A New Approach to Sequential Stopping for Stochastic Simulation

Jing Dong
Columbia University, New York, NY, 10027, jing.dong@gsb.columbia.edu

Peter W. Glynn
Stanford University, Stanford, CA, 94305 glynn@stanford.edu

In this paper, we solve the sequential stopping problem for a class of simulation problems in which variance estimation is difficult. We establish the asymptotic validity of sequential stopping procedures for estimators constructed using the sectioning (replication) methods with a fixed number of sections. The limiting distribution of the estimators at stopping times as the error size (the absolute error or the relative error) goes to 0 is characterized in closed form. This limiting distribution is *different* from the limiting distribution of the estimator constructed based on a fixed number of samples as the sample size goes to infinity, which indicates that we need a different scaling parameter when constructing the corresponding confidence intervals using the sequential stopping procedure. In particular, the scaling parameters we derived are *larger* than those suggested by the corresponding fixed sample procedure. We also investigate the empirical performance of our proposed sequential stopping algorithms through some simulation experiments.

Key words: stochastic simulation, sequential stopping

1. Introduction

Suppose that we wish to compute some performance measures α via simulation. Simulation-based procedures fall, roughly speaking, into two categories: *fixed sample-size procedures* and *sequential procedures*. In a fixed sample-size procedure, one decides a priori on the number of samples that will be generated to form an estimator for α, either based on knowledge of similar problems or on a set of initial "trail runs" to estimate the associated variance (thereby providing an estimate for the required sample size). On the other hand, in a sequential procedure, one continues drawing observations until some pre-specified level of absolute or relative precision has been reached. In most computational settings, sequential procedures are algorithmically more natural, and reflect the way practitioners run their simulations.

The most widely studied class of sequential algorithms assumes that the estimator $\alpha(t)$, based on a sample of size t, satisfies the central limit theorem (CLT), so that there exists $\sigma > 0$ for which

$$\alpha(t) \stackrel{D}{\approx} \alpha + \frac{\sigma}{\sqrt{t}} N(0,1) \tag{1.1}$$

when t is large, where $N(0,1)$ is a normal random variable (rv) having mean 0 and variance 1, and $\stackrel{D}{\approx}$ means "has approximately the same distribution as"; see A1 for a more rigorous formulation. Relation (1.1) implies that

$$\left[\alpha(t) - z\frac{\sigma}{\sqrt{t}}, \alpha(t) + z\frac{\sigma}{\sqrt{t}}\right]$$

is an approximate $100(1-\delta)\%$ confidence interval for α when t is large, provided that one chooses z so that $P(-z \leq N(0,1) \leq z) = 1-\delta$. To obtain a confidence interval having prescribed half-width ϵ, one then samples until $z\sigma/\sqrt{t} \leq \epsilon$. Of course, in practice, it is rarely the case that σ is known, so it must be estimated from the observed data by some estimator $s(t)$, say. This suggests that one continuously samples until $zs(t)/\sqrt{t} \leq \epsilon$. This type of sequential procedure was first studied by Chow and Robbins (1965), when $\alpha(t)$ is the sample mean of $\lfloor t \rfloor$ independent and identically distributed (iid) observations; in this setting, $s^2(t)$ is just the corresponding sample variance. Their analysis established that when the simulation is stopped in this way, then

$$P(\alpha \in [\alpha(T(\epsilon)) - \epsilon, \alpha(T(\epsilon)) + \epsilon]) \to 1-\delta \quad \text{as } \epsilon \downarrow 0,$$

where $T(\epsilon)$ is the (random) sample size at which $zs(t)/\sqrt{t}$ is first less than ϵ. Such a procedure is said to achieve *absolute precision* ϵ, with an asymptotic confidence level $1-\delta$ as $\epsilon \downarrow 0$. Similarly, a sequential procedure achieves *relative precision* ϵ, with an asymptotic confidence level of $1-\delta$ (based on stopping at sample size $T(\epsilon)$), if $P(\alpha \in [\alpha(T(\epsilon))(1-\epsilon), \alpha(T(\epsilon))(1+\epsilon)]) \to 1-\delta$ as $\epsilon \downarrow 0$ when $\alpha > 0$.

Glynn and Whitt (1992) extended the Chow-Robbins methodology to much more general simulation settings, in which $\alpha(t)$ either can not be expressed as a sample mean (e.g. sample quantiles) or $\alpha(t)$ is a sample mean constructed from dependent observations (e.g. a steady-state simulation). In particular, they showed that if $s(t)$ is *strongly consistent* for σ in the sense that

$$s(t) \to \sigma \text{ almost surely (a.s.) as } t \to \infty, \tag{1.2}$$

then Chow-Robbins type absolute and relative precision procedures exhibit the correct asymptotic confidence levels as $\epsilon \downarrow 0$ (under a slight enhancement of the CLT assumption, known as the functional central limit theorem (FCLT)). They further showed that such asymptotic coverage guarantees may fail to be valid when (1.2) is relaxed to, for example, the requirement of consistency for $s(t)$ (in which case one requires only that $s(t)$ converges to σ in probability as $t \to \infty$). Related literature on Chow-Robbins type procedures can be found in Lavenberg and Sauer (1977), Law and Carson (1979), Jones et al. (2006), and Bayraksan and Pierre-Louis (2012).

Unfortunately, there exist many simulation settings in which even consistent estimation of σ is either theoretically difficult or computationally inconvenient (let alone guaranteeing that the estimator is also strongly consistent). For example, in the steady-state simulation context, the quantity σ^2 is known as the time-average variance constant (TAVC). The TAVC is difficult to consistently estimate, because σ^2 reflects the complicated auto-correlation structure of the underlying stochastic process. In particular, when the underlying process is non-regenerative, the problem of consistently estimating the TAVC is an ongoing research area; see Alexopoulos et al. (2007), Wu (2009), Meterelliyoz et al. (2012). Even less is known about strong consistency of such estimators; see, however, Damerdji (1994). Computation of quantiles is an example of another setting in which estimating σ is challenging (in part because the formula for σ includes the density and density estimation is known to be both practically and theoretically difficult).

Our goal in this paper is to introduce an entirely new approach to sequentially estimating α to a given precision that applies in much greater generality than do the above Chow-Robbins type procedures. The method applies to steady-state simulations, quantile estimations, smooth functions of expectations, and can be used more generally as well.

Our idea stems from an alternative fixed sample size confidence interval procedure that avoids the need to consistently estimate σ. The method is known as the sectioning method or replication method. In particular, we replicate the estimator $m \geq 2$ independent times, thereby obtaining

m independent and identically distributed (iid) copies of the estimator $(\alpha(t) : t \geq 0)$, call them $\alpha_1(t), \ldots, \alpha_m(t)$. We can now form the grand mean of these estimators, namely

$$\bar{\alpha}_m(t) = \frac{1}{m} \sum_{i=1}^{m} \alpha_i(t),$$

and compute a standard deviation estimator from the m replications, namely

$$S_m(t) = \sqrt{\frac{1}{m-1} \sum_{i=1}^{m} (\alpha_i(t) - \bar{\alpha}_m(t))^2}.$$

In this case, (1.1) implies that

$$\frac{\sqrt{m}(\bar{\alpha}_m(t) - \alpha)}{S_m(t)} \overset{D}{\approx} t_{m-1} \tag{1.3}$$

when t is large, where t_{m-1} is a Student-t random variable (rv) with $m-1$ degrees of freedom. As a consequence, the fixed sample size confidence interval is

$$\left[\bar{\alpha}_m(t) - z_{\delta,m-1} \frac{S_m(t)}{\sqrt{m}}, \bar{\alpha}_m(t) + z_{\delta,m-1} \frac{S_m(t)}{\sqrt{m}} \right]$$

where $z_{\delta,m-1}$ is chosen so that $P(-z_{\delta,m-1} \leq t_{m-1} \leq z_{\delta,m-1}) = 1 - \delta$. This type of procedure is especially useful in simulation settings where it is difficult to estimate σ, such as in the steady-state simulation context.

If one proceeds as for Chow-Robbins, one then is led to sequential procedures for which one stops sampling at the time $T'(\epsilon)$ at which $z_{\delta,m-1} S_m(t)/\sqrt{m}$ is first less than ϵ. As we shall see later, this naive implementation fails, because

$$\frac{\sqrt{m}\,(\bar{\alpha}_m(T'(\epsilon)) - \alpha)}{S_m(T'(\epsilon))} \tag{1.4}$$

does not converge in distribution to the t_{m-1} as $\epsilon \downarrow 0$. (The corresponding limit theorem does hold in the Chow-Robbins setting, and is the fundamental explanation for the theoretical coverage results obtained there when $\epsilon \downarrow 0$ (Chow and Robbins 1965).) Empirical evidence of the failure of such sequential procedures was observed a number of years ago, in a related "batch means" setting; see Adam (1983). However, it turns out (1.4) does indeed converge to a limit as $\epsilon \downarrow 0$, albeit not the t_{m-1}. If one calibrates the confidence interval in terms of the correct limit distribution (rather

than the t_{m-1}), one gets an asymptotically valid sequential procedure. This is the key insight in this paper, and it leads to a flexible sequential procedure that can be used in simulation problems having difficult-to-estimate variances.

However, subtleties arise. For example, unlike the fixed sample-size procedure based on (1.3), in which the method is typically valid so long as $m \geq 2$, one must choose the number of replicated estimators, m, to be greater than or equal to 4 in the sequential setting. This constraint on m has to do with the transience of Bessel processes of dimension $m-1$, because it turns out that the Bessel process with dimension $m-1$ is intimately connected to the process-level behavior of our standard deviation estimator; see Section 2.

Our replication procedure is naturally and easily parallelizable, as each replicated estimator can be run on a different processor. In practice, the processors only need to periodically send their current estimated value to a central processor responsible for computing the current value of $S_m(t)$ and determining whether or not to stop. There is no fundamental requirement to store historical results, as the algorithm only checks against the current estimated values on each of the m processors. Finally, we note that in the presence of (1.1), $\bar{\alpha}_m(t)$ has approximately the same distribution as does $\alpha(mt)$, so it is no less asymptotically efficient than running the original estimation procedure m times longer.

In settings where one has a preference for estimating α via $\alpha(mt)$ (rather than $\bar{\alpha}_m(t)$), perhaps because of bias concerns, there is an alternative class of sequential procedures that avoids variance estimation. These methods are based on standardized time series (STS) ideas that have been previously studied in the steady-state simulation fixed sample-size context; see Schruben (1983) and Glynn and Iglehart (1990) for details. These STS-based sequential stopping procedures will be developed elsewhere.

2. Main Results

Given a quantity α to be estimated, we assume that there exists an estimator $(\alpha(t) : t > 0)$ which is consistent for α, in the sense that

$$\alpha(t) \Rightarrow \alpha \text{ as } t \to \infty, \tag{2.1}$$

where $\Rightarrow$ denotes weak convergence. The parameter t could correspond to either simulated time (e.g. the number of time units simulated) or the computational budget (e.g. the computer time expended). Because we are interested in sequential procedures to construct confidence intervals, we will need a stronger CLT-type assumption on $(\alpha(t) : t > 0)$:

A 1 **(Strong Approximation Assumption)**. *There exists (deterministic) constants* $\sigma \in (0, \infty)$ *and* $\gamma > 1/2$*, and a probability space supporting both* $(\alpha(t) : t \geq 0)$ *and a standard Brownian motion* $(B(t) : t \geq 0)$*, such that*

$$\alpha(t) = \alpha + \sigma \frac{B(t)}{t} + o(t^{-\gamma}) \text{ a.s.}$$

where $o(a(t))$ *is a term having the property that* $o(a(t))/a(t) \to 0$ *a.s. as* $t \to \infty$*.*

Note that A1 easily implies that

$$t^{1/2}(\alpha(t) - \alpha) \Rightarrow \sigma N(0, 1) \text{ as } t \to \infty,$$

so A1 guarantees the validity of the approximation (1.1) and (2.1). We shall argue in Section 3 that A1 holds in many different estimation contexts. In particular, A1 typically holds in the steady-state simulation context, the setting of smooth functions of expectations, sample-average approximations, and in the quantile estimation.

We now let $\alpha_i = (\alpha_i(t) : t > 0)$, $1 \leq i \leq m$, be m iid copies of $(\alpha(t) : t > 0)$. In view of A1, we may presume that there exist m independent standard Brownian motions, $B_1, B_2, \ldots, B_m$, for which

$$\alpha_i(t) = \alpha + \sigma \frac{B_i(t)}{t} + o(t^{-\gamma}) \text{ a.s. as } t \to \infty. \tag{2.2}$$

Put

$$\bar{B}_m(t) = \frac{1}{m} \sum_{i=1}^{m} B_i(t)$$

and

$$R_m(t) = \sqrt{\sum_{i=1}^{m} (B_i(t) - \bar{B}_m(t))^2}.$$

Note that (2.2) implies that

$$\bar{\alpha}_m(t) = \alpha + \sigma \frac{\bar{B}_m(t)}{t} + o(t^{-\gamma}) \text{ a.s.} \tag{2.3}$$

and

$$S_m(t) = \sigma \frac{1}{\sqrt{m-1}} \frac{R_m(t)}{t} + o(t^{-\gamma}) \text{ a.s..} \tag{2.4}$$

As discussed in the Introduction, Bessel processes emerge naturally in our sequential setting. Our next result provides the connection. Recall that $Z_d = (Z_d(t) : t \geq 0)$ is a *Bessel process of dimension d starting from the origin*, if $(Z_d(t) : t \geq 0) \stackrel{D}{=} ((\sum_{i=1}^{d} B_i^2(t))^{1/2} : t \geq 0)$, where $\stackrel{D}{=}$ denotes equality in distribution.

PROPOSITION 1. *$\bar{B}_m = (\bar{B}_m(t) : t \geq 0)$ and $R_m = (R_m(t) : t \geq 0)$ are independent stochastic processes. Furthermore, $\sqrt{m}\bar{B}_m \stackrel{D}{=} B$, and $R_m \stackrel{D}{=} Z_{m-1}$.*

We note that the fact that R_m is a Bessel process of dimension $m-1$ implies that

$$\begin{aligned} tS_m^2(t) &\Rightarrow \sigma^2 \frac{1}{m-1} \sum_{i=1}^{m-1} B_i^2(1) \text{ as } t \to \infty \\ &\stackrel{D}{=} \sigma^2 \frac{\chi_{m-1}^2}{m-1}, \end{aligned}$$

where χ_{m-1}^2 is a chi-squared rv with $m-1$ degrees of freedom. So, the appearance of the Bessel process z_{m-1} is consistent with the fact that for $m \geq 2$,

$$\frac{\sqrt{m}(\bar{\alpha}_m(t) - \alpha)}{S_m(t)} \Rightarrow t_{m-1} \text{ as } t \to \infty, \tag{2.5}$$

and hence is aligned with the existing fixed sample size limit theory for the sectioning method.

We now turn to the precise definition and analysis of the sequential stopping rule. Based on the fixed sample size procedure associated with (2.5), we are led to consider the family of sequential stopping rules (parameterized by $r > 0$) given by

$$\kappa_m'(\epsilon, r) = \inf \left\{ t > 0 : S_m(t)/\sqrt{m} < \epsilon/r \right\}.$$

Our goal is to show that there exists $r_\delta > 0$ so that the sequential procedure achieves absolute precision/half-width ϵ with asymptotic confidence level $1 - \delta$.

However, the problem with the above family of sequential procedures is that $\kappa'_m(\epsilon, r)$ could terminate the sampling process too early. For example, if $\alpha(t)$ is obtained by averaging discrete rv's, there is positive probability that the first few rv's generated take on identical values, thereby leading to $S_m(t) = 0$, and to early termination with an inaccurate answer. We therefore modify $\kappa'_m(\epsilon, r)$ to

$$\kappa_m(\epsilon, r) = \inf\left\{t > \epsilon^{-1/\beta} : \frac{S_m(t)}{\sqrt{m}} < \epsilon/r\right\},$$

where $\beta \in (1/2, \gamma)$ for γ defined in A1.

In view of (2.4),

$$\begin{aligned}
\kappa'_m(\epsilon, r) &\overset{D}{\approx} \inf\left\{t > 0 : m^{-1/2}\left(\frac{1}{m-1}\sum_{i=1}^{m}\left(\sigma\frac{B_i(t)}{t} - \sigma\frac{\bar{B}_m(t)}{t}\right)^2\right)^{1/2} < \epsilon/r\right\} \\
&= \inf\left\{t > 0 : \frac{\sigma r}{\epsilon\sqrt{m(m-1)}}\left(\sum_{i=1}^{m}\left(\frac{B_i(t)}{t} - \frac{\bar{B}_m(t)}{t}\right)^2\right)^{1/2} < 1\right\} \\
&= \eta_\epsilon^2 \inf\left\{t > 0 : \eta_\epsilon\left(\sum_{i=1}^{m}\left(\frac{B_i(\eta_\epsilon^2 t)}{\eta_\epsilon^2 t} - \frac{\bar{B}_m(\eta_\epsilon^2 t)}{\eta_\epsilon^2 t}\right)^2\right)^{1/2} < 1\right\} \text{ where } \eta_\epsilon = \tfrac{\sigma r}{\epsilon\sqrt{m(m-1)}} \\
&\overset{D}{=} \eta_\epsilon^2 \inf\left\{t > 0 : \left(\sum_{i=1}^{m}\left(\frac{B_i(t)}{t} - \frac{\bar{B}_m(t)}{t}\right)^2\right)^{1/2} < 1\right\} \\
&\overset{D}{=} \eta_\epsilon^2 \inf\{t > 0 : Z_{m-1}(t) < t\}
\end{aligned}$$

We therefore have the following proposition characterizing the limiting behavior of $\kappa_m(\epsilon, r)$.

PROPOSITION 2. *Under A1,*

$$\epsilon^2\kappa_m(\epsilon, r) \Rightarrow \frac{r^2\sigma^2}{m(m-1)}K_{m-1}(1) \text{ as } \epsilon \downarrow 0,$$

where $K_{m-1}(a) := \inf\left\{t > 0 : Z_{m-1}(t) < \frac{1}{a}t\right\}$.

Proposition 2 suggests that $\kappa_m(\epsilon, r)$ is of order ϵ^{-2}, and larger σ leads to larger $\kappa_m(\epsilon, r)$.

According to Pitman and Yor (1981) p. 335, for $m \geq 4$,

$$K_{m-1}(1) \overset{D}{=} \chi^2_{m-3}.$$

On the other hand,

$$\begin{aligned} K_d(1) &= \inf\{t>0 : z_d(t) < t\} \\ &\stackrel{D}{=} \inf\left\{t>0 : \sqrt{\sum_{i=1}^{d}\left(\frac{B_i(t)}{t}\right)^2} < 1\right\} \\ &= \inf\left\{1/u>0 : \sqrt{\sum_{i=1}^{d}\left(uB_i(1/u)\right)^2} < 1\right\} \\ &\stackrel{D}{=} \left(\sup\left\{t>0 : \sqrt{\sum_{i=1}^{d} B_i(t)^2} < 1\right\}\right)^{-1}, \end{aligned}$$

where the last equality follows from the fact that $(uB(1/u) : u \geq 0) \stackrel{D}{=} (B(u) : u \geq 0)$. Since the unit ball is recurrent for Brownian motion in dimensions 1 and 2, $K_d(1) = 0$ for $d = 1, 2$. This suggests that our sequential procedure is subject to early termination when $m \leq 3$.

From Proposition 2, we have

$$\kappa_m(\epsilon, r) \stackrel{D}{\approx} \epsilon^{-2}\frac{r^2\sigma^2}{m(m-1)}K_{m-1}(1).$$

Furthermore, one expects that $S_m(\kappa_m(\epsilon,r))/\sqrt{m}$ will be approximately of size ϵ/r (by the definition of $\kappa_m(\epsilon, r)$). Then for $m \geq 4$,

$$\frac{\sqrt{m}(\bar{\alpha}_m(\kappa_m(\epsilon,r)) - \alpha)}{S_m(\kappa_m(\epsilon,r))} \stackrel{D}{\approx} \frac{r}{\epsilon}\frac{\sigma\bar{B}_m\left(\epsilon^{-2}\frac{r^2\sigma^2}{m(m-1)}K_{m-1}(1)\right)}{\epsilon^{-2}\frac{r^2\sigma^2}{m(m-1)}K_{m-1}(1)} \stackrel{D}{=} \frac{B(K_{m-1}(1))}{K_{m-1}(1)/\sqrt{m-1}}.$$

Since $K_{m-1}(1)$ is independent of B, it follows (by conditioning on $K_{m-1}(1)$) that

$$\frac{B(K_{m-1}(1))/\sqrt{K_{m-1}(1)}}{\sqrt{K_{m-1}(1)}/\sqrt{m-1}} \stackrel{D}{=} \sqrt{m-1}\frac{N(0,1)}{\sqrt{K_{m-1}(1)}} \stackrel{D}{=} \sqrt{\frac{m-1}{m-3}}t_{m-3}.$$

where $K_{m-1}(1)$ is independent of the $N(0,1)$ rv.

It follows that for $u \geq 0$,

$$P\left(-u \leq \sqrt{m}(\bar{\alpha}_m(\kappa_m(\epsilon,r)) - \alpha) \leq u\right) \approx P\left(-u \leq \sqrt{\frac{m-1}{m-3}}t_{m-3} \leq u\right)$$

when ϵ is small. Hence, if we set $\tilde{u} = \sqrt{(m-1)/(m-3)}z_{\delta,m-3}$, we have that

$$P\left(\alpha \in \left[\bar{\alpha}_m(\kappa_m(\epsilon,r)) - \tilde{u}\frac{S_m(\kappa_m(\epsilon,r))}{\sqrt{m}}, \bar{\alpha}_m(\kappa_m(\epsilon,r)) - \tilde{u}\frac{S_m(\kappa_m(\epsilon,r))}{\sqrt{m}}\right]\right) \approx 1 - \delta$$

when ϵ is small. So if we select r_δ so that $r_\delta = \tilde{u} = \sqrt{(m-1)/(m-3)}z_{\delta,m-3}$, we conclude that

$$P\left(\alpha \in [\bar{\alpha}_m(\kappa_m(\epsilon,r)) - \epsilon, \bar{\alpha}_m(\kappa_m(\epsilon,r)) + \epsilon]\right) \approx 1-\delta$$

should hold when ϵ is small.

We can now state our main theorem.

THEOREM 1. *Suppose $m \geq 4$. Then under A1,*

$$\frac{\sqrt{m}(\bar{\alpha}_m(\kappa_m(\epsilon,r)) - \alpha)}{S_m(\kappa_m(\epsilon,r))} \Rightarrow \sqrt{\frac{m-1}{m-3}}t_{m-3} \text{ as } \epsilon \downarrow 0.$$

Furthermore, if we set $r_\delta = \sqrt{\frac{m-1}{m-3}}z_{\delta,m-3}$, then

$$P\left(\alpha \in [\bar{\alpha}_m(\kappa_m(\epsilon,r_\delta)) - \epsilon, \bar{\alpha}_m(\kappa_m(\epsilon,r_\delta)) + \epsilon]\right) \to 1-\delta \text{ as } \epsilon \downarrow 0.$$

The actual simulation algorithm goes as follows.

Algorithm 1

Input: the number of sections m, the error bound ϵ, the confidence level $100(1-\delta)\%$, and the time increment Δ

Output: a $100(1-\delta)\%$ confidence interval with half width ϵ

(i) Sample $X_i(t)$ for $0 \leq t \leq \epsilon^{-1/\beta}$ and $i = 1, 2, \ldots, m$. Initialize $T = \epsilon^{-1/\beta} + \Delta$.

(ii) Sample $X_i(t)$ between $T-\Delta$ and T. Update $\alpha_i(T)$, for $i = 1, 2, \ldots, m$, and calculate $\bar{\alpha}_m(T)$ and $S_m(T)$.

(i) If $r_\delta S_m(T)/\sqrt{m} \geq \epsilon$, set $T = T + \Delta$, go back to step (ii); otherwise, output $[\bar{\alpha}_m(T) - \epsilon\ ,\ \bar{\alpha}_m(T) + \epsilon]$ as the $100(1-\delta)\%$ confidence interval.

We now compare this sequential stopping rule to the fixed sample-size procedure based on (1.3). Note that for the fixed sample-size procedure the computational budget is fixed, but the actual half-width of the confidence interval is random. On the other hand, for the sequential stopping procedure, the computational cost is random, but the half-width of the confidence interval is fixed.

For the fixed sample-size procedure, if we choose the computational budget $\tau(\epsilon)$ according to (1.3) so that the expected half-width is ϵ, then

$$\epsilon = E\left[z_{\delta,m-1}\frac{S_m(\tau(\epsilon))}{\sqrt{m}}\right] \approx \frac{z_{\delta,m-1}}{\sqrt{m}}\frac{\sigma E[\sqrt{\chi^2_{m-1}}]}{\sqrt{(m-1)\tau(\epsilon)}}.$$

Thus,

$$\tau(\epsilon) \approx \frac{z^2_{\delta,m-1}\sigma^2 E[\sqrt{\chi^2_{m-1}}]^2}{\epsilon^2 m(m-1)} = \frac{2z^2_{\delta,m-1}\sigma^2\Gamma(m/2)^2}{\epsilon^2 m(m-1)\Gamma((m-1)/2)^2}$$

for ϵ small. The squared coefficient of variation, c^2, for the half-width of the corresponding confidence interval is

$$\begin{aligned} c^2 &= \frac{Var\left(z_{\delta,m-1}S_m(\tau(\epsilon))/\sqrt{m}\right)}{E\left[z_{\delta,m-1}S_m(\tau(\epsilon))/\sqrt{m}\right]^2} \\ &\approx \frac{E[\chi^2_{m-1}]}{E[\sqrt{\chi^2_{m-1}}]^2} - 1 = \frac{(m-1)\Gamma((m-1)/2)^2}{2\Gamma(m/2)^2} - 1 \end{aligned}$$

for ϵ small.

For the sequential confidence interval constructed in Algorithm 1, as it is designed to have half-width ϵ, the squared coefficient of variation for the half-width is zero. The expected amount of time for the sequential procedure to terminate is given by

$$E[\kappa_m(\epsilon, r_\delta)] \approx \frac{z^2_{\delta,m-3}\sigma^2}{\epsilon^2 m},$$

so that

$$\frac{E[\kappa_m(\epsilon, r_\delta)]}{\tau(\epsilon)} \approx \frac{z^2_{\delta,m-3}(m-1)\Gamma((m-1)/2)^2}{2z^2_{\delta,m-1}\Gamma(m/2)^2} > 1 \tag{2.6}$$

for ϵ small. We also note that

$$\mathrm{std}(\kappa_m(\epsilon, r_\delta)) \approx \frac{\sqrt{2}z^2_{\delta,m-3}\sigma^2}{\epsilon^2 m\sqrt{m-3}}$$

Lastly, note that as $z_{\delta,m-3}$ is decreasing in m, both $E[m\kappa_m(\epsilon, r_\delta)]$ and $\mathrm{std}(m\kappa_m(\epsilon, r_\delta))$ are decreasing in m for ϵ small.

We turn next to adapting our sequential stopping rule so as to achieve a given relative error; see Asmussen and Glynn (2007) p. 158. In particular, we define a new sequence of stopping times indexed by ϵ and r as

$$\tilde{\kappa}_m(\epsilon, r) := \inf\left\{t > 0 : S_m(t)/\sqrt{m} + a(t) < \epsilon|\bar{\alpha}_m(t)|/r\right\}.$$

Similar to the absolute precision case, if we find the correct scaling parameter r_δ, then we can terminate the simulation at $\tilde{\kappa}_m(\epsilon, r_\delta) = \inf\{t > 0 : (S_m(t)/\sqrt{m} + a(t)) < \epsilon\,|\bar{\alpha}_m(t)|\,/r_\delta\}$.

Following a similar line of analysis as in Proposition 2 and Theorem 1, we have the following convergence results respectively.

PROPOSITION 3. *If $\alpha \neq 0$ and A1 holds, then*

$$\epsilon^2\tilde{\kappa}_m(\epsilon, r) \Rightarrow \frac{r^2\sigma^2}{\alpha^2 m(m-1)}K_{m-1}(1) \text{ as } \epsilon \downarrow 0.$$

THEOREM 2. *Suppose $m \geq 4$. If $\alpha \neq 0$ and A1 holds, then*

$$\frac{\bar{\alpha}_m(\tilde{\kappa}_m(\epsilon)) - \alpha}{S_m(\tilde{\kappa}_m(\epsilon))/\sqrt{m}} \Rightarrow \sqrt{\frac{m-1}{m-3}}t_{m-3} \text{ as } \epsilon \downarrow 0.$$

Furthermore, if we set $r_\delta = \sqrt{\frac{m-1}{m-3}}z_{\delta,m-3}$, then

$$P\left(\alpha \in [\bar{\alpha}_m(\tilde{\kappa}_m(\epsilon, r_\delta))(1-\epsilon), \bar{\alpha}_m(\tilde{\kappa}_m(\epsilon, r_\delta))(1+\epsilon)]\right) \to 1-\delta \text{ as } \epsilon \downarrow 0.$$

In particular, we notice that we can use the same scaling parameter for the absolute precision and relative precision procedure. The actual simulation algorithm goes as follows.

Algorithm 2

Input: the number of sections m, the error bound ϵ, the confidence level $100(1-\delta)\%$, and the time increment Δ

Output: a $100(1-\delta)\%$ confidence interval with half width $\epsilon|\bar{Y}(t)|$

(i) Sample $X_i(t)$ for $0 \leq t \leq \epsilon^{-1/\beta}$ and $i = 1, 2, \ldots, m$. Initialize $T = \epsilon^{-1/\beta} + \Delta$.

(ii) Sample $X_i(t)$ between $T - \Delta$ and T. Update $\alpha_i(T)$ for $i = 1, 2, \ldots, m$, and calculate $\bar{\alpha}_m(T)$ and $S_m(T)$

(iii) If $r_\delta S_m(T)/\sqrt{m} \geq \epsilon|\bar{\alpha}_m(T)|$, set $T = T + \Delta$, go back to step (ii); otherwise, if $\bar{\alpha}_m(T) > 0$, output $[\bar{\alpha}_m(T)(1-\epsilon)\ ,\ \bar{\alpha}_m(T)(1+\epsilon)]$; if $\bar{\alpha}_m(T) < 0$, output $[\bar{\alpha}_m(T)(1+\epsilon)\ ,\ \bar{\alpha}_m(T)(1-\epsilon)]$ as the $100(1-\delta)\%$ confidence interval.

3. Examples for which the sectioning framework applies

In this section, we present some examples, for which A1 is satisfied, i.e. to which we can apply the sequential stopping sectioning framework.

We first review some important strong approximation results. The strong approximation, which is also known as the strong invariance principle, is a powerful tool in probability and statistics. We consider two important classes of processes here. One is the partial sums of random variables, $S(n) := \sum_{i=1}^{n} X(i)$. The other is the empirical process, $R(n,x) := \sum_{i=1}^{n} 1\{X(i) \leq x\}$. Let $b_1(n) = o(\sqrt{n})$ be some slowly increasing function. We also denote d as the dimension of $X(i)$. For the partial sums, we are interested in constructing an $\mathbb{R}^d$ valued Brownian motion with covariance Γ, after suitably enlarging the probability space, such that

$$|S(n) - n\mu - B(n)| = O(b(n)) \quad \text{a.s..} \tag{3.1}$$

For the empirical process, we are interested in constructing a centered Gaussian field $K_X(n,x)$ with covariance function $\Gamma_X(x,x',t,t') = \min\{t,t'\}\,(F(x \wedge x') - F(x)F(x'))$, after suitably enlarging the probability space, such that

$$\sup_{x \in \mathbb{R}^d} |R(n,x) - nF(x) - K_X(n,x)| = O(b(n)) \quad \text{a.s..} \tag{3.2}$$

When $X(k)$'s are iid real-valued random variables, Komlós et al. (1975, 1976) obtained the optimal Gaussian approximation for $S(n)$. In particular, if $E[\exp(t|X_1|)] < \infty$ for some $t > 0$, the optimal rate is $b(n) = \log n$, for $b(n)$ in (3.1); if $E[|X_1|^p] < \infty$, for $p > 2$, $b(n) = n^{1/p}$. They also showed that if $X(k)$'s are iid uniform on $[0,1]$, one can obtain (3.2) with $b(n) = (\log n)^2$.

There has since been a lot of work extending the results to the multidimensional case and weakly dependent sequences. For example, Einmahl (1989) and Zaitsev (1998) extended the strong

invariance principles to partial sums of independent vectors. Merlevede and Rio (2012) obtained strong invariance principle for partial sums of strictly stationary strongly mixing sequences of random variables. Dedecker et al. (2014) established the strong invariance principle for the empirical process associated to absolutely regular stationary sequences of random vectors. Berkes et al. (2014) studied strong approximation for stationary processes that are functions of iid innovations.

3.1. Example A: smooth functions of expectations

Estimating smooth functions of expectations arise a lot in statistics and simulation output analysis (see examples in Asmussen and Glynn (2007) Chapter III.3). Let $X(k)$'s be iid random vectors in $\mathbb{R}^d$. We are interested in computing $h(\mu)$, where $\mu = E[X(k)]$ and $h : \mathbb{R}^d \to \mathbb{R}$ is a "smooth" function. Let $\bar{X}_n$ denote the sample mean of $X(1), X(2), \ldots, X(n)$. Then

$$\alpha(n) = h(\bar{X}_n).$$

We next show that if we have a strong approximation for $\bar{X}_n$, then we can build from it a strong approximation for $h(\bar{X}_n)$.

LEMMA 1. *We assume that h differentiable and ∇h is Lipchitz continuous. If there exists a probability space supporting $\{X(k) : k \geq 0\}$ and a d-dimensional Brownian motion $\vec{B}$ with Covariance Γ, such that $\bar{X}_n = \mu + \vec{B}(n)/n + O(n^{-\gamma})$ a.s. for some $\gamma \in (1/2, 1)$, then there exists a standard Brownian motion B, such that*

$$h(\bar{X}_n) = h(\mu) + \sigma \frac{B(n)}{n} + O\left(n^{-\gamma}\right) \quad a.s.,$$

where $\sigma^2 = \nabla h(\mu) \Gamma \nabla h(\mu)^T$.

3.2. Example B: quantile and quantile process

Let X be a real-valued random variable with cumulative distribution function (cdf) $F(\cdot)$. For any $0 < p < 1$, the p-th quantile of X is defined as the generalized inverse function of F at p, i.e. $F^{-1}(p) = \inf\{x : F(x) \geq p\}$. Let $Q(n, p)$ denote the p-th sample quantile of X. If we write $X_{(n,1)} \leq X_{(n,2)} \leq \cdots \leq X_{(n,n)}$ as the ordered statistics of $X(1), X(2), \ldots, X(n)$, then $Q(n, p) = X_{(n,\lceil np \rceil)}$.

LEMMA 2. *For fixed $p \in (0,1)$, if F is twice differentiable in a neighborhood of $F^{-1}(p)$ and $f(F^{-1}(p)) := F'(F^{-1}(p)) > 0$, then we can construct a probability space supporting both $\{Q(n,p) : n \geq 0\}$ and a standard Brownian motion B, such that*

$$Q(n,p) = F^{-1}(p) + \sigma \frac{B(n)}{n} + o\left(n^{-\gamma}\right) \quad a.s.,$$

for any $1/2 < \gamma < 3/4$, where $\sigma = \sqrt{p(1-p)}/f(F^{-1}(p))$.

The difficulty in constructing confidence intervals for quantiles arises in the unknown value of $f(F^{-1}(p))$, and estimating the density function, $f(\cdot)$, is a costly task.

More generally, at the process level, if F is twice differentiable and $f := F' > 0$ on (a,b) where $a = \sup\{x : F(x) = 0\}$, $b = \inf\{x : F(x) = 1\}$, and in addition $\sup_{a<x<b} F(x)(1-F(x)) \left|\frac{f'(x)}{f(x)^2}\right| < \infty$, then Csörgö and Révész (1978) proved that one can construct a Kiefer process $\{K(t,p) : t \geq 0, 0 \leq p \leq 1\}$, after suitably enlarging the probability space, such that

$$\sup_{\delta_n \leq p \leq 1-\delta_n} \left| nQ(n,p) - nF^{-1}(p) - \frac{K(n,p)}{f(F^{-1}(p))} \right| = O\left((n \log\log n)^{1/4} (\log n)^{1/2}\right),$$

where $\delta_n = 25n^{-1}\log\log n$. From this, we can build strong approximation results for more general quantile related estimation problems, such as inter-quantile range.

3.3. Example C: steady-state simulation

For steady-state performance analysis problems, we are interested in estimating $\alpha = \psi(\pi) = \int h(x)d\pi(s)$, where π is the equilibrium distribution of X and h is a real-valued function defined on the state space of the stochastic process X. In this case,

$$\alpha(t) = \frac{1}{t}\int_0^t h(X(s))ds.$$

In this section, we focus on additive functionals of positive Harris recurrent process. For Harris recurrent chain under certain regularity conditions, we have (Asmussen and Glynn 2007)

$$\sqrt{t}(\hat{\alpha}(t) - \alpha) \Rightarrow N(0, \sigma^2),$$

where σ^2 is the TAVC. In particular, let $\rho_h(s) = Cov_\pi(h(X(0)), h(X(s)))$ and $\sigma_h^2 = Var_\pi(h(X(0))$. If $\int_0^\infty |\rho_h(s)|ds < \infty$ and $\sigma_h^2 < \infty$, then $\sigma^2 = 2\sigma_h^2 \int_0^\infty \rho_h(s)ds$.

We start by discussing the discrete time Harris chain. Consider $\{X(n) : n \geq 0\}$ defined on $(\Omega, \mathcal{F})$ a Harris recurrent chain. Then there exists an $\mathcal{F}$-measurable set $A \subset \Omega$, a positive constant $\lambda > 0$, an integer $l \geq 1$, and a probability measure ψ on $(\Omega, \mathcal{F})$ such that i) $P(\tau_A < \infty) = 1$, where $\tau_A = \inf\{n \geq 0 : X_n \in A\}$, and ii) $P_x(X_l \in B) \geq \lambda\psi(B)$ for all $x \in A$ and $B \in \mathcal{F}$. i) and ii) allows us to identify regenerative structures for X. In particular, we can define a sequence of infinitely many regenerative times $T_0 < T_1 < T_2 < \ldots$, such that for $Z_k = \sum_{i=T_{k-1}}^{T_k} h(X(i))$ and $\tau_k = T_k - T_{k-1}$, $\{(Z_k, \tau_k) : k \geq 1\}$ form a one-dependent stationary sequence (Glynn 2011).

When $l = 1$ in ii), (Z_k, τ_k)'s are iid. Csaki and Csorgo (1995) established the following strong invariance principle for partial sums of $h(X(i))$.

THEOREM 3 **(Csaki and Csorgo (1995))**. *If $E[\tau_k^p] < \infty$ for some $1 \leq p \leq 2$ and $E[Z_k^q] < \infty$ for some $q > 2$, then on a suitable probability space, one can construct $\{X(n) : n \geq 0\}$ together with a standard Brownian motion $\{B(t) : t \geq 0\}$, such that*

$$\left| \frac{1}{n} \sum_{i=1}^{n} h(X(i)) - \alpha - \frac{\sigma B(n)}{n} \right| = O\left(n^{-\tilde{\gamma}} \log n\right),$$

where $\tilde{\gamma} = \min\{1 - 1/(2p), 1 - 1/q\}$

When $\{X(n) : n \geq 0\}$ is a geometrically ergodic Markov chain and h is a bounded measurable function, Merlevede and Rio (2015) improves the strong approximation error bound to $O(\log n)$.

When $l > 1$, we can establish similar strong approximation results. We would like to point out that the rate that we established here is far from optimal.

THEOREM 4. *If $E[\tau_k^p] < \infty$ for some $p > 2$ and $E\left[\left(\sum_{i=T_{k-1}}^{T_k} |h(X(i)) - \alpha|\right)^q\right] < \infty$ for some $q > 2$, then on a suitable probability space, one can construct $\{X(n) : t \geq 0\}$ together with a standard Brownian motion $\{B(t) : t \geq 0\}$, such that*

$$\left| \frac{1}{n} \sum_{i=1}^{n} h(X(i)) - \alpha - \sigma \frac{B(n)}{n} \right| = O(n^{-\tilde{\gamma}} \log n).$$

where $\tilde{\gamma} = \min\{1 - 1/p, 1 - 1/q, 3/4\}$.

For continuous Harris processes, $\{X(t) : t \geq 0\}$, we can construct a regenerative structure using the following property. If $\{\Gamma_n : n \geq 0\}$ is the sequence of jump times of a unit rate Poisson process, independent of X, then $\{X(\Gamma_n) : n \geq 0\}$ is a discrete time Harris recurrent chain. We can then exploit the one-dependent regenerative structure of $\{X(\Gamma_n) : n \geq 0\}$ to construct one-dependent regenerative cycles for $\{X(t) : t \geq 0\}$ (Glynn 2011). In particular, we can find an infinitely sequence of regenerative times $T_0 < T_1 < T_2 < \ldots$, such that for $Z_k = \int_{T_{k-1}}^{T_k} h(X(s))ds$ and $\tau_k = T_k - T_{k-1}$, $\{(Z_k, \tau_k) : k \geq 1\}$ form a one-dependent sequence of identically distributed random vectors. We can then establish the following strong approximation results.

THEOREM 5. *If $E[\tau_k^p] < \infty$ for some $p > 2$ and $E[Z_k^q] < \infty$ for some $q > 2$, then on a suitable probability space, one can construct $\{X(t) : t \geq 0\}$ together with a standard Brownian motion $\{B(t) : t \geq 0\}$, such that*

$$\left| \frac{1}{t} \int_0^t h(X(s))ds - \alpha - \sigma \frac{B(t)}{t} \right| = O\left(t^{-\tilde{\gamma}} \log t\right),$$

where $\tilde{\gamma} = \min\{1 - 1/p, 1 - 1/q, 3/4\}$.

4. Simulation experiments

In this section, we demonstrate the performance of our sequential stopping procedures through some simulation experiments.

4.1. Quantile estimation

We apply Algorithm 1 to construct the 95% confidence interval (CI) for the p-quantile of an exponential random variable with unit rate. $\alpha(t)$ is the $\lceil pt \rceil$-th order statistics of t iid exponentials with rate 1. For this example, the quantiles can be calculated in closed form, i.e. the p-quantile is $\xi_p = -\log(1-p)$. We can also evaluate the asymptotic variance in closed form, i.e. $\sigma^2 = p/(1-p)$. Table 1 - 3 summarizes some of the simulation results. We compare the performance of the sequential stopping procedure to the fixed sample-size procedure. Note that as we can calculate σ, we can also calculate the required sample-size in the fixed sample-size procedure. For Algorithm 1, we set $\beta = 2.9/4$ and $\Delta = \epsilon^{-1}$. For both procedures, we report the coverage rates C_{seq} and C_{fix} based on 10^3 replications of the confidence interval. For the sequential procedure, we report $E[\kappa_m(\epsilon, r)]$

and $\text{std}(\kappa_m(\epsilon, r))$ based on 10^3 iid replications. For the fixed sample sample-size procedure, we report the expected half-width of the confidence interval, $E[w(\epsilon)]$, and the coefficient of variation of the half-width, $c(w(\epsilon))$, based on 10^3 iid replications.

Table 1 reports results for different values of ϵ. We observe that for ϵ reasonably small, e.g. $\epsilon \leq 0.1$, both procedure achieves the correct coverage rate, 0.95. For $\epsilon = 0.5$, both procedures do not quite achieve the desired coverage rate, this is because the process level dynamics is far from the dynamics of the limiting process when ϵ is large. The calculation in (2.6) suggests that $E[\kappa_{10}(\epsilon, r)]/\tau(\epsilon) \approx 1.15$. We indeed observe the ratio close to 1.15 in Table 1 for ϵ small.

Table 1 Performance of the sequential stopping procedure and the fixed sample-size procedure to achieve an absolute precision level ϵ with $m = 10$ sections: the 0.9-quantile of Exp(1)

ϵ	C_{seq}	$E[\kappa_{10}(\epsilon, r)]$	$\text{std}(\kappa_{10}(\epsilon, r))$	C_{fix}	$\tau(\epsilon)$	$E[w(\epsilon)]$	$c(w(\epsilon))$
0.5	0.968	26	12	0.923	18	0.53	0.28
0.1	0.956	516	265	0.949	436	0.10	0.24
0.05	0.955	2024	1091	0.949	1743	0.05	0.24
0.01	0.950	50172	26304	0.953	43573	0.01	0.24

Table 2 reports results for different quantiles p. We note that as p gets larger, the estimation problem becomes more difficult/costly. Thus a larger sample size is needed to achieve the desired level of accuracy. We observe that both procedures perform well across different problem instances. When $p = 0.8$, the sequential stopping procedure achieves a slightly higher coverage rate than 0.95. This is because the required sample size in this case is fairly small. Setting increment $\Delta = \epsilon^{-1} = 10$ in Algorithm 1 is too large for this problem instance. Specifically, when we set $\Delta = 2$ in Algorithm 1, the estimated coverage rate is 0.952.

Table 3 reports results for different choice of m. We note that as m increases, both $E[m\kappa_m(\epsilon, r)]$ and $m\tau(\epsilon)$ decreases. There's also less uncertainty in the sense that both $\text{std}(m\kappa_m(\epsilon, r))$ and $c(w(\epsilon))$ decreases as well. However, the decreasing effect is diminishing as m increases. When $m = 5$, C_{seq}

Table 2 **Performance of the sequential stopping procedure and the fixed sample-size procedure to achieve an absolute precision level $\epsilon = 0.1$ with $m = 10$ sections: the p-quantile of Exp(1)**

p	C_{seq}	$E[\kappa_{10}(\epsilon, r)]$	$\mathrm{std}(\kappa_{10}(\epsilon, r))$	C_{fix}	$\tau(\epsilon)$	$E[w(\epsilon)]$	$c(w(\epsilon))$
0.8	0.965	249	111	0.952	194	0.10	0.24
0.95	0.952	1090	582	0.949	920	0.10	0.24
0.99	0.949	5484	2887	0.947	4793	0.10	0.24

is larger than 0.95. This may be because when $m = 5$, there is too much variability in the underline procedure. When $m = 40$, $C_{fix} < 0.95$. However, if we reduce ϵ to 0.05 or 0.01, $C_{fix} \approx 0.95$. Overall, our suggestion is to choose m between 10 and 30

Table 3 **Performance of the sequential stopping procedure and the fixed sample-size procedure to achieve an absolute precision level $\epsilon = 0.1$ with m sections: the 0.9-quantile of Exp(1)**

m	C_{seq}	r	$E[m\kappa_m(\epsilon, r)]$	$\mathrm{std}(m\kappa_m(\epsilon, r))$	C_{fix}	$m\tau(\epsilon)$	$E[w(\epsilon)]$	$c(w(\epsilon))$
5	0.978	6.08	18214	16276	0.948	6135	0.10	0.37
20	0.951	2.23	4220	1357	0.953	3860	0.10	0.17
30	0.954	2.13	3978	1048	0.946	3720	0.10	0.14
40	0.950	2.08	3906	879	0.876	3640	0.10	0.12

We lastly provide some comments about the choice of the parameter β. We try different problem instances, and find the performance of the algorithm quite robust to different choices of the parameter β. Table 4 lists some of our simulation results. In Table 4, our goal is to construct the 95% confidence interval with half-width $\epsilon = 0.05$ for the 0.9-quantile for different distributions. We report the average coverage rate based on 10^3 independent replications of Algorithm 1 with $D = 10$. We note, for β ranges from close to 1/2 to close to 3/4, the estimated coverage rates are all very close to 0.95.

4.2. Steady-state simulation of M/M/1 queue

We consider an M/M/1 queue, which has a Poisson arrival process with rate λ, and iid exponential service times with mean $1/\mu$. Let $X(t)$ denote the number of customers in the system at time

Table 4 **Performance of the sequential stopping procedure to achieve an absolute precision level $\epsilon = 0.05$ with $m = 10$ sections: the 0.9-quantile of different distributions**

Distribution	$\beta = 2.1/4$	$\beta = 2.5/4$	$\beta = 2.9/4$
N(0,1)	0.965	0.945	0.952
Exp(1)	0.949	0.949	0.950
Pareto(1,3)	0.952	0.947	0.950

t. We are interested in estimating $\alpha = \lim_{T\to\infty} \int_0^T X(t)dt = \rho/(1-\rho)$ (the steady-state average number of people in the system), where $\rho = \lambda/\mu$, is called the traffic intensity. We apply Algorithm 1 to construct the 95% CI for α. The results are summarized in Table 5 & 6. The coverage rates, expectations and standard deviations are calculated based on 10^3 independent replications of the procedures.

Table 5 reports results for different values of ϵ. Similar to our quantile estimation examples, when $\epsilon = 0.5$, the estimation process dynamics still far from the limiting behavior. Thus, both procedures do not achieve the correct coverage rate, and the sequential stopping procedure suffers quite a bit from under-coverage. However, for reasonably small values of ϵ, e.g. $\epsilon \leq 0.1$, both procedures achieve the desired coverage rate.

Table 5 **Performance of the sequential stopping procedure to achieve an absolute precision level ϵ with $m = 10$ sections: Queue length (number of people in the system) process of an $M/M/1$ queue ($\lambda = 0.8$, $\mu = 1$)**

ϵ	C_{seq}	$E[\kappa_{10}(\epsilon, r)]$	$\text{std}(\kappa_{10}(\epsilon, r))$	C_{fix}	$\tau(\epsilon)$	$E[w(\epsilon)]$	$c(w(\epsilon))$
0.5	0.874	3.51×10^3	2.47×10^3	0.930	3.53×10^3	0.48	0.33
0.1	0.952	1.03×10^5	5.42×10^4	0.945	8.83×10^4	0.10	0.24
0.05	0.949	4.12×10^5	2.18×10^5	0.947	3.53×10^5	0.05	0.24
0.01	0.948	1.02×10^7	5.31×10^6	0.951	8.83×10^6	0.01	0.24

Table 2 reports results for different values of the traffic intensity ρ (this is achieved by varying the arrival rate λ). We note that as ρ gets larger, the estimation problem becomes more difficult/costly

as there is more correlation across time. Thus, a larger sample size is needed to achieve the desired level of accuracy. We observe that both procedures perform well across different problem instances.

Table 6 **Performance of the sequential stopping procedure to achieve an absolute precision level $\epsilon = 0.1$ with $m = 10$ sections: Queue length (number of people in the system) process of an $M/M/1$ queue ($\mu = 1$)**

ρ	C_{seq}	$E[\kappa_{10}(\epsilon, r)]$	$\text{std}(\kappa_{10}(\epsilon, r))$	C_{fix}	$\tau(\epsilon))$	$E[w(\epsilon)]$	$c(w(\epsilon))$
0.85	0.953	3.46×10^5	1.94×10^5	0.946	2.98×10^5	0.10	0.24
0.9	0.951	1.94×10^6	1.01×10^6	0.950	1.68×10^6	0.10	0.24
0.95	0.952	3.20×10^7	1.86×10^7	0.954	2.76×10^7	0.10	0.24

5. Comparison to resampling techniques

Traditionally, for simulation problems where variance estimation is difficult, resampling techniques has been applied either to i) construct strongly consistent variance estimator so that the original sequential stopping framework of Glynn and Whitt (1992) can be applied, or to ii) estimate the distribution of sample statistics so that we can develop corresponding sequential stopping procedures. Bootstrap (Efron 1979) and Jackknife (delete-d Jackknife)(Wu 1986) are among the mostly commonly used techniques. Take the bootstrap as an example. Let α denote the true quantity that we are interested in, and $T_n(X_1, \ldots, X_n)$ denote the estimator constructed based on n samples. We write the bootstrap variance as

$$\hat{\sigma}_n^{*2} := Var^*(T_n(X_1^*, \ldots, X_n^*)|X_1, X_2, \ldots, X_n),$$

where X_i^* are iid samples from the empirical distribution F_n conditional on $X_1, X_2, \ldots, X_n$. Then for the smooth functions of expectations problem and quantile estimation problem covered in §3, under mild moments and smoothness conditions, we have $\hat{\sigma}_n^{*2}$ is a strongly consistent estimator of $\sigma_n^2 = Var(T_n(X_1, \ldots, X_n))$, i.e. $\hat{\sigma}_n^{*2}/\sigma_n^2 \to 1$ a.s. as $n \to \infty$ (see Theorem 3.8 & 3.9 in Shao and Tu (1995)). Futhermore, let

$$\hat{p}_n(\epsilon) := P^*(|T_n(X_1^*, \ldots, X_n^*) - T_n(X_1, \ldots, X_n)| + n^{-\beta} < \epsilon)$$

and

$$N(\epsilon) := \inf\{n > 0 : \hat{p}_n(\epsilon) \geq 1 - \delta\}.$$

Then it is shown in Swanepoel et al. (1983) that

$$\lim_{\delta \to 0} P(T_{N(\epsilon)}\left(X_1, \ldots, X_{N(\epsilon)}\right) - \epsilon < \alpha < T_{N(\epsilon)}(X_1, \ldots, X_{N(\epsilon)}) + \epsilon\big) = 1 - \delta.$$

Compared to the sectioning framework we studied in this paper, the resampling techniques are much more computationally intensive (expensive). Take the delete-d Jackknife method as an example, to achieve strongly consistent variance estimator, at each step n, we need to conduct the sample statistics computation $\binom{n}{d_n}$ times for samples of size $n - d_n$ each. For the bootstrap method, if we can not calculate the distribution of bootstrap statistics in closed form, then we need to conduct Monte Carlo simulation based on the empirical distribution (sampling with replacement). This requires drawing nB samples at step n, and calculate B sample statistics, each based on samples of size n. To control the error in this Monte Carlo estimation step, we require B to be large as well. Even if we can calculate the distribution of bootstrap statistics in closed form, the calculation can be intensive (e.g. for quantiles, it scales as $n!$). In contrast, for the sectioning framework we used here, we often have an efficient way of updating the statistics within each section, and the cross-section calculation often involves only m aggregated statistics. Here, m is the number of sections, which is fixed and does not grow with n.

6. Conclusion

In this paper, we analyze the sequential stopping problem for a class of simulation problems in which variance estimation is difficult. Specifically, we prove that when applying sequential stopping rules to the sectioning method, by properly adjusting the scaling parameter, we will be able to construct asymptotically valid confidence intervals. Our numerical results confirms that our sequential stopping algorithms perform very well (achieves the correct coverage rate) when the precision level is reasonably small, e.g. $\epsilon \leq 0.1$.

The sequential stopping procedure for the sectioning method we developed is very easy to implement (Algorithm 1 & 2). It is naturally adapted to the parallel computing environment and

has minimum storage requirement for a lot of interesting applications. The scaling parameters only depend on the number of sections and the confidence level we want to achieve.

Appendix A: Proof of Section 2

Proof of Proposition 1 We note that for any $s,t\geq 0$, the independence of the B_i's implies that

$$Cov(\bar{B}_m(s), B_i(t)-\bar{B}_m(t)) = \frac{1}{m}Cov(B_i(s),B_i(t)) - \frac{1}{m^2}\sum_{j=1}^{m} Cov(B_j(s),B_j(t)) = 0.$$

Since $\bar{B}_m$ and $(B_i-\bar{B}_m, 1\leq i\leq m)$ are Gaussian processes, it follows that $(\bar{B}_m(s_1),\ldots,\bar{B}_m(s_n))$ is independent of $(B_i(t_j)-\bar{B}_m(t_j) : 1\leq i\leq m, 1\leq j\leq l)$ for any set of time indices $s_1,\ldots,s_n$, $t_1,\ldots,t_l$. Because the finite-dimensional distributions characterize probabilities on $C[0,\infty)$ (the space of continuous functions on $[0,\infty)$; see Billingsley (1999) p. 80), we conclude that $\bar{B}_m$ is independent of $(B_i-\bar{B}_m, 1\leq i\leq m)$. Since R_m is a (measurable) function of $(B_i-\bar{B}_m, 1\leq i\leq m)$, it follows that R_m is independent of $\bar{B}_m$.

Note that $\sqrt{m}\bar{B}_m$ is a Gaussian process that has the same mean and covariance function as B. Therefore, $\sqrt{m}\bar{B}_m \stackrel{D}{=} B$.

As for R_m, set $V(t)=R_m(t)^2$ and use Itô's formula, yielding

$$\begin{aligned}
dV(t) &= \sum_{i=1}^{m} d(B_i(t)-\bar{B}_m(t))^2 \\
&= 2\sum_{i=1}^{m}(B_i(t)-\bar{B}_m(t))d(B_i(t)-\bar{B}_m(t)) + \sum_{i=1}^{m} d(B_i(t)-\bar{B}_m(t))d(B_i(t)-\bar{B}_m(t)) \\
&= 2\sum_{i=1}^{m}(B_i(t)-\bar{B}_m(t))dB_i(t) + \sum_{i=1}^{m}(dB_i(t))^2 - 2\sum_{i=1}^{m}(dB_i(t))(d\bar{B}_m(t)) + m(d\bar{B}_m(t))^2 \\
&= 2\sum_{i=1}^{m}(B_i(t)-\bar{B}_m(t))dB_i(t) + mdt - 2dt + dt \\
&= 2\sum_{i=1}^{m}(B_i(t)-\bar{B}_m(t))dB_i(t) + (m-1)dt.
\end{aligned}$$

Set

$$W(t) = \sum_{i=1}^{m}\int_0^t \frac{(B_i(s)-\bar{B}(s))}{R_m(s)} I(R_m(s)>0)dB_i(s).$$

Observe that the exchangeability of $B_1,\ldots,B_m$ and the the fact that $R_m(s)>0$ a.s. for $s>0$ implies that

$$E\left[\left(\frac{B_i(s)-\bar{B}_m(s)}{R_m(s)}\right)^2 1(R_m(s)>0)\right] = \frac{1}{m}$$

for $s>0$. Therefore, $W=(W(t):t\geq 0)$ is a well-defined continuous square-integrable stochastic integral, and its quadratic variation is given by

$$[W,W](t) = \int_0^t \sum_{i=1}^{m}\left(\frac{B_i(s)-\bar{B}_m(s)}{R_m(s)}\right)^2 ds = t.$$

Consequently, W is a standard Brownian motion (Revuz and Yor (1999), p. 150), and

$$dV(t) = 2R_m(t)dW(t) + (m-1)dt = 2\sqrt{V(t)}dW(t) + (m-1)t.$$

This is the SDE satisfied by Z^2_{m-1} (Revuz and Yor (1999), p. 439), and hence $R_m \stackrel{D}{=} Z_{m-1}$. $\square$

Proof of Proposition 2 Let $\eta_\epsilon = r\sigma/(\epsilon\sqrt{m(m-1)})$. We also write

$$\lambda(t) = \frac{S_m(t)}{\sqrt{m}} - \sqrt{m(m-1)}\frac{R_m(t)}{t}.$$

Then under A1, we have $\lambda(t) = o(t^{-\gamma})$.

$$\begin{aligned}
\epsilon^2\kappa_m(\epsilon, r) &= \epsilon^2 \inf\left\{t : t > \epsilon^{-1/\beta}, \frac{S_m(t)}{\sqrt{m}} < \epsilon/r\right\} \\
&= \epsilon^2 \inf\left\{t : t > \epsilon^{-1/\beta}, \frac{r\sigma}{\epsilon\sqrt{m(m-1)}}\frac{R_m(t)}{t} + \frac{r}{\epsilon}\lambda(t) < 1\right\} \\
&= \epsilon^2 \inf\left\{\eta_\epsilon^2 t : \eta_\epsilon^2 t > \epsilon^{-1/\beta}, \eta_\epsilon\frac{R_m(\eta_\epsilon^2 t)}{\eta_\epsilon^2 t} + \frac{r}{\epsilon}\lambda(\eta_\epsilon^2 t) < 1\right\} \\
&\stackrel{D}{=} \epsilon^2\eta_\epsilon^2 \inf\left\{t : t > \epsilon^{-1/\beta}\eta_\epsilon^{-2}, \frac{R_m(t)}{t} + \frac{r}{\epsilon}\lambda(\eta_\epsilon^2 t) < 1\right\} \\
&= \frac{r^2\sigma^2}{m(m-1)} \inf\left\{t : t > \epsilon^{-1/\beta}\eta_\epsilon^{-2}, \frac{R_m(t)}{t} + \frac{r}{\epsilon}\lambda(\eta_\epsilon^2 t) < 1\right\} \\
&\Rightarrow \frac{r^2\sigma^2}{m(m-1)} \inf\left\{t : t > 0, \frac{R_m(t)}{t} < 1\right\} \text{ as } \epsilon \downarrow 0.
\end{aligned}$$

The last convergence follows as for $t \geq \epsilon^{-1/\beta}\eta_\epsilon^{-2}$, $\frac{r}{\epsilon}\lambda(\eta_\epsilon^2 t) \Rightarrow 0$ as $\epsilon \downarrow 0$. $\square$

Proof of Theorem 1 We first notice that under A1, for $i = 1, 2, \ldots, m$,

$$\frac{1}{\epsilon}\left(\hat{\alpha}_i(t/\epsilon) - \alpha\right) \Rightarrow \sigma\frac{B_i(t)}{t} \text{ in } D(0,\infty) \text{ as } \epsilon \downarrow 0.$$

Then by continuous mapping theorem

$$\frac{\frac{1}{\epsilon}\sqrt{m}\left(\bar{\alpha}(t/\epsilon^2) - \alpha\right)}{\frac{1}{\epsilon}S_m(t/\epsilon^2)} \Rightarrow \frac{\sqrt{m}\sigma\bar{B}_m(t)/t}{\frac{1}{\sqrt{m-1}}\sigma R_m(t)/t} \text{ in } D(0,\infty) \text{ as } \epsilon \downarrow 0. \tag{A.1}$$

We also notice that

$$\frac{\sqrt{m}\left(\bar{\alpha}_m(\kappa_m(\epsilon, r)) - \alpha\right)}{S_m(\kappa_m(\epsilon, r))} = \frac{\frac{1}{\epsilon}\sqrt{m}\left(\bar{\alpha}_m(\epsilon^2\kappa_m(\epsilon, r)/\epsilon^2) - \alpha\right)}{\frac{1}{\epsilon}S_m(\epsilon^2\kappa_m(\epsilon, r)/\epsilon^2)}. \tag{A.2}$$

We next make two observations that will allow us to apply continuous mapping theorem to (A.2). 1) The limiting process in (A.1) is continuous. 2) $\kappa_m(\epsilon/t, r)$ is increasing in t, and from the proof of Proposition 2,

$$\epsilon^2\kappa_m(\epsilon/t, r) \Rightarrow \frac{r^2\sigma^2}{m(m-1)}K_{m-1}(t) \text{ in } D(0,\infty) \text{ as } \epsilon \to 0,$$

where $K_{m-1}(t)$ is also increasing in t. Then the composition map is continuous (Theorem 3.1 in Whitt (1980)). By continuous mapping theorem, we have

$$\begin{aligned}
&\frac{\frac{1}{\epsilon}\sqrt{m}\left(\bar{\alpha}_m(\epsilon^2\kappa_m(\epsilon/t,r)/\epsilon^2)-\alpha\right)}{\frac{1}{\epsilon}S_m(\epsilon^2\kappa_m(\epsilon/t,r)/\epsilon^2)} \\
\Rightarrow\ &\frac{\sigma\sqrt{m}\bar{B}_m\left(\frac{r^2\sigma^2}{m(m-1)}K_{m-1}(t)\right)/\left(\frac{r^2\sigma^2}{m(m-1)}K_{m-1}(t)\right)}{\frac{1}{\sqrt{m-1}}\sigma R_m\left(\frac{r^2\sigma^2}{m(m-1)}K_{m-1}(t)\right)/\left(\frac{r^2\sigma^2}{m(m-1)}K_{m-1}(t)\right)} \text{ in } D(0,\infty) \text{ as } \epsilon\downarrow 0 \\
\stackrel{D}{=}\ &\frac{\sqrt{m}\bar{B}_m(K_{m-1}(t))}{\frac{1}{\sqrt{m-1}}R_m(K_{m-1}(t))} \\
=\ &\frac{\sqrt{m}\bar{B}_m(K_{m-1}(t))/\sqrt{K_{m-1}(t)}}{\frac{1}{\sqrt{m-1}}\sqrt{K_{m-1}(t)/t^2}}
\end{aligned}$$

Then, applying continuous mapping with the projection map at $t=1$, we have

$$\frac{\sqrt{m}\left(\bar{\alpha}_m(\kappa_m(\epsilon,r))-\alpha\right)}{S_m(\kappa_m(\epsilon,r))} \Rightarrow \frac{\sqrt{m}\bar{B}_m(K_{m-1}(1))/\sqrt{K_{m-1}(1)}}{\sqrt{K_{m-1}(1)}/\sqrt{m-1}} \text{ as } \epsilon\to\infty.$$

Since $K_{m-1}(1)$ is independent of $\sqrt{m}\bar{B}_m$, it follows (by conditioning on $K_{m-1}(1)$) that

$$\frac{\sqrt{m}\bar{B}_m(K_{m-1}(1))/\sqrt{K_{m-1}(1)}}{\sqrt{K_{m-1}(1)}/\sqrt{m-1}} \stackrel{D}{=} \sqrt{m-1}\frac{N(0,1)}{\sqrt{K_{m-1}(1)}},$$

where $K_{m-1}(1)$ is independent of the $N(0,1)$. Now as $K_{m-1}(1)\stackrel{D}{=}\chi^2_{m-3}$,

$$\sqrt{m-1}\frac{N(0,1)}{\sqrt{K_{m-1}(1)}} \stackrel{D}{=} \sqrt{\frac{m-1}{m-3}}t_{m-3}$$

Lastly,

$$\begin{aligned}
&P\left(|\bar{\alpha}(\kappa_m(\epsilon,r_\delta))-\alpha|<\epsilon\right) \\
=\ &P\left(\frac{1}{\epsilon}|\bar{\alpha}(\kappa_m(\epsilon,r_\delta))-\alpha|<1\right) \\
\to\ &P\left(\left|\frac{\sigma\bar{B}_m\left(\frac{r_\delta^2\sigma^2}{m(m-1)}K_{m-1}(1)\right)}{\frac{r_\delta^2\sigma^2}{m(m-1)}K_{m-1}(1)}\right|<1\right) \text{ as } \epsilon\downarrow 0 \\
=\ &P\left(\left|\frac{\sqrt{m}\bar{B}_m(K_{m-1})/\sqrt{K_{m-1}(1)}}{\sqrt{K_{m-1}(1)}/\sqrt{m-1}}\right|<r_\delta\right) \\
=\ &P\left(\sqrt{\frac{m-1}{m-3}}t_{m-3}<r_\delta\right)=1-\delta.
\end{aligned}$$

□

The proof of Proposition 3 follows similar lines of argument as the proof of Proposition 2. We only need to notice that under Assumption 1

$$\bar{\alpha}(t/\epsilon^2)\Rightarrow\alpha\mathcal{U}(t) \text{ in } D(0,\infty) \text{ as } \epsilon\to 0,$$

where $\mathcal{U}(t) = 1$ for $t \in \mathbb{R}$. We shall omit the rest of the proof to avoid repetition.

Similarly, the proof of Theorem 2 follows the same lines of analysis as the proof of Theorem 1. We shall omit it here.

Appendix B: Proof of Section 3

Proof of Lemma 1 Let K denote the Lipchitz constant of ∇h. We first notice that by the mean value theorem

$$\left| h(\bar{X}_n) - h(\mu) - \nabla h(\mu) \frac{\vec{B}(n)}{n} \right| \leq \left| \nabla h(\mu) \left(\bar{X}_n - \mu - \frac{\vec{B}(n)}{n} \right) \right| + K||\bar{X}_n - \mu||^2$$

Then by the strong approximation assumption on $\bar{X}_n$, we have

$$\left| \nabla h(\mu) \left(\bar{X}_n - \mu - \frac{\vec{B}(n)}{n} \right) \right| = O\left(n^{-\gamma}\right) \quad \text{a.s..}$$

By the law of iterated logarithm, we have

$$K||\bar{X}_n - \mu||^2 = O(n^{-1} \log\log n) \quad \text{a.s..}$$

Thus, $\left| h(\bar{X}_n) - h(\mu) - \nabla h(\mu) \frac{\vec{B}(n)}{n} \right| = O\left(n^{-\gamma}\right)$ a.s.. Let $B(n) = \frac{1}{\sigma} \nabla h(\mu) \vec{B}(n)$, then we have B is a standard Brownian motion. □

Proof of Lemma 2 Let $\xi_p = F^{-1}(p)$. Let $S_n = \sum_{i=1}^{n} 1\{X(i) < \xi_p\}$. Then using strong approximation (Komlós et al. 1975), we can construct a standard Brownian motion, after suitably enlarging the probability space, such that

$$|S_n - np - \sqrt{p(1-p)} B(n)| = O(\log n) \text{ a.s..}$$

By Bahadur representation Bahadur (1966) we also have

$$Q(n,p) = \xi_p - \frac{S_n/n - p}{f(\xi_p)} = O(n^{-3/4} (\log n)^{1/2} (\log\log n)^{1/4}) \text{ a.s..}$$

Thus, $Q(n,p) = \xi_p - \frac{\sqrt{p(1-p)}}{f(\xi_p)} B(n) + o(n^{-\gamma})$ for any $\gamma < 3/4$. □

The proofs of Theorem 4 and 5 follow exact the same line of arguments, which rely on the one-dependent regenerative cycles structure. We shall provide the proof of Theorem 5 here only.

Proof of Theorem 5 We first introduce a few notations. Let $\mu_1 = E[Z_1]$, $\sigma_1^2 = Var(Z_1)$, and $\mu_2 = E[\tau_1]$, $\sigma_2^2 = Var(\tau_1)$. Then we have $\alpha = \mu_1/\mu_2$ and

$$\sigma^2 = \frac{Var(Z_1 - \alpha\tau_1)}{E[\tau_1]} = \frac{\sigma_1^2 - 2\alpha Cov(Z_1, \tau_1) + \alpha^2 \sigma_2^2}{\mu_2}.$$

We also write $L(t) = \sup\{k > 0 : T_k \le t\}$. Let $\beta = Cov(Z_1, \tau_1)/Var(\tau_1)$. Then $Z_k - \mu_1 - \beta(\tau_k - \mu_2)$ and τ_k are uncorrelated. Let $\tilde{\sigma}^2 = Var(Z_k - \mu_1 - \beta(\tau_k - \mu_2))$ As $\{(Z_k - \mu_1 - \beta(\tau_k - \mu_2), \tau_k) : k \ge 0\}$ is a sequence of one-dependent stationary sequence, by Theorem 2.1 in Liu and Lin (2009), on a richer probability space, we can construct two independent Brownian motions B_1 and B_2, such that

$$\sup_{1\le k\le n} \left| \sum_{i=1}^{k} Z_i - \mu_1 k - \alpha(T_k - \mu_2 k) - \tilde{\sigma} B_1(k) \right| = o\left(n^{1/q}\right) \text{ a.s.}, \tag{B.1}$$

$$\sup_{1\le k\le n} |T_k - \sigma_2 B_2(k)| = o\left(n^{1/p}\right) \text{ a.s.}. \tag{B.2}$$

Based on (B.2) and Theorem 3.1 of Csörgö et al. (1987), on a richer probability space, we can construct a Brownian motion $\tilde{B}_2$, such that

$$\left| L(t) - \frac{t}{\mu_2} - \sigma_2 \mu_2^{3/2} \tilde{B}_2(t) \right| = o\left(t^{1/p}\right) \text{ a.s.}. \tag{B.3}$$

Notice that as B_2 is independent of B_1, $\tilde{B}_2$ is also independent of B_1. From the law of iterated logarithm of Brownian motion, we also have

$$L(t) = \frac{t}{\mu_2} + O\left(\sqrt{t \log\log t}\right) \text{ a.s.}. \tag{B.4}$$

Setting $n = L(t)$ in (B.1), we have

$$\left| \sum_{i=1}^{L(t)} Z_i - \mu_1 L(t) - \beta(T_{L(t)} - \mu_2 L(t)) - \tilde{\sigma} B_1(L(t)) \right| = o\left(t^{1/q}\right) \text{ a.s.} \tag{B.5}$$

We also notice that under the assumption of Theorem 5 (the moment conditions), using Borel-Cantelli Lemma, we can show that

$$\left| t\hat{\alpha}(t) - \alpha t - \left(\sum_{i=1}^{L(t)} Z_i - \alpha T_{L(t)} \right) \right| \le \int_{L(t)}^{L(t+1)} |h(X(s)) - \alpha| ds = O(t^{1/q} \log t) \text{ a.s.}, \tag{B.6}$$

$$|t - T_{L(t)}| = O(t^{1/p} \log t) \text{ a.s.}, \tag{B.7}$$

In addition, from (B.4) and Theorem 1.2.1 of Csörgö and Révész (1981), we have

$$|B_1(L(t)) - B_1(t/\mu_2)| = O\left(t^{1/4} \log t\right) \text{ a.s.}. \tag{B.8}$$

Combining (B.5),(B.6),(B.7), and (B.8) we have

$$|t\hat{\alpha}(t) - \alpha t - (\mu_1 - \beta\mu_2)(L(t) - t/\mu_2) - \tilde{\sigma} B_1(t/\mu_2)| = O\left(t^{\tilde{\gamma}} \log t\right), \tag{B.9}$$

where $\tilde{\gamma} = \max\{1/p, 1/q, 1/4\}$. Combining (B.9) and (B.3), we have

$$\left| t\hat{\alpha}(t) - \alpha t - (\mu_1 - \beta\mu_2) \frac{\sigma_2}{\mu_2^{3/2}} \tilde{B}_2(t) - \tilde{\sigma} B_1(t/\mu_2) \right| = O\left(t^{\tilde{\gamma}} \log t\right) \tag{B.10}$$

Set $B(t) = \frac{1}{\sigma}\left((\mu_1 - \beta\mu_2)\frac{\sigma_2}{\mu_2^{3/2}}\tilde{B}_2(t) - \tilde{\sigma}B_1(t/\mu_2)\right)$, we have $B(t)$ is a standard Brownian motion. In particular, we notice that

$$(\mu_1 - \beta\mu_2)^2\frac{\sigma_2^2}{\mu_2^3} + \frac{\tilde{\sigma}^2}{\mu_2} = \frac{\sigma_1^2 - 2\alpha Cov(Z_1, \tau_1) + \alpha^2\sigma_2^2}{\mu_2} = \sigma^2.$$

Then (B.10) can be rewritten as

$$|t\hat{\alpha}(t) - \alpha t - \sigma B(t)| = O\left(t^{\tilde{\gamma}}\log t\right).$$

□

References

Adam N (1983) Achieving a confidence interval for parameters estimated by simulation. *Management Science* 29(7).

Alexopoulos C, Argon N, Goldsman D, Tokol G, Wilson J (2007) Overlapping variance estimators for simulation. *Operations Research* 55:1090–1103.

Asmussen S, Glynn P (2007) *Stochastic Simulation: Algorithms and Analysis* (New York, USA: Springer).

Bahadur R (1966) A note on quantiles in large samples. *Annals of Mathematical Statistics* 37:577–580.

Bayraksan G, Pierre-Louis P (2012) Fixed-width sequential stopping rules for a class of stochastic programs. *SIAM Journal on Optimization* 22(4):1518–1548.

Berkes I, Liu W, Wu W (2014) Komlos-Major-Tusnady approximation under dependence. *The Annals of Probability* 42(2):794–817.

Billingsley P (1999) *Convergence of Probability Measures* (New York, USA: Wiley), 2 edition.

Chow Y, Robbins H (1965) On the asymptotic theory of fixed-width sequential confidence intervals for the mean. *Annals of Mathematical Statistics* 36:457–462.

Csaki E, Csorgo M (1995) On additive functionals of Markov chains. *Journal of Theoretical Probability* 8(4):905–919.

Csörgö M, Horvath L, Steinebach J (1987) Invariance principles for renewal processes. *The Annals of Probability* 15(4):1441–1460.

Csörgö M, Révész P (1978) Strong approximation of the quantile process. *The Annals of Statistics* 6(4):882–894.

Csörgö M, Révész P (1981) *Strong Approximation in Probability and Statistics* (New York, USA: Academic Press).

Damerdji H (1994) Strong consistency of the variance estimator in steady-state simulation output analysis. *Mathematics of Operations Research* 19:494–512.

Dedecker J, Merlevede F, Rio E (2014) Strong approximation of the empirical distribution function for absolutely regular sequences in $\mathbb{R}^d$. *Electron. J. Probab.* 19(9):1–56.

Efron B (1979) Bootstrap methods: another look at the jackknife. *Annals of Statistics* 7:1–26.

Einmahl U (1989) Extensions of results of Komlós, Major, and Tusnády to multivariate case. *Journal of Multivariate Analysis* 28:20–68.

Glynn P (2011) Wide-sense regeneration for Harris recurrent Markov processes: an open problem. *Queueing Systems* 68:305–311.

Glynn P, Iglehart D (1990) Simulation output analysis using standardized time series. *Mathematics of Operations Research* 15(1):1–16.

Glynn P, Whitt W (1992) The asymptotic validity of sequential stopping rules for stochastic simulation. *Annals of Applied Probability* 2(1):180–198.

Jones G, Haran M, Caffo B, Neath R (2006) Fixed-width output analysis for Markov Chain Monte Carlo. *Journal of the American Statistical Association* 101(476):1537–1547.

Komlós J, Major P, Tusnády G (1975) An approximation of partial sums of independent rv's and the sample DF I. *Z. Wahrscheinlichkeitstheorie verw. Gebiete* 32:111–131.

Komlós J, Major P, Tusnády G (1976) An approximation of partial sums of independent rv's and the sample DF II. *Z. Wahrscheinlichkeitstheorie verw. Gebiete* 34:33–58.

Lavenberg S, Sauer C (1977) Sequential stopping rules for the regenerative method of simulation. *IBM J. Res. Develop.* 21:545–558.

Law A, Carson J (1979) A sequential procedure for determining the length of a steady-state simulation. *Operations Research* 27(5):1011–1025.

Liu W, Lin Z (2009) Strong approximation for a class of stationary processes. *Stochastic Processes and their Applications* 119:249–280.

Merlevede F, Rio E (2012) Strong approximation of partial sums under dependence condtions with application to dynamical systems. *Stochastic Processes and their Applications* 122:386–417.

Merlevede F, Rio E (2015) Strong approximation for additive functionals of geometrically ergodic Markov chains. *Electron. J. Probab.* (14):1–27.

Meterelliyoz M, Alexopoulos C, Goldsman D (2012) Folded overlapping variance estimators for simulation. *European Journal of Operational Research* 220(1):135–146.

Pitman J, Yor M (1981) Bessel processes and infinitely divisible laws. *Stochastic Integrals*, 285–370, Lecture Notes in Mathematics (Berlin Heidelberge, Germany: Springer).

Revuz D, Yor M (1999) *Continuous Martingales and Brownian Motion* (Berlin Heidelberge, Germany: Springer), 3 edition.

Schruben L (1983) Confidence interval estimation using standardized time series. *Operations Research* 31(6).

Shao J, Tu D (1995) *The Jackknife and Bootstrap* (New York, USA: Springer).

Swanepoel J, van Wyk J, Venter J (1983) Fixed width confidence intervals based on bootstrap procedures. *Sequential Anal.* 2:289–310.

Whitt W (1980) Some useful functionals for functional central limit theorems. *Mathematics of Operations Research* 5(1).

Wu C (1986) Jackknife, bootstrap and other resampling methods in regression analysis (with discussions). *Annals of Statistics* 14:1261–1350.

Wu W (2009) Recursive estimation of time-average variance constants. *The Annals of Applied Probability* 19(4):1529–1552.

Zaitsev A (1998) Multidimensional version of the results of Komlós, Major and Tusnády for vectors with finite exponential moments. *ESAIM Probab. Statist.* 2:41–108.