THE ASYMPTOTIC EFFICIENCY OF SIMULATION ESTIMATORS

by

Peter W. Glynn

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

and

Ward Whitt

AT&T Bell Laboratories
Murray Hill, NJ 07974-2070

October 17, 1986

Revision: January 17, 1991

To appear in *Operations Research*

ABSTRACT

A decision-theoretic framework is proposed for evaluating the efficiency of simulation estimators. The framework includes the cost of obtaining the estimate as well as the cost of acting based on the estimate. The cost of obtaining the estimate and the estimate itself are represented as realizations of jointly distributed stochastic processes. In this context, the efficiency of a simulation estimator based on a given computational budget is defined as the reciprocal of the risk (the overall expected cost). This framework is appealing philosophically, but it is often difficult to apply in practice (e.g., to compare the efficiency of two different estimators) because only rarely can the efficiency associated with a given computational budget be calculated. However, a useful practical framework emerges in a large sample context when we consider the limiting behavior as the computational budget increases. A limit theorem established for this model supports and extends a fairly well known efficiency principle, proposed by Hammersley and Handscomb (1964), p. 22: ‘‘The efficiency of a Monte Carlo process may be taken as inversely proportional to the product of the sampling variance and the amount of labour expended in obtaining this estimate.’’

Subject Classification: simulation, efficiency: definitions and asymptotic theory; simulation, statistical analysis: asymptotic efficiency; statistics, estimation: efficiency of simulation estimators.

In this paper we develop a framework for evaluating the efficiency of alternative simulation estimators. Our goal is to effectively capture the interplay between the variability of an estimator and the computational effort required to calculate it. We begin by developing what we regard as a philosophically appealing decision-theoretic model of estimation with a budget constraint. This model with a finite budget constraint may prove to be directly useful, but unfortunately the efficiency of an estimator is usually difficult to calculate. However, we obtain a useful framework in a large sample context by considering the limiting behavior as the computational budget increases. Thus, our primary focus is on asymptotic efficiency.

We believe that our asymptotic efficiency framework provides an effective means for comparing EITs (efficiency improvement techniques). Our analysis also supports replacing the classical notion of VRT (variance reduction technique) by EIT. An example in which the efficiency may be improved with higher variance occurs when we estimate the mean sojourn time in a queueing system: The sample variance is usually less if we use a direct sample mean than if we use an indirect estimator based on the number in system and $L = \lambda W$; see Glynn and Whitt (1989). However, it nevertheless may be more efficient to use the indirect estimator; see Nozari and Whitt (1988), p. 313. Another example in which an EIT is associated with higher sample variance occurs when estimating expected discounted costs; see Fox and Glynn (1989a).

In many simulation settings, our asymptotic efficiency framework provides theoretical support for an efficiency principle proposed without much discussion by Hammersley and Handscomb (1964), pp. 22, 51:

> The efficiency of a Monte Carlo process may be taken as inversely proportional to the product of the sampling variance and the amount of labour expended in obtaining this estimate. (1)

This efficiency principle is also cited on p. 35 of Bratley, Fox, and Schrage (1987) and p. 279 of Wilson (1985). (The present paper is an extensive revision of Glynn and Whitt (1986), where we first discussed (1).)

Efficiency principle (1) can be considered intuitively reasonable. However, as a byproduct of our analysis, we will see that in several different estimation settings that this criterion is not appropriate and, in fact, leads to incorrect conclusions. Nevertheless, in most simulation estimation problems, efficiency principle (1) does apply. In the context of such problems, our paper makes several contributions: First, we describe an appropriate domain of applicability for (1). Second, we give a precise interpretation to the terms ‘‘sampling variance’’ and ‘‘amount of labour expended’’. In addition, in (1) the ‘‘amount of labour expended’’ is apparently considered deterministic. Our analysis extends the principle to the setting in which the amount of labour expended is itself stochastic, which is typical of most simulations.

The rest of this paper is organized as follows. In Section 1 we introduce the decision-theoretic framework for estimation with a budget constraint. In Section 2 we introduce the concept of asymptotic efficiency of an estimator. In Section 3 we present a random-time-change limit theorem that provides the basis for characterizing the asymptotic efficiency of an estimator. The remaining sections are primarily devoted to examples illustrating how the asymptotic efficiency framework can be applied, but there also are some new asymptotic efficiency results for specific estimators.

In Section 4 we describe the canonical case, in which the asymptotic efficiency is consistent with (1), and discuss five examples. In Sections 5 and 6 we discuss examples in which (1) needs to be modified, because there is a non-canonical estimator convergence rate. Section 5 focuses on subcanonical convergence rates, while Section 6 focuses on supercanonical convergence rates. The examples of subcanonical convergence rate discussed in Section 5 are the Kiefer-Wolfowitz (1952) stochastic approximation algorithm, for which we draw on results of Ruppert (1982); a recursive variant of a derivative estimator discussed by Zazanis and

Suri (1986); other recursive estimators related to the replication schemes for limiting expectations in Fox and Glynn (1989b); and long-range dependency as discussed by Cox (1984). Supercanonical convergence rates are less likely to occur; the one example in Section 6 is a Monte Carlo integration rotation estimator, which is a variant of a rotation estimator in Fishman and Huang (1983).

In Section 7 we discuss independent replications together with other estimation procedures, and show that independent replications typically cause the efficiency to improve, remain unchanged or get worse, respectively, when the estimator convergence rate is subcanonical, canonical or supercanonical. Finally, we present all proofs in Section 8.

1. Efficiency with a Budget Constraint

Our decision-theoretic model for simulation estimation has eight elements:

(i) an unknown *parameter* α,

(ii) a *loss function* $L(a)$, a real-valued function specifying the loss associated with estimating α by a,

(iii) the *experiment*, a stochastic process $(Y, C) \equiv \{[Y(t), C(t)] : t \geq t_0\}$, $t_0 \geq 0$, with t representing simulated time (*not* computer time), $Y(t)$ representing the (simulated) *time-dependent estimator* of α and $C(t)$ representing the *cost of obtaining the estimator* $Y(t)$,

(iv) a *budget constraint* c,

(v) the realized *length of the experiment*, $T(c) = \sup\{t \geq 0 : C(t) \leq c\}$,

(vi) the *budget-constrained estimator* $Y(T(c))$,

(vii) the *risk function* $R(c) = EL(Y(T(c)))$,

(viii) the *efficiency* $e(c) = 1/R(c)$.

Our goal is to estimate the parameter α in (i). We assume that the parameter α is a real number, but the same ideas apply more generally.

We regard estimation as a special case of decision making under uncertainty, so we use the decision-theoretic framework advocated by Wald (1950), Savage (1954), and others; e.g., see Chapter 1 of Ferguson (1967). Of course, the loss function L in (ii) is actually a function of α as well as a, which may be important in a decision-theoretic analysis (e.g., in a Bayesian analysis using a prior on α), but we do not emphasize this aspect. We assume that L is nonnegative with $L(\alpha) = 0$. The classical squared error loss function arises when $L(a) = (a - \alpha)^2$, but we do not restrict attention to this case.

We have represented costs and benefits in two ways: via the loss function L and the cost process C. There are of course many different kinds of costs and benefits that might be considered. Many of these can easily be incorporated in L or C, but some cannot, e.g., the cost of the analyst's time; see p. 279 of Wilson (1985). Also, unexpected benefits beyond the original goals are often realized from simulation experiments. However, it is not our purpose to try to examine all costs and benefits in detail. We believe that the relatively simple two-cost framework above captures essential features for developing a useful efficiency principle, especially for evaluating alternative EITs.

Basic to our approach is the formulation of key features of the experiment as a stochastic process. In (iii) we have represented the estimator $Y(t)$ and the cost of generating that estimator $C(t)$ as jointly distributed stochastic processes. We refer to $Y(t)$ as the *estimation process* and $C(t)$ as the *cost process*. For example, Y might be a sample mean process; i.e., there might be another process Z such that $Y(t) = t^{-1} \int_0^t Z(s)\, ds$ for $t > 0$. We typically think of the cost $C(t)$ as being simply computer time, but there could be other cost components as well. We assume that the sample paths of C are nondecreasing nonnegative real-valued functions of t that are unbounded above. The randomness is appropriate because the cost associated with a given portion of the experiment is indeed often random. For example, when a sequential stopping procedure is used to terminate a simulation, the total number of observations generated will be random. It is important that we make *no assumption about the joint distribution* of Y and C, which in many applications will be quite complicated.

The experiment is assumed to evolve in ‘‘simulation time’’ t, but t could be any natural measure of the length of the experiment. We assume that the realized length of the experiment is $T(c)$ in (v), the ‘‘time’’ when the budget c in (iv) is exhausted. The final budget-constrained estimator is then $Y(T(c))$ in (vi). For example, in a regenerative simulation, t (for t integer) might represent the number of regenerative cycles, while the cost $C(t)$ is the random effort required to generate those cycles. The actual estimator $Y(T(c))$ then is based on the random number $T(c)$ of cycles achieved under the computational budget c.

Our realized length of the experiment $T(c)$ in (v) is the largest time such that the computational budget is not exceeded. The model could also be applied with other stopping rules. For example, an alternative rule that might be used if we wanted to determine $T(c)$ prior to performing the simulation would be to use the expected cost $EC(t)$ in (v) instead of the realized cost $C(t)$. Other alternative stopping rules could consider $EL(Y(t))$ as well $C(t)$. However, in this paper we restrict attention to (v).

We define the *efficiency* of the experiment for a given computational budget c as the reciprocal of the risk $R(c)$ in (vii). One experiment is said to be *more efficient* than another if its efficiency is greater. Of course, direct comparisons of this sort are usually difficult to make, because the efficiency is usually difficult to calculate.

2. Asymptotic Efficiency

Our goal now is to turn the philosophically appealing model of Section 1 into a practical basis for evaluating estimators by considering the asymptotic behavior as $c \to \infty$. The resulting notion of asymptotic efficiency is thus applicable only in a large sample context, but large samples are typical of most simulation experiments.

A concept of asymptotic efficiency emerges naturally from the notion of efficiency in Section 1. In particular, we say that one estimator is *more asymptotically efficient* than another if it is more efficient for all sufficiently large c. It is significant that when we focus on asymptotic efficiency, the analysis typically simplifies greatly. Then only the central-limit-theorem behavior of the estimation process Y and the law-of-large-numbers behavior of the cost process C matter; see §3. Moreover, the specific loss function often ceases to matter.

We establish conditions, which are often verifiable, under which

$$\lim_{c \to \infty} c^r R(c) = v^{-1} \tag{2}$$

for positive constants r and v. The pair (r, v) is our proposed characterization of asymptotic efficiency. We call r the *asymptotic efficiency rate* and v the *asymptotic efficiency value*. To compare two experiments with *asymptotic efficiency parameter pairs* (r_1, v_1) and (r_2, v_2), we use a *lexicographic criterion*. We say estimator 1 is more asymptotically efficient than estimator 2 if $r_1 > r_2$ or if $r_1 = r_2$ and $v_1 > v_2$. If $r_1 > r_2$, then we say that estimator 1 has a *more asymptotically efficient rate*. If $r_1 = r_2$, then we say that the *asymptotic relative efficiency* (ARE) of estimator 1 compared to estimator 2 is v_1 / v_2.

Note that the lexicographic criterion is consistent with the previous definition, i.e., estimator 1 is more asymptotically efficient than estimator 2 with the lexicographic criterion if and only if estimator 1 is more efficient than estimator 2 for all sufficiently large c.

It remains to show how the basic elements (L, Y, C) in the model of Section 1 determine asymptotic efficiency parameters r and v in (2). For this, we exploit a random-time-change limit theorem.

3. The Supporting Limit Theorem

We now establish a limit theorem for the budget constrained estimation process $Y(T(c))$ that provides a basis for characterizing the asymptotic efficiency of the experiment (Y, C). At first, we do not consider the loss function L.

Our key assumption for the estimation process Y is a functional central limit theorem (FCLT). For this purpose, let $D \equiv D((0, \infty), R)$ be the set of real-valued functions on the open interval $(0, \infty)$ that are right-continuous with left limits, endowed with the standard Skorohod J_1 topology, and let $\Rightarrow$ denote weak convergence (convergence in distribution); see Billingsley (1968), Ethier and Kurtz (1986) and Whitt (1980). (We use the open interval excluding 0 to avoid unimportant problems near the origin in estimators such as $t^{-1} \int_0^t Z(s)\, ds$.) For each $\varepsilon > 0$, let $=_\varepsilon \equiv \{ =_\varepsilon(t) : t > 0 \}$ be the random element of D defined by

$$=_\varepsilon(t) = \varepsilon^{-\gamma} [Y(t/\varepsilon) - \alpha] , \quad t > 0 , \tag{3}$$

for a positive constant γ. (We assume that Y is a random element of D.) We will assume that $=_\varepsilon \Rightarrow =$ in D as $\varepsilon \to 0$ for some limit process $=$, and write

$$\varepsilon^{-\gamma} [Y(t/\varepsilon) - \alpha] \Rightarrow =(t) \quad \text{in } D \quad \text{as } \varepsilon \to 0 . \tag{4}$$

For practical purposes, the FCLT (4) is essentially equivalent to the ordinary CLT in R obtained by focusing on a single t in (4), say $t = 1$, but an ordinary CLT is, technically speaking, slightly weaker than a FCLT. (See Example 1 of Glynn and Whitt (1988).) An easy consequence of (4) is that $Y(t) \xrightarrow{p} \alpha$ as $t \to \infty$, where $\xrightarrow{p}$ denotes convergence in probability. It is significant that (4) holds in great generality, so that this assumption is indeed typically satisfied.

In most cases, $\gamma = 1/2$ (the canonical convergence rate) and $= (t)$ is $\sigma t^{-1} B(t)$, where $B(t)$ is standard (zero-drift and unit variance) Brownian motion, so that $= (1)$ has a zero-mean Gaussian distribution with variance σ^2, but there are other possibilities.

We also will assume limiting behavior for the cost process C, but it is significant that we need to know much less about C. In particular, we only assume a simple stochastic growth condition corresponding to an ordinary strong law of large numbers (SLLN) for C; i.e., we will assume that

$$t^{-\beta} C(t) \to \lambda^{-\beta} \quad \text{w.p.1 as } t \to \infty \ , \tag{5}$$

where β is a positive constant. Typically $\beta = 1$, but we give examples in which $\beta \neq 1$; see Example 6.1.

It is significant that we do not directly assume anything about the joint behavior of Y and C. It turns out that we are able to establish the joint limiting behavior for Y and C, and thus the limiting behavior for the final budget-constrained estimator $Y(T(c))$, from these assumptions alone. Let $\stackrel{d}{=}$ denote equality in distribution.

Theorem 1. If the FCLT (4) holds for the estimation process $Y(t)$ with the limit process $= (t)$ being continuous at t w.p.1 for each t and the SLLN (5) holds for the cost process $C(t)$, then a FCLT holds for the budget-constrained estimation process $Y(T(c))$, i.e.,

$$c^{\gamma/\beta} [Y(T(ct) - \alpha] \Rightarrow = (\lambda\, t^{1/\beta}) \quad \text{in } D \quad \text{as } c \to \infty \ , \tag{6}$$

the associated CLT holds, i.e.,

$$c^{\gamma/\beta} [Y(T(c)) - \alpha] \Rightarrow = (\lambda) \stackrel{d}{=} \lambda^{-\gamma} = (1) \quad \text{in } R \quad \text{as } c \to \infty \ , \tag{7}$$

and the associated WLLN holds, i.e.,

$$Y(T(c)) \stackrel{p}{\to} \alpha \qquad as \ c \to \infty \ . \tag{8}$$

To prove Theorem 1, we first relate the SLLN for $C(t)$ in (5) to an associated SLLN for $T(c)$

in §1(v). This result is well known for $\beta = 1$, and essentially the same proof works for $\beta \neq 1$.

Lemma 1. Let λ and β be strictly positive constants. Then $t^{-\beta} C(t) \to \lambda^{-\beta}$ w.p.1 as $t \to \infty$ if and only if $c^{-1/\beta} T(c) \to \lambda$ w.p.1 as $c \to \infty$.

Next, as in Theorem 4 of Glynn and Whitt (1988), we note that the ordinary SLLN for $T(c)$ established in Lemma 1 is actually equivalent to a FSLLN (a functional version).

Lemma 2. If $c^{-1/\beta} T(c) \to \lambda$ w.p.1, then $c^{-1/\beta} T(ct) \to \lambda t^{1/\beta}$ w.p.1 in $D([0,\infty), R)$, i.e., $\sup_{0 \leq t \leq T} \{ | c^{-1/\beta} T(ct) - \lambda t^{1/\beta} | \} \to 0$ w.p.1 as $c \to \infty$ for all T.

Finally, we apply the continuous mapping theorem with the composition map, as in Section 17 of Billingsley and Section 3 of Whitt (1980); see Section 8 for the details.

Remarks (3.1) For our applications, we only use the CLT (7), but the FCLT (6) can be useful as well. The FCLT condition (4) is needed in Theorem 1 even to get the CLT (7); see Example 4 of Glynn and Whitt (1988). We could work with the CLT version of (4) instead of the FCLT if we added extra conditions, such as independence or the Anscombe (1952) condition; see p. 15 of Gut (1988). The Anscombe condition is closely related to the tightness associated with the FCLT; see p. 55 of Billingsley.

(3.2) Typically the limit process $= (t)$ in (4) is $\sigma t^{-1} B(t)$, where $B(t)$ is standard Brownian motion, which has continuous sample paths, but we do not require that the sample paths of $=$ be continuous. For example, B could be replaced by a stable process, which occurs as the limit for normalized partial sums of i.i.d. random variables when the random variables have infinite variance. For applications, see Mandelbrot (1963) and Fama (1963).

(3.3) In the vast majority of cases $= (1) \stackrel{d}{=} \sigma N(0,1)$, i.e., $= (1)$ has a centered Gaussian distribution with variance σ^2. In this case, the practical implication of the CLT (7) is that

$$Y(T(c)) \approx \alpha + \frac{\lambda^{-\gamma}\sigma}{c^{\gamma/\beta}} N(0,1) \text{ for large } c \; . \tag{9}$$

If all candidate experiments satisfy (9), then to achieve high asymptotic efficiency, without considering any loss functions, it is natural to first maximize the convergence rate γ/β and then minimize $\lambda^{-\gamma}\sigma$. Indeed, given the approximation (9), $|Y(T(c)) - \alpha|$, $[Y(T(c)) - \alpha]^+$ and $[Y(T(c)) - \alpha]^-$, where $x^+ = \max\{x,0\}$ and $x^- = -\min\{x,0\}$, are all minimized in the stochastic order sense with this criterion. Indeed, this is true provided the estimators satisfy $=_i(1) \stackrel{d}{=} \sigma_i Z$ for a fixed random variable Z.

(3.4) In (4) and (6) we consider limits as $\varepsilon \to 0$ and $c \to \infty$. This is not different in any essential way from considering limits involving a sequence, e.g., $\varepsilon = 1/n$ for n integer; see p. 16 of Billingsley. ■

We now consider the asymptotic behavior of the risk associated with a large class of loss functions. Unlike Remark 3.3, we now do not require that $=(1) \stackrel{d}{=} \sigma N(0,1)$. In order to get convergence of moments from convergence of distribution, we assume uniform integrability; there are many sufficient conditions; see p. 32 of Billingsley and Sections I.7,8 and II.5 of Gut (1988). In practice, we would rarely worry about this technical condition. (Uniform integrability has been checked in some cases, e.g., Examples 4.1 and 4.2.) A simple sufficient condition is for the loss function to be bounded.

Corollary 1. In addition to the assumptions of Theorem 1, suppose that the loss function L has two continuous derivatives with $L'(\alpha) = 0$ and $L''(\alpha) > 0$ and $\{c^{2\gamma/\beta} L(Y(T(c))) : c > 1\}$ is uniformly integrable. Then

$$\lim_{c \to \infty} c^{2\gamma/\beta} R(c) = 2^{-1} L''(\alpha) \lambda^{-2\gamma} E[=(1)^2] \; , \tag{10}$$

so that the asymptotic efficiency parameters are

$$r = \frac{2\gamma}{\beta} \quad \text{and} \quad v = \frac{2}{L''(\alpha)} \, \frac{1}{\lambda^{-2\gamma} E[= (1)^2]} \; . \tag{11}$$

Remarks (3.5) If L is twice differentiable and α is a strict local minimizer of L, then the conditions $L'(\alpha) = 0$ and $L''(\alpha) > 0$ must be satisfied.

(3.6) Suppose that we compare two estimators satisfying the assumptions of Corollary 1 with common β and γ, but with subcanonical convergence rate $\gamma < 1/2$ so that $r < 1$. Moreover, suppose that estimator 1's asymptotic mean square error is half that of estimator 2 (i.e., $E =_1 (1)^2 = E =_2 (1)^2/2$), while the cost rate is twice that of estimator 2 (i.e., $\lambda_1^{-1} = 2\lambda_2^{-1}$), then the asymptotic efficiency value of estimator 1 is greater than that of estimator 2 by a factor of $2^{1-2\gamma}$. This, of course, is inconsistent with the Hammersley-Handscomb efficiency criterion (which was clearly formulated with $\gamma = 1/2$ in mind). This analysis indicates that the variability of the estimator tends to be more important when there is subcanonical convergence. Thus, in the trade-off between variance reduction and (possible) additional computational complexity, variance reduction usually is top priority when $\gamma < 1/2$. ■

It is significant that *the form of the loss function does not affect asymptotic efficiency under the assumptions of Corollary 1*, because L appears in r and v only through the constant multiple $L''(\alpha)$ in v. In other words, if two candidate estimators satisfy the assumptions of Corollary 1, then our lexicographic efficiency criterion provides a ranking that is independent of the specific form of the loss function. Specifically, estimator 1 is more asymptotically efficient than estimator 2 if $\gamma_1/\beta_1 > \gamma_2/\beta_2$ or if $\gamma_1/\beta_1 = \gamma_2/\beta_2$ and $\lambda_2^{-2\gamma_2} E[=_2 (1)^2] > \lambda_1^{-2\gamma_1} E[Y_1(1)^2]$; note that the criterion is independent of L. Moreover, note that (11) is consistent with (9) when $= (1) \overset{d}{=} \sigma\, N(0, 1)$. Formulas (9) and (11) complement each other, because (9) applies to loss functions beyond those considered in Corollary 1, whereas (11) applies to non-Gaussian distributions (and distributions not all related by a scale transformation).

Of course, the loss function need not satisfy the conditions of Corollary 1, but other cases can be treated in a similar way. For example, here is another natural case.

Corollary 2. In addition to the conditions of Theorem 1, suppose that $L(a) = |a - \alpha|^p$ for $p > 0$. If $\{c^{p\gamma/\beta}|Y(T(c)) - \alpha|^p : c \geq 1\}$ is uniformly integrable, then

$$\lim_{c \to \infty} c^{p\gamma/\beta} R(c) = \lambda^{-p\gamma} E[|=(1)|^p] \ , \tag{12}$$

so that the asymptotic efficiency parameters are

$$r = \frac{p\gamma}{\beta} \text{ and } v = \frac{1}{\lambda^{-p\gamma} E[|=(1)|^p]} \ . \tag{13}$$

In general, (13) is not consistent with (11), so that the form of the loss function can matter. However, when β is fixed and $=(1) \stackrel{d}{=} \sigma N(0,1)$, (13) is fully consistent with (9) and (11), and $E[|=(1)|^p] = \sigma^p E[|N(0,1)|^p]$.

Remark (3.7) In general, the cost process $C(t)$ can affect the asymptotic efficiency rate r through β, but *in the canonical case of* $\beta = 1$, *the asymptotic efficiency rate is determined solely by the estimator convergence rate* γ. Hence, to achieve maximum asymptotic efficiency when $\beta = 1$, the first objective is to maximize the estimator convergence rate γ. Then, among those estimators with maximum estimator convergence rate, we want to maximize the asymptotic efficiency value. ■

Obviously the cost process usually grows linearly, so that $\beta = 1$ in (5). However, other cases do arise, as is illustrated here in Example 6.1. The following variant of the SLLN for martingale differences is a basis for establishing the required nonlinear SLLN for $C(t)$ when $C(t) = \sum_{i=1}^{\lfloor t \rfloor} \tau_i$, $t \geq 0$, where $\lfloor t \rfloor$ is the greatest integer less than or equal to t. We allow the variables τ_i to be dependent as well as non-identically distributed, but most of our applications are under the extra condition of independence. (For an exception, see Example 4.5.)

Theorem 2. Let $\{\tau_n : n \geq 1\}$ be a sequence of real-valued r.v.'s on an underlying probability space $(\Omega, \wedge, P)$ and let $\{\wedge_n : n \geq 1\}$ be an increasing sequence of sub-σ-fields of $\wedge$ such that τ_n is measurable with respect to $\wedge_n$. If

$$n^{-b} E(\tau_n | \wedge_{n-1}) \rightarrow a \quad \text{w.p.1} \quad as\ n \rightarrow \infty$$

and

$$n^{-d} \operatorname{Var}[\tau_n - E(\tau_n | \wedge_{n-1})] \rightarrow c \quad as\ n \rightarrow \infty$$

with $b > -1$ and $d < 2b + 1$, then

$$n^{-b-1} \sum_{i=1}^{n} \tau_i \rightarrow a/(1 + b) \quad \text{w.p.1} \quad as\ n \rightarrow \infty\ .$$

4. The Canonical Case

The canonical case arises when the conditions of Corollary 1 to Theorem 1 hold with $\gamma = 1/2$, $\beta = 1$ and $= (t) \overset{\mathrm{d}}{=} t^{-1} \sigma B(t)$ for each t where B is standard Brownian motion, so that $= (1) \overset{\mathrm{d}}{=} \sigma N(0,1)$, i.e., $= (1)$ is distributed as a zero-mean Gaussian distribution with variance σ^2. From (11), the asymptotic efficiency rate is then $r = 1$ and all interest centers on the asymptotic value, which is

$$v = \frac{K}{\lambda^{-1} \sigma^2} \qquad \text{for} \qquad K = \frac{2}{L''(\alpha)} \tag{14}$$

and λ^{-1} the cost rate in (5). We interpret (14) as support for (1):

Asymptotic Efficiency Principle. *In the canonical case, the asymptotic efficiency value v may be taken as inversely proportional to the product of the sampling variance rate* σ^2 *and the cost rate* λ^{-1}.

This interpretation for the cost rate λ^{-1} is clear from (5). For σ^2, note that under the conditions of Corollary 1 to Theorem 1 (which includes uniform integrability) that

t Var $Y(t) \to \sigma^2$ as $t \to \infty$, so that Var $Y(t) \approx \sigma^2/t$. In this setting, we can take the asymptotic value v as being inversely proportional to the product of the sampling variance Var $Y(t) = \sigma^2/t + o(t^{-1})$ and the cost $C(t) = \lambda^{-1}t + o(t)$.

The rest of this paper is primarily devoted to examples. The five examples in this section all produce the canonical case, for which the asymptotic efficiency principle above and (1) are appropriate.

Example 4.1. Independent Replications. Suppose that α can be represented as $\alpha = EX$ for some random variable (r.v.) X. (X might correspond to the number of customers served in a queue during the time interval $[a, b]$.) Suppose that α can be estimated by the sample mean $\bar{X}_n = n^{-1} \sum_{i=1}^{n} X_i$, where $X_1, X_2, \ldots$ are i.i.d. copies of X. Then the estimation process here is $Y(t) = \bar{X}_{\lfloor t \rfloor}$, $t \geq 1$ (where again $\lfloor t \rfloor$ is the greatest integer less than or equal to t).

Let τ_i be the amount of computer time required to generate X_i. Assuming that we can disregard the amount of computer time required to initialize the simulation and compute $Y(t)$ from the X_i's (which is often, but not always, appropriate), we let the cost process be $C(t) = \sum_{i=1}^{\lfloor t \rfloor} \tau_i$. It seems reasonable to assume that the τ_i's are positive i.i.d. r.v.'s. For most applications, τ_i will indeed be random. For example, in any algorithm in which acceptance/rejection is used as a variate generation technique, τ_i will be random.

If $0 < \sigma^2 < \infty$, where $\sigma^2 = \text{Var } X$, then the FCLT (4) holds for $Y(t)$ with $\gamma = 1/2$ with $= (t) = t^{-1}\sigma B(t)$ with B being standard Brownian motion, by Donsker's theorem, p. 137 of Billingsley (1968). If $0 < E\tau_1 < \infty$, then the SLLN (5) holds for $C(t)$ with $\beta = 1$. Hence, the conditions of Theorem 1 hold for the canonical case. Moreover, for this example, these assumptions also imply the uniform integrability needed for Corollary 1; see pp. 32, 54 of Gut (1988).

Example 4.2. Functions of Mean Vectors and Regenerative Simulation. Let $X \equiv [X(1), \ldots, X(d)]$ be an R^d-valued random vector with $\mu = EX$. Suppose that $\alpha = g(\mu)$ for some known smooth function $g : R^d \to R$. In this case the estimation process is $Y(t) = g(\bar{X}_{\lfloor t \rfloor})$, $t \geq 1$, where $\bar{X}_n$ is the sample mean of i.i.d. random vectors distributed as X. This estimation process arises with *ratio estimators;* then $d = 2$, $g(x_1, x_2) = x_1 / x_2$ and $\alpha = EX(1)/EX(2)$. A ratio estimator is often used with the *regenerative simulation method* to calculate the steady-state mean of a real-valued regenerative process Z. Then $X(1)$ is the integral of Z over a regenerative cycle and $X(2)$ is the duration of the cycle.

Let the cost of generating the i^{th} cycle be τ_i and let the cost process $C(t)$ be defined as in Example 4.1. For example, if the computational time to generate the cycle can be regarded as approximately proportional to the length of the cycle, then $\tau_i \stackrel{d}{=} \xi X(2)$ for some constant ξ, which is of course typically random.

The following theorem establishes the FCLT condition in Theorem 1, without requiring that the random vectors X_i actually be i.i.d. (We apply this extension in Example 4.4 below.)

Theorem 3. Suppose that $\gamma > 0$, $\mu \in R^d$ and

$$\varepsilon^{-\gamma} [\psi_\varepsilon(t) - \mu] \Rightarrow \psi(t) \text{ in } D \text{ as } \varepsilon \to 0 ,$$

where ψ_ε and ψ are random elements of D. If $g : R^d \to R^1$ is continuously differentiable in a neighborhood of μ, then

$$\varepsilon^{-\gamma} [g(\psi_\varepsilon(t)) - g(\mu)] \Rightarrow \nabla g(\mu) \psi(t) \quad \text{in } D \quad \text{as } \varepsilon \to 0 .$$

Returning to our example, we assume that $E|X|^2 < \infty$, where $|\cdot|$ is the Euclidean norm in R^d. Then the multivariate version of Donsker's theorem implies that

$$n^{-1/2} \Big(\sum_{i=1}^{\lfloor nt \rfloor} X_i - \mu \Big) \Rightarrow \Gamma^{1/2} B(t) \quad \text{in } D \quad \text{as } n \to \infty ,$$

where B is a standard Brownian motion in R^d (with d mutually independent 1-dimensional marginal standard Brownian motions) and Γ is the covariance matrix of X. The matrix $\Gamma^{1/2}$ is not uniquely specified by Γ, but may be taken as the lower triangular matrix obtained by Cholesky factorization; see p. 84 of Feller (1971) and p. 165 of Bratley, Fox and Schrage. Hence,

$$\varepsilon^{-1/2}\left[\frac{1}{\lfloor t/\varepsilon \rfloor}\sum_{i=1}^{\lfloor t/\varepsilon \rfloor} X_i - \mu\right] \Rightarrow \Gamma^{1/2} B(t)/t \quad \text{in } D \quad \text{as } \varepsilon \to 0 \ .$$

Finally, by Theorem 3, if g is continuously differentiable in the neighborhood of μ, then $=_\varepsilon \Rightarrow =$ in D where $=(t) = \nabla g(\mu)\Gamma^{1/2} B(t)/t$. The remaining conditions in Theorem 1 and Corollary 1 hold as in Example 4.1. (For the uniform integrability, it suffices to treat the marginals separately.) Then we have the canonical case, i.e., the limits in (7) and (8) with $\gamma = 1/2$, $\beta = 1$, $\lambda^{-1} = E\tau_1$, $=(1) \stackrel{\mathrm{d}}{=} \sigma N(0,1)$ and

$$\sigma^2 = \nabla g(\mu)\Gamma\nabla g(\mu)^t \ . \tag{15}$$

Example 4.3. Steady-State Means. Suppose that α is the steady-state mean of a real-valued stochastic process $X \equiv \{X(t) : t \geq 0\}$, and suppose that we intend to estimate α with the sample mean process $Y(t) = \bar{X}(t) = t^{-1}\int_0^t X(s)\,ds$, $t > 0$. Thus, we assume a FCLT for the cumulative process associated with X, i.e.,

$$\varepsilon^{1/2}\left[\int_0^{t/\varepsilon} X(s)\,ds - \alpha t/\varepsilon\right] \Rightarrow \sigma B(t) \quad \text{in } D \quad \text{as } \varepsilon \to 0 \tag{16}$$

which, as in Example 4.2, immediately implies $=_\varepsilon \Rightarrow =$ as $\varepsilon \to 0$ where $=(t) = \sigma B(t)/t$. A variety of different assumptions on the structure of X give rise to such a FCLT. For example, there are FCLTs of the form (16) when X is stationary and satisfies a mixing condition (e.g., Section 20 of Billingsley), when X is regenerative (e.g., Glynn and Whitt (1987)) and when X is a martingale (e.g., Chapter 7 of Ethier and Kurtz (1986)). The great variety of very robust hypotheses which lead to FCLTs of the form (16) lead us to view (16) as a very general assumption, which can be expected to hold for virtually all ‘‘real world’’ steady-state simulations.

Remark (4.1) Suppose that $X(t) \Rightarrow X(\infty)$ as $t \to \infty$. It is important to note that it is typically *not* the case that $\sigma^2 = \text{Var } X(\infty)$ for σ in (16). The constant σ^2 in (16) is the *time-average variance constant* of X, which reflects the correlation structure of X. In particular, if X is a uniformly integrable stationary stochastic process having an integrable covariance function, then

$$\sigma^2 = 2\int_0^\infty \text{cov}(X(0), X(t))\, dt . \tag{17}$$

Formulas for σ^2 when X is a function of a Markov process appear in Glynn (1984), Whitt (1991) and references cited there. The time-average variance constant σ^2 is difficult to estimate. Consequently, much attention has been devoted in the simulation literature to its estimation; see Section 3.3 of Bratley, Fox, and Schrage (1987). ■

Turning to the process $C(t)$, we assume a SLLN of the form (5) holds with $\beta = 1$. As in the case of assumption (16), a wide variety of steady-state simulations possess behavior that is characterized by such a SLLN. For example, suppose that the process X takes the form $X(t) = f(Z(t))$ for some real-valued function f. One then simulates X by simulating Z. It seems reasonable to assume that $C(t) = \int_0^t h(Z(s))\, ds$ for some nonnegative real-valued h. If Z is a positive recurrent regenerative process, then (5) is known to hold with $\beta = 1$ under suitable moment conditions. Similarly, if $Z(t)$ is stationary and ergodic, then so is $h(Z(t))$, so that we obtain (5) with $\beta = 1$ if $Eh(Z(t)) < \infty$. As with (16), we view (5) as a relatively mild regularity hypothesis on the steady-state simulation. In this example the constant λ can be viewed as the rate at which simulation time is produced as a function of computation time.

Example 4.4. Functions of Steady-State Means. As a generalization of Examples 4.2 and 4.3, suppose that $\alpha = g(\mu)$ where μ is the steady-state mean of an R^d-valued process X, and that $Y(t) = g(\overline{X}(t))$ where $\overline{X}(t)$ is the sample mean. This kind of estimator arises in calculating the steady-state conditional probability

$$P(X(t) \in A | X(t) \in B) = P(X(t) \in AB)/P(X(t) \in B)$$

and in estimating the steady-state variance. (In this case, g is again the ratio functional $g(x_1, x_2) = x_1/x_2$.) We can combine a d-dimensional analog of (16) with Theorem 3 to obtain the desired FCLT $=_\varepsilon \Rightarrow =$ as $\varepsilon \to 0$. As in Example 4.2, if $=(1) = \nabla g(\mu)\Gamma^{1/2}B(t)/t$, then σ^2 is given by (15). The cost can be treated as in Example 4.3.

Remark (4.2) As in Remark 4.1, here Γ is not the covariance of the steady-state variable $X(\infty)$. For a reasonably behaved stationary process, Γ can be represented as

$$\Gamma = \int_0^\infty E(X(0) - \mu)^t (X(s) - \mu)\, ds + \int_0^\infty E(X(s) - \mu)^t (X(0) - \mu)\, ds \ .$$

As in the scalar case, Γ is hard to estimate.

Example 4.5. The Robbins-Monro Stochastic Approximation Algorithm. To depart from the familiar sample mean setting, we now briefly consider the Robbins-Monro (1951) stochastic approximation algorithm (denoted by RM) which is finding application in simulation; see Wasan (1969), Kushner and Clark (1978), and Glynn (1986). Our goal is to find the parameter $\alpha = \theta^*$ that minimizes a smooth function $\beta(\theta)$. We assume that there exist random variables $Z(\theta)$ such that $\beta'(\theta) = EZ(\theta)$. To calculate θ^*, we use the RM algorithm $\theta_{n+1} = \theta_n - c_n X_{n+1}$, $n \geq 0$, where $\{c_n : n \geq 0\}$ is a sequence of deterministic nonnegative constants and X_{n+1} is independently generated, conditional on θ_n, i.e.,

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

Then the estimation process is defined by $Y(t) = \theta_{\lfloor t \rfloor}$, $t \geq 0$. We assume that the time τ_{n+1} required to calculate θ_{n+1} from θ_n has a conditional cdf

$$P(\tau_{n+1} \leq t | \tau_0, \theta_0, \ldots, \tau_n, \theta_n) = F_{\theta_n}(t)$$

for some family of cdf's $F_\theta(t)$ indexed by θ. Then $C(t) = \sum_{i=1}^{\lfloor t \rfloor} \tau_i$, $t \geq 0$.

Now assume that $c_n = c/n$ for $c > 0$ and that β is continuously differentiable with $c\beta'(\theta^*) > 1/2$. Kersting (1977) and Ruppert (1982) have shown that under mild additional regularity assumptions that $=_\varepsilon \Rightarrow =$ as $\varepsilon \to 0$, as needed for Theorem 1, with $\gamma = 1/2$ and

$$= (t) = \sigma t^{-(D+1)} B(t^{2D+1}) \stackrel{d}{=} \sigma t^{-1} B(t) \;, \tag{18}$$

where $D = c\beta'(\theta^*) - 1$, $\sigma^2 = c^2 \kappa^2 (2D + 1)^{-1}$ and $\kappa^2 = \text{Var}\, Z(\theta^*)$. Hence, together with (5), (18) implies that the FCLT (6) holds.

To establish (5) with $\beta = 1$ we can apply Theorem 2. For this purpose, let

$$\lambda^{-1}(\theta) = \int_0^\infty t dF_\theta(t) \quad \text{and} \quad \sigma^2(\theta) = \int_0^\infty [t - \lambda^{-1}(\theta)]^2 dF_\theta(t) \;. \tag{19}$$

If $\sup\{\sigma^2(\theta) : \theta \in R\} < \infty$ and λ^{-1} is continuous in the neighborhood of θ^*, then the conditions of Theorem 2 hold with $b = 0$ and $d = 1$, because of the well known convergence $\theta_n \to \theta^*$ w.p.1. Hence, $t^{-1} C(t) \to \lambda^{-1}$ w.p.1 as $t \to \infty$.

5. Subcanonical Estimator Convergence Rates

In this section we consider examples in which FCLTs hold for the estimation process, but with a rate $\gamma < 1/2$. Hence, the cost rate λ^{-1} appears in the asymptotic efficiency value v in (11) raised to the power $2\gamma < 1$, so that principle (1) needs to be modified as indicated in Remark 3.6. These examples with subcanonical convergence rates are leading candidates for VRTs.

Example 5.1. The Kiefer-Wolfowitz Stochastic Approximation Algorithm. Unlike the RM stochastic approximation algorithm in Example 4.5, the Kiefer-Wolfowitz (1952) stochastic approximation algorithm (denoted by KW) yields a subcanonical estimator convergence rate. The subcanonical convergence rate occurs because now we must estimate derivatives with finite differences. As before, our goal is to find a parameter $\alpha = \theta^*$ that minimizes a smooth function $\beta(\theta)$. Now we assume that $\beta(\theta)$ can be represented as $\beta(\theta) = EZ(\theta)$. Successive estimates of θ^* are $\theta_{n+1} = \theta_n - c_n X_{n+1}$ where $\{c_n : n \geq 0\}$ is a sequence of deterministic constants and

X_{n+1} is independently generated conditional on θ_n, i.e.,

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

where $Z(\theta_n + h_{n+1})$ and $Z(\theta_n - h_{n+1})$ are independently generated. As in Example 4.5, the estimation process is $Y(t) = \theta_{\lfloor t \rfloor}$, $t \geq 0$. Suppose that the constants c_n and h_n are chosen to be of the form $c_n = cn^{-1}$ and $h_n = hn^{-1/3}$ for $c > 0$ and $h > 0$. Assume that β is three times continuously differentiable on R and that θ^* is the unique solution of $\beta'(\theta) = 0$. We further require that c satisfy $c\beta''(\theta^*) > 1/3$. Then Ruppert (1982) shows that, under mild additional regularity conditions,

$$n^{1/3}(Y(nt) - \alpha) \Rightarrow \sigma\, t^{-b}\, B(t^{2A+1}) \quad \text{in } D \quad \text{as } n \to \infty\ , \tag{20}$$

where $b = c\beta''(\theta^*)$, $A = b - 5/6$, $\sigma^2 = c^2\kappa^2/(2A + 1)(4h^2)$, $\kappa^2 = 2 \text{ Var } Z(\theta^*)$, and B is a standard Brownian motion.

The cost process $C(t)$ can be treated as in Example 4.5, so that under the regularity conditions there, (5) holds with $\beta = 1$. Combining (5), (20) and Theorem 1 we obtain

$$c^{1/3}(Y(T(c)) - \alpha) \Rightarrow \sigma\lambda^{-b}B(\lambda^{2A+1}) \stackrel{d}{=} \sigma\lambda^{-1/3}N(0,1)\ . \tag{21}$$

The limit in (21) is a centered Gaussian as in the examples of Section 4, but the asymptotic efficiency rate in (11) is $r = 2/3$ and the asymptotic efficiency value v is inversely proportional to $\lambda^{-2/3}\sigma^2$. The non-canonical estimator convergence rate leads to the cost rate λ^{-1} in v in (11) being raised to the power $2\gamma \neq 1$.

Example 5.2. Recursive Derivative Estimators. Suppose that our goal is to estimate $\alpha = \beta'(\theta_0)$ where $\beta(\theta)$ is a smooth function of θ which can be represented as $\beta(\theta) = EZ(\theta)$ for each θ in an open interval about θ_0. We can estimate α via the sample mean $\overline{X}_n = n^{-1} \sum_{k=1}^{n} X_k$, where the X_k are independently generated, with X_k being the random forward difference

$$X_k = [Z_k(\theta_0 + h_k) - Z_k(\theta_0)]/h_k \ , \ k \geq 1 \ ,$$

and $Z_k(\theta_0 + h_k)$ and $Z_k(\theta_0)$ are independently generated. The resulting estimation process is $Y(t) = \bar{X}_{\lfloor t \rfloor}$, $t \geq 0$. This estimator is a recursive version of a derivative estimator studied by Zazanis and Suri (1986). Their estimator is $n^{-1} \sum_{k=1}^{n} X_{k,n}$ where $X_{k,n} = [Z_k(\theta_0 + h_n) - Z_k(\theta_0)]/h_n$. In contrast to Zazanis and Suri's estimator, note that we can easily compute our $\bar{X}_{n+1}$ from $\bar{X}_n$ by setting

$$\bar{X}_{n+1} = (n\bar{X}_n + X_{n+1})/(n + 1) \ ,$$

but our estimator is harder to analyze because the random variables X_k are not identically distributed. Thus this example is not a special case of Example 4.1.

Suppose that $\tau(\theta)$ is the computer time required to calculate $Z(\theta)$ and ϕ_k is the time required to compute X_k from $Z_k(\theta_0 + h_k)$ and $Z_k(\theta_0)$. Then a reasonable approximation for the cost process might be

$$C(t) = \sum_{k=1}^{\lfloor t \rfloor} (\tau_k(\theta_0 + h_k) + \tau_k(\theta_0) + \phi_k) \ , \quad t \geq 0 \ ,$$

where the r.v.'s $\tau_k(\theta_0 + h_k)$ and $\tau_k(\theta_0)$ are independently generated. As in Example 4.5, it is possible to impose conditions so that we can apply Theorem 2 to obtain (5) with $\beta = 1$.

In order to apply Theorem 1 to characterize the asymptotic efficiency, we establish a FCLT for the estimation process Y. The limit process is of particular interest because it is *not* centered. Hence, the approach to asymptotic efficiency in Remark 3.3 is not possible.

Theorem 4. Suppose that $Z(\theta) \Rightarrow Z(\theta_0)$ as $\theta \to \theta_0$ and that $\{Z(\theta)^2 : \theta_0 - \varepsilon \leq \theta \leq \theta_0 + \varepsilon\}$ is uniformly integrable. If β is twice continuously differentiable in $(\theta_0 - \varepsilon, \ \theta_0 + \varepsilon)$ and $h_k = hk^{-1/4}$ with $h > 0$, then

$$\varepsilon^{-1/4}(Y(t/\varepsilon) - \alpha) \Rightarrow \frac{\kappa B(t^{3/2})}{t} + \frac{\eta}{t^{1/4}} \text{ in } D \text{ as } \varepsilon \to 0 ,$$

where $\kappa^2 = 4\text{Var } Z(\theta_0)/3h^2$, $\eta = 2\beta''(\theta_0)h/3$ and $B(t)$ is standard Brownian motion.

Hence, under the conditions of Theorem 4 and (5),

$$c^{1/4}(Y(T(c)) - \alpha) \Rightarrow \frac{\kappa B(\lambda^{3/2})}{\lambda} + \frac{\eta}{\lambda^{1/4}} \stackrel{d}{=} \lambda^{-1/4} N(\eta, \kappa^2) \quad \text{in } R \quad \text{as } c \to \infty$$

and, under the extra conditions of Corollary 1 to Theorem 1,

$$\lim_{c \to \infty} c^{1/2} R(c) = \frac{L''(\alpha)}{2} \lambda^{-1/2}(\kappa^2 + \eta^2) . \tag{22}$$

Now the asymptotic efficiency principle in Section 4 needs to be modified in three ways: First, the asymptotic efficiency rate in (11) is $r = 1/2$ instead of $r = 1$; second, the asymptotic cost rate λ^{-1} appears in the value v in (11) in the form $\lambda^{-1/2}$ instead of λ^{-1}; and, third, the variance has to be replaced by the second moment.

An important question that arises in this setting is the choice of the constant h that determines the difference increment $h_k = h\, k^{-1/4}$ used in the k^{th} finite-difference approximation X_k. For example, in the setting of (22), we want to minimize the second moment of the limiting normal distribution,

$$\kappa^2 + \eta^2 = \frac{4 \text{ Var } Z(\theta_0)}{3\, h^2} + \frac{4\beta''(\theta_0)^2 h^2}{9} . \tag{23}$$

By differentiating, we see that the value of h that minimizes (23) is

$$h^* = \frac{3 \text{ Var } Z(\theta_0)}{\beta''(\theta_0)^2} . \tag{24}$$

This analysis based on (22) is equivalent to using a squared error loss function. If, instead, the loss function were $L(a) = |a - \alpha|^p$ for $p \neq 2$, then we would want to minimize the pth absolute moment of the limiting Gaussian distribution, which typically leads to a different minimizing value h^*. Thus, when the loss function does not satisfy the conditions of Corollary 1 to Theorem 1 and the limiting distribution = (1) is not centered Gaussian, the form of the loss function can affect asymptotic efficiency.

Example 5.3. More General Recursive Estimators. We now consider a generalization of Example 5.2 that includes certain replication schemes for limiting expectations in Fox and Glynn (1989b). Suppose that the parameter α can be represented as the limit of EX_n for a sequence of random variables $\{X_n : n \geq 1\}$. The proposed estimator is the sample mean $\bar{X}_n$ where the r.v.'s X_i are taken to be independent (but typically not identically distributed). Then the estimation process is $Y(t) = \bar{X}_{\lfloor t \rfloor}$, $t \geq 1$.

As in Example 5.2, let $\hat{X}_k = X_k - EX_k$. We characterize the limiting behavior by relating the asymptotic behavior of $E\hat{X}_n^2$ and $EX_n - \alpha$ as $n \to \infty$. There are three cases, one of which involves a non-centered Gaussian limit, as in Example 5.2.

Theorem 5. Suppose that $\{n^{2\gamma-1}\hat{X}_n^2 : n \geq 1\}$ is uniformly integrable and $n^{2\gamma-1}E\hat{X}_n^2 \to \sigma^2$ as $n \to \infty$ for $0 < \gamma \leq 1/2$. Suppose that $n^{\eta}(EX_n - \alpha) \to b$ as $n \to \infty$, $0 < \eta < 1$.

(1) If $\eta > \gamma$, then

$$\varepsilon^{-\gamma}(Y(t/\varepsilon) - \alpha) \Rightarrow \left[\frac{\sigma^2}{2-2\gamma}\right]^{1/2} \frac{B(t^{2-2\gamma})}{t} \quad \text{in } D \quad \text{as } \varepsilon \to 0 .$$

(b) If $\eta = \gamma$, then,

$$\varepsilon^{-\gamma}(Y(t/\varepsilon) - \alpha) \Rightarrow \left[\frac{\sigma^2}{2-2\gamma}\right]^{1/2} \frac{B(t^{2-2\gamma})}{t} + \frac{bt^{-\gamma}}{1-\gamma} \quad \text{in } D \quad \text{as } \varepsilon \to 0 .$$

(c) If $\eta < \gamma$, then

$$\varepsilon^{-\eta}(Y(t/\varepsilon) - \alpha) \xrightarrow{p} \frac{bt^{1-\eta}}{1-\eta} \quad \text{in } D \quad \text{as } \varepsilon \to 0 .$$

The following corollary describes the combination of Theorem 5 with a SLLN for the cost process. Motivated by Fox and Glynn (1989b), we allow nonlinear growth.

Corollary. In addition to the assumptions of Theorem 5, suppose that the cost of generating X_n is τ_n where $\{\tau_n\}$ is an independent sequence with $n^{-\rho} E\tau_n \to a$ and $n^{-\beta}$ Var $\tau_n \to d$ where $\rho > -1, a > 0, \beta < 2\rho + 1$. Then

$$n^{-\rho-1} \sum_{i=1}^{n} \tau_i \to \frac{a}{1+\rho} \quad \text{w.p.1} \quad \text{as } n \to \infty \ .$$

(a) If $\eta > \gamma$, then

$$c^{\gamma/(1+\rho)}(Y(T(c)) - \alpha) \Rightarrow \left[\frac{\sigma^2}{2-2\gamma}\right]^{1/2} \left[\frac{a}{1+\rho}\right]^{\gamma/(1+\rho)} N(0,1) \quad \text{in } R \quad \text{as } c \to \infty \ .$$

(b) If $\eta = \gamma$, then

$$c^{\gamma/(1+\rho)}(Y(T(c)) - \alpha) \Rightarrow \left[\frac{\sigma^2}{2-2\gamma}\right]^{1/2} \left[\frac{a}{1+\rho}\right]^{\gamma/(1+\rho)} N(0,1)$$
$$+ \left[\frac{a}{1+\rho}\right]^{\gamma/(1+\rho)} \frac{b}{1-\gamma} \quad \text{in } R \quad \text{as } c \to \infty \ .$$

(c) If $\eta < \gamma$, then

$$c^{\eta/(1+\rho)}(Y(T(c)) - \alpha) \Rightarrow \left[\frac{a}{1+\rho}\right]^{\eta/(1+\rho)} \frac{b}{1-\eta} \quad \text{in } R \quad \text{as } c \to \infty \ .$$

Typically $\gamma < 1/2$, so that $\gamma/(1+\rho) < 1/2$ and the convergence rate of both $Y(t)$ and $Y(T(c))$ is subcanonical. However, Theorem 5 and its Corollary also cover the canonical convergence when $\gamma = 1/2$ and $\rho = 0$.

Although not stated in the full generality of the results in Fox and Glynn (1989b), because the results there permit non-polynomial growth rates, Theorem 5 and its Corollary provides improvements by treating recursively defined estimators and allowing the τ_k's to be random.

Example 5.4. Long-Range Dependence. Subcanonical estimator convergence rates also can arise in the estimation of steady-state means as in Example 4.3 when there is long-range dependence in the underlying stochastic process X. Instead of (16), a FCLT may hold with $\gamma < 1/2$. For examples of long-range dependence, see Mandelbrot (1977), Taqqu (1982), Cox (1984), Vervaat (1985) and references cited there.

6. Supercanonical Estimator Convergence Rates

In this section we give an example of an estimator having a supercanonical convergence rate. This estimator may have some intrinsic interest, but we present it primarily to illustrate how supercanonical convergence rates can arise. Moreover, the example is interesting because some thought is required to determine just what the convergence rate is.

In particular, we consider a variant of a rotation estimator proposed by Fishman and Huang (1983). We explain the supercanonical convergence by showing the connection to numerical integration using the rectangular rule, as discussed on p. 53 of Davis and Rabinowitz (1984). This numerical integration, being deterministic, is not bound by the canonical convergence rate associated with a stochastic CLT. Indeed, for numerical integration alone, $\gamma = 1$ in (4).

The example we consider is only one-dimensional. Of course, deterministic numerical integration tends to perform less well as the dimension increases; e.g., deterministic schemes with cubic grid lattices tend to have convergence rate $\gamma = 1/d$ in dimension d; e.g., see Cheng and Davenport (1989).

Example 6.1. Monte Carlo Integration with Rotation. Our goal is to estimate $\alpha = \int_0^1 f(x)\,dx$. Note that $\alpha = Ef(U)$ where U is uniformly distributed on $[0,1]$ and that $U \oplus x$ is also uniformly distributed on $[0,1]$ for any x, where $\oplus$ denotes addition modulo one. Hence, we obtain an unbiased *recursive rotation estimator* by setting $Y(t) = Y_{\lfloor t \rfloor}$, where

$$Y_n = \frac{2}{n(n+1)} \sum_{k=1}^{n} X_k \ , \quad n \geq 1 \ , \tag{25}$$

and

$$X_k = \sum_{j=0}^{k-1} f\left[U_k \oplus \frac{j}{k} \right] \ , \quad k \geq 1 \ . \tag{26}$$

As with Examples 4.5, 5.1 and 5.2, recursive estimators are useful for developing sequential procedures.

The asymptotic behavior of X_k alone as $k \to \infty$ can be regarded as a stochastic analog of a well known theorem for Riemann sum (rectangular rule) approximation of Riemann integrals; see p. 53 of Davis and Rabinowitz (1984).

Theorem 6. Suppose that the derivative f' of f exists, is bounded and is Riemann integrable. Then

$$(X_k - k\alpha) \Rightarrow (f(1) - f(0))(U - \frac{1}{2}) \quad \text{in } R \text{ as } k \to \infty$$

and $\{|X_k - k\alpha|^p : k \geq 1\}$ is uniformly integrable for all $p > 0$ so that

$$E(|X_k - k\alpha|^p) \to |f(1) - f(0)|^p \, E[|U - \frac{1}{2}|^p] \quad \text{as } k \to \infty \ . \tag{27}$$

Now we obtain the FCLT for the estimator Y. Let $\lceil x \rceil$ be the smallest integer greater than x. (This is nonstandard notation; for x integer, $\lfloor x \rfloor = x$ and $\lceil x \rceil = x + 1$.)

Theorem 7. Under the conditions of Theorem 6 with X_k in (26),

$$n^{-1/2} \left[\sum_{k=1}^{\lfloor nt \rfloor} X_k - \lfloor nt \rfloor \lceil nt \rceil \frac{\alpha}{2} \right] \Rightarrow \frac{|f(1) - f(0)|}{2\sqrt{3}} \ B(t) \text{ in } D([0,\infty), R) \text{ as } n \to \infty \ ,$$

so that

$$\varepsilon^{-3/2}(Y(t/\varepsilon) - \alpha) \Rightarrow \frac{|f(1) - f(0)|}{\sqrt{3}} \frac{B(t)}{t^2} \quad \text{in } D((0,\infty), R) \quad \text{as } \varepsilon \to 0 .$$

Let τ_k be the r.v. representing the time to generate X_k. Since k function evaluations are required to generate X_k, it is reasonable to assume that $E\tau_k = ak + b$ and Var $\tau_k = dk + e$ for nonnegative constants a, b, d and e. (We expect d and e to be small, but they could be large. For example, with Monte Carlo Bayesian updating, some observations may be infeasible under the prior, so that the integrand is zero without further calculation, whereas an integrand associated with a feasible observation may involve a complicated computation.) Then, by Theorem 2,

$$n^{-2} \sum_{k=1}^{n} \tau_k \to a \text{ w.p.1 as } n \to \infty . \tag{28}$$

Corollary. With (28) and the FCLT for Y in Theorem 7,

$$c^{3/4}(Y(T(ct)) - \alpha) \Rightarrow \frac{|f(1) - f(0)|}{\sqrt{3}} \frac{B(\sqrt{t/a})}{t/a} \quad \text{in } D \quad \text{as } c \to \infty ,$$

$$c^{3/4}(Y(T(c)) - \alpha) \Rightarrow \frac{a^{3/4}|f(1) - f(0)|}{\sqrt{3}} N(0,1) \quad \text{in } R \text{ as } c \to \infty$$

and, assuming a loss function as in Corollary 1 to Theorem 1 plus uniform integrability,

$$c^{3/2} R(c) \to \frac{a^{3/2}[f(1) - f(0)]^2}{3} \frac{L''(\alpha)}{2} \text{ as } c \to \infty .$$

7. Independent Replications Together with Other Estimators

We now consider independent replications together with other estimators. Let the framework in Section 1 be modified by having the experiment in (iii) be a k-tuple $\{(Y^i, C^i) : 1 \le i \le k\}$ of independent stochastic processes. (The superscript is an index, not a power.) Let the overall cost process be $C(t) = C^1(t) + \ldots + C^k(t)$. Suppose that we also combine the observations by averaging in the usual way, i.e., by using the estimator $Y(t) = [Y^1(t) + \ldots + Y^k(t)]/k$.

We are thinking of the number k of replications being fixed with the length of the experiment being indexed by t. The final estimator is then

$$k^{-1} Y(T(c)) = [Y^1(T(c)) + \ldots + Y^k(T(c))]/k \tag{29}$$

where $T(c)$ is defined as before.

With this modified framework, it is of interest to know how independent replications affect the efficiency of an experiment. Hammersley and Handscomb (1964), p. 51, assert that independent replications do not alter the efficiency. We show that this is the case with a centered Gaussian limit if and only if $\gamma/\beta = 1/2$ in (6).

Theorem 8. (a) If the conditions of Theorem 1 hold for each i, then

$$\varepsilon^{-\gamma}(Y_\varepsilon(t) - \alpha) \Rightarrow k^{-1}[{=}^1(t) + \ldots + {=}^k(t)] \quad \text{in } D \text{ as } \varepsilon \to 0 .$$

If, in addition, (5) holds for each i, then

$$t^{-\beta} C(t) \to k\lambda^{-\beta} \quad \text{w.p.1} \quad \text{as } t \to \infty ,$$

so that

$$c^{-1/\beta} T(c) \to k^{-1/\beta} \lambda \quad \text{w.p.1} \quad \text{as } c \to \infty ,$$

$$c^{\gamma/\beta}[Y(T(ct)) - \alpha] \Rightarrow k^{-1}[{=}^1(k^{-1/\beta}\lambda t^{1/\beta}) + \ldots + {=}^k(k^{-1/\beta}\lambda t^{1/\beta})] \quad \text{in } D \quad \text{as } c \to \infty$$

and

$$c^{\gamma/\beta}[Y(T(c)) - \alpha] \Rightarrow k^{(\gamma/\beta)-1}\lambda^{-\gamma} ({=}^1(1) + \ldots + {=}^k(1)) \quad \text{in } R \quad \text{as } c \to \infty .$$

(b) If, in addition, ${=}^i(1) \overset{d}{=} N(\mu,\sigma^2)$, then ${=}^1(1) + \ldots + {=}^k(1) \overset{d}{=} N(k\mu, k\sigma^2)$ and

$$c^{\gamma/\beta}[Y(T(c)) - \alpha] \Rightarrow k^{\left[\frac{\gamma}{\beta} - \frac{1}{2}\right]} \lambda^{-\gamma} \left[\mu + N(0,1)\right] \quad \text{in } R \quad \text{as } c \to \infty .$$

At least when $\mu = 0$, Theorem 8(b) implies that independent replications cause the efficiency to get better, remain unchanged or get worse, respectively, when $\gamma/\beta < 1/2$, $\gamma/\beta = 1/2$ or $\gamma/\beta > 1/2$; i.e., if $\mu = 0$ under the assumptions of Corollary 1 to Theorem 1, then the asymptotic efficiency parameters are

$$r = \frac{2\gamma}{\beta} \quad \text{and} \quad v = \frac{2}{L''(\alpha)} \lambda^{2\gamma} \sigma^{-2} k^{-((2\gamma/\beta)-1)} . \tag{30}$$

From (30), we see that the asymptotic efficiency rate r is unchanged by using k independent replications, but the asymptotic efficiency value is higher, the same or lower when $\gamma/\beta < 1/2$, $\gamma/\beta = 1/2$ or $\gamma/\beta > 1/2$. For example, the asymptotic efficiency value is increased by using multiple replications with the Kiefer-Wolfowitz stochastic approximation algorithm in Section 5, which has a subcanonical convergence rate. (There $\gamma = 1/3$ and $\beta = 1$, so that $\gamma/\beta < 1/2$.)

8. Proofs

Proof of Lemma 1. Suppose that $t^{-\beta} C(t) \to \lambda^{-\beta}$ w.p.1 as $t \to \infty$. Then $T(c) \to \infty$ w.p.1 as $c \to \infty$. Since

$$C(T(c) - 1) \leq c \leq C(T(c)) \text{ for all } c , \tag{30}$$

we can divide through by $T(c)^\beta$ in (30) and let $c \to \infty$ to get $c^{-1} T(c)^\beta \to \lambda^\beta$ as $c \to \infty$. Finally take β^{th} roots. Starting with $c^{-1/\beta} T(c) \to \lambda$ w.p.1 as $c \to \infty$, use

$$T(C(t) - 1) \leq t \leq T(C(t)) \text{ for all } t$$

and reason similarly.

Proof of Lemma 2. We actually establish a stronger result.

Lemma 3. Let $X(t)$ be a random element of $D([0,\infty), R)$. If $t^{-\beta} X(t) \to \lambda$ w.p.1 as $t \to \infty$ for some $\beta > 0$, then for each $T > 0$

$$\sup_{0 \leq t \leq T} \{ | \varepsilon^\beta X(t/\varepsilon) - \lambda t^\beta | \} \to 0 \text{ w.p.1 as } \varepsilon \to 0 .$$

Proof. For arbitrary $\delta > 0$, choose t_0 so that $|t^{-\beta} X(t) - \lambda| < \delta$ for all $t > t_0$, which we can do by the assumed w.p.1 convergence. Then considering the supremum separately over $[0, \varepsilon t_0]$ and $[\varepsilon t_0, T]$, we obtain

$$\sup_{0 \le t \le T} \{ |\varepsilon^\beta X(t/\varepsilon) - \lambda t^\beta | \} \le \varepsilon^\beta \sup_{0 \le t \le t_0} \{ |X(t)| \} + \lambda (\varepsilon t_0)^\beta$$

$$+ \sup_{\varepsilon t_0 \le t \le T} \{ t^\beta \left| \frac{T(t/\varepsilon)}{(t/\varepsilon)^\beta} - \lambda \right| \}$$

$$\le \varepsilon^\beta \sup_{0 \le t \le t_0} \{ |X(t)| \} + \lambda (\varepsilon t_0)^\beta + T^\beta \delta \ .$$

Since X is a random element of D, $\sup_{0 \le t \le t_0} \{ |X(t)| \} < \infty$. Finally, let $\varepsilon \to 0$ and then let $\delta \to 0$ to obtain the desired result.

Proof of Theorem 1. By Lemmas 1 and 2, the SLLN (5) for $C(t)$ implies a FSLLN for $T(c)$. By (4) and Theorem 4.4 of Billingsley (1968),

$$(c^{\gamma/\beta} [Y(c^{1/\beta} t) - \alpha] \ , \ c^{-1/\beta} T(ct)) \ \Rightarrow \ (=(t), \ \lambda t^{1/\beta}) \tag{31}$$

in $D((0,\infty), R) \times D([0,\infty), R)$ as $c \to \infty$. Now apply the continuous mapping theorem (Theorem 5.1 of Billingsley) with the composition map, which is continuous because $\lambda t^{1/\beta}$ is continuous and strictly increasing; see Theorem 3.1 of Whitt (1980). We must also make sure that the range of $c^{-1/\beta} T(ct)$ is contained in the domain of $c^{\gamma/\beta} [Y(c^{1/\beta} t) - \alpha]$. Since we are establishing convergence in $D((0,\infty), R)$, it suffices to establish convergence in $D([t_1,\infty), R)$ for all $t_1 > 0$; see Section 2 of Whitt (1980). For any given t_1, choose t_0 such that $\lambda t_1^{1/\beta} > t_0$, so that $c^{-1/\beta} T(ct) > t_0$ for $t \ge t_1$ w.p.1 for all sufficiently large c. Since the limit process $=$ is continuous at t w.p.1 for each t, the convergence in (31) also holds on $D([t_0,\infty) \times D([t_1,\infty))$. Then replace $c^{-1/\beta} T(ct)$ by $\max\{ t_0, c^{1/\beta} T(ct) \}$ on $[t_1,\infty)$ and apply Theorem 4.1 of Billingsley to show that the composition argument above remains valid on $D([t_0,\infty), R) \times D([t_1, \infty), R)$. This yields the desired weak convergence (4) in $D([t_1, \infty), R)$. Since t_1 was arbitrary, we obtain convergence in $D((0,\infty), R)$. To establish (7) from (6), use the continuous mapping theorem with the projection map, p. 121 of Billingsley. To see that $= (\lambda) \overset{d}{=} \lambda^{-\gamma} = (1)$, note that $=_{\varepsilon/\lambda}(t) = \lambda^{\gamma} =_{\varepsilon}(\lambda t)$ in (3) and let $\varepsilon \to 0$. ■

Proof of Corollary 1. Using Taylor's theorem, we expand $L[Y(T(c))]$ about α to get

$$L[Y(T(c))] = 2^{-1} L''(\xi_c)[Y(T(c)) - \alpha]^2 ,$$

where ξ_c falls between $Y(T(c))$ and α. Note that $\xi_c \overset{p}{\to} a$ as $c \to \infty$ by (8) and $L''(\xi_c) \overset{p}{\to} L''(\alpha)$ by the assumed continuity of L''. Hence $c^{2\gamma/\beta} L(Y(T(c))) \Rightarrow = (\lambda t^{1/\beta})^2$ in D as $c \to \infty$. Finally, (10) holds by uniform integrability.

Proof of Corollary 2. Since $L[Y(T(c))] = |Y(T(c)) - \alpha|^p$,

$$c^{p\gamma/\beta} L[Y(T(c))] \Rightarrow \lambda^{-p\gamma} |= (1)|^p \text{ in } R \text{ as } c \to \infty$$

from Theorem 1 and the continuous mapping theorem. Hence, (12) follows from the assumed uniform integrability. ■

Proof of Theorem 2. We show that

$$n^{-b-1} \sum_{i=1}^{n} [\tau_i - E(\tau_i | \wedge_{i-1})] \to 0 \text{ w.p.1 as } n \to \infty$$

and

$$n^{-b-1} \sum_{i=1}^{n} E(\tau_i | \wedge_{i-1}) \to a/(1+b) \quad \text{w.p.1 as } n \to \infty .$$

The first limit follows from the SLLN for martingale differences, p. 243 of Feller (1971), because under the stated conditions

$$\sum_{n=1}^{\infty} n^{-(2b+2)} \operatorname{Var} [\tau_n - E(\tau_n | \wedge_{n-1})] < \infty .$$

The second limit follows from the following lemma.

Lemma 4. If $\{a_k : k \geq 1\}$ is a sequence of real numbers such that $k^{-b} a_k \to a$ as $k \to \infty$ for $b > -1$, then

$$\lim_{n \to \infty} n^{-1-b} \sum_{k=1}^{n} a_k = a/(1+b) \ .$$

Proof. Suppose that $a > 0$. (Similar arguments apply for $a = 0$ and $a < 0$.) Fix an arbitrary $\varepsilon > 0$. Let k_0 be such that $a(1-\varepsilon)k^b \leq a_k \leq a(1+\varepsilon)k^b$ for all $k \geq k_0$. Then, for all $n \geq k_0$,

$$-n^{-1-b} \sum_{j=1}^{k_0} \left[|a_j| + aj^b \right] + a(1-\varepsilon) n^{-1-b} \sum_{j=1}^{n} j^b$$
$$\leq n^{-1-b} \sum_{j=1}^{n} a_j \leq n^{-1-b} \sum_{j=1}^{k_0} |a_j| + a(1+\varepsilon) n^{-1-b} \sum_{j=1}^{n} j^b \ .$$

Since $n^{-1-b} \sum_{j=1}^{n} j^b = n^{-1} \sum_{j=1}^{n} (j/n)^b$ is a Riemann sum approximation to the integral $\int_0^1 x^b \, dx = (1+b)^{-1}$, we can let $n \to \infty$ and then $\varepsilon \to 0$ to obtain the desired result.

Proof of Theorem 3. We use the assumed weak convergence and the Skorohod representation theorem, see Whitt (1980), to construct versions such that $\varepsilon^{-\gamma}[\psi_\varepsilon(\lambda_\varepsilon(t)) - \mu] \to \psi(t)$ and $\lambda_\varepsilon(t) \to t$ uniformly over the interval $[a, b]$ w.p.1 where λ_ε are the homeomorphisms of $(0, \infty)$ associated with the Skorohod J_1 topology. Then we use the continuous differentiability of g to expand $g(\psi_\varepsilon(\lambda_\varepsilon(t)))$ as

$$\varepsilon^{-\gamma}[g(\psi_\varepsilon(\lambda_\varepsilon(t))) - g(\mu)] = \nabla g(v_\varepsilon(t)) \, \varepsilon^{-\gamma}[\psi_\varepsilon(\lambda_\varepsilon(t)) - \mu] \tag{32}$$

where $v_\varepsilon(t)$ is on the line segment joining $\psi_\varepsilon(\lambda_\varepsilon(t))$ and μ, $a \leq t \leq b$. This implies the desired conclusion.

Proof of Theorem 4. This is a special case of Theorem 5(b) with $\gamma = \eta = 1/4$.

Proof of Theorem 5. We can express $t(Y(t/\varepsilon) - \alpha)$, except for a factor of $(t/\varepsilon)/\lfloor t/\varepsilon \rfloor$, as

$$\varepsilon \sum_{i=1}^{\lfloor t/\varepsilon \rfloor} X_i - \alpha t = \varepsilon \sum_{i=1}^{\lfloor t/\varepsilon \rfloor} \hat{X}_i + \varepsilon \sum_{i=1}^{\lfloor t/\varepsilon \rfloor} (EX_i - \alpha) + (\varepsilon \lfloor t/\varepsilon \rfloor - t)\alpha . \tag{33}$$

To treat the first term in (33), we apply the martingale FCLT (part b) on p. 340 of Ethier and Kurtz. For this purpose, note that

$$M_\varepsilon(t) \equiv \varepsilon^{1-\gamma} \sum_{i=1}^{\lfloor t/\varepsilon \rfloor} \hat{X}_i , \quad t \ge 0 ,$$

is a martingale with quadratic variation process

$$A_\varepsilon(t) = \varepsilon^{2-2\gamma} \sum_{i=1}^{\lfloor t/\varepsilon \rfloor} E\hat{X}_i^2 , \quad t \ge 0 .$$

By Lemma 4, since $n^{2\gamma-1} E\hat{X}_n^2 \to \sigma^2$ for $2\gamma - 1 > -1$,

$$A_\varepsilon(t) \to \frac{\sigma^2 t^{2-2\gamma}}{2-2\gamma} \quad \text{as } \varepsilon \to 0 .$$

By Lemma 3, this convergence is uniform on bounded intervals. Furthermore, for any $\delta > 0$,

$$\begin{aligned} \Delta_\varepsilon(T) \equiv E[\sup_{0 \le t \le T} |M_\varepsilon(t) - M_\varepsilon(t-)|^2] &= \varepsilon^{2-2\gamma} E[\max_{1 \le i \le T/\varepsilon} \hat{X}_i^2] \\ &\le \varepsilon^{2-2\gamma} \sum_{i=1}^{\lfloor T/\varepsilon \rfloor} E\hat{X}_i^2 \\ &\le \delta + \sum_{i=1}^{\lfloor T/\varepsilon \rfloor} \varepsilon^{2-2\gamma} E[\hat{X}_i^2 ; \ \hat{X}_i^2 > \delta \varepsilon^{2\gamma-2}] \\ &\le \delta + \varepsilon \sum_{k=1}^{\lfloor T/\varepsilon \rfloor} (k/T)^{2\gamma-1} E[\hat{X}_k^2 ; \ \hat{X}_k^2 > \delta (k/T)^{1-2\gamma} \varepsilon^{-1}] , \end{aligned}$$

but since $\{n^{2\gamma-1} \hat{X}_n^2\}$ is uniformly integrable, $k^{2\gamma-1} E[\hat{X}_k^2 ; \ \hat{X}_k^2 > \delta (k/T)^{1-2\gamma} \varepsilon^{-1}] \to 0$ uniformly in k as $\varepsilon \to 0$. Since δ was arbitrary, we conclude that $\Delta_\varepsilon(T) \to 0$ as $\varepsilon \to 0$. Hence,

$$M_\varepsilon(t) \Rightarrow [\sigma^2/(2-2\gamma)]^{1/2} B(t^{2-2\gamma}) \quad \text{in } D \quad \text{as } \varepsilon \to 0 . \tag{34}$$

Turning to the second term in (33), we apply Lemma 4 to deduce that

$$(\varepsilon/t)^{1-\eta} \sum_{i=1}^{\lfloor t/\varepsilon \rfloor} (EX_i - \alpha) \to \frac{b}{1-\eta} .$$

By Lemma 3,

$$\varepsilon^{1-\eta} \sum_{i=1}^{\lfloor t/\varepsilon \rfloor} (EX_i - \alpha) \to \frac{bt^{1-\eta}}{1-\eta} \quad \text{in } D \quad \text{as } \varepsilon \to 0 \tag{35}$$

Finally, we have the three results (a), (b) and (c) by combining (34) and (35).

Proof of the Corollary to Theorem 5. The SLLN for $C(t) = n^{-\rho-1} \sum_{i=1}^{\lfloor t \rfloor} t_i$ follows from Theorem 2 with $\beta = 1 + \rho$ and $\lambda^{-\beta} = a/(1 + \rho)$. The rest follows from Theorems 1 and 5.

Proof of Theorem 6. If $f(0) = f(1)$, then $f(x \oplus c)$ is continuous for all c and $(X_n - n\alpha) \to 0$ w.p.1 by (2.1.10) on p. 53 of Davis and Rabinowitz (1984). (The argument there remains valid if f is differentiable everywhere except at one point where left and right derivatives exist.) Hence, add a linear function to f to make $f(0) = f(1)$ and apply Lemma 5 below to the linear function. (Of course, both the integral and the approximating sum are additive for the two functions.) For the uniform integrability, note that the bounds in (2.1.8) on p. 53 of Davis and Rabinowitz apply uniformly to the translation point U.

Lemma 5. Suppose that $f(x) = ax$, $0 \le x \le 1$. Then

$$n^{-1} X_n \stackrel{\mathrm{d}}{=} \frac{a}{2} + \frac{a}{n} \left[U - \frac{1}{2} \right] .$$

Proof. Note that

$$n^{-1} X_n \stackrel{\mathrm{d}}{=} \sum_{k=0}^{n-1} a \left[\frac{k}{n} + \frac{U}{n} \right] \frac{1}{n} = \frac{a}{2} + \frac{a}{n} \left(U - \frac{1}{2} \right) .$$

Proof of Theorem 7. Since the sequence of uniforms $\{U_k\}$ is i.i.d., $\{X_k : k \geq 1\}$ is an independent sequence with $EX_k = k\alpha$. As in the proof of Theorem 5, we apply the martingale FCLT on p. 340 of Ethier and Kurtz (1986). With $\hat{X}_k = X_k - k\alpha$,

$$M_n(t) = 2\frac{n^{1/2}}{n+1} \sum_{k=1}^{\lfloor nt \rfloor} \hat{X}_k$$

is a martingale and the associated quadratic variation process

$$A_n(t) = \frac{4n}{(n+1)^2} \sum_{k=1}^{\lfloor nt \rfloor} \mathrm{Var}\, \hat{X}_k \to \frac{(f(1) - f(0))^2}{3}\, t \quad \text{as } n \to \infty ,$$

using the moment convergence established in Theorem 6. Also, for $\varepsilon > 0$,

$$E[\sup_{0 \leq t \leq T} |M_n(t) - M_n(t-)|^2] \leq \varepsilon + \sum_{k=1}^{\lfloor nt \rfloor} 4\frac{n}{(n+1)^2}\, E\{\hat{X}_k^2 ; \quad \hat{X}_k^2 > \varepsilon (n+1)^2\, n^{-1}\} .$$

When we let $n \to \infty$, the second term goes to zero by the uniform integrability derived in Theorem 6. Letting $\varepsilon \downarrow 0$, we find that $E[\sup_{0 \leq t \leq T} |M_n(t) - M_n(t-)|^2] = 0$. Hence, we have the FCLT for M_n. Since

$$\frac{(\lfloor nt \rfloor \lceil nt \rceil)}{2}(Y(t) - \alpha) = \left[\sum_{k=1}^{\lfloor nt \rfloor} X_k - \lfloor nt \rfloor \lceil nt \rceil \frac{\alpha}{2} \right] ,$$

we obtain the FCLT (4).

Proof of Theorem 8. The FCLT for Y holds by Theorems 3.2 and 5.1 of Billingsley. The rest is elementary.

Acknowledgments. Peter Glynn's research was begun at the University of Wisconsin-Madison with support from the National Science Foundation under grant ECS-8404809 and the U.S. Army under contract DAAG-29-80-C-0041. It was completed at Stanford University with support from the U.S. Army under contract DAAL-03-88-K-0063. The authors thank Bennett L. Fox and the referees for helpful comments.

REFERENCES

ANSCOMBE, F. J. 1952. Large-sample theory of sequential estimation. *Proc. Camb. Phil. Soc.* 48, 600-607.

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

BRATLEY, P., B. L. FOX, and L. E. SCHRAGE. 1987. *A Guide to Simulation*, Second ed., Springer-Verlag, New York.

CHENG, R. C. H. and T. Davenport. 1989. The problem of dimensionality in stratified sampling. *Management Sci.* 35, 1278-1296.

COX, D. R. 1984. Long-range dependence: a review. In *Statistics: An Appraisal*, eds. H. A. David and H. T. David, Iowa State University Press, Ames, Iowa, 55-74.

DAVIS, P. J. and P. RABINOWITZ. 1984. *Methods of Numerical Integration*, second ed., Academic Press, New York.

ETHIER, S. N. and T. G. KURTZ. 1986. *Characterization and Approximation of Markov Processes*. Wiley, New York.

FAMA, E. F. 1963. Mandelbrot and the stable Paretian hypothesis. *J. of Business* 36, 420-429 (Reprinted in *The Random Character of Stock Market Prices.*, P. H. Cootner (ed.), M.I.T. Press, 1964).

FELLER, W. 1971. *An Introduction to Probability Theory and Its Applications*, Vol. II, Second ed., Wiley, New York.

FERGUSON, T. S. 1967. *Mathematical Statistics, A Decision Theoretic Approach*, Academic, New York.

FISHMAN, G. S. and B. D. HUANG. 1983. Antithetic variates revisited. *Commun. ACM* 26, 964-971.

FOX, B. L. and P. W. GLYNN. 1989a. Simulating discounted costs. *Management Sci.* 35, 1297-1315.

FOX, B. L. and P. W. GLYNN. 1989b. Replication schemes for limiting expectations. *Prob. Engr. Inf. Sci.* 3, 299-318.

GLYNN, P. W. 1984. Some asymptotic formulas for Markov chains with applications to simulation. *J. Statist. Comput. Simul.* 19, 97-112.

GLYNN, P. W. 1986. Stochastic approximation for Monte Carlo optimization. *Proceedings of 1986 Winter Simulation Conference*.

GLYNN, P. W. and W. WHITT. 1986. The efficiency of simulation estimators. AT&T Bell Laboratories, Holmdel, NJ.

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

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

GLYNN, P. W. and W. WHITT. 1989. Indirect estimation via $L = \lambda W$. *Opns. Res.* 37, 82-103.

GUT, A. 1988. *Stopped Random Walks*, Springer-Verlag, New York.

HAMMERSLEY, J. M. and D. C. HANDSCOMB. 1964. *Monte Carlo Methods*. Methuen, London.

KERSTING, G. 1977. Almost sure approximation of the Robbins-Monro process by sums of independent random variables. *Ann. Probab. 5*, 954-965.

KIEFER, J. and J. WOLFOWITZ. 1952. Stochastic estimation of the maximum of a regression function. *Ann. Math. Statist.* 23, 462-466.

KUSHNER, H. J. and D. S. CLARK. 1978. *Stochastic Approximation Methods for Constrained and Unconstrained Systems,* Springer-Verlag, New York.

MANDELBROT, B. B. 1963. The variation of certain speculative prices. *J. of Business* 36, 394-419 (Reprinted in *The Random Character of Stock Market Prices.*, P. H. Cootner (ed.), M.I.T. Press, 1964).

MANDELBROT, B. B. 1977. *Fractals, Form, Chance and Dimension,* W. H. Freeman, San Francisco.

NOZARI, A. and W. WHITT. 1988. Estimating average production intervals using inventory measurements: Little's law for partially observable processes. *Opns. Res.* 36, 308-323.

ROBBINS, H. and S. MONRO. 1951. A stochastic approximation method. *Ann. Math. Statist.* 22, 400-407.

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

SAVAGE, L. J. 1954. *The Foundations of Statistics*, Wiley, New York.

TAQQU, M. S. 1982. Self-similar processes and related ultraviolet and infrared catastrophes. *Colloq. Math. Soc. Janos Bolyai* 27. In *Random Fields*, eds. Fritz, Lebowitz and Szasz, North-Holland Amsterdam, 1057-1096.

VERVAAT, W. 1985. Sample path properties of self-similar processes with stationary increments. *Ann. Probab.* 13, 1-27.

WALD, A. 1950. *Statistical Decision Functions*, Wiley, New York.

WASAN, M. T. 1969. *Stochastic Approximation*, Cambridge University Press, New York.

WHITT, W. 1980. Some useful functions for functional limit theorems. *Math. Opns. Res.* 5, 67-85.

WHITT, W. 1991. Asymptotic formulas for Markov processes with applications to simulation. *Opns. Res.* (revision under review).

WILSON, J. R. 1985. Variance reduction techniques for digital simulation. *Amer. J. Math. Management Sci.* 4, 277-312.

ZAZANIS, M. and R. SURI. 1986. Comparison of perturbation analysis with conventional sensitivity estimates for stochastic systems, Department of Industrial Engineering, University of Wisconsin at Madison.