THE SKOROKHOD REPRESENTATION

STANLEY SAWYER

1. **Introduction.** This is an expository and survey article about a certain embedding technique which is very useful for proving limit theorems in probability and statistics. It is also of interest because, when it can be applied, it usually provides a very illuminating proof. The main purpose here is to discuss what is known about the method and its applications; in illustration we prove two basic limit theorems using the technique. There is also one original result, in §4.

The Skorokhod representation is an embedding of a random variable $X(\omega)$ (or a sequence of random variables) in a Brownian motion. Here, a Brownian motion is a collection of random variables $\{b(s, \omega) : 0 \leq s < \infty \}$ such that (i) the differences $b(t) - b(s)$ are normally distributed with mean zero and variance $t - s$, with $b(0) = 0$, (ii) non-overlapping increments $\{b(t_j) - b(s_j)\}$ (i.e., the intervals (s_j, t_j) are disjoint) are (statistically) independent, and (iii) with probability one, the "sample paths" $b = b(s) = b(s, \omega)$ are continuous functions of s. If $\{b^i(s) : i = 1, 2, 3\}$ are three independent Brownian motions, $(x_0 + b^1(s), y_0 + b^2(s), z_0 + b^3(s))$ is a representation of, for example, the position of a small particle of dust or pollen in a glass of water. See Breiman [2], Freedman [15], or Ito-McKean [19] for the mathematical background and Einstein [9] or Nelson [29] for the connections with nature.

If $X = X(\omega)$ is a random variable, the representations of X considered here will be of the following form. We find a "stopping time" $T = T(\omega)$ (a particular non-negative random variable) such that $b(T(\omega), \omega) \cong X(\omega)$; i.e., $P[b(T(\omega)) \leq \lambda] = P[X(\omega) \leq \lambda]$ for all λ. A stopping time is a random time $T(\omega)$ whose occurrence depends on $b(s)$ only up to that time; i.e., such that for all t, the event $\{\omega : T(\omega) \leq t\} \in B_t$, where B_t is the σ-algebra of events generated by the random variables $b(s, \omega)$ for $0 \leq s \leq t$. Here we shall allow $T(\omega)$ to depend, in addition, on one or more random variables which are independent of the process $\{b(s) : 0 \leq s < \infty \}$.

Two questions arise at this point. First, why is such a representation useful; second, do such representations in fact exist for a general random variable X. To answer the first question first,

2. **Applications of the Embedding.** Assume we have a stopping

---

time $T(\omega)$ representing a random variable X in the above sense, and consider the process $b_2(s, \omega) = b(s + T(\omega), \omega) - b(T(\omega), \omega)$. Since $T(\omega)$ is a stopping time, it is standard that $b_2(s)$ is also a Brownian motion in the sense of §1, and $\{b_2(s)\}$ is independent of both $T(\omega)$ and $b(T(\omega), \omega)$. Suppose we now have a second stopping time $T_2(\omega)$ for $b_2(s)$ for which $b_2(T_2(\omega)) \cong X_2(\omega)$, where X_2 is a second random variable. Then T_2 and $b_2(T_2)$ are independent of $b(T)$ and T, and if X and X_2 were independent to begin with,

$$X + X_2 \cong b_2(T_2) + b(T) = (b(T_2 + T) - b(T)) + b(T)$$
$$\cong b(T + T_2).$$

Now, $T(\omega) + T_2(\omega)$ is also a stopping time for $b(s, \omega)$, and the Brownian motion $b_3(s) = b(s + T + T_2) - b(T + T_2)$ is independent of anything depending on $b(s, \omega)$ for $s \leqq (T + T_2)(\omega)$ (in particular T, $b(T)$, T_2, and $b_2(T_2)$).

Hence if we are given a set of independent random variables $\{X_1, X_2, \cdots, X_n\}$ (with $X = X_1$), and follow this procedure for n steps, we obtain n random times $\{T_1, T_2, T_3, \cdots, T_n\}$ $(T = T_1)$, which are themselves independent random variables, such that

$$(1) \qquad \begin{aligned} &\{X_1 + X_2 + X_3 + \cdots + X_k : 1 \leqq k \leqq n\} \\ &\cong \{b(T_1 + T_2 + \cdots + T_k) : 1 \leqq k \leqq n\}. \end{aligned}$$

I.e., $b[(T_1 + T_2 + \cdots + T_k(\omega), \omega)]$ $(1 \leqq k \leqq n)$ is an exact statistical replica of the partial sums $X_1 + X_2 + \cdots + X_k$.

The representation (1) can be used to obtain information on the behavior of sums of independent random variables. For definiteness, assume the $\{X_k\}$ are identically distributed (i.e., $P[X_k \leqq \lambda] = P[X_1 \leqq \lambda]$) and we have a uniform procedure for generating the $\{T_k\}$. Then the $\{T_k\}$ will also be identically distributed. If the T_k have a finite mean, then by the strong law of large numbers:

$$(2) \qquad \lim_{n \to \infty} (T_1(\omega) + T_2(\omega) + \cdots + T_n(\omega))/n = E(T_1) \text{ a.s.,}$$

and, in particular, $T_1(\omega) + T_2(\omega) + \cdots + T_n(\omega) = O(n)$ a.s. Now, for Brownian motion, it is fairly easy to compute ([2, p. 266], [19, p. 36])

$$(3) \qquad b(t, \omega) = O((t \log \log t)^{1/2}) \text{ a.s.} \qquad \text{as } t \to \infty.$$

Hence by (2)

$$b(T_1(\omega) + T_2(\omega) + \cdots + T_n(\omega), \omega) = O((n \log \log n)^{1/2}) \text{ a.s.}$$

The above gives the probability of a certain event whose probability

can be reduced to expressions involving the finite-dimensional joint distributions of the $b(T_1 + T_2 + \cdots + T_n)$. Hence we also have, by (1),

(4) $X_1(\omega) + X_2(\omega) + \cdots + X_n(\omega) = O((n \log \log n)^{1/2})$ a.s.

Thus we have established half of the so-called law of the iterated logarithm, using mainly the simple calculation (3). If the X_k have mean zero and a finite second moment, the expression on the left-hand side of (4) has (for fixed n) mean zero and mean deviation from zero of order $n^{1/2}$. The relation (4) is an analogous result which holds for fixed ω. As we will see below, the factor "$\log \log n$" cannot be removed.

The relation (3) can be sharpened to

$$\limsup_{t \to \infty} \frac{b(t, \omega)}{(2t \log \log t)^{1/2}} = 1 \quad \text{a.s.}$$

and we also have the estimates

$$\max_{0 \leq s \leq Tn} |b(s + U_{n-1}) - b(U_{n-1})| = \sigma((n)^{1/2}) \quad \text{a.s.}$$

where $U_n = T_1 + T_2 + \cdots + T_n$. This follows from the Borel-Cantelli lemma and the martingale inequality (Doob [6, p. 317, 376]):

$$E(\max_{0 \leq s \leq T} b(s, \omega)^2) \leq 4Eb(T(\omega))^2 < \infty$$

which holds whenever T is a stopping time with $E(T) < \infty$. Assuming $E(T_1) = 1$ in (2), we obtain as before (from (1))

$$\limsup_{n \to \infty} \frac{X_1(\omega) + X_2(\omega) + \cdots + X_n(\omega)}{(2n \log \log n)^{1/2}} = 1 \quad \text{a.s.}$$

which is the full law of the iterated logarithm.

The usual form of this theorem suggests that any random variable $X(\omega)$ with a representing stopping time $T(\omega)$ with $E(T) < \infty$ should satisfy $E(X) = 0$ and $E(X^2) = E(T) < \infty$; this is in fact correct. Conversely, any mean-zero finite-variance random variable $X(\omega)$ has such a stopping time $T(\omega)$, as we shall see in §4.

This is an argument apparently first due to Strassen [45]. See Breiman [3], Freedman [15, Ch. 1], Strassen [46, 47], and Stouts [44] for similar applications and generalizations.

With a little extra work, we obtain another type of theorem. Dividing (1) by $n^{1/2}$,

(5) $$\frac{X_1 + X_2 + \cdots + X_n}{n^{1/2}} \cong \frac{1}{n^{1/2}} b(T_1 + T_2 + \cdots + T_n).$$

Now, for any constant time T,

(6) $$\frac{1}{n^{1/2}} b(T) \cong b(T/n),$$

since both random variables in (6) are normally distributed with mean zero and variance T/n. If we could use this in (5), we would have a direct route from the strong law of large numbers for the $\{T_k\}$ to the Central Limit Theorem for the $\{X_k\}$.

However, (6) does not hold in general for non-constant T. For example, if $T(\omega)$ is the first exit time of $b(s, \omega)$ from the interval $(-1, +1)$, then $b(T(\omega))/n^{1/2} = \pm 1/n^{1/2}$, while $b(T(\omega)/n, \omega)$ will have some continuous distribution. We can, however, do the following. For each N, let $b'(s, \omega)$ be the Brownian motion $n^{1/2}b(s/n, \omega)$, and let $T'(\omega)$ be a stopping time for b' such that $b'(T'(\omega)) \cong b(T(\omega)) \cong X(\omega)$ and $T(\omega) \cong T'(\omega)$. Then since $b'(A)/n^{1/2} = b(A/n)$ for all $A > 0$

$$b(T)/n^{1/2} \cong b'(T')/n^{1/2} = b(T'/n).$$

Applying this technique in (5), and arguing as in (1), we obtain

(7) $$\frac{X_1 + X_2 + \cdots + X_n}{n^{1/2}} \cong b\left(\frac{T'_1 + T'_2 + \cdots + T'_n}{n}\right).$$

The difficulty now is that T'_k depends on n. However, for fixed n, the T'_k are independent with the same distribution as T_1, while by (2) (with $E(T_1) = 1$)

$$\lim_{n \to \infty} P\left[-\epsilon < \frac{T_1 + T_2 + \cdots + T_n}{n} < \epsilon\right] = 1$$

for all $\epsilon > 0$. Since this probability is insensitive as to whether we use T_k or T'_k, the limit is the same for either T_k or T'_k. Hence for all constants $\lambda, \delta > 0, \epsilon > 0$,

(8) $$P\left[b\left(\frac{T'_1 + T'_2 + \cdots + T'_n}{n}\right) \leq \lambda\right] \leq P[b(1) \leq \lambda + \delta] + E$$

$$\geq P[b(1) \leq \lambda - \delta] - E$$

where

$$E = P \left[\left| b \left(\frac{T'_1 + T'_2 + \cdots + T'_n}{n} \right) - b(1) \right| > \delta \right]$$

$$\leqq P[\max_{|s| \leqq \epsilon} |b(1 + s) - b(1)| > \delta]$$

$$+ P \left(\left| \frac{T'_1 + T'_2 + \cdots + T'_n}{n} - 1 \right| > \epsilon \right)$$

$$= A + B.$$

Now $b_1(s) = b(1 + s) - b(1)$, $b_2(s) = b(1) - b(1 - s)(0 < s \leqq 1)$, and $b_3(s) = b(s\epsilon)/\epsilon^{1/2}$ are all Brownian motions in the sense of § 1; hence

$$A \leqq 2P[\max_{s \leqq \epsilon} |b(s)| > \delta] = 2P(\max_{0 \leqq s \leqq 1} |b(s\epsilon)| > \delta)$$

$$= 2P(\max_{0 \leqq s \leqq 1} |b(s)| > \delta/\epsilon^{1/2}).$$

Since $b = b(s)$ is continuous in s a.s., $\max_{0 \leqq s \leqq 1}|b(s)| < \infty$ a.s., and also $P(b(1) \leqq \lambda \pm \delta) = P(b(1) \leqq \lambda) + O(\delta)$. Hence by (7) and (8)

$$\lim_{n \to \infty} P \left[\frac{X_1 + X_2 + \cdots + X_n}{n^{1/2}} < \lambda \right]$$

(9)
$$= \lim_{n \to \infty} P \left[b \left(\frac{T'_1 + T'_2 + \cdots + T'_n}{n} \right) \leqq \lambda \right]$$

$$= P(b(1) \leqq \lambda) = \frac{1}{(2\pi)^{1/2}} \int_{-\infty}^{\lambda} e^{-1/2x^2} dx$$

and we have proven the Central Limit Theorem.

Of course, one should not get too excited about a new proof of the Central Limit Theorem. However, one can get a powerful generalization of (9) very cheaply. Let $\Phi(f)$ be a functional defined for all *sample paths*:

$$s_n(t) = s(k, n), (k - 1)/n \leqq t < k/n, 0 \leqq t \leqq 1,$$

$$s(k, n) = (X_1 + X_2 + \cdots + X_k)/n^{1/2}, 1 \leqq k \leqq n + 1.$$

Specifically, $\Phi(f)$ is a measurable function defined on all piecewise continuous functions $f(s)$ on $[0, 1]$. Then, one can use the representation (7) much as before (see [2, Ch. 13], [15, p. 76]; the approach of F. Knight for coin-tossing random variables constructs the Brownian motion as a by-product) to prove:

THEOREM (DONSKER). *If $\Phi(f)$ is continuous with respect to uniform convergence at almost every Brownian path $f = b(s, \omega)$ $(0 \leqq s \leqq 1)$, and if the X_k are independent and identically distributed with mean zero and variance one, then*

$$(10) \qquad \lim_{n \to \infty} P[\Phi(s_n(\,\cdot\,)) \leqq \lambda] = W(\lambda) = P[\Phi(b(\,\cdot\,)) \leqq \lambda]$$

at every value of λ at which $W(\lambda)$ is continuous.

Examples of Φ for which Donsker's theorem would apply are

$\quad$ (i) $\Phi(f) = f(1),$

$\quad$ (ii) $\Phi(f) = 1 \qquad$ if $\max_{0 \leqq t \leqq 1} |f(t)| \leqq 1,$

$\qquad\qquad\quad = 0 \qquad$ otherwise,

$\quad$ (iii) $\Phi(f) = \max_{0 \leqq t \leqq 1} (f(t) - At),$

$(11) \quad$ (iv) $\Phi(f) = \displaystyle\int_0^1 f(s)^2 \, ds,$

$\quad$ (v) $\Phi(f) = \displaystyle\int_0^1 (f(s) - sf(1))^4 \, ds,$

$\quad$ (vi) $\Phi(f) = 1 \quad$ if $a(t) < f(t) < c(t)$ for $0 \leqq t \leqq 1,$

$\qquad\qquad\quad = 0 \qquad$ otherwise; $a(t), c(t)$ smooth,

$\quad$ (vii) $\Phi(f) = \text{meas } \{t : f(t) > 0\} - \text{meas } \{t : f(t) < 0\}.$

Example (i) above is the Central Limit Theorem; (ii, iii) give two versions of the Gambler's Ruin problem. Example (iv), combined with (10), gives the limit theorem

$$(12) \qquad \lim_{n \to \infty} P\left[\sum_1^n s_k^2 \leqq \lambda n^2 \right] = P\left[\int_0^1 b(s)^2 \, ds \leqq \lambda \right]$$

where $s_k = X_1 + X_2 + \cdots + X_k$, which was apparently first due to Kac and Erdös [11]. Example (vi) is the "Khinchin problem", which has motivated most of the Russian work in this area (Nagaev [28], Skorokhod [41], Prokhorov [31]), and has applications for sequential tests in statistics. The function $\Phi(f)$ here (and in (ii)) is indeed continuous at almost all Brownian paths; this amounts to the statement that a Brownian path $y = b(s, \omega)$ is tangent to one of the curves $y = a(s)$ or $y = c(s)$ with probability zero. Similarly continuity

in (vii) results from the fact $P[\text{meas}\{s : b(s, \omega) = 0\} = 0] = 1$; this example combined with (10) gives one of the "Arc-Sine" laws, so-called because the function $W(\lambda)$ in this case involves the Arc-Sine function ([12, 15]).

The term "Invariance Principle" is often used for Donsker's theorem; this refers to the fact that the limiting distribution in (10) or (12) does not depend on the specific distribution of the X_k. In the same sense, of course, the Central Limit Theorem and the strong law of large numbers are Invariance Principles.

Müller [26] has an alternate form of (10) where $C[0, 1]$ is replaced by the set D of all $x(s) \in C[0, \infty]$ for which $x(s) = \sigma(s)$ as $s \to \infty$. As an application, with the functional

$$\Phi(f) = \sup\{t : |f(t)| > At\},$$

he obtains a limit theorem giving the asymptotic distribution of the *last* integer k for which $|s_k/k| > A/n^{1/2}$. Note that by (3), almost all $f = b(s, \omega)$ are in the domain D of Müller's theorem.

3. **Rates of Convergence, and Applications.** While this representation technique was not the original (or even the second) proof of Donsker's theorem, it does have the advantage that it lends itself to explicit error estimates in (10). In fact, these "rate of convergence" results appear to be the main use of the Skorokhod representation at the present time.

For definiteness, assume that the functional $\Phi(f)$ in (10) satisfies

(13)
    (i) $|\Phi(f) - \Phi(g)| \leqq C \displaystyle\sup_{0 \leqq s \leqq 1} |f(s) - g(s)|$

    (ii) The distribution $W(\lambda) = P[\Phi(b(\ \cdot\ )) \leqq \lambda]$ has a bounded density.

The first rate of convergence result in this context was due to Prokhorov [31, §5], who showed

$$\text{(14)} \qquad \sup_{-\infty < \lambda < +\infty} |P[\Phi(s_n(\ \cdot\ )) \leqq \lambda] - W(\lambda)| = O\left(\frac{(\log n)^a}{n^{1/8}}\right)$$

for some $a > 0$, given $E(|X_k|^3) < \infty$. Prokhorov's result is actually stated as a rate of convergence for the Lévy distance between measures on $C[0, 1]$; see Dudley [8] for a survey of results in this more abstract context. Later Rosenkrantz [34] duplicated Prokhorov's result using the Skorokhod representation, and improved the error to $O((\log n)^{1/2}/n^{1/5})$ if $E(X_k^4) < \infty$. Rosenkrantz' method was to replace the estimate for E in (8) by

$$P \left[\max_k \left| b \left(\frac{T_1' + T_2' + \cdots + T_k'}{n} \right) - b \left(\frac{k}{n} \right) \right| > A \frac{(\log n)^{1/2}}{n^{1/5}} \right]$$

$$= O(1/n^{1/2}).$$

Subsequently several authors have improved this rate to

(15) $$O \left(\frac{(\log n)^a}{n^{1/4}} \right)$$

under various moment conditions on X_k (see Dudley [8, §5], Rosen-krantz (unpublished notes), and Fraser [14]; the first two assuming $E(|X_k|^5) < \infty$). Dudley conjectures that (15) is the best possible rate of convergence for the Lévy distance; it probably is also the best possible rate for a general $\Phi(f)$ satisfying (13) even for bounded X_k.

(*Added in Proof*: A recent preprint of three Hungarian mathematicians, Komlós, Major and Tusnády [56], obtains the rate $O((\log n)/n^{1/2})$ in (14) for $\{X_k\}$ with a moment-generating function and an L^p-density (some $p > 1$). Their method, however, is different from the one described here.)

It is clear that the direct technique of §2 cannot be used to improve the power of n in (15). For, if the T'_k are as in (7) (and §4), one can show (Sawyer [38])

$$\lim_{n \to \infty} P \left[b \left(\frac{T'_1 + T'_2 + \cdots + T'_n}{n} \right) - b(1) \leq \lambda c / n^{1/4} \right] = G(\lambda)$$

for a non-degenerate distribution $G(\lambda)$, where $c \cong E(X_k^4)^{1/4}$. Thus, in (8), if $\delta = \sigma(1/n^{1/4})$, $E \to 1$.

For a general Φ in (10), or even for the Central Limit Theorem, $O(1/n^{1/2})$ is a definite upper bound for the rate of convergence (for a general X_k). For, if the X_k are coin-tossing random variables (i.e., $P(X_k = 1) = P(X_k = -1) = \frac{1}{2}$), $P(s_n(1) \leq \lambda)$ has a jump at $\lambda = 0$ of magnitude

$$\binom{2n}{n}(1/2^{2n}) \sim 1/(\pi n)^{1/2}$$

and so cannot be (uniformly) closer to $P[b(1) \leq \lambda]$ than $1/2(\pi n)^{1/2}$. If $E(|X_k|^3) < \infty$, however, the Berry-Esséen theorem (Feller [12, p. 515]) does yield a uniform rate of convergence of $O(1/n^{1/2})$ in (9), which is then best possible. Nagaev [28] gives the same uniform rate for the Gambler's Ruin and Khinchin problems (see [11] ii, iii, iv), also if $E(|X_k|^3) < \infty$. The work of Nagaev is considerably more difficult and runs to over 40 pages in print.

There are other classes of Φ for which one can get sharper results than (15) using the Skorokhod representation. Assume the X_k have mean zero, variance one, and satisfy

$$(16) \qquad\qquad E(X_k^\sigma) < \infty .$$

Then, for a class of Φ of the form $\Phi(f) = \int_0^1 H(s, f(s))\, ds$ one obtains

$$(17) \qquad\qquad O\left(\frac{(\log n)^a}{n^{1/2}}\right), \quad a = a(\alpha),$$

as a uniform error in the Invariance Principle (10) (see [38], [52]). In particular, given (16), this rate prevails in (12) ($H(s, x) = x^2$). Recently Nienstadt [52] has obtained a similar result for the stochastic integral. For $\{X_k\}$ satisfying (16) and $f(x) \in C^1(R)$ of slow growth, he obtains the uniform rate (17) for the convergence of the distribution function of

$$\sum_1^n \frac{f[s(k, n)] + f[s(k + 1, n)]}{2} \frac{X_{k+1}}{n^{1/2}}$$

to its limit, where $s(k, n)$ is as before and the limiting distribution has a bounded density. (If $f(x) \geqq \epsilon > 0$, this is automatic.) The same techniques of proof also go through for the Khinchin problem, obtaining of course a weaker result than Nagaev's.

There have been many applications of the Skorokhod representation to problems in statistics, often non-parametric problems connected with the sample distribution function (sample d.f.). Here, if $\{X_1, X_2, \cdots, X_n\}$ are n independent observations of a source which is assumed to be uniformly distributed in $[0, 1]$, the sample d.f. is

$$F_n(s, \omega) = (1/n)\ \mathrm{card}\ \{k : X_k(\omega) \leqq s\}.$$

which defines a distribution function for each ω. The "order statistics" are the jumps of $F_n(s, \omega)$; i.e., the X_k arranged according to size. Historically, one of the first uses of the Invariance Principle (see [2]) was to obtain the limiting distribution of $\max_{0 \leqq s \leqq 1} n^{1/2}|F_n(s, \omega) - s|$. Using the Skorokhod representation, Rosenkrantz [35] (see also Müller [27], [8, §7]) obtained an error estimate for the limiting distribution of $n \int_0^1 (F_n(s, \omega) - s)^2\, ds$; and later, with O'Reilly [30], a similar rate of convergence for linear combinations of order statistics which extend some results of Shorack [40]. Kiefer [20] studied the sample quantile processes (the inverse of $F_n(s, \omega)$) using the Skorokhod representation, and later [21] joint distributions (over various n) of the sample d.f. Specifically, he obtained an embedding

$$t(F_{[t]}(s, \omega) - s) \cong A(s, T(s, t, \omega))(\omega)$$

where $A(s, t)$ is a particular Gaussian process with a two-dimensional time discovered by Müller ([27], [8]). For large t, $T(s, t, \omega) \cong t$, and rate of convergence theorems of the form

$$\lim_{n \to \infty} n^{1/2} t(F_{[nt]}(s, \omega) - s) \cong A(s, t)(0 \leqq s, t \leqq 1)$$

(using the fact $A(s, tT) \cong T^{1/2}A(s, t)$ for constant $T > 0$) follow. (See also [27]).

See Brillinger [4], Dudley [8], Hall [49], and the bibliography of Kiefer [21] for other applications in statistics.

Another application of these methods is a uniform estimate of the difference between a (classical) harmonic function and the corresponding function with the same boundary values for a random walk, see Fraser [13]. Kennedy [50, 51] applies the Skorokhod representation to obtain a diffusion approximation for "queues in heavy traffic", which denotes a queue with average service time very close to the average interarrival time, considered over a long period of time. The approximation is to a Brownian motion with constant drift. Other classes of Invariance Principles can be found in the book [43]; see also Griego et al. [48]. For some more classical applications, see Skorokhod [41, Ch. 7] and Sawyer [37].

The examples given so far have been applications of embeddings of partial sums of independent random variables in Brownian motion. The same techniques work for martingales; Heyde and Brown [18] have an inequality which can be used to generalize (14) to discrete parameter martingales. Any right-continuous martingale can be embedded in Brownian motion using a right-continuous process of stopping times; also, one can embed random variables (given a moment condition) in stable processes as well as in Brownian motion (see Monroe [24, 25]). Hall [49] embeds certain submartingales in a Brownian motion with constant drift and gives applications to Wald's sequential ratio probability test (SPRT) in statistics. There is also an application of the Skorokhod representation for Markov chains; Skorokhod [42] uses his embedding for martingales to prove a Khinchin-like theorem for Markov chains converging to a diffusion process. Skorokhod's approach seems to be very promising but apparently has never been followed up.

For recent Russian work in this area, see [53]–[55], in particular [54], and their references.

4. **The embedding.** We now consider the second question raised

in the introduction; namely, for an arbitrary random variable $X(\omega)$ with mean zero and finite variance, to find a stopping time $T(\omega)$ such that $b(T(\omega), \omega) \cong X(\omega)$. There are actually many ways of doing this. For example, one can always set $b(T(\omega), \omega) = X'(\omega)$, where $X'(\omega)$ and $b(s)$ are independent, $T(\omega) = \inf\{s : b(s, \omega) = X'(\omega)\}$ and $X'(\omega) \cong X(\omega)$ [17]. But then $E(T) = \infty$ whenever $E(|X|) > 0$, and this representation is too singular to be useful.

The first reasonable representation, historically, was devised by Skorokhod [41] for $X(\omega)$ with a continuous distribution; i.e., such that $F(\lambda) = P[X(\omega) \leq \lambda]$ is a continuous function of λ. His representation can be extended to $X(\omega)$ with a general distribution in any of several ways; see Breiman [2], Freedman [15], or Sawyer [39]. The simplest representation for a general $X(\omega)$ seems to be one due to Breiman [3], and is as follows. Let $X(\omega)$ be a random variable with $E(|X|) < \infty$, $E(X) = 0$, and $q^{-1} = \frac{1}{2} E(|X|) > 0$. We now define a bivariate distribution in the plane, concentrated in the second quadrant, by

$$G(du \; x \; dv) = q(v - u)\chi_{[u \leq 0 < v]}(u, v)F(du)F(dv)$$

where $F(du) = P[X(\omega) \in du]$ and $\chi_{[u \leq 0 < v]}(u, v)$ is the characteristic function. It is easily checked that $\iint G(du \times dv) = 1$, and hence G does define a probability distribution in the plane. Let $b(s, \omega)$ be a Brownian motion in the sense of §1, and let (U, V) be a random vector with distribution

$$P[(U, V) \in du \times dv] = G(du \times dv)$$

which is independent of $\{b(s, \omega) : 0 \leq s < \infty \}$. If need be, (U, V) can be defined by taking the Cartesian product of the probability space of $b(s, \omega)$ with a copy of the plane and taking a product measure on this Cartesian product (i.e., by "extending" the probability space). Finally, for $-\infty < c \leq 0 < d < \infty$, set

(18)
$$T(c, d)(\omega) = \inf\{s : b(s, \omega) \in (-\infty, c] \cup [d, \infty)\}$$
$$T(\omega) = T(U, V)(\omega).$$

Thus $T(\omega)$ is a randomized $T(c, d)$. I now claim $b(T(\omega)) \cong X(\omega)$; i.e. $P(b(T(\omega)) \leq \lambda) = F(\lambda)$ for all λ. First, note that for all ω $b[T(c, d)(\omega), \omega] = c$ or d, and ([2], [41])

$$P[b[T(c, d)] = c] = \frac{d}{d - c}, P[b[T(c, d)] = d] = \frac{|c|}{d - c}.$$

Then, if $\lambda < 0$, by conditioning on (U, V),

$$P[b(T(\omega)) \leqq \lambda] = \int\int P[b[T(u, v)] \leqq \lambda]\, G(du\, dv)$$

$$= \int_0^\infty \int_{-\infty}^\lambda P[b(T(u, v)) = u]\, q(v - u) F(du) F(dv)$$

$$= q \int_{-\infty}^\lambda \int_0^\infty \frac{v}{v - u} (v - u) F(dv) F(du)$$

$$= q \int_0^\infty v F(dv) F(\lambda) = F(\lambda)$$

since $\int_0^\infty v F(dv) = E(X^+) = \frac{1}{2} E(|X|)$. The same calculation also goes through if $\lambda \geqq 0$, and hence $b(T(\omega)) \cong X(\omega)$. Since $E[T(c, d)] = |cd|$, we have by a similar calculation

$$E(T) = E(-UV) = \int_{-\infty}^{+\infty} v^2 F(dv) = E(X^2).$$

The representation (18) has the disadvantage that it requires "external randomization" (and possible extension of the probability space) at each step. Thus, the event $\{\omega : T(\omega) \leqq t\}$ depends on (U, V) as well as $b(s, \omega)$, and, in (1), $\{\omega : T_1(\omega) + T_2(\omega) + \cdots + T_n(\omega) \leqq t\}$ depends on not only $b(s, \omega)(0 \leqq s \leqq t)$, but on n outside random variables (U_k, V_k) as well. Skorokhod's original representation was also of this form. It is of course conceptually simpler to have $T(\omega)$ depend only on $b(s, \omega)$ and not on outside random variables. This can also be done, in at least three different ways (Dubins [7], Root [32], Monroe [24]); Hall [17] also has a way of converting any representation of type (18) into one in which $T(\omega)$ depends only on $b(s, \omega)$ by a process he calls "internal randomization". However all of these representations are definitely less explicit than (18), which makes it more difficult to obtain estimates on the $T(\omega)$.

The proofs of most of the theorems in §3 involving the Skorokhod representation depend on estimating the moments of T in terms of X^2; i.e., on inequalities

(19) $$\alpha_p E(|b(T)|^{2p}) \leqq E(T^p) \leqq A_p E(|b(T)|^{2p})$$

(in particular the upper estimates) for $p > \frac{1}{2}$ and

(20) $$A_p = O((cp)^p) \quad \text{as } p \to \infty, \text{ some } c > 0.$$

These estimates can be obtained fairly easily for $T(\omega)$ of (18) (see [34], [39]). Millar [23] has obtained (19) for a general stopping time $T(\omega)$ of Brownian motion under the assumption

(21) $$E(T^p) < \infty.$$

Some condition on $T(\omega)$ is necessary here, as the stopping time $T(1) =$ inf$\{s : b(s) = 1\}$ indicates; it can be dispensed with in (18) since in that case $T(\omega)$ is a randomized $T(c, d)$ for which (21) holds for all p. Later methods of derivation of (19) show that (20) holds for a general $T(\omega)$ satisfying (21) with $c = 1$ ([1], [16], [36]; [1] assumes only $E(T) < \infty$).

All of the obvious counterexamples of (19) require $T(\omega)$ involving one-sided boundaries for Brownian motion, and hence $T(\omega)$ for which $E(T(\omega)^{1/2}) = \infty$. It is interesting that the simple condition

(22) $$E(T^{1/2}) < \infty$$

by itself is sufficient for (19). For, (22) implies that the martingale $\{b(T(\omega) \wedge t)\}$ is uniformly integrable (Burkholder and Gundy [5]; here $s \wedge t = \min\{s, t\}$). Hence by a standard result for uniformly integrable martingales (Doob [6, p. 324])

$$E(|b(T(\omega) \wedge t)|^a) \leqq E(|b(T(\omega))|^a), \ a \geqq 1$$

(which follows from Jensen's inequality). Hence by (19) for bounded $T(\omega)$,

$$E[(T(\omega) \wedge t)^p] \leqq A_p E(|b(T(\omega) \wedge t)|^{2p}) \leqq A_p E(|b(T(\omega))|^{2p}).$$

Letting $t \to \infty$, we obtain the upper inequality of (19), given $E(T^{1/2}) < \infty$; the lower inequality follows from Fatou's theorem. Other proofs of the sufficiency of (22) are due to Louis Gordon and Don Burkholder; the above is a modification of Burkholder's proof.

For generalization of (19) for Brownian motion and arbitrary martingales, see Burkholder and Gundy [5] and also Rosencrans and Sawyer, [33]. Athreya and Kurtz [1] prove (19) for a class of processes with stationary independent increments (with $E(T^p)$ replaced by $E(T^p) \wedge E(T)$ in the lower inequality).

Other properties of $T(\omega)$ can also be used for rate of convergence results as in §3. First, we remark that the stopping time $T(\omega)$ of (18) usually does not have a moment-generating function. For, if $X(\omega)$ is, e.g., symmetric unbounded, then for any finite M

$$P[T(\omega) \geqq t] \geqq \rho P[T(-M, M) \geqq t] \sim C \exp(-ct/M^2) \text{ as } t \to \infty$$

(see equation (27) below) and $E(\exp(\beta T)) = \infty$ for every $\beta < 0$. On the other hand, for any $a < \infty$, the condition

(23) $$E(\exp(\alpha|X|^a)) < \infty, \text{ some } \alpha > 0,$$

is necessary and sufficient for

(24) $\qquad E(\exp(\beta T^\delta)) < \infty$, some $\beta > 0$,

where $\delta = a/(2 + a)$ (Sawyer [39]). This result, in combination with the inequalities in [33], can be used to provide an alternate proof of the results mentioned after inequality (16) in §3.

Half of the equivalence of (23) and (24) above follows from (19) and (20) and hence holds for an arbitrary stopping time $(b(T) \cong X)$; whether the other direction is valid if, for example, T is unbounded or $a \gtreqqless 2$ (or some appropriate conditions are imposed) is an open question.

If $X(\omega)$ is semi-bounded, one can prove an asymptotic form of this equivalence (see also [39]).

THEOREM. *Assume* $X(\omega) \gtreqqless -M$, *and let* $T(\omega)$ *be given by* (18). *Assume also*

(25) $\qquad P[X(\omega) \gtreqqless t] \sim Ct^c \exp(-\alpha t^a)$ *as* $t \to \infty$

for C, α *positive and some* $a \gtreqqless 0$. *Let* $\delta = a/(2 + a)$ *and* $n = [a]$ *(the greatest integer function). Then*

(26) $P[T(\omega) \gtreqqless s] \sim C's^d \exp(-\beta s^\delta) \int_{-M}^{0} \exp\left(\sum_{1}^{n} C_k(-u)^k s^{\delta_k} \right) F(du)$

where $F(\lambda) = P(X \lesseqqgtr \lambda)$, $d = \frac{1}{2}\delta + (1 - \delta)c/2$, $\delta_k = (a - k)/(2 + a)$, $\beta > 0$, *and* C_k *are constants. (The integral does not appear if* $a \lesseqqgtr 1$*).*

EXAMPLE. If $X = Y - 1$, where $P(Y > t) = e^{-t}$, then

$$P(T(\omega) \gtreqqless s) \sim Cs^{1/6} \exp(-\beta s^{1/3}).$$

For $a = 0$ (in which case $c < -1$, since $E(|X|) < \infty$) the theorem can also be obtained from Monroe [25, Theorem 12].

PROOF. For all c, d, (Feller [12, p. 330])

(27) $\quad P[T(c, d) \gtreqqless s] = \dfrac{4}{\pi} \sum_{0}^{\infty} \dfrac{\sin(2n + 1)B}{2n + 1} \exp\left(- \dfrac{(2n + 1)^2\pi^2 s}{2(d - c)^2} \right)$

where $B = |c|\pi/(d - c)$. Hence by (18)

$$P[T(\omega) \gtreqqless s] = \sum_{0}^{\infty} I_n(s),$$

where

$$I_n(s) = q\frac{49}{\pi}\int_{-M}^{0}\int_{0}^{\infty}\frac{\sin(2n+1)B}{2n+1}(v-u)\exp(-A/(v-u)^2)F(dv)F(du)$$

and $A = (2n+1)^2\pi^2/2$. Now, any region $0 \leqq v \leqq C$ contributes at most exponentially to the integral $I_n(s)$, and

$$(28) \qquad I_n(s) \sim q\int_{-M}^{0}|u|\int_{0}^{\infty}\exp(-A/(v-u)^2)F(dv)F(du)$$

where the right-hand side of (28) is also an upper bound for $I_n(s)$. Computing the integrand

$$J(u, A) = \int_{0}^{\infty}\exp(-A/(v-u)^2)F(dv) = E(e^{-AY}\chi_{[X>0]})$$

$$= A\int_{0}^{\infty}e^{-Ay}P(Y < y, X > 0)\, dy$$

where $Y = 1/(X + |u|)^2$. Hence

$$(29) \qquad J(u, A) = A\int_{0}^{\infty}e^{-Ay}P(X > y^{-1/2} - |u|, X > 0)\, dy.$$

On the other hand, by (25), for large t,

$$(30) \qquad \begin{aligned} P[X \geqq t + u] &\sim P[X \geqq t]\exp(-\alpha t^a[(1 + u/t)^a - 1])\\ &\sim P[X \geqq t]\exp(-\alpha t^a[au/t + \cdots + C_n u^n/t^n]) \end{aligned}$$

since $n = [a]$. Thus if $a < 1$,

$$P[X \geqq t + u] \sim P[X \geqq t]$$

uniformly for bounded u, and as $A \to \infty$

$$I_n(s) \sim A\int_{0}^{\infty}e^{-Ay}P[X \geqq y^{-1/2}]\, dy$$

$$(31) \qquad \sim CA\int_{0}^{\infty}e^{-Ay}\exp(-\alpha y^{-a/2})y^{-c/2}\, dy$$

$$= CA^{d+\delta/2}\int_{0}^{\infty}[e^{-z}\exp(-\alpha z^{-a/2})]^{A^\delta}z^{-c/2}\, dz$$

where $z = yA^{1-\delta}$. Thus by Laplace's method (Erdélyi [10])

$$(32) \qquad I_n(s) \sim CA^d\exp(-\beta A^\delta)$$

as $A = (2n + 1)^2\pi^2 s/2 \to \infty$. Now, the expression on the right-hand side of (28) depends only on $A \cong n^2 s$ and is an upper bound for $I_n(s)$; hence the series $\sum I_n(s)$ converges uniformly and the asymptotic relation (32) can be summed. The first term of the series is dominant asymptotically, and we obtain (26) (for $a < 1$). If $a \geqq 1$, we substitute (30) into (29) and proceed as before as in (31); the relation (26) in this case also results from Laplace's method although the calculations become more tedious.

References

1. K. Athreya and T. Kurtz, *A Generalization of Dynkin's Identity and Some Applications*, Annals of Prob. 1 (1973), 570–579.

2. L. Breiman, *Probability*, Addison-Wesley, Reading, Mass., 1968.

3. ———, *On the Tail Behavior of Sums of Independent Random Variables*, Z. Wahrscheinlichkeitstheorie und Verw. Gebiete 9 (1967), 20–25.

4. D. Brillinger, *An Asymptotic Representation of the Sample Distribution Function*, Bull. Amer. Math. Soc. 75 (1969), 545–547.

5. D. Burkholder and R. Gundy, *Extrapolation and Interpolation of Quasi-linear Operators on Martingales*, Acta Math. 124 (1970), 249–304.

6. J. Doob, *Stochastic Processes*, J. Wiley & Sons, New York, 1953.

7. L. Dubins, *On a Theorem of Skorokhod*, Annals of Math. Statist. 39 (1968), 2094–2097.

8. R. Dudley, *Speeds of Metric Probability Convergence*, Z. Wahrscheinlichkeitstheorie und Verw. Gebiete 22 (1972), 323–332.

9. A. Einstein, *Investigations on the Theory of the Brownian Movement*, R. Fürth, ed., Dover, New York, 1956.

10. A. Erdélyi, *Asymptotic Expansions*, Dover, New York, 1956.

11. P. Erdös and M. Kac, *On Certain Limit Theorems of the Theory of Probability*, Bull. Amer. Math. Soc. 52 (1946), 292–302.

12. W. Feller, *An Introduction to Probability Theory and Its Applications*, Volume II, John Wiley and Sons, New York, 1966.

13. D. Fraser, *Rates of Convergence for a Probabilistic Solution to the Dirichlet Problem*, Pacific Jour. Math 42 (1972), 657–665.

14. ———, *The Rate of Convergence of a Random Walk to Brownian Motion*, Annals of Prob. 1 (1973), 699–701.

15. D. Freedman, *Brownian Motion and Diffusion*, Holden-Day, San Francisco, 1971.

16. L. Gordon, *An Equivalent to the Martingale Square Function Inequality*, Annals of Math. Statist. 43 (1972), 1927–1934.

17. W. Hall, *On the Skorokhod Embedding Theorem*, Technical Report (1968), Dept. of Statistics, Stanford University.

18. C. Heyde and B. Brown, *On the Departure From Normality of a Certain Class of Martingales*, Annals of Math. Statist. 41 (1970), 2161–2165.

19. K. Ito and H. P. McKean, Jr., *Diffusion Processes and Their Sample Paths*, Springer-Verlag, Berlin, 1965.

20. J. Kiefer, *Old and New Methods for Studying Order Statistics and Sample Quantiles*, in Nonparametric Techniques in Statistical Inference, M. L. Puri, ed., Cambridge Univ. Press, Cambridge, 1970.

21. ——, *Skorokhod Embeddings for Multi-variate Random Variables, and the Sample DF*, Z. Wahrscheinlichkeitstheorie und Verw. Gebiete **24** (1972), 1–35.

22. F. Knight, *On the Random Walk and Brownian Motion*, Trans. Amer. Math. Soc. **103** (1962), 218–228.

23. P. Millar, *Martingale Integrals*, Trans. Amer. Math. Soc. **133** (1968), 145–166.

24. I. Monroe, *Using Additive Functionals to Embed Preassigned Distributions in Symmetric Stable Processes*, Trans. Amer. Math. Soc. **163** (1972), 131–146.

25. ——, *On Embedding Right Continuous Martingales In Brownian Motion*, Annals of Math. Statist. **43** (1972), 1293–1311.

26. D. Müller, *Verteilungs-Invarianzprinzipien für das Starke Gesetz der Grossen Zahl*, Z. Wahrscheinlichkeitstheorie und Verw. Gebiete **10** (1968), 173–192.

27. ——, *On Glivenko-Cantelli Convergence*, Z. Wahrscheinlichkeitstheorie und Verw. Gebiete **16** (1970), 195–210.

28. S. V. Nagaev, *On the Speed of Convergence in a Boundary Problem I, II*, Theory of Prob. and its Appl. **15** (1970), 163–186 and 419–441. (Russian Translation).

29. E. Nelson, *Dynamical Theories of Brownian Motion*, Princeton University Press, Princeton, New Jersey, 1967.

30. N. O'Reilly and W. Rosenkrantz, *Application of the Skorokhod Representation to Rates of Convergence for Linear Combinations of Order Statistics*, Annals of Math. Statist. **43** (1972), 1204–1212.

31. Yu. Prokhorov, *Convergence of Random Processes and Limit Theorems in Probability Theory*, Theory Prob. Appl. **1** (1956), 157–214 (Russian Translation).

32. H. D. Root, *The Existence of Certain Stopping Times on Brownian Motion*, Annals of Math. Statist. **40** (1969), 715–718.

33. S. Rosencrans and S. Sawyer, *An Extremal Property of Independent Random Variables*, Proc. Amer. Math. Soc. **36** (1972), 552–556.

34. W. Rosenkrantz, *On Rates of Convergence for the Invariance Principle*, Trans. Amer. Math. Soc. **129** (1967), 542–552.

35. ——, *A Rate of Convergence for the von Mises Statistic*, Trans. Amer. Math. Soc. **139** (1969), 329–337.

36. W. Rosenkrantz and S. Sawyer, *An Elementary Derivation of the Moment Inequalities for Skorokhod Stopping Times*, unpub. manuscript; see Notices Amer. Math. Soc. **19** (1972), pA226. (And also Root [32]).

37. S. Sawyer, *A Uniform Rate of Convergence for the Maximum Absolute Value of Partial Sums in Probability*, Comm. Pure Appl. Math. **20** (1967), 647–658.

38. ——, *Rates of Convergence for Some Functionals in Probability*, Annals of Math. Statist. **43** (1972), 273–284.

39. ——, *A Remark on the Skorokhod Representation*, Z. Wahrscheinlichkeitstheorie und Verw. Gebiete **23** (1972), 67–74.

40. G. Shorack, *Asymptotic Normality of Linear Combinations of Order Statistics*, Annals of Math. Statist. **40** (1969), 2041–2050.

41. A. Skorokhod, *Studies in the Theory of Random Processes*, Addison-Wesley, Reading, Mass., 1965, Chapter 7. (Russian Translation; original publ. Kiev University 1961).

42. ——, *A Limit Theorem for Homogeneous Markov Chains*, Theory Prob. Appl. **8** (1963), 67–75. (Russian Translation).

43. A. Skorokhod and N. Slobodenyuk, *Limit Theorems for Random Walks*, Naukova Dumka, Kiev, 1970. (In Russian).

44. W. Stouts, *The Hartman-Wintner Law of the Iterated Logarithm for Martingales*, Annals of Math. Statist. 41 (1970), 2158–2160.

45. V. Strassen, *An Invariance Principle for the Law of the Iterated Logarithm*, Z. Wahrscheinlichkeitstheorie und Verw. Gebiete 3 (1964), 211–226.

46. ――――, *A Converse to the Law of the Iterated Logarithm*, Z. Wahrscheinlichkeitstheorie und Verw. Gebiete 4 (1965), 265–268.

47.――――, *Almost Sure Behavior of Sums of Independent Random Variables and Martingales*, Proc. Fifth Berkeley Symposium, Vol. II, Part 1 (1967), 315–343.

48. R. Griego, D. Heath, and A. Ruiz-Moncayo, *Almost sure convergence of Uniform Transport Processes to Brownian motion*, Annals of Math. Statist. 42 (1971), 1129–1131.

49. W. J. Hall, *Embedding submartingales in Wiener processes with drift, with Applications to Sequential Analysis*, Jour. Applied Prob. 6 (1969), 612–632.

50. D. P. Kennedy, *Estimates of the Rates of Convergence in Limit Theorems for the First Passage Times of Random Walks*, Annals of Math. Statist. 43 (1972), 2090–2094.

51. ――――, *Rates of Convergence for Queues in Heavy Traffic I, II*, Adv. Applied Prob. 4 (1972), 357–391.

52. H. Nienstadt, *Rates of Convergence for Approximations to Stochastic Integrals*, Thesis, Yeshiva University, 1974.

53. T. V. Arak and V. B. Nevzorov, *Some Estimates for the Maximum Cumulative Sum of Independent Random Variables*, Theory Prob. Appl. 18 (1973), 402–405. (English translation.)

54. A. A. Borovkov, *On Rates of Convergence in the Invariance Principle*, Theory Prob. Appl. 18 (1973), 217–234. (English translation.)

55. A. A. Novikov, *On Stopping Times for the Wiener Process*, Theory Prob. Appl. 16 (1971), 458–465. (English translation.)

56. J. Komlós, P. Major and G. Tusnády, *An Approximation of Partial Sums of Independent RV's, and the Sample DF, I*, Preprint of the Mathematical Institute of the Hungarian Academy of Sciences, Budapest, 1974.

BELFER GRADUATE SCHOOL, YESHIVA UNIVERSITY, NEW YORK, N.Y. 10033.