (1972) 45–64.
[3] P.J. Bickel and M.J. Wichura, Convergence criteria for multiparameter stochastic processes and some applications, Ann. Math. Statist. 42 (1971) 1656–1670.
[4] P.J. Bickel and J.A. Yahav, Renewal theory in the plane, Ann. Math. Statist. 36 (1965) 946–955.
[5] P. Billingsley, Convergence of probability measures (Wiley, New York, 1968).
[6] P. Billingsley, Weak Convergence of Measures: Applications in Probability (SIAM, Philadelphia, Pa., 1971).
[7] N.H. Bingham, Limit theorems for occupation times of Markov processes, Z. Wahrscheinlichkeitstheorie Verw. Gebiete 17 (1971) 1–22.
[8] A.A. Borovkov, Convergence of weakly dependent processes to the Wiener process, Theor. Probab. Appl. 12 (1967) 159–186 (English Transl.).
[9] A.A. Borovkov, On convergence to diffusion processes, Theor. Probab. Appl. 12 (1967) 405–431 (English Transl.).
[10] A.A. Borovkov, Theorems on the convergence to Markov diffusion processes, Z. Wahrscheinlichkeitstheorie Verw. Gebiete 16 (1970) 47–76.
[11] K.L. Chung, Markov Chains with Stationary Transition Probabilities (Springer, Berlin, 1960).
[12] E. Çinlar, Superposition of point processes, in: P.A.W. Lewis, ed., Stochastic Point Processes: Statistical Analysis, Theory and Applications (Wiley, New York, 1972) 549–606.
[13] M. Donsker, An invariance principle for certain probability limit theorems, Mem. Am. Math. Soc 6 (1951).
[14] R.M. Dudley, Distances of probability measures and random variables, Ann. Math. Statist. 39 (1968) 1563–1572.

[15] R.M. Dudley, Speeds of metric probability convergence, Z. Wahrscheinlichkeitstheorie Verw. Gebiete 22 (1972) 323–332.
[16] M. Dwass, Extremal processes, Ann. Math. Statist. 35 (1964) 1718–1725.
[17] M. Dwass, Extremal processes II, Illinois J. Math. 10 (1966) 381–391.
[18] R. Farrell, Limit theorems for stopped random walks III, Ann. Math. Statist. 37 (1966) 1510–1527.
[19] P. Franken, A refinement of the limit theorem for the superposition of independent renewal processes, Theor. Probab. Appl. 8 (1963) 320–328 (English Transl.).
[20] D.A. Freedman, Some invariance principles for functionals of a Markov chain, Ann. Math. Statist. 38 (1967) 1–7.
[21] I.I. Gikhman and A.V. Skorohod, Introduction to the Theory of Random Processes (Saunders, Philadelphia, Pa., 1969) (English Transl.).
[22] B.V. Gnedenko, Sur la distribution de terme maximum d'une série aléatoire, Ann. Math. 33 (1943) 423–453.
[23] B. Grigelionis, On the convergence of sums of random step processes to a Poisson process, Theor. Probab. Appl. 8 (1963) 177–182 (English Transl.).
[24] D.L. Iglehart, Weak convergence of probability measures on product spaces with applications to sums of random vectors, Tech. Rept. No. 109, Dept. of Operations Res., Stanford Univ., Stanford, Calif. (1968).
[25] D.L. Iglehart, Weak convergence of compound stochastic processes I, Stochastic Processes Appl. 1 (1973) 11–31, 185–186.
[26] D.L. Iglehart, Functional central limit theorems for random walks conditioned to stay positive, Ann. Probab. (1974), to appear.
[27] D.L. Iglehart and W. Whitt, Multiple channel queues in heavy traffic I, Adv. Appl. Probab. 2 (1970) 150–177.
[28] D.L. Iglehart and W. Whitt, The equivalence of functional central limit theorems for counting processes and associated partial sums, Ann. Math. Statist. 42 (1971) 1372–1378.
[29] K. Îto and H.P. McKean, Diffusion Processes and their Sample Paths (Springer, Berlin, 1965).
[30] P. Jagers, On the weak convergence of superpositions of point processes, Z. Wahrscheinlichkeitstheorie Verw. Gebiete 22 (1972) 1–7.
[31] D.P. Kennedy, Weak convergence for the superposition and thinning of point processes. Tech. Rept. No. 11, Dept. of Operations Res., Stanford Univ., Stanford, Calif. (1970).
[32] D.P. Kennedy, A functional central limit theorem for k-dimensional renewal theory, Ann. Math. Statist. 42 (1971) 376–380.
[33] D.P. Kennedy, Rates of convergence for queues in heavy traffic I, Adv. Appl. Probab. 4 (1972) 357–381.
[34] D.P. Kennedy, Rates of convergence for queues in heavy traffic II; Sequences of queueing systems, Adv. Appl. Probab. 4 (1972) 382–391.
[35] A. Khintchine, Mathematical Methods in the Theory of Queueing (Griffin, London, 1960) (English Transl.).
[36] J. Lamperti, On extreme order statistics, Ann. Math. Statist. 35 (1964) 1726–1737.
[37] T.M. Liggett, An invariance principle for conditioned sums of independent random variables, J. Math. Mech. 18 (1968) 559–570.
[38] T.M. Liggett, Weak convergence of conditioned sums of independent random vectors, Trans. Am. Math. Soc. 152 (1970) 195–213.
[39] T.M. Liggett, On convergent diffusions: the densities and the conditioned processes, Indiana Univ. Math. J. 3 (1970) 265–279.
[40] T.M. Liggett, Convergence of sums of random variables conditioned on a future change of sign, Ann. Math. Statist. 41 (1970) 1978–1982.
[41] R.M. Loynes, Extreme values in uniformly mixing stationary stochastic processes, Ann. Math. Statist. 36 (1965) 993–999.

[42] D. Miller, On the asymptotic behavior of regenerative processes and functionals of regenerative processes, Ph.D. Thesis, Cornell Univ., Ithaca, N.Y. (1971).
[43] M.F. Norman, Markov Processes and Learning Models (Academic Press, New York, 1972).
[44] C. Palm, Intensitätsschwankungen im Fernsprechverkehr, Ericsson Techniks 44 (1943) 1–189.
[45] Ju. Prohorov, Convergence of random processes and limit theorems in probability theory, Theor. Probab. Appl. 1 (1956) 157–214 (English Transl.).
[46] R. Pyke, Applications of almost surely convergent constructions of weakly convergent processes, Proc. Intern. Symp. on Information Theory (Springer, Berlin, 1968) 187–200.
[47] R. Pyke and R. Schaufle, Limit theorems for Markov renewal Processes, Ann. Math. Statist. 35 (1964) 1746–1764.
[48] B. Rosén, On the central limit theorem for sums of independent random variables, Z. Wahrscheinlichkeitstheorie Verw. Gebiete 7 (1967) 48–82.
[49] B. Rosén, On asymptotic normality of sums of dependent random vectors, Z. Wahrscheinlichkeitstheorie Verw. Gebiete 7 (1967) 95–102.
[50] W.A. Rosenkrantz, On rates of convergence for the invariance principle, Trans. Am. Math. Soc. 129 (1967) 542–552.
[51] W.A. Rosenkrantz, A rate of convergence for the Von Mises statistic, Trans. Am. Math. Soc. 139 (1969) 329–337.
[52] A.V. Skorohod, Limit theorems for stochastic processes, Theor. Probab. Appl. 1 (1956) 262–290 (English Transl.).
[53] A.V. Skorohod, Limit theorems for stochastic processes with independent increments, Theor. Probab. Appl. 2 (1957) 138–171 (English Transl.).
[54] A.V. Skorohod, Limit theorems for Markov processes. Theor. Probab. Appl. 3 (1958) 202–246 (English Transl.).
[55] A.V. Skorohod, Studies in the Theory of Random Processes (Addison-Wesley, Reading, Mass., 1965) (English Transl.).
[56] C.J. Stone, Limit theorems for birth and death and diffusion processes, Ph.D. Thesis, Stanford Univ., Stanford, Calif. (1961).
[57] C.J. Stone, Limit theorems for random walks, birth and death processes, and diffusion processes, Illinois J. Math. 7 (1963) 638–660.
[58] C.J. Stone, Weak convergence of stochastic processes defined on semi-infinite time intervals, Proc. Am. Math. Soc. 14 (1963) 694–696.
[59] M. Straf, A general Skorohod space and its application to the weak convergence of stochastic processes with several parameters, Ph.D. Thesis, Univ. of Chicago, Chicago, Ill. (1969).
[60] M. Straf, Weak convergence of stochastic processes with several parameters, in: Proc. 6th Berkeley Symp. Vol. II (1972) 187–221.
[61] H.F. Trotter, Approximations of semi-groups of operators, Pacific J. Math. 8 (1958) 887–919.
[62] W. Vervaat, Success epochs in Bernoulli trials, Ph.D. Thesis, Univ. of Amsterdam, Amsterdam (1972).
[63] R. Welsch, Limit laws for extreme order statistics from strong-mixing processes, Ann. Math. Statist. 43 (1972) 439–446.
[64] W. Whitt, Weak convergence theorems for queues in heavy traffic, Ph.D. Thesis, Cornell Univ., Ithaca, N.Y. (1968) (Tech. Rept. No.2, Dept. of Operations Res., Stanford Univ., Stanford, Calif.).
[65] W. Whitt, Weak convergence of probability measures on the function space $C[0, \infty)$, Ann. Math. Statist. 41 (1970) 939–944.
[66] W. Whitt, Stochastic Abelian and Tauberian theorems, Z. Wahrscheinlichkeitstheorie Verw. Gebiete 22 (1972) 251–267.
[67] M.J. Wichura, On the construction of almost uniformly convergent random variables with given weakly convergent image laws, Ann. Math. Statist. 41 (1970) 284–291.