The Annals of Applied Probability
1992, Vol. 2, No. 1, 180–198

THE ASYMPTOTIC VALIDITY OF SEQUENTIAL STOPPING RULES FOR STOCHASTIC SIMULATIONS

BY PETER W. GLYNN[1] AND WARD WHITT

Stanford University and AT & T Bell Laboratories

We establish general conditions for the asymptotic validity of sequential stopping rules to achieve fixed-volume confidence sets for simulation estimators of vector-valued parameters. The asymptotic validity occurs as the prescribed volume of the confidence set approaches 0. There are two requirements: a functional central limit theorem for the estimation process and strong consistency (with-probability-1 convergence) for the variance or "scaling matrix" estimator. Applications are given for: sample means of i.i.d. random variables and random vectors, nonlinear functions of such sample means, jackknifing, Kiefer–Wolfowitz and Robbins–Monro stochastic approximation and both regenerative and nonregenerative steady-state simulation.

1. Introduction. The runlength of a stochastic simulation is typically determined by one of two methods. The first method is to assign the runlength prior to performing the simulation. This is usually done by specifying either the amount of simulation time to be generated or the amount of computer time to be expended. The principal disadvantage of this approach is that the posterior precision of the estimator may not be appropriate. Since the volume of the confidence set (the width of a confidence interval in one dimension) is unknown in advance, the volume may be too large to be of practical use (meaning that the preassigned runlength was too small) or too small (meaning that computational resources were wasted in refining the estimator beyond the level of accuracy required).

The second method is a sequential stopping procedure; that is, we let the simulation run until the volume of a confidence set achieves a prescribed value. This avoids the problems associated with preassigned runlengths, but new difficulties are introduced because the runlength is now randomly determined. The first difficulty is that we no longer have direct control of the amount of simulation time to be generated or the amount of computer time to be expended. Consequently, the runlength may turn out to be much longer than we want. On the other hand, it is possible that the runlength may turn out to be inappropriately short. This creates certain statistical difficulties that

---

Received December 1989; revised January 1991.

[1]Research supported by the U.S. Army Research Office under contract DAAL-03-88-K-0063 and by a Natural Sciences and Engineering Research Council of Canada grant.

AMS 1980 *subject classifications.* Primary 68U20, 62L12; secondary 65C05, 62L15, 60G40, 60F17.

Key words and phrases. Stochastic simulation, sequential estimation, fixed-width confidence intervals, sequential stopping rules, functional central limit theorems, strong consistency, variance estimators.

can compromise the accuracy of such procedures. For example, it is known that in many statistical settings, the point estimator and the variance estimator are positively correlated. Since the volume of a confidence set is typically determined by the variance estimator, this suggests that the confidence set volume will tend to be small when the point estimator is small. Consequently, the resulting sequential procedure will tend to terminate early in situations in which the point estimator is too small, leading to possibly significant coverage problems for the confidence sets. Nevertheless, sequential stopping rules are of interest because of the possibility of automatically obtaining prescribed precision.

Various sequential stopping rules for simulation estimators have been proposed and investigated empirically. Among these are sequential procedures involving: batch means in Law and Carson (1979) and Law and Kelton (1982), regenerative simulation in Fishman (1977) and Lavenberg and Sauer (1977) and spectral methods in Heidelberger and Welch (1981a, b, 1983); see pages 81, 92, 97 and 103 of Bratley, Fox and Schrage (1987) for an overview. Unfortunately, however, the empirical evidence is not entirely encouraging. Evidently, care must be taken in the design and implementation of sequential procedures to avoid inappropriate early termination. On the positive side, the sequential procedures do tend to perform well when the run lengths are relatively long, which is achieved in part by having a suitably small prescribed volume for the confidence set.

The observed good performance with small prescribed confidence set volumes is consistent with the classical asymptotic theory of sequential procedures for obtaining fixed-width confidence intervals for the mean of independent and identically distributed (i.i.d.) real-valued random variables; see Anscombe (1952, 1953), Chow and Robbins (1965), Starr (1966), Nadas (1969), Chapter 7 of Siegmund (1985), Section 8.8 of Wetherill and Glazebrook (1986) and Chapter 5 of Govindarajulu (1987). This asymptotic theory establishes that the sequential procedure is indeed asymptotically valid as the prescribed width of the confidence interval approaches 0 (and the resulting run length approaches ∞). This classical asymptotic theory provides a theoretical basis for confidence in sequential procedures, but it is not directly relevant to most simulation estimators, because the classical asymptotic theory is for i.i.d. random variables. The classical theory does apply relatively directly to regenerative simulations, as was shown by Lavenberg and Sauer (1977), but there evidently is not yet any asymptotic theory for nonregenerative steady-state simulation estimators.

The purpose of this paper is to fill the gap. We provide general conditions for the asymptotic validity of sequential stopping rules for a large class of simulation estimators. The main conditions are that the estimation process obey a functional central limit theorem (FCLT) and that there be a strongly consistent estimator for the asymptotic variance of the estimator. (We also treat d-dimensional parameters; then the asymptotic variance should be replaced by an asymptotic covariance matrix or, equivalently, by an associated "scaling" matrix; see Section 2.) Alternatively, for the variance estimator it suffices for it to satisfy a functional weak law of large numbers (FWLLN), which is often obtained as a consequence of an FCLT. The strong consistency (w.p.1 convergence) or the FWLLN for the variance estimator are important; we provide a counterexample in Section 4 showing that asymptotic validity need not hold with only weak consistency (ordinary one-dimensional in-probability convergence). Indeed, the conditions here are natural for the random-time-change limit theorems upon which the proofs depend; see, for example, Richter (1965), Section 17 of Billingsley (1968), Sections 3 and 5 of Whitt (1980), Glynn and Whitt (1988), and Gut (1988).

The rest of this paper is organized as follows. Section 2 provides limit theorems which guarantee that the coverage of a sequential procedure converges to the desired level when the prescribed volume of the confidence set is shrunk to 0. Section 3 contains applications of the limit theorems to various estimation settings. We give conditions under which sequential stopping rules are valid for a variety of estimation problems not previously considered: estimation of a nonlinear function of means, nonregenerative steady-state simulation and jackknife estimators. Section 4 contains the counterexample when the variance estimator is only weakly consistent. Finally, Section 5 contains all proofs.

2. The framework and main limit theorems. Our goal is to estimate a parameter $\alpha \in \mathbb{R}^d$. We assume that there exists an $\mathbb{R}^d$-valued stochastic process $Y \equiv \{Y(t)\colon t \geq 0\}$ called the *estimation process* for which $Y(t) \Rightarrow \alpha$ as $t \to \infty$, where $\Rightarrow$ denotes convergence in distribution (which here coincides with convergence in probability because α is deterministic). Actually, we shall need to require that the estimation process satisfy a stronger hypothesis, in particular, a functional central limit theorem (FCLT); in most applications, it will effectively amount to assuming that the estimation process satisfies an ordinary central limit theorem (CLT).

Let $D(0,\infty)$ be the space of right-continuous $\mathbb{R}^d$-valued functions with left limits on the open interval $(0,\infty)$, endowed with the standard J_1 topology; see Ethier and Kurtz (1986) or Whitt (1980). We work with $D(0,\infty)$ rather than $D[0,\infty)$ in order to avoid having to deal with possible singularities in Y at the origin $t = 0$; for example, in estimators such as $Y(t) = t^{-1}\int_0^t Z(s)\,ds$. At continuous limits, convergence in $D(0,\infty)$ is equivalent to uniform convergence over compact subintervals of $(0,\infty)$. We assume that Y has sample paths in $D(0,\infty)$ and consider the family of scaled processes

$$\mathscr{Y}_\varepsilon(t) = \varepsilon^{-\gamma}(Y(t/\varepsilon) - \alpha), \qquad t > 0,$$

in $D(0,\infty)$ for $\varepsilon > 0$. We assume that:

(2.1) There exists a nonsingular $d \times d$ matrix Γ, a constant $\gamma > 0$ and an $\mathbb{R}^d$-valued process $\mathscr{Y} \equiv \{\mathscr{Y}(t)\colon t > 0\}$ that is continuous at t w.p.1 for all t such that $\mathscr{Y}_\varepsilon \Rightarrow \Gamma\mathscr{Y}$ in $D(0,\infty)$ as $\varepsilon \downarrow 0$, which we denote by

$$\mathscr{Y}_\varepsilon(t) \equiv \varepsilon^{-\gamma}(Y(t/\varepsilon) - \alpha) \Rightarrow \Gamma\mathscr{Y}(t) \quad \text{in } D(0,\infty) \text{ as } \varepsilon \downarrow 0.$$

Typically the convergence in (2.1) occurs with $\gamma = 1/2$ and the limit process $\mathscr{Y}$ takes the form $\mathscr{Y}(t) = B(t)/t$, where B is standard Brownian motion (which is composed of d independent one-dimensional Brownian motions, each having zero drift and unit diffusion coefficient), but these are not required. Note that assumption (2.1) guarantees that

$$Y(t) - \alpha = t^{-\gamma}\mathscr{Y}_{1/t}(1) \Rightarrow 0 \cdot \Gamma\mathscr{Y}(1) = 0 \quad \text{as } t \to \infty;$$

that is, $Y(t) \Rightarrow \alpha$ as $t \to \infty$, so that $Y(t)$ is a weakly consistent estimator of α.

First, suppose that we preassign the amount of simulation time t to be generated by the computer. To obtain an approximate $100(1 - \delta)\%$ confidence set for α, we assume that there exists a bounded set A for which

$$(2.2) \qquad P\{\mathscr{Y}(1) \in A\} = 1 - \delta \quad \text{and} \quad P\{\mathscr{Y}(1) \in \partial A\} = 0,$$

where ∂A is the boundary of A. Then let $\tilde{C}(t) = Y(t) - t^{-\gamma}\Gamma A$, where we use the notation $z + QA$ to denote the set $\{x \in \mathbb{R}^d\colon \exists y \in A \text{ such that } x = z + Qy\}$ for $z \in \mathbb{R}^d$ and $d \times d$ matrices Q. The following proposition shows that $\tilde{C}(t)$ achieves the nominal coverage level as the sample size t is permitted to go to ∞.

PROPOSITION 1. *Under* (2.1) *and* (2.2), $P\{\alpha \in \tilde{C}(t)\} \to 1 - \delta$ *as* $t \to \infty$.

Of course, in applications, Γ is typically unknown so that it must be estimated. Assume that there exists an estimator $\Gamma(t)$ which is weakly consistent; that is, $\Gamma(t) \Rightarrow \Gamma$ as $t \to \infty$. Let $C(t) = Y(t) - t^{-\gamma}\Gamma(t)A$.

PROPOSITION 2. *If* $\Gamma(t) \Rightarrow \Gamma$ *as* $t \to \infty$ *under* (2.1) *and* (2.2), *then* $P\{\alpha \in C(t)\} \to 1 - \delta$ *as* $t \to \infty$.

Thus the confidence set $C(t)$ yields a procedure which is both implementable and provides an asymptotically valid region.

REMARK 2.1. Propositions 1 and 2 actually require only the CLT version of (2.1); that is, $\mathscr{Y}_\varepsilon(1) \Rightarrow \mathscr{Y}(1)$ in $\mathbb{R}^d$ as $\varepsilon \to 0$; see the proofs in Section 5.

We turn now to a discussion of sequential stopping procedures. For a generic (measurable) set $B \subseteq \mathbb{R}^d$, let $m(B)$ denote the d-dimensional volume (Lebesgue measure) of the set. Of course, when $d = 1$ and B is an interval, $m(B)$ is just the length of the interval. We first consider the case in which the procedure terminates when the dth root of the volume of the confidence region $C(t)$ drops below a prescribed level ε. [It is natural to use the dth root, because $m(xB)^{1/d} = xm(B)^{1/d}$ for a scalar x.] We call such a procedure an *absolute-precision sequential stopping rule*. For such a rule, the time $\tilde{T}(\varepsilon)$ at which the simulation terminates execution is defined by

$$\tilde{T}(\varepsilon) = \inf\left\{t \geq 0\colon m(C(t))^{1/d} \leq \varepsilon\right\}.$$

Actually, this stopping rule needs to be modified, because $\tilde{T}(\varepsilon)$ can terminate much too early if the estimator $\Gamma(t)$ is badly behaved for small t. To see this, suppose that $P(\Gamma(1) = 0,\ m(C(t)) = 1,\ 0 \le t < 1) = 1$. In this case, $\tilde{T}(\varepsilon) = 1$ for $\varepsilon < 1$, so $C(\tilde{T}(\varepsilon)) = Y(1)$ for $\varepsilon < 1$. Hence, in this example, $P(\alpha \in C(\tilde{T}(\varepsilon))) = P(\alpha = Y(1))$ for $\varepsilon < 1$. Hence convergence of the coverage probability of the region $C(T(\varepsilon))$ to the nominal level $1 - \delta$ does *not* occur when we let $\varepsilon \downarrow 0$.

In order for the asymptotic theory associated with (2.1) to be relevant to the sequential stopping problem, it is necessary that $T(\varepsilon) \to \infty$ as $\varepsilon \downarrow 0$. In other words, small values of the precision constant ε need to correspond to large values of simulation time. We can force the termination time to behave in this way if we inflate the volume $m(C(t))$ slightly. Let $a(t)$ be a strictly positive function that decreases monotonically to 0 as $t \to \infty$ and satisfies $a(t) = o(t^{-\gamma})$. Set

$$T_1(\varepsilon) = \inf\left\{t \ge 0\colon m(C(t))^{1/d} + a(t) \le \varepsilon\right\} \tag{2.3}$$

and

$$t_1(\varepsilon) = \inf\{t \ge 0\colon a(t) \le \varepsilon\}.$$

Note that $T_1(\varepsilon) \ge t_1(\varepsilon) \to \infty$ as $\varepsilon \downarrow 0$. Thus the early termination associated with $\tilde{T}(\varepsilon)$ is prevented by the stopping rule $T_1(\varepsilon)$. In practice, one might use a priori analytical estimates of required simulation run lengths, as in Whitt (1989, 1992), to determine the function $a(t)$; for example, $a(t) = 1$ for $t \le t_0$ and $a(t) = \varepsilon^2 t^{-2\gamma}$ for $t \ge t_0$; but we do not examine specific procedures here.

Throughout the rest of this paper, we assume that $m(A) > 0$ for A in (2.2). The next two theorems provide our main asymptotic results about sequential stopping rules for stochastic simulations. The first theorem shows that under the relatively mild assumption (2.1), strong consistency of the estimator $\Gamma(t)$ suffices to guarantee that the absolute-precision stopping rule $T_1(\varepsilon)$ has a variety of desirable asymptotic properties, including asymptotic validity.

THEOREM 1. *Suppose that* (2.1) *and* (2.2) *hold. If* $\Gamma(t) \to \Gamma$ *w.p.*1 *as* $t \to \infty$, *then as* $t \to \infty$ *or* $\varepsilon \to 0$:

(i) $t^{\gamma}[m(C(t))^{1/d} + a(t)] \to m(\Gamma A)^{1/d}$ *w.p.*1,
(ii) $\varepsilon^{1/\gamma} T_1(\varepsilon) \to m(\Gamma A)^{1/\gamma d}$ *w.p.*1,
(iii) $\varepsilon^{-1} m(C(T_1(\varepsilon)))^{1/d} \to 1$ *w.p.*1,
(iv) $\varepsilon^{-1}[Y(T_1(\varepsilon)) - \alpha] \Rightarrow m(\Gamma A)^{-1/d}\Gamma\mathscr{Y}(1)$ *in* $\mathbb{R}^d$,
(v) $P(\alpha \in C(T_1(\varepsilon))) \to 1 - \delta$ (*asymptotic validity*).

Part (i) shows that w.p.1 both the dth root of the volume and the inflated dth root of the volume of the confidence set at time t are asymptotically equal to $t^{-\gamma} m(\Gamma A)^{1/d}$ to first order as $t \to \infty$. Part (ii) shows that w.p.1 the termination time $T_1(\varepsilon)$ is asymptotically equal to $d(\varepsilon) = \varepsilon^{-1/\alpha} m(\Gamma A)^{1/\gamma d}$ to first order as $\varepsilon \to 0$. Note that $d(\varepsilon)$ is precisely that deterministic time at which $m(\tilde{C}(d(\varepsilon)))^{1/d} = \varepsilon$. Of course, the time $d(\varepsilon)$ is not directly implementable because Γ is presumed unknown. Part (iii) shows that w.p.1 the dth root of the volume of the final realized confidence set, $m(C(T_1(\varepsilon)))$, is asymptotically equal to the prescribed value ε to first order as $\varepsilon \to 0$. Part (iv) is a random-time-change CLT that serves to establish the asymptotic validity in (v). Thus the sequential stopping rule $T_1(\varepsilon)$ provides the same asymptotic behavior as $d(\varepsilon)$, despite the fact that $T_1(\varepsilon)$ needs to estimate Γ.

REMARK 2.2. The FCLT (2.1) instead of the ordinary CLT is needed to establish the random-time-change CLT in part (iv) of Theorem 1; see, for example, Example 4 of Glynn and Whitt (1988).

Our next result shows that we can replace the strong consistency of $\Gamma(t)$ in Theorem 1 with a functional weak law of large numbers (FWLLN). An FWLLN is easily obtained as a corollary to an FCLT. Since an ordinary strong law of large numbers (SLLN) is equivalent to an FSLLN [see Theorem 4 of Glynn and Whitt (1988) or Lemma 3 of Glynn and Whitt (1992)], which in turn implies an FWLLN, we also obtain the following results under the strong consistency assumption of Theorem 1, too. Moreover, the $\Rightarrow$ convergence in (i)–(iii) can then be replaced with w.p.1. However, an FWLLN need not imply a SLLN [see Example 2 of Glynn and Whitt (1988)], so that the condition of Theorem 2 is actually more general. We obtain convergence statements with $t = 1$ paralleling those of Theorem 1 by simply applying the continuous mapping theorem with the projection map at $t = 1$. In Section 4 we show that we cannot simply assume an ordinary WLLN, that is, that $\Gamma(t)$ is weakly consistent. Let $=_d$ denote equality in distribution.

THEOREM 2. *Suppose that* (2.1) *and* (2.2) *hold. If* $\Gamma(t/\varepsilon) \Rightarrow \Gamma$ *in* $D(0,\infty)$ *with range* $\mathbb{R}^{d^2}$ *as* $\varepsilon \to 0$, *then as* $\varepsilon \to 0$:

(i) $\varepsilon^{-\gamma}[m(C(t/\varepsilon))^{1/d} + a(t/\varepsilon)] \Rightarrow t^{-\gamma} m(\Gamma A)^{1/d}$ *in* $D(0,\infty)$,

(ii) $\varepsilon^{1/\gamma} T_1(\varepsilon/t) \Rightarrow t^{1/\gamma} m(\Gamma A)^{1/\gamma d}$ *in* $D(0,\infty)$,

(iii) $\varepsilon^{-1} m(C(T_1(\varepsilon/t)))^{1/d} \Rightarrow t^{-1}$ *in* $D(0,\infty)$,

(iv) $\varepsilon^{-1}[Y(T_1(\varepsilon/t)) - \alpha] \Rightarrow \Gamma\mathscr{Y}(t^{1/\gamma} m(\Gamma A)^{1/\gamma d}) =_d m(\Gamma A)^{-1/d}\Gamma\mathscr{Y}(t^{1/\gamma})$ *in* $D(0,\infty)$,

(v) $P(\alpha \in C(T_1(\varepsilon))) \to 1 - \delta$ (*asymptotic validity*).

Our next theorem considers a variant of the stopping rule $T_1(\varepsilon)$ known as a *relative-precision sequential stopping rule*. The basic idea here is that the simulation should terminate when the dth root of the volume of the confidence region is less than an εth fraction of the norm of the parameter α. Since $Y(t)$ is an estimator for α, this suggests replacing $T_1(\varepsilon)$ with

$$T_2(\varepsilon) = \inf\left\{t \ge 0: m(C(t))^{1/d} + a(t) \le \varepsilon |Y(t)|\right\}. \tag{2.4}$$

The next theorem shows that such relative-precision stopping rules have analogous asymptotic behavior to that exhibited by absolute-precision stopping rules. Note that $T_2(\varepsilon)$ behaves asymptotically like $T_1(|\alpha|\varepsilon)$, as one would expect. For the analog of Theorem 1, it is important that $Y(t)$ now be a strongly consistent estimator of α. This is a reasonable condition, but it does not follow from (2.1); see, for example, Example 2 of Glynn and Whitt (1988).

THEOREM 3. *Suppose that* (2.1) *and* (2.2) *hold. If* $Y(t) \to \alpha$ *and* $\Gamma(t) \to \Gamma$ *w.p.*1 *as* $t \to \infty$, *where* $|\alpha| > 0$, *then as* $t \to \infty$ *and* $\varepsilon \to 0$:

(i) $t^{\gamma}[m(C(t))^{1/d} + a(t)]/|Y(t)| \to |\alpha|^{-1} m(\Gamma A)^{1/d}$ *w.p.*1,
(ii) $\varepsilon^{1/\gamma} T_2(\varepsilon) \to m(\Gamma A)^{1/\gamma d} |\alpha|^{-1/\gamma}$ *w.p.*1,
(iii) $\varepsilon^{-1} m(C(T_2(\varepsilon)))^{1/d} \to |\alpha|$ *w.p.*1,
(iv) $\varepsilon^{-1}[Y(T_2(\varepsilon)) - \alpha] \Rightarrow |\alpha| m(\Gamma A)^{-1/d} \Gamma \mathscr{Y}(1)$ *in* $\mathbb{R}^d$,
(v) $P(\alpha \in C(T_2(\varepsilon))) \to 1 - \delta$ *(asymptotic validity)*.

The proof of Theorem 3 is a minor modification of the proof of Theorem 1, and is therefore omitted. There is also an analog of Theorem 2 for the relative-precision sequential stopping rules, but we do not state it. Nothing beyond the assumptions of Theorem 2 is needed except $|\alpha| > 0$.

3. Examples. In this section, we discuss the implications of Theorems 1 and 3 in a variety of different estimation settings. As we shall see, assumption (2.1) is a mild hypothesis that is satisfied in virtually all practical contexts. Given the presence of such an FCLT, the asymptotic validity of a sequential stopping rule basically depends upon the availability of a strongly consistent estimator for the scaling matrix Γ that appears in (2.1). (Alternatively we could have an FWLLN, as in Theorem 2.)

EXAMPLE 1 (The sample mean of i.i.d. random variables). Suppose that α can be represented as $\alpha = EX$ for some real-valued r.v. X. For example, α might correspond to the expected number of customers served in a queue over the time interval $[0, T]$. Then α can be estimated by generating i.i.d. replicates $X_1, X_2, \dots$ of the r.v. X; the resulting estimator for α is then the sample mean $\overline{X}_n = n^{-1}\sum_{i=1}^{n} X_i$. The corresponding estimation process is $Y(t) = \overline{X}_{\lfloor t \rfloor}$, where $\lfloor t \rfloor$ is the greatest integer less than t and $\overline{X}_0 = 0$. If $EX^2 < \infty$, Donsker's theorem asserts that (2.1) holds with $\gamma = 1/2$, $\Gamma = \sigma$, where $\sigma^2 = \operatorname{var} X$, and $\mathscr{Y}(t) = B(t)/t$, where $B(t)$ is standard (zero-drift, unit-diffusion-coefficient) Brownian motion; see Section 16 of Billingsley (1968). Note that $\mathscr{Y}(1) =_d N(0,1)$. The typical choice for the set A in this setting is the interval $[-z(\delta), z(\delta)]$, where $z(\delta)$ is chosen to satisfy $P\{N(0,1) \le z(\delta)\} = 1 - \delta/2$. Of course, it is well known that

$$\Gamma_n \equiv \left[\frac{1}{n-1}\sum_{i=1}^{n}\left(X_i - \overline{X}_n\right)^2\right]^{1/2} \to \sigma \quad \text{w.p.1 as } n \to \infty.$$

Suppose that $\sigma^2 \to 0$. Setting $\Gamma(t) = \Gamma_{\lfloor t \rfloor}$, we have the strong consistency required by Theorems 1 and 3. Hence both the absolute- and relative-precision stopping rules $T_1(\varepsilon)$ and $T_2(\varepsilon)$ are asymptotically valid for this example when the precision constant ε shrinks to 0. In this setting, Theorems 1 and 3 reproduce the results of Chow and Robbins (1965), Starr (1966) and Nadas (1969). Implementation considerations are discussed in Law, Kelton and Koenig (1981).

EXAMPLE 2 (The sample mean of i.i.d. random vectors). Now we consider the case in which α can be represented as $\alpha = EX$, where X is $\mathbb{R}^d$-valued. Assume that $E|X|^2 < \infty$. As in Example 1, we can estimate α via the sample mean $\overline{X}_n = n^{-1}\sum_{i=1}^n X_i$, where X_i's are i.i.d copies of X. Setting $Y(t) = \overline{X}_{\lfloor t \rfloor}$, we obtain (2.1) from the d-dimensional version of Donsker's theorem [see, e.g., Glynn and Whitt (1987)], where $\mathscr{Y}(t) = B(t)/t$, B is d-dimensional standard Brownian motion (composed of d independent one-dimensional standard Brownian motions) and $\Gamma\Gamma^t$ is the covariance matrix C of X. We assume that C is positive definite. Note that $\mathscr{Y}(1) = B(1) =_d N(0, 1)$. In this d-dimensional setting, we can assume that A is the d-sphere $\{x\colon |x| \le w(\delta)\}$, where $w(\delta)$ is chosen so that

$$P\{|N(0, I)|^2 \le w^2(\delta)\} = P\{\mathscr{X}_d^2 \le w^2(\delta)\} = 1 - \delta,$$

with $\mathscr{X}_d^2$ being a chi-squared r.v. with d degrees of freedom. Let

$$C_n = \frac{1}{n}\sum_{i=1}^n X_i X_i^t - \overline{X}_n \overline{X}_n^t$$

(writing all d-vectors as column vectors). Then $C_n \to C$ a.s. as $n \to \infty$. Let Γ_n be obtained by taking the Cholesky factorization of C_n, so that Γ_n is a lower triangular matrix such that $C_n = \Gamma_n \Gamma_n^t$; see pages 164 and 165 of Bratley, Fox and Schrage (1987). It follows that $\Gamma_n \to \Gamma$ w.p.1 as $n \to \infty$, since Cholesky factors are continuous at positive definite matrices. Setting $\Gamma(t) = \Gamma_{\lfloor t \rfloor}$, we again have the strong consistency required by Theorems 1 and 3. Thus we have proved that the absolute- and relative-precision stopping rules $T_1(\varepsilon)$ and $T_2(\varepsilon)$ are asymptotically valid for sequential stopping of multiple performance measure stochastic simulations.

In this setting, Theorems 1 and 3 reproduce results by Gleser (1965), Albert (1966) and Srivastava (1967); see Section 5.5 of Govindarajulu (1987).

EXAMPLE 3 (Functions of sample means). Let X be an $\mathbb{R}^d$-valued random vector and let $\mu = EX$. Suppose that α can be represented as $\alpha = g(\mu)$ for some (known) real-valued function $g\colon \mathbb{R}^d \to \mathbb{R}$. An example of this occurs in the ratio estimation setting, in which $d = 2$ and $g(x, y) = x/y$. Because the steady state of a regenerative stochastic process can be expressed as a ratio of two means, this estimation setting subsumes that of regenerative steady-state simulation. Of course, this observation lies at the heart of the regenerative method of steady-state simulation; see, for example, Crane and Lemoine (1977).

In this nonlinear setting, we estimate α via $Y(t) = g(\overline{X}_{\lfloor t \rfloor})$, where X_i are i.i.d. random vectors as in Example 2. Suppose that $E|X|^2 < \infty$ and that g is continuously differentiable in a neighborhood of μ. In addition, we require that $\nabla g(\mu) \neq 0$ and that the covariance matrix C of X is positive definite. Then Theorem 3 of Glynn and Whitt (1992) implies that (2.1) holds with $\gamma = 1/2$, $\mathscr{Y}(t) = B(t)/t$ and $\Gamma = \sigma$ as in Example 1, but with $\sigma = (\nabla g(\mu)^t C \, \nabla g(\mu))^{1/2}$.

Let C_n be defined as in Example 2 and note that $[\nabla g(Y(t))^t C_{\lfloor t \rfloor} \nabla g(Y(t))]^{1/2} \to \sigma$ w.p.1 as $t \to \infty$. Hence we have the strong consistency required for the application of Theorems 1 and 3. As a consequence, we are assured that the stopping rules $T_1(\varepsilon)$ and $T_2(\varepsilon)$ are asymptotically valid in this estimation setting. In particular, in the regenerative simulation setting, we recover the asymptotic theory developed by Lavenberg and Sauer (1977).

EXAMPLE 4 (The jackknife). Consider the estimation problem of Example 3 in which our goal is to estimate $\alpha = g(\mu)$, where μ can be expressed as $\mu = EX$ and g is real-valued. One practical difficulty with the estimator suggested in Example 3 is that it tends to be significantly affected by bias problems induced by the presence of the nonlinearity in g. One way to address the small-sample bias problem that this nonlinearity creates is to jackknife the estimator. Specifically, let $\alpha(n) = g(\overline{X}_n)$ and, for $1 \le i \le n$, let

$$\overline{X}_{in} = \frac{1}{n-1} \sum_{\substack{j=1 \\ j \neq i}}^{n} X_j, \qquad \alpha_i(n) = g\left(\overline{X}_{in}\right), \tag{3.1}$$

$$\tilde{\alpha}_i(n) = n\alpha(n) - (n-1)\alpha_i(n).$$

Then the estimator $Y_n = n^{-1}\sum_{i=1}^{n} \tilde{\alpha}_i(n)$ is the *jackknife estimator* of α. Let $Y(t) = Y_{\lfloor t \rfloor}$. It is shown in Glynn and Heidelberger (1989) that if $E|X|^3 < \infty$ and g is twice continuously differentiable in a neighborhood of μ, then (2.1) holds where σ and $\mathscr{Y}(t)$ are as in Example 3. Since the form of the FCLT (2.1) is the same as for Example 3, the jackknife has the same asymptotic efficiency as the estimator of Example 3. However, as argued in Miller (1974), the jackknife estimator typically possesses superior small-sample bias properties.

Two estimators for the scaling constant $\sigma = [\nabla g(\mu)^t C \, \nabla g(\mu)]^{1/2}$ are possible. One approach is to use the estimator $\sigma(t) = [\nabla g(Y(t))^t C_{\lfloor t \rfloor} \nabla g(Y(t))]^{1/2}$ suggested in Example 3. Theorem 4(i) of Glynn and Heidelberger (1989) shows that $Y(t) \to \alpha$ w.p.1 as $t \to \infty$, under the conditions stated here. Since $C_n \to C$ w.p.1, it follows that $\sigma(t) \to \sigma$ w.p.1 as $t \to \infty$. Hence sequential stopping procedures based on the jackknife point estimator and the "variance" estimator $\sigma^2(t)$ are asymptotically valid by Theorems 1 and 3, provided that $\sigma^2 > 0$.

An alternative estimator for the scaling constant σ is given by the jackknife variance estimator $\sigma_J(t)$:

$$\sigma_J(t) = \left(\frac{1}{\lfloor t \rfloor} \sum_{i=1}^{\lfloor t \rfloor} \left(\tilde{\alpha}_i(\lfloor t \rfloor) - Y(t)\right)^2 \right)^{1/2}.$$

Although it is known that $\sigma_J^2(t) \Rightarrow \sigma^2$ as $t \to \infty$ under suitable regularity conditions [see, e.g., Miller (1964), (1974)], we need convergence w.p.1 in order to satisfy the hypothesis of Theorems 1 and 3. We therefore establish the following result.

THEOREM 4. *If g is continuously differentiable in a neighborhood of μ and $E|X|^2 < \infty$, then $\sigma_J^2(t) \to \sigma^2 = \nabla g(\mu)^t C \ \nabla g(\mu)$ w.p.1 as $t \to \infty$.*

Thus the sequential stopping rules $T_1(\varepsilon)$ and $T_2(\varepsilon)$ may be applied to jackknife point estimators in conjunction with the jackknifed variance estimator $\sigma_J^2(t)$.

EXAMPLE 5 (A steady-state mean). Suppose that our goal is to estimate the steady-state mean vector α of an $\mathbb{R}^d$-valued stochastic process $X = \{X(t): t \geq 0\}$. We assume that X satisfies an FCLT, namely,

$$\varepsilon^{-1/2}\left(\int_0^{t/\varepsilon} X(s)\,ds - t\alpha\right) \Rightarrow \Gamma B(t) \quad \text{in } D(0,\infty) \text{ as } \varepsilon \downarrow 0, \tag{3.2}$$

where B is a standard $\mathbb{R}^d$-valued Brownian motion. It is easily shown that (3.2) implies that

$$Y(t) \equiv t^{-1}\int_0^t X(s)\,ds \Rightarrow \alpha \quad \text{as } t \to \infty.$$

Hence (3.2) implies that the centering vector α appearing in (3.2) is indeed the steady-state mean of X. Another easy consequence of (3.2) is that (2.1) holds with $\gamma = 1/2$ and $\mathscr{Y}(t) = B(t)/t$.

It turns out that (3.2) is typically satisfied for most "real-world" steady-state simulations. In particular, a great variety of different assumptions on the structure of the process X give rise to FCLTs of the form (3.2). For example, such FCLTs hold when X is regenerative and satisfies suitable moment conditions [see Glynn and Whitt (1987)], or when X is a martingale process or when X satisfies appropriate mixing conditions [see Chapter 7 of Ethier and Kurtz (1986)], or when there is appropriate positive dependence in the X process [specifically, when the $X(t)$'s are associated; see Newman and Wright (1981)].

The primary difficulty in applying Theorems 1–3 arises in the construction of a process $\Gamma(t)$ such that $\Gamma(t) \to \Gamma$ w.p.1 as $t \to \infty$ or $\Gamma(t/\varepsilon) \to \Gamma$ in $D(0,\infty)$ as $\varepsilon \downarrow 0$. Since $\Gamma\Gamma^t$ is the covariance matrix of the limiting Brownian motion, this is equivalent to the construction of a strongly consistent estimator $C(t)$ for the *time-average covariance matrix* $C = \Gamma\Gamma^t$ of X. In general, this is known to be a challenging problem.

Suppose that X is regenerative, with regeneration times $0 = \tau_0 < \tau_1 < \tau_2 < \cdots$. Suppose that $E(\int_{\tau_1}^{\tau_2}|X(s) - \alpha|\,ds)^2 < \infty$ and that $E(\tau_2 - \tau_1) < \infty$. Let $N(t) = \max\{n \geq 0: \tau_n \leq t\}$. Then it is easily proved that

$$C(t) = \frac{1}{t}\sum_{i=1}^{N(t)} \int_{\tau_{i-1}}^{\tau_i} [X(s) - Y(t)][X(s) - Y(t)]^t\,ds$$

is strongly consistent for C, where $C = \Gamma\Gamma^t$ and Γ is the scaling matrix appearing in (3.2). Thus when X is regenerative, the sequential stopping rules $T_1(\varepsilon)$ and $T_2(\varepsilon)$ are asymptotically valid. Of course, when X is scalar, we already established this result in Example 3.

For nonregenerative processes, less is known about the strong consistency of estimators $C(t)$ for the steady-state covariance matrix. However, Glynn and Iglehart (1988) and Damerdji (1989a, b) have recently used strong approximation techniques to establish strong consistency for a broad class of estimators for C. Thus Theorems 1 and 3 prove that these estimators do indeed lead to asymptotically valid sequential procedures.

Our theory for this example provides theoretical support complementing previous work by Fishman (1977), Law and Carson (1979) and Law and Kelton (1982) which develop specific empirically based sequential stopping rules for steady-state simulations.

EXAMPLE 6 (Kiefer–Wolfowitz stochastic approximation). This example is interesting, in part, because it illustrates that the FCLT (2.1) can hold for the estimator with a subcanonical convergence rate; in particular, here $\gamma = 1/3$. For other examples of noncanonical estimator convergence rates, see Fox and Glynn (1989) and Sections 5 and 6 of Glynn and Whitt (1992). Suppose that we are given a real-valued smooth function $\beta(\theta)$, which can be represented as $\beta(\theta) = EZ(\theta)$. Assume that our goal is to compute the parameter $\alpha \equiv \theta^*$ minimizing β. If θ is a scalar, we can apply the following Kiefer–Wolfowitz stochastic approximation algorithm:

$$\theta_{n+1} = \theta_n - c_n X_{n+1},$$

where $\{c_n: n \geq 0\}$ is a sequence of (deterministic) nonnegative constants,

$$P(X_{n+1} \in A | \theta_0, X_0, \dots, \theta_n, X_n)$$
$$= P([Z(\theta_0 + h_{n+1}) - Z(\theta_0 - h_{n+1})]/2h_{n+1} \in A),$$

$Z(\theta_0 + h_{n+1})$ and $Z(\theta_0 - h_{n+1})$ are generated independently of one another and $\{h_n: n \geq 1\}$ is another sequence of deterministic constants. Suppose that $c_n = cn^{-1}$ and $h_n = hn^{-1/3}$, $c, h > 0$. Let $Y(t) = \theta_{\lfloor t \rfloor}$. Then Ruppert (1982) shows that under suitable regularity conditions, (2.1) holds with $\gamma = 1/3$, $\Gamma = \kappa$, $\mathscr{Y}(t) = t^{-b}B(t^{2\eta+1})$, B is a standard Brownian motion, $b = c\beta''(\theta^*)$, $\eta = b - 5/6$, $\kappa^2 = c^2\sigma^2/(2\eta + 1)(4h^2)$ and $\sigma^2 = 2\operatorname{var} Z(\theta^*)$.

The construction of a strongly consistent estimator for $\Gamma \equiv \kappa$ involves more work. For some directions on how to obtain such an estimator, see page 189 of Venter (1967).

EXAMPLE 7 (Robbins–Monro stochastic approximation). As in Example 6, suppose that our goal is to estimate the minimizer θ^* is a smooth function β:

$\mathbb{R} \to \mathbb{R}$. However, we assume here that we can represent the derivative β' as an expectation; that is, there exists a process $Z(\theta)$ such that $\beta'(\theta) = EZ(\theta)$. [In Example 6 we assumed only that the function values $\beta(\theta)$ could be represented as expectations.] To calculate θ^* in this setting, we can use the Robbins–Monro stochastic approximation algorithm:

$$\theta_{n+1} = \theta_n - c_n X_{n+1},$$

where $\{c_n: n \geq 0\}$ is a sequence of (deterministic) nonnegative constants and

$$P(X_{n+1} \in A | \theta_0, X_0, \ldots, \theta_n, X_n) = P(Z(\theta_n) \in A).$$

Suppose that our estimator is $Y(t) = \theta_{\lfloor t \rfloor}$ and $c_n = cn^{-1}$ with $c > 0$. Then Ruppert (1982) has shown that under suitable regularity hypotheses, the FCLT (2.1) holds with $\gamma = 1/2$, $\Gamma = \kappa$, $\mathscr{Y}(t) = t^{-(D+1)}B(t^{2D+1})$, $D = c\beta'(\theta^*) - 1$, $\kappa^2 = c^2\sigma^2(2D+1)^{-1}$, $\sigma^2 = \operatorname{var} Z(\theta^*)$ and B is a standard Brownian motion.

Construction of a strongly consistent estimator for $\Gamma \equiv \kappa$ follows from results established by Venter (1967). When this estimator is used, the sequential stopping rule $T_1(\varepsilon)$ reduces to one studied by McLeish (1976). Our analysis of the rule $T_2(\varepsilon)$ seems to be new.

4. A Counterexample for weak consistency. We have developed a framework to analyze the asymptotic behavior of sequential stopping rules. Our analysis shows that a sequential stopping rule is asymptotically well behaved if the estimation process satisfies an FCLT as in (2.1) and if there exists a w.p.1 or FWLLN limit for the estimator for the scaling matrix Γ appearing in (2.1). The examples of Section 3 strongly suggest that (2.1) is typically satisfied in applications, but strong consistency or FWLLN consistency of the estimator of Γ is often more difficult to verify. As a consequence, it is natural to ask whether weak consistency [i.e., $\Gamma(t) \Rightarrow \Gamma$ as $t \to \infty$] is enough.

Unfortunately, weak consistency is not enough. The difficulty is in establishing the in-probability analog of Theorem 1(ii). If $\varepsilon^{1/\gamma}T_1(\varepsilon) \Rightarrow m(\Gamma A)^{1/\gamma d}$, then parts (iv) and (v) of Theorem 1 would hold by the argument of Theorem 17.1 of Billingsley (1968). In a proof of the in-probability analog of Theorem 1(ii), we cannot conclude that $T_1(\varepsilon)^\gamma V(T_1(\varepsilon)) \Rightarrow m(\Gamma A)^{1/d}$ when $T_1(\varepsilon) \Rightarrow \infty$ as $\varepsilon \downarrow 0$ and $t^\gamma V(t) \Rightarrow m(\Gamma A)^{1/d}$ as $t \to \infty$; see Richter (1965) and pages 10–15 of Gut (1988) for counterexamples and discussion.

We now give a direct counterexample. Consider Example 1 and the process $\Gamma(t)$ defined there. Let N be a unit rate Poisson process independent of $\{X_i: i \geq 1\}$ and let $T_1, T_2, \ldots$ be the jump times of the process N. Suppose that

$$\tilde{\Gamma}_{(t)} = \begin{cases} \Gamma(t), & t \notin \bigcup_{n=1}^{\infty} [T_n, T_n + 1/n), \\ 0, & t \in \bigcup_{n=1}^{\infty} [T_n, T_n + 1/n). \end{cases} \tag{4.1}$$

Then

$$P(\tilde{\Gamma}(t) \neq \Gamma(t)) = P\left(t \in \left[T_{N(t)}, T_{N(t)} + \frac{1}{N(t)}\right]\right)$$
$$\leq P(t - T_{N(t)} \leq \varepsilon) + P\left(N(t) \leq \frac{1}{\varepsilon}\right)$$

for ε arbitrary. Letting $t \to \infty$, we find that $\limsup_{t\to\infty} P(\tilde{\Gamma}(t) \leq \Gamma(t)) = 1 - \exp(-\varepsilon)$ (recall that the equilibrium age distribution of N is exponential with mean 1). Since ε was arbitrary, it follows that $P(\tilde{\Gamma}(t) \neq \Gamma(t)) \to 0$ as $t \to \infty$. Then it is evident that $\tilde{\Gamma}(t) \Rightarrow \sigma$ as $t \to \infty$, since $\Gamma(t) \to \sigma$ w.p.1 as $t \to \infty$. Hence $\tilde{\Gamma}(t)$ is weakly consistent for σ.

Now, in the setting of Example 1 using $\tilde{\Gamma}(t)$,

$$\tilde{T}_1(\varepsilon) = \inf\left\{t \geq 0 \colon z(\delta)\left(\frac{\tilde{\Gamma}(t)}{\sqrt{t}} + a(t)\right) \leq \varepsilon\right\}.$$

Put $a(t) = 1/t$. Then clearly $z(\delta)(\tilde{\Gamma}(s)/\sqrt{s} + 1/s) \geq z(\delta)/t$ and $s \leq t$, so $\tilde{T}_1(z(\delta)/t) \geq t$. On the other hand, $\tilde{\Gamma}(T_{N(t)+1}) = 0$, so $\tilde{T}_1(z(\delta)/t) \leq T_{N(t)+1}$. By the SLLN, $t^{-1}T_{N(t)+1} \to 1$ w.p.1 as $t \to \infty$. Hence $\tilde{T}_1(z(\delta)/t) \sim t$ w.p.1 as $t \to \infty$. Thus the stopping rule is asymptotically independent of the scaling constant Γ. As a consequence, formation of asymptotically valid confidence intervals is impossible. In fact, even the asymptotic scaling of the rule is incorrect. It is well known that for estimation problems of the type described in Example 1, the amount of simulation time required to obtain an absolute precision of order ε is of order ε^{-2}, whereas the stopping rule $\tilde{T}_1(\varepsilon)$ based on $\tilde{\Gamma}(t)$ in (4.1) yields a termination time of order ε^{-1}.

5. Proofs.

Proof of Proposition 1. Since Γ is nonsingular,

$$P(\alpha \in \tilde{C}(t)) = P(\Gamma^{-1}t^{\gamma}(Y(t) - \alpha) \in A),$$

but

$$\Gamma^{-1}t^{\gamma}(Y(t) - \alpha) = \Gamma^{-1}\mathscr{Y}_{1/t}(1) \Rightarrow \Gamma^{-1}\Gamma\mathscr{Y}(1) = \mathscr{Y}(1) \quad \text{as } t \to \infty.$$

Since $P(\mathscr{Y}(1) \in \partial A) = 0$, it follows that

$$P(\Gamma^{-1}t^{\gamma}(Y(t) - \alpha) \in A) \to P(\mathscr{Y}(1) \in A) = 1 - \delta \text{ as } t \to \infty;$$

see Theorem 2.1 of Billingsley (1968). □

Proof of Proposition 2. By (2.1) and Theorem 4.4 of Billingsley (1968),

$$[\Gamma(t), t^{\gamma}(Y(t) - \alpha)] \Rightarrow [\Gamma, \Gamma\mathscr{Y}(1)] \quad \text{as } t \to \infty.$$

Then we can apply the continuous mapping theorem, Theorem 5.1 of Billingsley (1968), to deduce that

$$\Gamma(t)^{-1}t^{\gamma}(Y(t)-\alpha) \Rightarrow \Gamma^{-1}\Gamma\mathscr{Y}(1) = \mathscr{Y}(1) \quad \text{as } t \to \infty.$$

To apply the continuous mapping theorem, note that matrix inversion is continuous at all nonsingular limits. (To see this, note the determinant is a continuous function of the matrix elements.) The rest of the proof is identical to that of Proposition 1. □

Proof of Theorem 1. Let $V(t) = m(C(t))^{1/d} + a(t)$. By the spatial invariance and the scaling properties of Lebesgue measure m,

$$m(Y(t) - t^{-\gamma}\Gamma(t)A)^{1/d} = m(-t^{-\gamma}\Gamma(t)A)^{1/d} = t^{-\gamma}m(\Gamma(t)A)^{1/d}.$$

Since A is a bounded set, $\Gamma(t)A$ is contained in a bounded set for all sufficiently large t w.p.1. It then follows from the bounded convergence theorem that $m(\Gamma(t)A)^{1/d} \to m(\Gamma A)^{1/d}$ w.p.1 as $t \to \infty$. Recalling that $a(t) = o(t^{-\gamma})$, we conclude that

$$t^{\gamma}V(t) \to m(\Gamma A)^{1/d} > 0 \quad \text{w.p.1 as } t \to \infty, \tag{5.1}$$

since $m(A) > 0$ and Γ is nonsingular. This establishes part (i). By definition of $T_1(\varepsilon)$, $V(T_1(\varepsilon) - 1) > \varepsilon$ and there exists a random variable $Z(\varepsilon)$ with $0 \le Z(\varepsilon) \le 1$ such that $V(T_1(\varepsilon) + Z(\varepsilon)) \le \varepsilon$. [Note that $V(t)$ is not necessarily monotone.] By (5.1) and the fact that $T_1(\varepsilon) \to \infty$ w.p.1 as $\varepsilon \downarrow 0$,

$$\limsup_{\varepsilon \downarrow 0} \varepsilon T_1(\varepsilon)^{\gamma} \le \limsup_{\varepsilon \downarrow 0} T_1(\varepsilon)^{\gamma} V(T_1(\varepsilon) - 1) = m(\Gamma A)^{1/d} \quad \text{w.p.1}.$$

Similarly,

$$\liminf_{\varepsilon \downarrow 0} \varepsilon T_1(\varepsilon)^{\gamma} \ge \liminf_{\varepsilon \downarrow 0} T_1(\varepsilon)^{\gamma} V(T_1(\varepsilon) + Z(\varepsilon)) = m(\Gamma A)^{1/d} \quad \text{w.p.1}.$$

This proves part (ii).

For part (iii), note that $m(C(T_1(\varepsilon)))^{1/d} = T_1(\varepsilon)^{-\gamma} m(\Gamma(T_1(\varepsilon))A)^{1/d}$ and recall that $m(\Gamma(t)A) \to m(\Gamma A)$ w.p.1 as $t \to \infty$, so that $m(\Gamma(T_1(\varepsilon))A) \to m(\Gamma A)$ w.p.1. as $\varepsilon \to 0$. By part (ii), $\varepsilon^{-1}T_1(\varepsilon)^{-\gamma} \to m(\Gamma A)^{-1/d}$. Hence

$$\begin{aligned}\varepsilon^{-1}m(C(T_1(\varepsilon)))^{1/d} &= \varepsilon^{-1}T_1(\varepsilon)^{-\gamma}m(\Gamma(T_1(\varepsilon))A)^{1/d}\\ &\to m(\Gamma A)^{-1/d}m(\Gamma A)^{1/d} = 1.\end{aligned}$$

To obtain part (iv), let $\beta = m(\Gamma A)^{1/d}$ and set $\tau_\varepsilon(t) = T_1(\varepsilon)\varepsilon^{1/\gamma}\beta^{-1/\gamma}t$, $t \ge 0$. Since $\tau_\varepsilon \Rightarrow e$ as $\varepsilon \downarrow 0$, where $e(t) = t$, it follows from the FCLT (2.1) and a standard random-time-change argument, page 144 of Billingsley (1968), that

$$\mathscr{Y}_{\varepsilon^{1/\gamma}\beta^{-1/\gamma}}(\tau_\varepsilon(1)) \Rightarrow \Gamma\mathscr{Y}(e(1)) = \Gamma\mathscr{Y}(1) \quad \text{as } \varepsilon \downarrow 0, \tag{5.2}$$

where

$$\mathscr{Y}_{\varepsilon^{1/\gamma}\beta^{-1/\gamma}}(\tau_\varepsilon(1)) = \beta\varepsilon^{-1}(Y(T_1(\varepsilon)) - \alpha). \tag{5.3}$$

To establish part (v), note that

$$
\begin{aligned}
&P\big(\alpha \in C(T_1(\varepsilon))\big)\\
&\quad = P\big(Y(T_1(\varepsilon)) - \alpha \in T_1(\varepsilon)^{-\gamma}\Gamma(T_1(\varepsilon))A\big)\\
&\quad = P\big(\alpha\beta^{-1}\varepsilon T_1(\varepsilon)^{\gamma}\Gamma(T_1(\varepsilon))^{-1}\beta\varepsilon^{-1}\big[Y(T_1(\varepsilon)) - \alpha\big] \in A;\\
&\qquad\qquad \det\big(\Gamma(T_1(\varepsilon)) \neq 0\big)\\
&\qquad + P\big(Y(T_1(\varepsilon)) - \alpha \in T_1(\varepsilon)^{-\gamma}\Gamma(T_1(\varepsilon))A; \det\big(\Gamma(T_1(\varepsilon))\big) = 0\big).
\end{aligned}
\tag{5.4}
$$

Since $P(\det(\Gamma(T_1(\varepsilon))) = 0) \to 0$ as $\varepsilon \to 0$, the second term converges to 0. By a convergence-together argument, Theorem 4.1 of Billingsley (1968), the first term has the same limit as $P(\beta\Gamma^{-1}\varepsilon^{-1}[Y(T_1(\varepsilon)) - \alpha] \in A)$, which is $P(\mathscr{Y}(1) \in A) = 1 - \delta$ by part (iv) because $P(\partial A) = 0$; see Theorem 2.1 of Billingsley (1968) and (2.2). To establish the convergence together, note that $P(\det(\Gamma(T_1(\varepsilon))) \neq 0) \to 1$ as $\varepsilon \to 0$ and $\beta^{-1}\varepsilon T_1(\varepsilon)^{\gamma}\Gamma(T_1(\varepsilon))^{-1} \to \beta^{-1}\beta\Gamma^{-1}$ w.p.1 as $\varepsilon \downarrow 0$. □

Proof of Theorem 2. For part (i), modify the proof of part (i) of Theorem 1. First, apply the assumed convergence of $\Gamma(t/\varepsilon)$ in $D(0,\infty)$ as $\varepsilon \downarrow 0$ to obtain $m(\Gamma(t/\varepsilon)A)^{1/d} \Rightarrow m(\Gamma A)^{1/d}$ in $D(0,\infty)$ as $\varepsilon \downarrow 0$. For this purpose, use the w.p.1 convergence representation of convergence in probability; see, for example, page 68 of Whitt (1980). (Consider a sequence $\{X_n\}$ of random elements of a separable metric space. Then $X_n \to_p X$ if and only if every subsequence of $\{X_n\}$ has a further subsequence converging w.p.1 to X.) Then, for any w.p.1 convergent subsequence of $\{\Gamma(t/\varepsilon)\colon 0 < \varepsilon < 1\}$ converging to Γ, note that the intersection and union of $\Gamma(t/\varepsilon)A$ over $t \in [t_0, t_1]$ for $0 < t_0 < t_1$ both approach ΓA w.p.1 as $\varepsilon \to 0$. [Since the limit is a continuous function, convergence in $D(0,\infty)$ is equivalent to uniform convergence on bounded intervals.] Then apply the standard bounded convergence theorem with the w.p.1 convergence to get $m(\Gamma(t/\varepsilon)A)^{1/d} \to m(\Gamma A)^{1/d}$ for this subsequence, uniformly for $t \in [t_0, t_1]$. This yields w.p.1 convergence in $D(0,\infty)$. Finally, since the same limit is obtained for every w.p.1 convergent subsequence, we have the desired convergence in probability. Since $\varepsilon^{-\gamma}a(t/\varepsilon) \to 0$ uniformly in t for $t > t_0 > 0$, we obtain the following analog of (5.1):

$$
(\varepsilon/t)^{-\gamma}V(t/\varepsilon) \Rightarrow m(\Gamma A)^{1/d} \quad \text{in } D(0,\infty) \text{ as } \varepsilon \to 0,
$$

which is equivalent to what is to be proved.

For part (ii) apply the continuous mapping theorem, Theorem 5.1 of Billingsley (1968), with the inverse mapping; see Section 7 of Whitt (1980). The inverse map there is $x^{-1}(t) = \inf\{s \geq 0\colon x(s) > t\}$, $t > 0$, but the results also apply to first passage times of the form (2.3). Note that

$$
\begin{aligned}
\varepsilon T(\varepsilon^{\gamma}/t) &= \inf\{s \geq 0\colon \varepsilon^{-\gamma}V(s/\varepsilon) \leq t^{-1}\}\\
&\Rightarrow \inf\{s \geq 0\colon s^{-\gamma}m(\Gamma A)^{1/d} \leq t^{-1}\} = t^{1/\gamma}m(\Gamma A)^{1/\gamma d},
\end{aligned}
$$

so that the same limit holds for $\varepsilon^{1/\gamma}T(\varepsilon/t)$.

For part (iii), apply the continuous mapping theorem with the composition map, using Theorem 3.1 of Whitt (1980), with parts (i) and (ii) above. In particular, use

$$\varepsilon^{-1} m\big(C(t/\varepsilon^{1/\gamma})\big) \Rightarrow t^{-\gamma} m(\Gamma A)^{1/d} \quad \text{in } D(0,\infty)$$

with part (ii) to obtain part (iii). Similarly, use the continuous mapping theorem with the composition map to obtain part (iv), now drawing on (2.1) and part (ii). To obtain the equivalent forms of the limit, note that $\mathscr{Y}_{\varepsilon/c}(t) = c^{\gamma}\mathscr{Y}_{\varepsilon}(ct)$, so that $\Gamma\mathscr{Y}(t) =_d c^{\gamma}\Gamma\mathscr{Y}(ct)$ for any scalar c. Finally, to obtain part (v), apply the continuous mapping theorem with the projection map at $t = 1$ to obtain the ordinary CLT in part (iv) of Theorem 1. Then apply the argument for the proof of part (v) in Theorem 1. □

PROOF OF THEOREM 4. It is shown in Glynn and Heidelberger (1989) that $\max\{|\overline{X}_{in} - \overline{X}_n|: 1 \le i \le n\} \to 0$ w.p.1 as $n \to \infty$. Hence, in expanding $g(\overline{X}_{in})$ in a Taylor series about $\overline{X}_n$, namely,

$$g(\overline{X}_{in}) = g(\overline{X}_n) + \nabla g(\xi_{in})^t(\overline{X}_{in} - \overline{X}_n),$$

we may assert that $\max_{1\le i\le n}|\xi_{in} - \mu| \to 0$ w.p.1 as $n \to \infty$. Then

$$\begin{aligned}
\tilde{\alpha}_i(n) - Y_n &= \big[n\alpha(n) - (n-1)\alpha_i(n)\big] - n^{-1}\sum_{j=1}^{n}\big[n\alpha(n) - (n-1)\alpha_j(n)\big] \\
&= (n-1)n^{-1}\sum_{j=1}^{n}\alpha_j(n) - (n-1)\alpha_i(n) \\
&= (n-1)n^{-1}\sum_{j=1}^{n}\Big[g(\overline{X}_n) + \nabla g(\xi_{jn})^t(\overline{X}_{jn} - \overline{X}_n)\Big] \\
&\quad -(n-1)\Big[g(\overline{X}_n) + \nabla g(\xi_{in})^t(\overline{X}_{in} - \overline{X}_n)\Big] \\
&= (n-1)n^{-1}\sum_{j=1}^{n}\nabla g(\xi_{jn})^t(\overline{X}_{jn} - \overline{X}_n) \\
&\quad -(n-1)\nabla g(\xi_{in})^t(\overline{X}_{in} - \overline{X}_n).
\end{aligned}$$

However,

$$\begin{aligned}
(n-1)\big(\overline{X}_{jn} - \overline{X}_n\big) &= (n-1)\overline{X}_{jn} - n\overline{X}_n + \overline{X}_n \\
&= \sum_{\substack{k=1\\k\neq j}}^{n} X_k - \sum_{k=1}^{n} X_k + \overline{X}_n = \overline{X}_n - X_j,
\end{aligned}$$

so that

$$(5.5)\quad \tilde{\alpha}_i(n) - Y_n = n^{-1}\sum_{j=1}^{n}\nabla g(\xi_{jn})^t\big(\overline{X}_n - X_j\big) - \nabla g(\xi_{in})^t\big(\overline{X}_n - X_i\big).$$

The first term on the right-hand side of (5.5) may be written as

$$n^{-1}\nabla g(\mu)^t \sum_{j=1}^n \left(\bar{X}_n - X_j\right) + n^{-1} \sum_{j=1}^n \left(\nabla g(\xi_{jn})^t - \nabla g(\mu)^t\right)\left(\bar{X}_n - X_j\right)$$

$$= n^{-1} \sum_{j=1}^n \left(\nabla g(\xi_{jn}) - \nabla g(\mu)\right)^t\left(\bar{X}_n - X_j\right) \equiv \beta_n.$$

Let $M_n = \max_{1\le j\le n}|\nabla g(\xi_{jn}) - \nabla g(\mu)|$. Then

$$|\beta_n| \le M_n n^{-1} \sum_{j=1}^n \left(|\bar{X}_n| + |X_j|\right) \to 0 \cdot (|\mu| + E|X|) = 0 \quad \text{w.p.1 as } n \to \infty.$$

Thus

$$\tilde{\alpha}_i(n) - Y_n = \nabla g(\mu)^t\left(X_i - \bar{X}_n\right) + \left[\nabla g(\xi_{in}) - \nabla g(\mu)\right]^t\left(X_i - \bar{X}_n\right) + \beta_n.$$

Let $\beta_{in} = \nabla g(\xi_{in}) - \nabla g(\mu)$ and $W_{in} = X_i - \bar{X}_n$. Then

$$\begin{aligned}\sigma_J^2(n) &= \nabla g(\mu)^t \frac{1}{n}\sum_{i=1}^n W_{in}W_{in}^t \nabla g(\mu) + \frac{1}{n}\sum_{i=1}^n \beta_{in}^t W_{in}W_{in}^t\beta_{in} \\ &\quad + \beta_n^2 + 2\beta_n \frac{1}{n}\sum_{i=1}^n \beta_{in}^t W_{in} + 2\nabla g(\mu)^t \frac{1}{n}\sum_{i=1}^n W_{in}W_{in}^t\beta_{in}.\end{aligned} \tag{5.6}$$

Also,

$$\begin{aligned}\frac{1}{n}\sum_{i=1}^n |W_{in}W_{in}^t| &\le \frac{1}{n}\sum_{i=1}^n \left[|X_iX_i^t| + |\bar{X}_nX_i^t| + |X_i\bar{X}_n^t| + |\bar{X}_n\bar{X}_n^t|\right] \\ &\to E\left[|XX^t| + |\mu X^t| + |X\mu^t| + |\mu\mu^t|\right] \quad \text{w.p.1 as } n \to \infty\end{aligned} \tag{5.7}$$

and

$$\frac{1}{n}\sum_{i=1}^n |W_{in}| \le \frac{1}{n}\sum_{i=1}^n \left[|X_i| + |\bar{X}_n|\right] \to E|X| + |\mu| \quad \text{w.p.1 as } n \to \infty. \tag{5.8}$$

Since $M_n \to 0$ w.p.1 as $n \to \infty$, it follows that $|\beta_{in}| \to 0$ uniformly in i w.p.1. Hence it is evident from (5.7) and (5.8) that all the terms in (5.6) involving β_{in} and β_n converge to 0 w.p.1, but the first term on the right-hand side of (5.6) clearly converges to $\sigma^2 = \nabla g(\mu)^t C\, \nabla g(\mu)$, completing the proof. □

Acknowledgment. The first author's major contribution to this paper was made while visiting the Department of Mathematics and Statistics at Carleton University, Ottawa, Canada. He wishes to thank the department for its hospitality.

REFERENCES

ALBERT, A. (1966). Fixed size confidence ellipsoids for linear regression parameters. *Ann. Math. Statist.* **37** 1602–1630.

ANSCOMBE, F. J. (1952). Large sample theory for sequential estimation. *Proc. Cambridge Philos. Soc.* **48** 600–607.

ANSCOMBE, F. J. (1953). Sequential estimation. *J. Roy. Statist. Soc. Ser. B* **15** 1–21.

BILLINGSLEY, P. (1968). *Convergence of Probability Measures*. Wiley, New York.

BRATLEY, P., FOX, B. L. and SCHRAGE, L. E. (1987). *A Guide to Simulation*, 2nd ed. Springer, New York.

CHOW, Y. S. and ROBBINS, H. (1965). On the asymptotic theory of fixed-width sequential confidence intervals for the mean. *Ann. Math. Statist.* **36** 457–462.

CRANE, M. A. and LEMOINE, A. J. (1977). *An Introduction to the Regenerative Method for Simulation Analysis. Lecture Notes in Control and Inform. Sci.*. Springer, New York.

DAMERDJI, H. (1989a). Strong consistency of the spectral variance estimator. Unpublished manuscript, Dept. Operations Research, Stanford Univ.

DAMERDJI, H. (1989b). Strong consistency of the variance estimator in steady-state simulation output analysis. Unpublished manuscript, Dept. Operations Research, Stanford Univ.

ETHIER, S. N. and KURTZ, T. G. (1986). *Markov Processes: Characterization and Convergence*. Wiley, New York.

FISHMAN, G. S. (1977). Achieving specific accuracy in simulation output analysis. *Comm. ACM* **20** 310–315.

FOX, B. L. and GLYNN, P. W. (1989). Replication schemes for limiting expectations. *Probab. Engr. Inform. Sci.* **3** 299–318.

GLESER, L. J. (1965). On the asymptotic theory of fixed-size confidence bounds for linear regression parameters. *Ann. Math. Statist.* **36** 463–467. [Correction note: **37** (1966) 1053–1055].

GLYNN, P. W. and HEIDELBERGER, P. (1989). Jackknifing under a budget constraint. Unpublished manuscript, Dept. Operations Research, Stanford Univ.

GLYNN, P. W. and IGLEHART, D. L. (1988). A new class of strongly consistent variance estimators for a steady-state simulation. *Stochastic Process. Appl.* **16** 71–80.

GLYNN, P. W. and WHITT, W. (1987). Sufficient conditions for functional-limit-theorem versions of $L = \lambda W$. *Queueing Systems* **1** 279–287.

GLYNN, P. W. and WHITT, W. (1988). Ordinary CLT and WLLN versions of $L = \lambda W$. *Math. Oper. Res.* **13** 674–692.

GLYNN, P. W. and WHITT, W. (1992). The asymptotic efficiency of simulation estimators. *Oper. Res.* To appear.

GOVINDARAJULU, Z. (1987). *The Sequential Statistical Analysis of Hypothesis Testing, Point and Interval Estimation, and Decision Theory*, 2nd ed. American Sciences Press, Columbus, Oh.

GUT, A. (1988). *Stopped Random Walks*. Springer, New York.

HEIDELBERGER, P. and WELCH, P. D. (1981a). A spectral method for confidence interval generation and run length control in simulations. *Comm. ACM* **24** 233–245.

HEIDELBERGER, P. and WELCH, P. D. (1981b). Adaptive spectral methods for simulation output analysis. *IBM J. Res. Develop.* **25** 860–876.

HEIDELBERGER, P. and WELCH, P. D. (1983). Simulation run length control in the presence of an initial transient. *Oper. Res.* **31** 1109–1144.

LAVENBERG, S. S. and SAUER, C. H. (1977). Sequential stopping rules for the regenerative method of simulation. *IBM J. Res. Develop.* **21** 545–558.

LAW, A. M. and CARSON, J. S. (1979). A sequential procedure for determining the length of a steady-state simulation. *Oper. Res.* **27** 1011–1025.

LAW, A. M. and KELTON, W. D. (1982). Confidence intervals for steady-state simulation. II. A survey of sequential procedures. *Management Sci.* **28** 550–562.

LAW, A. M., KELTON, W. D. and KOENIG, L. W. (1981). Relative width sequential confidence intervals for the mean. *Comm. Statist. B—Simulation Comput.* **10** 29–39.

McLeish, D. L. (1976). Functional and random central limit theorems for the Robbins–Monro process. *J. Appl. Probab.* **13** 148–154.

Miller, R. G. (1964). A trustworthy jackknife. *Ann. Math. Statist.* **35** 1594–1605.

Miller, R. G. (1974). The jackknife—a review. *Biometrika* **61** 1–15.

Nadas, A. (1969). An extension of a theorem of Chow and Robbins on sequential confidence intervals for the mean. *Ann. Math. Statist.* **40** 667–671.

Newman, C. M. and Wright, A. L. (1981). An invariance principle for certain dependent sequences. *Ann. Probab.* **9** 671–675.

Richter, W. (1965). Limit theorems for sequences of random variables with sequences of random indices. *Theory Probab. Appl.* **10** 74–84.

Ruppert, D. (1982). Almost sure approximations to the Robbins–Monro and Kiefer–Wolfowitz processes with dependent noise. *Ann. Probab.* **10** 178–187.

Siegmund, D. (1985). *Sequential Analysis*. Springer, New York.

Srivastava, M. S. (1967). On fixed-width confidence bounds for regression parameters and the mean vector. *J. Roy. Statist. Soc. Ser. B* **29** 132–140.

Starr, N. (1966). The performance of a sequential procedure for the fixed-width interval estimation of the mean. *Ann. Math. Statist.* **37** 36–50.

Venter, J. H. (1967). An extension of the Robbins–Monro procedure. *Ann. Math. Statist.* **38** 181–190.

Wetherill, G. B. and Glazebrook, K. D. (1986). *Sequential Methods in Statistics*, 3rd ed. Chapman and Hall, London.

Whitt, W. (1980). Some useful functions for functional limit theorems. *Math. Oper. Res.* **5** 67–85.

Whitt, W. (1989). Planning queueing simulations. *Management Sci.* **35** 1341–1366.

Whitt, W. (1992). Asymptotic formulas for Markov processes with application to simulation. *Oper. Res.* To appear.

Department of Operations Research
Stanford University
Stanford, California 94305-4022

AT & T Bell Laboratories
600 Mountain Avenue
Murray Hill, New Jersey 07974-0636