A New Approach to Sequential Stopping for Stochastic Simulation

Jing Dong
Northwestern University, Evanston, IL, 60208, jing.dong@northwestern.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 Student-t distribution. 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 \( \alpha \) via simulation. Simulation-based estimators fall, roughly speaking, into two categories: fixed sample size procedure and sequential procedure. In a fixed sample size procedure, one decides a priori on the number of samples that will be generated to form an estimator for \( \alpha \), 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.
The most widely studied class of sequential algorithms assumes that the estimator \( \alpha(t) \), based on a sample size \( t \), satisfies the central limit theorem (CLT), so that there exists \( \sigma > 0 \) for which

\[
\alpha(t) \overset{D}{\approx} \alpha + \frac{\sigma}{\sqrt{t}} N(0, 1)
\]

when \( t \) is large, where \( N(0, 1) \) is a normal random variable (rv) having mean 0 and variance 1, and \( \overset{D}{\approx} \) means “has approximately the same distribution as”; see §2 for a more rigorous formulation.

Relation (1) implies that

\[
\left[ \alpha(t) - \frac{z_{\delta/2} \sigma}{\sqrt{t}}, \alpha(t) + \frac{z_{\delta/2} \sigma}{\sqrt{t}} \right]
\]

is an approximate 100(1 - \( \delta \))% confidence interval for \( \alpha \) when \( t \) is large, provided that one choses \( z_{\delta/2} \) so that \( P(-z_{\delta/2} \leq N(0, 1) \leq z_{\delta/2}) = 1 - \delta \). To obtain a confidence interval for prescribed half-width \( \epsilon \), one then samples until \( z_{\delta/2} \sigma/\sqrt{t} \leq \epsilon \). Of course, in practice, it will never be the case that \( \sigma \) is known, so it must be estimated from the observed data by some estimator \( S(t) \), say. This suggests that one continuous sampling until \( z_{\delta/2} S(t)/\sqrt{t} \leq \epsilon \). This type of sequential procedures was first studied by Chow and Robbins (1965), when \( \alpha(t) \) is a 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

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

where \( T(\epsilon) \) is the (random) sample size at which \( z_{\delta/2} S(t)/\sqrt{t} \) is first less than \( \epsilon \). Such a procedure is said to achieve absolute precision \( \epsilon \), with an asymptotic confidence level of \( 1 - \delta \) as \( \epsilon \downarrow 0 \). Similarly, a sequential procedure achieves relative precision 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) extend 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 quantile) 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 \( \sigma \) in the setting that

\[
S(t) \to \sigma
\]
almost surely (a.s.) as \( t \to \infty \), then Chow-Robbins type absolute and relative precision procedures exhibit the correct asymptotic coverage 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 (2) is relaxed to, for example, the requirement of consistency for \( S(t) \) (in which case one requires only that \( S(t) \) converges to \( \sigma \) in probability as \( t \to \infty \)). Related literature on Chow-Robbins type processes 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 \( \sigma \) 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 \( \sigma^2 \) is known as the time-average variance constant (TAVC). The TAVC is difficulty to consistently estimate, because \( \sigma^2 \) 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), Meterelliyou et al. (2012). Even less is known about strong consistency of such estimators; see however Damerdji (1994). Computation of quantiles and conditional value-at-risk (CVaR) are examples of other settings in which estimating \( \sigma \) is challenging (in part because the formula for \( \sigma \) includes the density and density estimation is known to be both practically and theoretically difficult).

Our goal in this paper is to introduce a entirely new approach to sequentially estimating \( \alpha \) to a given precision that apply in much greater generality than do the above Chow-Robbins type procedures. The method apply to steady-state simulations, quantile computations, and CVaR estimations, 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 \( \sigma \). In particular, the sectioning (replication method), run \( m \)
independent replications of stochastic process to time \( t \). One then computes the mean of each section (replication) and constructs the associated sample variance estimator \( S^2_m(t) \) from the \( m \) section means. In this case,
\[
\frac{\sqrt{m}(\alpha(t) - \alpha)}{S_m(t)} \overset{p}{\to} t_{m-1}
\]
in great generality when \( t \) is large, where \( t_{m-1} \) is a Student-t rv with \( m - 1 \) degrees of freedom. Note that \( m \) is fixed in this analysis. The approximation (3) holds even when the simulation output is auto-correlated, so that it applies to steady-state simulations. As a consequence, the fixed sample size confidence interval is
\[
\left[ \alpha(t) - t_{\delta/2,m-1} S_m(t)/\sqrt{m}, \alpha(t) + t_{\delta/2,m-1} S_m(t)/\sqrt{m} \right]
\]
where \( t_{\delta/2,m-1} \) is chosen so that \( P(-t_{\delta/2,m-1} < t_{m-1} < t_{\delta/2,m-1}) = 1 - \delta \).

If one proceeds as for Chow-Robbins, one then is led to sequential procedures for which one stops sampling when \( t_{\delta/2,m-1} S_m(t)/\sqrt{m} \leq \epsilon \). As we shall see later, this naive implementation fails, because
\[
\frac{\sqrt{m}(\alpha(T(\epsilon)) - \alpha)}{S_m(T(\epsilon))}
\]
does not converge in distribution to \( t_{m-1} \) 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 \( \epsilon \downarrow 0 \).) Empirical evidence of the failures of such sequential procedures was observed a number of years ago; see Adam (1983). However, it turns out (4) does indeed converge to a limit as \( \epsilon \downarrow 0 \), albeit not a Student-t rv. If one calibrate the confidence interval in terms of the correct limit distribution (rather than the \( t_{\delta/2,m-1} \) value associated with \( t_{m-1} \)), one gets an asymptotically valid sequential procedure.

The sectioning framework is naturally parallelizable, as each section can be run on a different processor. We run the sequential checking in a synchronized fashion, and the unsynchronized implementation would be an interesting future research. For a fairly large class of problems, the framework also has minimum storage requirement, as we only need to store and update the sample
mean for each section. These are advantages over other cancellation methods such as standardized
time series (Glynn and Iglehart 1990) (e.g. batch means method with \( m \) bathes, where we divide
a single long simulation run into \( m \) sections).

The following notations are used throughout the text. \( \mathbb{R} \) denotes the Euclidean space and \( \mathbb{R}^+ \)
denotes the nonnegative subspace. \( C[0,t] \) denotes the space of continuous real-valued functions
on \([0,t]\) endowed with the uniform topology. \( C[0,\infty) \) denotes the space of continuous real-valued
functions on \([0,\infty)\) endowed with the uniform norm on compact time intervals. Let \( D(0,\infty) \) denote
the space of right continuous real-valued functions with left limits on the interval \((0,\infty)\), endowed
with the Skorohod \( J_1 \) topology. Let \( I \) denote the identity map from \( \mathbb{R}^+ \) to \( \mathbb{R}^+ \), i.e. \( I(t) = t \) for \( t \geq 0 \).

2. The sectioning method

Let \( X = \{X(t) : t \geq 0\} \) be a real-valued stochastic process, which represents the output of a sim-
ulation run. We can handle discrete-time processes by setting \( X(t) = X(\lfloor t \rfloor) \). Let \( F \) denote the
stationary distribution of \( X \), assuming its existence. When \( X(k) \)'s i.i.d. samples, \( F \) is the distri-
bution of \( X(k) \), \( k = 1, 2, \ldots \).

We are interested in estimating the quantity \( \alpha = \psi(F) \), where \( \psi \) is a real-valued functional on the
space of probability distributions. For example the expectation can be written as \( \psi(F) = \int xF(dx) \).
Other examples of \( \psi(F) \) include smooth functions of the expectation, i.e. \( \psi(F) = h(\int xF(dx)) \),
quantiles (also known as value at risk), i.e. the \( p \)-quantile of \( F \) is defined as \( \psi(F) = \inf\{x : F(x) \geq p\} \),
etc.

Let \( \hat{F}(t, \cdot) \) denote the empirical distribution based on \( \{X(s) : 0 \leq s \leq t\} \), i.e.

\[
\hat{F}(t, x) := \frac{1}{t} \int_0^t 1\{X(s) \leq x\}ds
\]

We impose the following assumption on \( X \) and \( \psi \).

**Assumption 1.** There exists a finite constant \( \sigma > 0 \), such that

\[
\frac{t}{\epsilon} \left( \psi \left( \frac{\hat{F}(t/\epsilon^2)}{\epsilon^2} \right) - \alpha \right) \Rightarrow \sigma B(t)
\]

in \( D(0,\infty) \) as \( \epsilon \to 0 \), where \( B \) is a standard Brownian motion.
Remark 1. Assumption 1 implies that the estimator satisfies a CLT, i.e. \( \sqrt{n} \left( \psi \left( \hat{F}(n) \right) - \alpha \right) \Rightarrow N(0,1) \) as \( n \to \infty \) as we can fixed \( t = 1 \) and set \( n = \epsilon^{-2} \). In particular, Assumption 1, which is known as the Functional Central Limit Theorem (FCLT), is a slightly stronger assumption than the CLT. We will establish the FCLT for some interesting applications in §4.

Under Assumption 1, we have

\[
\psi \left( \hat{F}(t/\epsilon^2) \right) \Rightarrow \alpha \text{ as } \epsilon \to 0,
\]

i.e. \( \psi \left( \hat{F}(t) \right) \) is a consistent estimator of \( \alpha \). The challenge lies in assessing the quality of the estimator arises from the unknown constant \( \sigma \).

For the sectioning method with \( m \) sections, we simulate \( m \) independent runs of \( X \). We denote the \( i \)-th run as \( X_i = \{ X_i(t) : t \geq 0 \} \), for \( i = 1,2,\ldots,m \). Let \( \hat{F}_i \) denote the empirical distribution based on the \( i \)-th run for \( i = 1,2,\ldots,m \). At time \( t \), we output \( \psi(\hat{F}_i(t)) \) from the \( i \)-th section (replication).

We also write \( \hat{F}_m \) as the empirical distribution based on the aggregated data from \( m \) sections, i.e.

\[
\hat{F}_m(t,x) := \frac{1}{mt} \sum_{i=1}^{m} \int_{0}^{t} 1\{X_i(s) \leq x\} ds = \frac{1}{m} \sum_{i=1}^{m} \hat{F}_i(t,x).
\]

We make the following assumption about the relationship between \( \psi \left( \hat{F}_m(t) \right) \) and \( \psi \left( \hat{F}_i(t) \right) \)’s for \( i = 1,\ldots,m \).

Assumption 2.

\[
\frac{t}{\epsilon} \left( \psi \left( \hat{F}_m(t/\epsilon^2) \right) - \frac{1}{m} \sum_{i=1}^{m} \psi \left( \hat{F}_i(t/\epsilon^2) \right) \right) \Rightarrow 0
\]

in \( D(0,\infty) \) as \( \epsilon \to 0 \).

Remark 2. If we are interested in estimating the expectation, i.e. \( \psi(F) = \int x dF(x) \), then

\[
\psi \left( \hat{F}_m(t) \right) = \frac{1}{mt} \sum_{i=1}^{m} \int_{0}^{t} X_i(s) ds = \frac{1}{m} \sum_{i=1}^{m} \psi \left( \hat{F}_i(t/\epsilon^2) \right).
\]

In a lot of other applications, Assumption 1 & 2 are satisfied simultaneously (see examples in §4.1 & 4.2).
Under Assumption 1 & 2 we can use either $\frac{1}{m} \sum_{i=1}^{m} \psi(\hat{F}_i(t))$ (the average of output from the $m$ sections) or $\psi(\hat{F}_0^m(t))$ as an estimator of $\alpha$. We would prefer the later because it in general has less bias with a finite sample size.

To estimate the error size, we write

$$
\Gamma_{\psi}(t, \hat{F}) := \sqrt{\frac{1}{m(m-1)} \sum_{i=1}^{m} \left( \psi(\hat{F}_{i}(t)) - \psi(\hat{F}_0^m(t)) \right)^2}
$$

Before we present the sequential stopping procedure, we shall first introduce some preliminary results about the sectioning method.

### 2.1. Basic results about the sectioning method

The sectioning method uses cancellation idea, in which the key is to cancel out the unknown variance parameter, which appears as a common factor in the numerator and the denominator of the estimator. In particular, we write the sectioning estimator as the ratio of $\psi(\hat{F}_0^m(t)) - \alpha$ and $\Gamma_{\psi}(t, \hat{F})$. The following proposition characterize the asymptotic distribution of the scaled ratio estimator.

**Proposition 1.** Under Assumption 1 & 2, we have

$$
\frac{\psi(\hat{F}_0^m(t/e^2)) - \alpha}{\Gamma_{\psi}(t/e^2, \hat{F})} \Rightarrow \frac{\bar{B}(t)}{\sqrt{\frac{1}{m(m-1)} \sum_{i=1}^{m} (B_i(t) - \bar{B}(t))^2}}
$$

as $e \to 0$, where $B_i$'s, $i = 1, 2, \ldots, m$, are independent Brownian motion and $\bar{B}(t) = \frac{1}{m} \sum_{i=1}^{m} B_i(t)$.

For fixed value of $t$,

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

Let $\nu$ denote the $(1 - \delta/2)$-quantile of $t_{m-1}$. If we generate $t$ samples for each section (each run is of length $t$), then we can construct the asymptotic 100(1 - $\delta$)% confidence interval as

$$
[\psi(\hat{F}_0^m(t)) - \nu \Gamma_{\psi}(t, \hat{F}), \psi(\hat{F}_0^m(t)) + \nu \Gamma_{\psi}(t, \hat{F})],
$$

i.e. $\lim_{t \to \infty} P\left( \psi(\hat{F}_0^m(t)) - \nu \Gamma_{\psi}(t, \hat{F}) < \alpha < \psi(\hat{F}_0^m(t)) + \nu \Gamma_{\psi}(t, \hat{F}) \right) = 1 - \delta$. Note that here the total number of samples generated is $mt$. 

3. Sequential stopping for the sectioning method

For the sequential stopping procedure, to achieve an absolute precision level, we consider a sequence of stopping times, indexed by $\epsilon$,

$$\kappa_0(\epsilon) := \inf \left\{ t > 0 : \Gamma_\psi(t, \hat{F}) < \epsilon \right\}.$$  

If we find the correct scaling parameter $u$, then $\kappa_0(\epsilon/u) := \inf \left\{ t > 0 : u \Gamma_\psi(t, \hat{F}) < \epsilon \right\}$ would be the time at which the simulation terminates, i.e. it is the time when the width of the corresponding confidence interval is less than $\epsilon$, where $\epsilon$ is some user-specified precision level. However, we notice that $\kappa_0(\epsilon)$ can terminate much too early if $\Gamma_\psi(t, \hat{F})$ is badly behaved for small $t$. For example, when applying the sectioning method to estimate the steady-state expectation of a discrete time Markov chain $X$, if $X_i(0) = 0$ for $i = 1, 2, \ldots, m$, then $\Gamma_\psi(t, \hat{F}) = 0$ for $t < 1$. To avoid this early termination issue, we next introduce a modification of the stopping times as in (Glynn and Whitt 1992). We first notice from the proof of Proposition 1 that $\sqrt{t} \Gamma_\psi(t, \hat{F}) \Rightarrow \sqrt{\chi_{m-1}^2/(m(m-1))}$ as $t \to \infty$. This suggests that $\Gamma_\psi(t, \hat{F})$ decays roughly at rate $t^{-1/2}$, which is consistent with the decay rate of sample standard deviation. Thus, we let $a(t)$ be a strictly positive function that decreases monotonically to 0 as $t \to \infty$, and satisfies $a(t) = O(t^{-\gamma})$ for $\gamma > 1/2$, i.e. it decays at a faster rate than $t^{-1/2}$. We then define

$$\kappa_\psi(\epsilon) := \inf \left\{ t > 0 : \Gamma_\psi(t, \hat{F}) + a(t) < \epsilon \right\}.$$  

We notice that $\kappa_0(\epsilon) := \inf \{ t > 0 : a(t) < \epsilon \} \to \infty$ as $\epsilon \to 0$ and $\kappa_\psi(\epsilon) \geq \kappa_0(\epsilon)$.

To find the appropriate scaling parameter, our task is to characterize the limiting distribution of

$$\lim_{\epsilon \to 0} \frac{\psi \left( \hat{F}_0^m(\kappa_\psi(\epsilon)) \right) - \alpha}{\Gamma_\psi(\kappa_\psi(\epsilon), \hat{F})}.$$  

This task is divided into two steps, where we shall characterize the limiting distribution of $\kappa_\psi(\epsilon)$ first.

We define

$$K(\sigma) := \inf \left\{ t > 0 : \sigma \sqrt{\frac{1}{m(m-1)} \sum_{i=1}^m (B_i(t) - \bar{B}(t))^2} < 1 \right\}$$

where $B_i$'s, $i = 1, 2, \ldots, m$, are independent Brownian motions and $\bar{B} = \frac{1}{m} \sum_{i=1}^m B_i$. 
Theorem 1. Under Assumption 1 and 2,

\[ 2^2 \kappa_\psi(\epsilon) \Rightarrow K(\sigma) \text{ as } \epsilon \to 0. \]

Theorem 1 suggests that \( \kappa_\psi(\epsilon) \) is \( O_p(\epsilon^{-2}) \). This is consistent with the convergence rate of Monte Carlo methods in general. We also notice that

\[
K(\sigma) = \frac{1}{\sigma^2} \inf \left\{ t > 0 : \sigma \sqrt{\frac{1}{m(m-1)t^2 \sum_{i=1}^{m} (B_i(t) - \bar{B}(t))^2}} < 1 \right\}
\]

\[
= \inf \left\{ t > 0 : \frac{1}{m(m-1)(\sigma^2 t)^2 \sum_{i=1}^{m} (B_i(\sigma^2 t) - \bar{B}(\sigma^2 t))^2} < 1 \right\}
\]

\[
\overset{d}{=} \inf \left\{ t > 0 : \frac{1}{m(m-1)t^2 \sum_{i=1}^{m} (B_i(t) - \bar{B}(t))^2} < 1 \right\}
\]

\[= K(1), \]

indicating that \( \kappa_\psi(\epsilon) \) is stochastically larger for larger values of \( \sigma \). We next take a closer look at \( K(1) \).

Lemma 1. (1) When \( m = 2 \), \( K(1) = 0 \).

(2) When \( m = 3 \), \( K(1) = 0 \).

(3) When \( m \geq 4 \), \( K(1) \) follows a Gamma distribution with shape parameter \( \gamma = (m - 3)/2 \) and rate \( \lambda = m(m - 1)/2 \).

Having \( K(1) = 0 \) is problematic as it will leads to an infinite scaling parameter (This will become apparent from Theorem 2). To overcome this issue, we shall restrict to \( m \geq 4 \). Notice that when sample size \( t \) is fixed, we only require \( m \geq 2 \). When applying sequential stopping procedure, we need to put more restrictions on the number of sections, i.e. \( m \geq 4 \).

Based on Theorem 1 and Lemma 1, the following theorem characterizes the limiting distribution of the sectioning estimator.

Theorem 2. Under Assumption 1 and 2, for \( m \geq 4 \),

\[
\frac{\psi \left( F_0^m (\kappa_\psi(\epsilon)) \right) - \alpha}{\Gamma_\psi(\kappa_\psi(\epsilon), \bar{F})} \Rightarrow \frac{\sigma \bar{B}(K(\sigma))}{K(\sigma)} \overset{d}{=} \frac{Z}{\sqrt{mK(1)}}
\]

as \( \epsilon \to 0 \), where \( Z \sim N(0,1) \) is independent of \( K(1) \).
Theorem 2 provides us with the right scaling parameter when constructing asymptotically valid confidence intervals under the sequential stopping procedure. In particular, if one wants to achieve a confidence level of \(100(1 - \delta)\%\), one can select \(u\) such that \(P(|Z/\sqrt{mK(1)}| \leq u) = 1 - \delta\). Notice that the independence of \(Z\) and \(K(1)\) ensures that the distribution of \(Z/\sqrt{mK(1)}\) is symmetric around zero. Then the asymptotic \(100(1 - \delta)\%\) confident interval with an absolute precision level \(\epsilon\) is constructed as

\[
\left[ \psi \left( \hat{F}_0^m(\kappa_\psi(\epsilon/u)) \right) - \epsilon, \psi \left( \hat{F}_0^m(\kappa_\psi(\epsilon/u)) \right) + \epsilon \right].
\]

Specifically, we have

\[
P \left( \psi \left( \hat{F}_0^m(\kappa_\psi(\epsilon/u)) \right) - \epsilon \leq \alpha \leq \psi \left( \hat{F}_0^m(\kappa_\psi(\epsilon/u)) \right) + \epsilon \right)
= P \left( \left| \frac{\psi \left( \hat{F}_0^m(\kappa_\psi(\epsilon/u)) \right) - \alpha}{\epsilon/u} \right| < u \right)
= P \left( \left| \frac{\psi \left( \hat{F}_0^m(\kappa_\psi(\epsilon/u)) \right) - \alpha}{\Gamma_\psi(\kappa_\psi(\epsilon/u), \hat{F})} \right| < u \right)
\rightarrow P(|Z/\sqrt{mK(1)}| < u) = 1 - \delta\) as \(\epsilon \to 0\).

The actual simulation algorithm goes as follows.

**Algorithm 1**

**Input:** the number of sections \(m\), the error bound \(\epsilon\), the scaling parameter \(u\) (see Table 1 for some of the commonly used scaling parameters), and the step size \(\Delta\). We choose \(a(t) = t^{-1}\).

**Output:** a \(100(1 - \delta)\%\) confidence interval with half width \(\epsilon\).

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

(ii) Sample \(X_i(t)\) between \(T - \Delta\) and \(T\). Calculate \(\psi \left( \hat{F}_i(T) \right)\), for \(i = 1, 2, \ldots, m\), and \(\psi \left( \hat{F}_0^m(T) \right)\)

\[
\Gamma = \sqrt{\frac{1}{m(m-1)} \sum_{i=1}^{m} \left( \psi \left( \hat{F}_i(T) \right) - \psi \left( \hat{F}_0^m(T) \right) \right)^2}
\]
(iii) If \( u \times (\Gamma + a(T)) \geq \epsilon \), set \( T = T + \Delta \), go back to step (ii); otherwise, output 
\[
[\psi(\hat{F}_0^m(T)) - \epsilon, \psi(\hat{F}_0^m(T)) + \epsilon]
\] as the 100(1 - \delta)% confidence interval.

Remark 3. In Algorithm 1, we pick \( a(t) = t^{-1} \). As for \( t < \epsilon^{-1} \), \( a(t) > \epsilon \), we start by sampling \( X(t) \) for \( 0 \leq t \leq \epsilon^{-1} \).

3.1. The distribution of \( Z/\sqrt{mK(1)} \)

For fixed \( t \), we have for \( m \geq 2 \),
\[
\frac{\psi(\hat{F}_0^m(t/\epsilon^2)) - \alpha}{\Gamma_\psi(t/\epsilon^2, \tilde{F})} \Rightarrow \frac{Z}{\sqrt{\chi^2_{m-1}/(m-1)}} \text{ as } \epsilon \to 0,
\]
where \( Z \) is independent of \( \chi^2_{m-1} \); while for \( \kappa_\psi(\epsilon) \), we have for \( m \geq 4 \),
\[
\frac{\psi(\hat{F}_0^m(\kappa_\psi(\epsilon))) - \alpha}{\Gamma_\psi(\kappa_\psi(\epsilon), \tilde{F})} \Rightarrow \frac{Z}{\sqrt{mK(1)}} \text{ as } \epsilon \to 0.
\]

The distribution of \( mK(1) \) is different from the distribution of \( \chi^2_{m-1}/(m-1) \) for \( m \geq 4 \). In particular, from Lemma 1, \( mK(1) \sim \text{Gamma}((m-3)/2, (m-1)/2) \). From the connection of Gamma distribution and Chi-square distribution, we have \( \chi^2_{m-1}/(m-1) \sim \text{Gamma}((m-1)/2, (m-1)/2) \). Then \( mK(1) \) has a lighter tail than \( \chi^2_{m-1}/(m-1) \). Thus, \( Z/\sqrt{mK(1)} \) has a heavier tail than \( t_{m-1} \), which indicates that for the same level of confidence, the sequential stopping procedure would require a larger multiplier to the variance estimator, \( \Gamma_\psi \), than the fixed-budget procedures. We interpret this as the price we pay to get control over the size of the error (width of the confidence interval).

As \( m \) increases, \( Z/\sqrt{mK(1)} \) has a lighter tail. In particular, we notice that
\[
\sqrt{\frac{m}{m-1} \frac{1}{\beta^2} \sum_{k=1}^{m} \left( B_k \left( \frac{t}{m} \right) - \Bar{B} \left( \frac{t}{m} \right) \right)^2} \to \frac{1}{\beta} \text{ as } m \to \infty.
\]
If we define \( \infty K^\infty(1) := \inf \{ t : \sqrt{1/\beta} < 1 \} = 1 \), then \( Z/\sqrt{\infty K^\infty(1)} \sim N(0,1) \).

Table 1 lists some sample percentile of \( Z/\sqrt{mK(1)} \) (with the corresponding 95% confidence interval), which can be used as the corresponding multiplier when applying the sectioning method with \( m \) sections.
Table 1 Sample percentile of $Z/\sqrt{mK(1)}$ (10^6 i.i.d. samples)

<table>
<thead>
<tr>
<th>m</th>
<th>0.9</th>
<th>0.925</th>
<th>0.95</th>
<th>0.975</th>
<th>0.995</th>
</tr>
</thead>
<tbody>
<tr>
<td>10</td>
<td>1.603 ± 0.001</td>
<td>1.833 ± 0.001</td>
<td>2.148 ± 0.002</td>
<td>2.680 ± 0.002</td>
<td>3.963 ± 0.005</td>
</tr>
<tr>
<td>15</td>
<td>1.465 ± 0.001</td>
<td>1.661 ± 0.001</td>
<td>1.925 ± 0.002</td>
<td>2.352 ± 0.002</td>
<td>3.298 ± 0.003</td>
</tr>
<tr>
<td>20</td>
<td>1.409 ± 0.001</td>
<td>1.593 ± 0.001</td>
<td>1.840 ± 0.001</td>
<td>2.230 ± 0.002</td>
<td>3.064 ± 0.003</td>
</tr>
<tr>
<td>25</td>
<td>1.380 ± 0.001</td>
<td>1.558 ± 0.001</td>
<td>1.792 ± 0.001</td>
<td>2.164 ± 0.002</td>
<td>2.945 ± 0.003</td>
</tr>
<tr>
<td>30</td>
<td>1.361 ± 0.001</td>
<td>1.536 ± 0.001</td>
<td>1.765 ± 0.001</td>
<td>2.126 ± 0.002</td>
<td>2.872 ± 0.002</td>
</tr>
</tbody>
</table>

3.2. Relative error

We have analyzed the sequential stopping rules to achieve an absolute precision $\epsilon$. There is another commonly used precision criteria called relative error (Asmussen and Glynn 2007), where we want to achieve a precision level relative to the value of the quantity of interests. In particular, we define a new sequence of stopping times indexed by $\epsilon$

$$\tilde{\kappa}_\psi(\epsilon) := \inf \left\{ t > 0 : \Gamma_\psi(t, \hat{F}) + a(t) < \epsilon \left| \psi \left( \hat{F}_0^m(t) \right) \right| \right\}.$$  

Similar to the absolute precision case, if we find the correct scaling parameter $u$, then we can terminate the simulations at $\tilde{\kappa}_\psi(\epsilon/u) = \inf \left\{ t > 0 : u \left( \Gamma_\psi(t, \hat{F}) + a(t) \right) < \epsilon \left| \psi \left( \hat{F}_0^m(t) \right) \right| \right\}$.

The following theorem characterized the scaled limiting distribution of $\tilde{\kappa}_\psi(\epsilon)$.

**Theorem 3.** *Under Assumption 1 and 2,*

$$\epsilon^2 \tilde{\kappa}_\psi(\epsilon) \Rightarrow K(\sigma/|\alpha|) \text{ as } \epsilon \rightarrow 0.$$  

Similar to $\kappa_\psi(\epsilon)$ in the absolute precision formulation, $\tilde{\kappa}_\psi(\epsilon)$ is $O_p(\epsilon^{-2})$ and is stochastically larger for larger values of $\sigma$. Unlike the absolute precision case, $\tilde{\kappa}_\psi(\epsilon)$ is also influenced by the value of $|\alpha|$. In particular, it is stochastically larger for smaller values of $|\alpha|$.

The following theorem establishes the limiting distribution of $(\psi(\tilde{\kappa}_\psi(\epsilon), \hat{F}) - \alpha)/\Gamma_\psi(\tilde{\kappa}_\psi(\epsilon), \hat{F})$ as $\epsilon \rightarrow 0$, which is the same as the limiting distribution of $(\psi(\kappa_\psi(\epsilon), \hat{F}) - \alpha)/\Gamma_\psi(\kappa_\psi(\epsilon), \hat{F})$ established in Theorem 2. This suggests we can use the same scaling parameter for the two formulations (absolute v.s. relative).
Theorem 4. Under Assumption 1 and 2, for \( m \geq 4 \),

\[
\psi(\hat{\kappa}_0(\epsilon), \hat{F}) - \frac{\alpha}{\Gamma(\hat{\kappa}_0(\epsilon), \hat{F})} \Rightarrow \frac{\sigma \bar{B}(K(\sigma/|\alpha|))}{|\alpha| K(\sigma/|\alpha|)} \equiv \frac{Z}{\sqrt{mK(1)}}
\]

as \( \epsilon \to 0 \), where \( Z \sim N(0,1) \) is independent of \( K(1) \).

From Theorem 4, we have, when applying the sequential stopping procedure, we can select \( u \) such that \( P(|Z/\sqrt{mK(1)}| \leq u) = 1 - \delta \). Then the 100(1 - \delta)% asymptotic confidence interval with a relative precision level \( \epsilon \) is constructed as

\[
\left[ \psi(\hat{F}_0^m(\hat{\kappa}_0(\epsilon/u))) - \epsilon \psi(\hat{F}_0^m(\hat{\kappa}_0(\epsilon/u))), \psi(\hat{F}_0^m(\hat{\kappa}_0(\epsilon/u))) + \epsilon \psi(\hat{F}_0^m(\hat{\kappa}_0(\epsilon/u))) \right].
\]

The actual simulation algorithm goes as follows.

Algorithm 2

Input: the number of sections \( m \), the error bound \( \epsilon \), the scaling parameter \( u \) (see Table 1 for some of the commonly used scaling parameters) and the step size \( \Delta \). We choose \( a(t) = t^{-1} \)

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

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

(ii) Sample \( X_i(t) \) between \( T - \Delta \) and \( T \). Calculate \( \psi(\hat{F}_i(T)) \), for \( i = 1, 2, \ldots, m \), and \( \psi(\hat{F}_0^m(T)) \)

\[
\Gamma = \sqrt{\frac{1}{m(m-1)} \sum_{i=1}^{m} \left( \psi(\hat{F}_i(T)) - \psi(\hat{F}_0^m(T)) \right)^2}
\]

(iii) If \( u \times (\Gamma + a(T)) \geq \epsilon \psi(\hat{F}_0^m(T)) \), set \( T = T + \Delta \), go back to step (ii); otherwise, output

\[
\left[ \psi(\hat{F}_0^m(T)) - \epsilon \psi(\hat{F}_0^m(T)), \psi(\hat{F}_0^m(T)) + \epsilon \psi(\hat{F}_0^m(T)) \right]
\]
as the 100(1 - \delta)% confidence interval.

4. Examples for which the sectioning framework applies

In this section, we present some examples, for which Assumption 1 & 2 are satisfied, so that we can apply the sequential stopping sectioning framework.
4.1. 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 i.i.d. random variables. We are interested in computing $h(\mu)$, where $\mu = E[X(k)]$ and $h$ is a smooth function. In this case, $\psi(F) = h(\int xF(dx))$. We next show that under proper smoothness assumptions of $h$, Assumption 1 & 2 are satisfied.

**Theorem 5.** Assume that $h$ is differentiable and $h'$ is Lipchitz continuous at $\mu$. We also assume that $\text{Var}(X(k)) = \sigma_0^2 < \infty$.

$$\frac{t}{\epsilon}(h(\bar{Y}(t/\epsilon^2)) - h(\mu)) \Rightarrow \sigma_0 h'(\mu)B(t)$$

in $D(0, \infty)$ as $\epsilon \to 0$.

The value of $h'(\mu)$ is in general unknown and estimating it would be another smooth function of expectation problem.

In the sectioning framework, we generate $m$ independent copies of $X = \{X(t) : t \geq 0\}$, which are denote as $X_i$ for $i = 1, 2, \ldots, m$. We also denote $\bar{Y}_i(t) := \frac{1}{t} \int_0^t X(s)ds$, for $i = 1, 2, \ldots, m$, and $\bar{Y}_0^m(t) := \frac{1}{m} \sum_{i=1}^m \bar{Y}_i(t)$. The following lemma follows directly from the proof of Theorem 5.

**Lemma 2.** Under the assumptions of Theorem 5,

$$\frac{t}{\epsilon} \left( h(\bar{Y}_0^m(t/\epsilon)) \right) \to \frac{1}{m} \sum_{i=1}^m h(\bar{Y}_i(t/\epsilon^2))$$

in $D(0, \infty)$ as $\epsilon \to 0$.

4.2. Example B: quantiles

Let $X$ be a real-valued random variable with CDF $F(\cdot)$. For a real-valued function $G$, we define the generalized inverse function as $G^{-1}(y) = \inf\{x : G(x) \geq y\}$. For any $0 < p < 1$, the $p$-th quantile of $X$ is then defined as $\xi_p = F^{-1}(p)$. Let $\hat{\xi}_{i,p} := \hat{F}^{-1}(t,p)$ denote the 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 $X(k)$’s, then $\hat{\xi}_{n,p} = X_{(n,\lfloor np \rfloor)}$. For $t \in (n, n+1)$, we define $\xi_{t,p} = \hat{\xi}_{n,p} + (n + 1 - t)(\xi_{n+1,p} - \hat{\xi}_{n,p})$ using linear interpolation.
Bahadur representation draws the connection between \( \hat{\xi}_{n,p} \) and \( \hat{F}(n, \xi_p) \) (Bahadur 1966). Specifically if \( F \) is differentiable in some neighborhood of \( \xi_p \) and the density function \( f(\xi_p) = dF(\xi_p)/d\xi_p \in (0, \infty) \), then

\[
\hat{\xi}_{n,p} = \xi_p - \frac{\hat{F}(n, \xi_p) - p}{f(\xi_p)} + O\left( n^{-3/4} (\log n)^{1/2} (\log \log n)^{1/4} \right) \text{ a.s.}
\]

Building on Bahadur representation, we next show that Assumption 1 & 2 are satisfied.

**Theorem 6.** If \( F \) is differentiable in some neighborhood of \( \xi_p \) and the density function \( f(\xi_p) = dF(\xi_p)/d\xi_p \in (0, \infty) \), then

\[
\frac{t}{\epsilon} \left( \hat{\xi}_{i/\epsilon^2, p} - \xi_p \right) \Rightarrow \sigma B(t)
\]

in \( D(0, \infty) \) as \( \epsilon \to 0 \), where \( \sigma = \sqrt{p(1-p)/f(\xi_p)} \).

**Remark 4.** It is possible to extend Theorem 6 to some other stationary sequence of random variables (Sen 1972, Wu 2005).

The difficulty in constructing confidence intervals for quantiles arises in the unknown value of \( f(\xi_p) \), and estimating the density function, \( f(\cdot) \), is a costly task.

When implementing the sectioning framework, we generate \( m \) independent sequence of \( X = \{ X(k) : k \geq 0 \} \). Let \( \hat{\xi}_{n,p}^i \) denote the sample quantile of the \( i \)-th section based on \( \{ X_i(1), \ldots, X_i(n) \} \). We also denote \( \hat{\xi}_{mn,p}^0 \) as the sample quantile based on \( \{ X_i(k) : i = 1, \ldots, m, k = 1, \ldots n \} \). The following lemma follows from the proof of Theorem 6.

**Lemma 3.** Under the assumption of Theorem 6,

\[
\frac{t}{\epsilon} \left( \xi_{0i/\epsilon^2, p}^0 - \frac{1}{m} \sum_{i=1}^m \hat{\xi}_{i/\epsilon^2, p}^i \right) \Rightarrow 0.
\]

in \( D(0, \infty) \) as \( \epsilon \to 0 \).

**4.3. Example C: Steady-state simulation**

For steady-state performance analysis problems, we are interested in estimating \( \alpha = \psi(F) = \int h(x)dF(s) \), where \( F \) 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,

\[
\psi(\hat{F}(t)) = \frac{1}{t} \int_0^t h(X(s))ds.
\]
A variety of stochastic processes satisfy Assumption 1. These include $\phi$-mixing processes and regenerative processes (Glynn and Iglehart 1990). As we explained in Remark 2, Assumption 2 is automatically satisfied when $\psi(F)$ denote the expectation operator.

In steady-state simulation, $\sigma^2$, which is known as the time-average variance constant, is in general difficulty to consistently estimate, because it reflects the complicated auto-correlation structure of the underlying stochastic process.

5. Simulation experiments

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

5.1. Quantile estimation

We apply Algorithm 1 to construct the 95% confidence interval (CI) for the 0.8-quantile of an Exponential random variable with unit rate. We set $\Delta = \epsilon^{-1}$. The quantile can be calculated in closed form. Specifically, $\xi_{0.8} = 1.6094$. Fix $m = 10$, we apply Algorithm 1 with different values of $\epsilon$. Table 2 summarizes the results. When $m = 10$, the scaling parameter is $u = 2.680$, which is the 0.975-quantile of $Z/\sqrt{K(1)}$ (see Table 1), whereas the 0.975-quantile of $t_9$ (a student-t distribution with 9 degrees of freedom) is 2.262. The coverage rates and $E[\kappa_{\psi}(\epsilon)]$’s are calculated based on $10^3$ independent applications of the algorithm. We also report the corresponding 95% CIs. “The coverage rate for $t_{m-1}$” shows the coverage rate if we plug in the “wrong” scaling parameter suggested by the corresponding Student-t distribution. We notice that because the quantity we are interested in is quite small, when $\epsilon = 0.5$, both scaling parameter achieves very good performance but the coverage rate is much larger than 0.95. As $\epsilon$ gets smaller, plugging in the “right” scaling parameter becomes important. Specifically, we always achieve around 95% coverage with correct scaling parameter, while if we use the scaling parameter suggested by $t_9$, the coverage rate eventually suffers undercoverage. We also notice that $\epsilon^2 E[\kappa_{\psi}(\epsilon)]$’s are of about the same value for different values of $\epsilon$, as suggested by Lemma 1.

Table 3 summarize the results for the relative accuracy formulation, where we apply Algorithm 2 to construct the CIs. The observations are similar to the absolute precision case.
Table 2  Performance of the sequential stopping procedure to achieve an absolute precision level $\epsilon$ with $m = 10$

<table>
<thead>
<tr>
<th>$\epsilon$</th>
<th>Coverage rate</th>
<th>$E[\kappa_\psi(\epsilon)]$</th>
<th>$\epsilon^2 E[\kappa_\psi(\epsilon)]$</th>
<th>Coverage rate for $t_{m-1}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.5</td>
<td>0.994 ± 0.005</td>
<td>(1.27 ± 0.03) × 10</td>
<td>3.19</td>
<td>0.995 ± 0.005</td>
</tr>
<tr>
<td>0.1</td>
<td>0.950 ± 0.014</td>
<td>(2.25 ± 0.07) × 10^2</td>
<td>2.25</td>
<td>0.941 ± 0.015</td>
</tr>
<tr>
<td>0.05</td>
<td>0.950 ± 0.014</td>
<td>(8.80 ± 0.27) × 10^2</td>
<td>2.20</td>
<td>0.924 ± 0.016</td>
</tr>
<tr>
<td>0.01</td>
<td>0.953 ± 0.013</td>
<td>(2.06 ± 0.07) × 10^4</td>
<td>2.06</td>
<td>0.904 ± 0.018</td>
</tr>
</tbody>
</table>

Table 3  Performance of the sequential stopping procedure to achieve a relative precision level $\epsilon$ with $m = 10$

<table>
<thead>
<tr>
<th>$\epsilon$</th>
<th>Coverage rate</th>
<th>$E[\kappa_\psi(\epsilon)]$</th>
<th>$\epsilon^2 E[\kappa_\psi(\epsilon)]$</th>
<th>Coverage rate for $t_{m-1}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.5</td>
<td>0.950 ± 0.014</td>
<td>4.13 ± 0.23</td>
<td>1.03</td>
<td>0.934 ± 0.015</td>
</tr>
<tr>
<td>0.1</td>
<td>0.952 ± 0.013</td>
<td>(1.01 ± 0.03) × 10^2</td>
<td>1.01</td>
<td>0.915 ± 0.017</td>
</tr>
<tr>
<td>0.05</td>
<td>0.954 ± 0.013</td>
<td>(3.84 ± 0.11) × 10^2</td>
<td>0.96</td>
<td>0.927 ± 0.016</td>
</tr>
<tr>
<td>0.01</td>
<td>0.949 ± 0.014</td>
<td>(8.99 ± 0.27) × 10^3</td>
<td>0.90</td>
<td>0.919 ± 0.017</td>
</tr>
</tbody>
</table>

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

We consider an M/M/1 queue, which has a Poisson arrival process with rate $\lambda$, and independent identically distributed exponential service times with mean $1/\mu$. Let $X(t)$ denote the number of customers in the system at time $t$. 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 $\alpha$ for different values of $m$ and $\epsilon$. The results are summarized in Table 4 & 5. We set the scaling parameter $u = 2.667$ for $m = 10$ and $u = 2.219$ for $m = 20$ (the value of these quantiles can be found in Table 1). The statistics in Table 4 are calculated based on $10^3$ independent applications of Algorithm 1. We observe that with the “right” scaling parameter, we achieve the claimed coverage rate (confidence level) when $\epsilon$ is small enough and plugging in the wrong scaling parameter (suggested by $t_{m-1}$) will lead to undercoverage. As $m$
increases from 20 to 30, the undercoveraged caused by plugging in $t_{\delta/2,m-1}$ is less severe. We also notice that for fixed $m$, $\epsilon^2E[\kappa_\psi(\epsilon)]$’s are of about the same value. As $m$ increases (from 10 to 20), the simulation cost $mE[\kappa_\psi(\epsilon)]$ decreases.

Table 4  Performance of the sequential stopping procedure to achieve an absolute precision level $\epsilon$ with $m$ sections: Queue length (number of people in the system) process of an $M/M/1$ queue ($\lambda = 0.8, \mu = 1, m = 10$)

<table>
<thead>
<tr>
<th>$\epsilon$</th>
<th>Coverage rate</th>
<th>$E[\kappa_\psi(\epsilon)]$</th>
<th>$\epsilon^2E[\kappa_\psi(\epsilon)]$</th>
<th>Coverage rate for $t_{m-1}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.5</td>
<td>0.851 ± 0.022</td>
<td>(3.27 ± 0.15) x 10^3</td>
<td>8.19 x 10^2</td>
<td>0.766 ± 0.026</td>
</tr>
<tr>
<td>0.1</td>
<td>0.947 ± 0.014</td>
<td>(9.65 ± 0.31) x 10^4</td>
<td>9.65 x 10^2</td>
<td>0.916 ± 0.017</td>
</tr>
<tr>
<td>0.05</td>
<td>0.951 ± 0.013</td>
<td>(3.82 ± 0.13) x 10^5</td>
<td>9.56 x 10^2</td>
<td>0.919 ± 0.017</td>
</tr>
<tr>
<td>0.01</td>
<td>0.948 ± 0.014</td>
<td>(9.53 ± 0.31) x 10^6</td>
<td>9.53 x 10^2</td>
<td>0.902 ± 0.018</td>
</tr>
</tbody>
</table>

Table 5  Performance of the sequential stopping procedure to achieve an absolute precision level $\epsilon$ with $m$ sections: Queue length (number of people in the system) process of an $M/M/1$ queue ($\lambda = 0.8, \mu = 1, m = 20$)

<table>
<thead>
<tr>
<th>$\epsilon$</th>
<th>Coverage rate</th>
<th>$E[\kappa_\psi(\epsilon)]$</th>
<th>$\epsilon^2E[\kappa_\psi(\epsilon)]$</th>
<th>Coverage rate for $t_{m-1}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.5</td>
<td>0.798 ± 0.025</td>
<td>(1.15 ± 0.05) x 10^3</td>
<td>2.89 x 10^2</td>
<td>0.764 ± 0.026</td>
</tr>
<tr>
<td>0.1</td>
<td>0.951 ± 0.013</td>
<td>(3.84 ± 0.08) x 10^4</td>
<td>3.84 x 10^2</td>
<td>0.931 ± 0.016</td>
</tr>
<tr>
<td>0.05</td>
<td>0.955 ± 0.013</td>
<td>(1.53 ± 0.03) x 10^5</td>
<td>3.83 x 10^2</td>
<td>0.937 ± 0.015</td>
</tr>
<tr>
<td>0.01</td>
<td>0.947 ± 0.014</td>
<td>(3.94 ± 0.08) x 10^6</td>
<td>3.94 x 10^2</td>
<td>0.935 ± 0.015</td>
</tr>
</tbody>
</table>

We also tested Algorithm 1 for systems with different values of traffic intensity $\rho$. The results are summarized in Table 6. We make two observations here. The first one is that the coverage rates are around 95% for all the traffic intensity values tested. This is in contrast to ABATCH and ASPS (Steiger et al. 2005) type of procedure, where the performance deteriorate dramatically as $\rho$ increases. This is because our algorithm does not involve any statistical tests for correlation. The second one is that as $\rho$ increases, $mE[\kappa_\psi(\epsilon)]$ increases. This is expected, as $E[\kappa_\psi(\epsilon)]$ is increasing in $\sigma$. 
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 ($\lambda = \rho, \mu = 1$)

<table>
<thead>
<tr>
<th>$\rho$</th>
<th>Coverage rate</th>
<th>$mE[\kappa_\psi(\epsilon)]$</th>
<th>Coverage rate for $t_{0.025,9}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.8</td>
<td>0.947 ± 0.014</td>
<td>$(1.05 \pm 0.05) \times 10^6$</td>
<td>0.916 ± 0.017</td>
</tr>
<tr>
<td>0.85</td>
<td>0.951 ± 0.013</td>
<td>$(3.73 \pm 0.13) \times 10^6$</td>
<td>0.842 ± 0.023</td>
</tr>
<tr>
<td>0.9</td>
<td>0.949 ± 0.014</td>
<td>$(2.00 \pm 0.09) \times 10^7$</td>
<td>0.843 ± 0.023</td>
</tr>
<tr>
<td>0.95</td>
<td>0.952 ± 0.013</td>
<td>$(3.44 \pm 0.13) \times 10^8$</td>
<td>0.818 ± 0.024</td>
</tr>
</tbody>
</table>

6. Comparison to resampling techniques and concluding remarks

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), Jackknife (delete-d Jackknife)(Wu 1986) and Subsampling (Politis et al. 2001) are among the mostly commonly used techniques. Take the bootstrap as an example. Let $\theta$ 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 i.i.d. samples from 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 §4, 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)). Moreover, let

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

and

$$N(\epsilon) := \inf\{n \geq z_{1-\delta}/2 : \hat{p}_n(\epsilon) \geq 1 - \delta\}.$$
Then it is shown in Swanepoel et al. (1983) that

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

Thus, we could use bootstrap to find the probability of converge of a fixed width confidence interval of the form $[T_n(X_1, \ldots, X_n) - \epsilon, T_n(X_1, \ldots, X_n) + \epsilon]$, sequentially for each $n$, until we find an appropriate sample size that achieves a coverage probability of $(1 - \delta)$. Similar results would hold for delete-d Jackknife where the “delete” sample size $d_n \to \infty$ as $n \to \infty$ (Wu 1986), and subsampling where the subset size $b_n$ grows to infinity with the sample size $n$, but at a much slower rate i.e. $\lim_{n \to \infty} b_n/n = 0$ (Politis et al. 2001). In general, subsampling works for a more general class of estimation problems as its corresponding asymptotic results require much less smoothness conditions on $\psi(F)$ (Politis et al. 2001), where $\psi$ is a real-valued functional on the space of probability distributions as introduced in §2.

Compared to the sectioning framework we studied in this paper, the resampling techniques are much more computationally intensive (expensive). Take the subsampling method as an example, at each step $n$, we need to conduct the sample statistics computation $\binom{n}{b_n}$ times for samples of size $b_n$ 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 involves $n!$). In contrast, for the sectioning framework we used here, we often have a recursive way of updating the statistics such as sample mean and sample variance within each section, and the cross-section calculation often involves only $m$ aggregated statistics. Here, $m$ is the number of sections and it does not grow with $n$. 
To conclude, 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. We characterize the distribution leading to the scaling parameters in closed form, and provide a table (Table 1) that can be used to find the corresponding scaling parameters.

Appendix A: Proof of Section 2 & 3

Proof of Proposition 1 Let $h(t,x_1,x_2,\ldots,x_m) = \frac{1}{m} \sum_{i=1}^{m} x_i(t)$. As $\frac{t}{\epsilon} \left( \psi \left( \hat{F}_i(t/\epsilon^2) \right) - \alpha \right) \Rightarrow \sigma B_i(t)$ in $D(0,\infty)$ and $P(\sigma B_i \in D(h)) = 0$, applying the continuous mapping theorem, we have

$$\frac{1}{m} \sum_{i=1}^{m} \frac{t}{\epsilon} \left( \psi \left( \hat{F}_i(t/(me^2)) \right) - \alpha \right) \Rightarrow \sigma \tilde{B}(t) \text{ as } \epsilon \to 0.$$

As $\frac{1}{\epsilon} \left( \psi \left( \hat{F}_0^m(t/\epsilon^2) \right) - \frac{1}{m} \sum_{i=1}^{m} \psi \left( \hat{F}_i(t/(me^2)) \right) \right) \Rightarrow 0$ as $\epsilon \to 0$, we have

$$\frac{t}{\epsilon} \left( \psi \left( \hat{F}_0^m(t/\epsilon^2) \right) - \alpha \right) \Rightarrow \sigma \tilde{B}(t) \text{ as } \epsilon \to 0.$$

Using again the continuous mapping theorem, we have

$$\frac{\psi \left( \hat{F}_0^m(t/\epsilon^2) \right) - \alpha}{\Gamma_{\psi} \left( t/\epsilon^2, \hat{F} \right)} = \frac{\frac{t}{\epsilon} \left( \psi \left( \hat{F}_0^m(t/\epsilon^2) \right) - \alpha \right)}{\sqrt{\frac{1}{m(m-1)} \sum_{i=1}^{m} \left( \frac{t}{\epsilon} \left( \psi \left( \hat{F}_i(t/\epsilon^2) \right) - \alpha \right) - \frac{t}{\epsilon} \left( \psi \left( \hat{F}_0^m(t/\epsilon^2) \right) - \alpha \right) \right)^2}} \Rightarrow \frac{\tilde{B}(t)}{\sqrt{\frac{1}{m(m-1)} \sum_{i=1}^{m} \left( B_i(t) - B(t) \right)^2}} \text{ as } \epsilon \to 0$$

$\square$
Before we prove Theorem 1, we prove an auxiliary lemma (Lemma 4) that characterize the limiting variance estimation process. This lemma will also be useful in characterizing the distribution of $K(\sigma)$ (the proof of Lemma 1).

Let
\[ S(t) = \sqrt{\sum_{i=1}^{m} (B_i(t) - \bar{B}(t))^2} \]
for $t > 0$.

**Lemma 4.** \{\( S(t) : t > 0 \)\} is a Bessel process of dimension \((m-1)\). When \( m > 2 \), \( S(t) \) solves the SDE
\[ dS(t) = \frac{m-2}{2S(t)} dt + dW(t); \]
when \( m = 2 \),
\[ dS(t) = dW(t) + dL_0(t) \]
where \( W(t) \) is a standard Brownian motion and \( L_0(t) \) is the local time at zero of \( W(t) \).

**Proof.** We denote \( \vec{B} = (B_1, B_2, \ldots, B_m) \) and \( R(t) = \sum_{i=1}^{m} (B_i(t) - \bar{B}(t))^2 \). Then
\[ dR(t) = 2 \sum_{i=1}^{m} (B_i(t) - \bar{B}(t)) dB_i(t) + (m-1) dt \]
As \( W(t) := \sum_{i=1}^{m} \int_{0}^{t} \frac{(B_i(u) - \bar{B}(u))}{\sqrt{\sum_{j=1}^{m} (B_j(u) - \bar{B}(u))^2}} \left\{ \sum_{j=1}^{m} (B_j(u) - B(u))^2 \neq 0 \right\} dB_i(u) \) is a continuous local martingale and \([W, W]_t = t\), we have \( W(t) \) is a standard Brownian motion (Jeanblanc et al. 2009). Thus,
\[ dR(t) = 2 \sqrt{R(t)} dW(t) + (m-1) dt. \]
Following Revuz and Yor (1999), when \( m > 2 \),
\[ dS(t) = \frac{m-2}{2S(t)} dt + dW(t); \]
when \( m = 2 \),
\[ dS(t) = dW(t) + dL_0(t). \]

\[ \square \]

**Proof of Theorem 1** We first notice that
\[ e^2 \kappa_\psi(\epsilon) \]
\[ = e^2 \inf \left\{ t > 0 : \sqrt{\frac{1}{m(m-1)} \sum_{i=1}^{m} \left( \psi \left( \hat{F}_i(t) - \hat{F}_0^n(t) \right) \right)^2 + a(t) < \epsilon} \right\} \]
\[
= \inf \left\{ t > 0 : \frac{1}{\varepsilon} \sqrt{\frac{1}{m(m-1)} \sum_{i=1}^{m} \left( \frac{t}{\varepsilon} \left( \bar{F}_i(t/\varepsilon^2) - \alpha \right) - \frac{t}{\varepsilon} \left( \bar{F}_0^m(t/\varepsilon^2) - \alpha \right) \right)^2 + \frac{1}{\varepsilon} a \left( t/\varepsilon^2 \right) < 1 \right\}
\]

= \inf \left\{ t > 0 : \sqrt{\frac{1}{m(m-1)t^2} \sum_{i=1}^{m} \left( \frac{t}{\varepsilon} \left( \bar{F}_i(t/\varepsilon^2) - \alpha \right) - \frac{t}{\varepsilon} \left( \bar{F}_0^m(t/\varepsilon^2) - \alpha \right) \right)^2 + \frac{1}{\varepsilon} a \left( t/\varepsilon^2 \right) < 1 \right\}.

Let \( \mathcal{T}(Y)(b) = \inf \{ t > 0 : Y(t) < b \} \), denoting the mapping induced by first passage time. As

\[
\sqrt{\frac{1}{m(m-1)t^2} \sum_{i=1}^{m} \left( \frac{t}{\varepsilon} \left( \bar{F}_i(t/\varepsilon^2) - \alpha \right) - \frac{t}{\varepsilon} \left( \bar{F}_0^m(t/\varepsilon^2) - \alpha \right) \right)^2 + \frac{1}{\varepsilon} a \left( t/\varepsilon^2 \right)} \Rightarrow \sqrt{\frac{\sigma^2}{m(m-1)t^2} \sum_{i=1}^{m} (B_i(t) - \bar{B}(t))^2 + 0} \text{ in } D(0, \infty) \text{ as } \varepsilon \to \infty,
\]

and the first passage time is a continuous mapping in \( M_1 \) (Whitt 2002), we have

\[
\mathcal{T} \left( \left\{ \sqrt{\frac{1}{m(m-1)t^2} \sum_{i=1}^{m} \left( \frac{t}{\varepsilon} \left( \bar{F}_i(t/\varepsilon^2) - \alpha \right) - \frac{t}{\varepsilon} \left( \bar{F}_0^m(t/\varepsilon^2) - \alpha \right) \right)^2} + \frac{1}{\varepsilon} a \left( t/\varepsilon^2 \right) : t \geq 0 \right\} \right) \Rightarrow \mathcal{T} \left( \left\{ \sigma \sqrt{\frac{1}{m(m-1)t^2} \sum_{i=1}^{m} (B_i(t) - \bar{B}(t))^2 : t > 0} \right\} \right)
\]

in \( D(0, \infty) \) when endowed with \( M_1 \) topology as \( \varepsilon \to \infty \). From Lemma 4, we have

\[
\sqrt{\frac{1}{m(m-1)t^2} \sum_{i=1}^{m} (B_i(t) - \bar{B}(t))^2} = \frac{d}{t} \frac{S(t)}{\sqrt{m(m-1)}},
\]

where \( S(t) \) is a Bessel process, which is a Feller process. Thus, we have the convergence of the first passage time process at point 1 (Whitt 2002), i.e.

\[
\inf \left\{ t > 0 : \sqrt{\frac{1}{m(m-1)t^2} \sum_{i=1}^{m} \left( \frac{t}{\varepsilon} \left( \bar{F}_i(t/\varepsilon^2) - \alpha \right) - \frac{t}{\varepsilon} \left( \bar{F}_0^m(t/\varepsilon^2) - \alpha \right) \right)^2 + \frac{1}{\varepsilon} a \left( t/\varepsilon^2 \right) < 1 \right\} \Rightarrow \inf \left\{ t > 0 : \sigma \sqrt{\frac{1}{m(m-1)t^2} \sum_{i=1}^{m} (B_i(t) - \bar{B}(t))^2} < 1 \right\} \text{ as } \varepsilon \to 0.
\]

\( \square \)

**Proof of Lemma 1** By the reversibility of Bessel process, we have

\[
K(1) = \inf \left\{ t > 0 : S(t) < \sqrt{m(m-1)t} \right\} = \inf \left\{ t > 0 : t S(1/t) < \sqrt{m(m-1)} \right\} = \inf \left\{ t > 0 : S(1/t) < \sqrt{m(m-1)} \right\} = \frac{1}{\sup \left\{ t > 0 : S(t) < \sqrt{m(m-1)} \right\}}.
\]
When \( m = 2 \), \( S(t) \) hits zero at arbitrarily large times (Revuz and Yor 1999). Therefore, \( K(1) = 0 \).

When \( m = 3 \), \( \lim_{t \to \infty} S(t) = 0 \) (Revuz and Yor 1999). Therefore, \( K(1) = 0 \).

When \( m \geq 4 \), it follows from the results in Section 8 of Pitman and Yor (1981) that \( K(1) \) has density

\[
 f_{K(1)}(t) = \frac{1}{\Gamma(\gamma)} \lambda^\gamma t^{\gamma-1} e^{-\lambda t}
\]

where \( \gamma = (m - 3)/2 \) and \( \lambda = m(m - 1)/2 \).

\[
\square
\]

**Proof of Theorem 2**

We first notice that

\[
\frac{\psi(\hat{F}_n(\kappa_\psi(\epsilon))) - \alpha}{\Gamma_\psi(\kappa_\psi(\epsilon), F)} = \frac{\epsilon^2 \kappa_\psi(\epsilon) \left( \psi(\hat{F}_n(\epsilon^2 \kappa_\psi(\epsilon)/\epsilon^2)) - \alpha \right)}{\epsilon^2 \kappa_\psi(\epsilon) \Gamma_\psi(\epsilon^2 \kappa_\psi(\epsilon)/\epsilon^2, F)}
\]

By Functional Central Limit Theorem (see Proof of Proposition 1), and standard random-time-change argument, and Convergence Together Theorem (Billingsley 1999), we have

\[
\frac{\epsilon^2 \kappa_\psi(\epsilon) \left( \psi(\hat{F}_n(\epsilon^2 \kappa_\psi(\epsilon)/\epsilon^2)) - \alpha \right)}{\epsilon^2 \kappa_\psi(\epsilon) \Gamma_\psi(\epsilon^2 \kappa_\psi(\epsilon)/\epsilon^2, F)} \Rightarrow \frac{\sigma \tilde{B}(K(\sigma))}{\sqrt{K(\sigma) - \tilde{B}(K(\sigma))}}
\]

\[
= \frac{\sigma \tilde{B}(K(\sigma))}{K(\sigma)} = \frac{\tilde{B}(K(\sigma))/\sqrt{K(\sigma)}}{\sqrt{K(\sigma)/\sigma^2}}
\]

We next show that \( \tilde{B}(K(\sigma))/\sqrt{K(\sigma)} \) is independent of \( K(\sigma) \) and \( \tilde{B}(K(\sigma))/\sqrt{K(\sigma)} \sim N(0, 1/m) \). We first define a sequence of discretized versions of \( K(\sigma) \). Let \( \Delta_n = 10^{-n} \) and define

\[
K_n(\sigma) = \min\{k\Delta_n : k\Delta_n > K(\sigma)\}
\]

Then \( K_n(\sigma) \downarrow K(\sigma) \) almost surely as \( n \to \infty \) (Revuz and Yor 1999). For \( K_n(\sigma) \), we have for any \( z \in \mathbb{R} \),

\[
P \left( \tilde{B}(K_n(\sigma))/\sqrt{K_n(\sigma)} \leq z, K_n(\sigma) = k\Delta_n \right)
\]

\[
= P \left( \tilde{B}(K_n(\sigma))/\sqrt{K_n(\sigma)} \leq z | K_n(\sigma) = k\Delta_n \right) P(K_n(\sigma) = k\Delta_n)
\]

\[
= P \left( \tilde{B}(k\Delta_n)/\sqrt{k}\Delta_n \leq z | \inf_{0 < t < (k-1)\Delta_n} \sigma \sqrt{\frac{1}{m(m-1)t^2} \sum_{i=1}^{m} (B_i(t) - \tilde{B}(t))^2} \geq 1 \right)
\]

\[
= \inf_{(k-1)\Delta_n \leq t < k\Delta_n} \sigma \sqrt{\frac{1}{m(m-1)t^2} \sum_{i=1}^{m} (B_i(t) - \tilde{B}(t))^2 < 1} \times P(K_n(\sigma) = k\Delta_n)
\]
As \( B(t) \) is independent of \( \sum_{i=1}^{m} (B_i(t) - \bar{B}(t))^2 \), and \( B(k\Delta_n) - \bar{B}(t) \) is independent of \( \sum_{i=1}^{m} (B_i(t) - \bar{B}(t))^2 \) for \( t \leq k\Delta_n \), \( B(k\Delta_n)/\sqrt{k\Delta_n} \) is independent of \( \sum_{i=1}^{m} (B_i(t) - \bar{B}(t))^2 \) for \( t \leq k\Delta_n \),

\[
(5) = P \left( \frac{B(k\Delta_n)}{\sqrt{k\Delta_n}} \leq z \right) P(K_n(\sigma) = k\Delta_n) = P(Z/\sqrt{m} \leq z) P(K_n(\sigma) = k\Delta_n)
\]

where \( Z \sim N(0,1) \). Thus, \( B(K_n(\sigma))/\sqrt{K_n(\sigma)} \) follows \( N(0,1/m) \), and is independent of \( K_n(\sigma) \). As \( B(t)/\sqrt{t} \) is continuous in \( t \),

\[
\left( K_n(\sigma), B(K_n(\sigma))/\sqrt{K_n(\sigma)} \right) \Rightarrow \left( K(\sigma), B(K(\sigma))/\sqrt{K(\sigma)} \right) \quad \text{as} \quad n \to \infty.
\]

Then \( B(K(\sigma))/\sqrt{K(\sigma)} \) is distributed as \( N(0,1/m) \), and for any \( z \in \mathbb{R}, \ t \in \mathbb{R}^+ \),

\[
P \left( \frac{B(K(\sigma))}{\sqrt{K(\sigma)}} \leq z, K(\sigma) \leq t \right) = \lim_{n \to \infty} P \left( \frac{B(K_n(\sigma))}{\sqrt{K_n(\sigma)}} \leq z, K_n(\sigma) \leq t \right) = \lim_{n \to \infty} P(Z/\sqrt{m} \leq z) P(K_n(\sigma) \leq t) = P(Z/\sqrt{m} \leq z) P(K(\sigma) \leq t)
\]

Therefore, \( B(K(\sigma))/\sqrt{K(\sigma)} \) is independent of \( K(\sigma) \).

We also notice that \( K(\sigma)/\sigma^2 \overset{d}{=} K(1) \) Thus,

\[
\frac{B(K(\sigma))/\sqrt{K(\sigma)}}{\sqrt{K(\sigma)/\sigma^2}} \overset{d}{=} \frac{Z}{\sqrt{mK(1)}}.
\]

\( \Box \)

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

\[
\psi \left( \hat{F}(t/\epsilon^2) \right) \Rightarrow a\mathcal{U}(t) \text{ in } D(0,\infty) \text{ as } \epsilon \to 0,
\]

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 4 is very similar to the proof of Theorem 2. We shall omit it here.

**Appendix B: Proof of Section 4**

_Proof of Theorem 5_ We first notice that \( \tilde{\xi}(\tilde{Y}(t/\epsilon^2)) - \mu \Rightarrow \sigma_0 B(t) \) in \( D(0,\infty) \) as \( \epsilon \to 0 \) by Donsker’s theorem. We next prove the results by showing that for any \( T > 0 \),

\[
\sup_{0 \leq t \leq T} \left| \frac{1}{\epsilon} \left( h(\tilde{Y}(t/\epsilon^2)) - h(\mu) \right) - h'(\mu) \frac{t}{\epsilon} (\tilde{Y}(t/\epsilon^2) - \mu) \right| \to 0
\]
in probability as $\epsilon \to 0$.

Define a sequence of $t_\epsilon$ such that

$$t_\epsilon \to \infty \text{ and } \epsilon^2 t_\epsilon \to 0 \text{ as } \epsilon \to 0$$

We first notice that

$$\sup_{0 \leq t \leq T} \left| \frac{t}{\epsilon} (h(\bar{Y}(t/\epsilon^2)) - h(\mu)) - h'(|\mu|) \frac{t}{\epsilon} (\bar{Y}(t/\epsilon^2) - \mu) \right|$$

$$\leq \sup_{0 \leq s \leq \epsilon^2 t_\epsilon} \left| \frac{s}{\epsilon} (h(\bar{Y}(s/\epsilon^2)) - h(\mu)) \right| + \sup_{0 \leq s \leq \epsilon^2 t_\epsilon} \left| h'(\mu) \frac{s}{\epsilon} (\bar{Y}(s/\epsilon^2) - \mu) \right|$$

$$+ \sup_{\epsilon^2 t_\epsilon \leq s \leq T} \left| \frac{s}{\epsilon} (h(\bar{Y}(s/\epsilon^2)) - h(\mu)) - h'(\mu) \frac{s}{\epsilon} (\bar{Y}(s/\epsilon^2) - \mu) \right|$$

(6)

We then show that each of the three parts on the right hand side of (6) converges to zero in probability as $\epsilon \to 0$.

For the first part,

$$\sup_{0 \leq s \leq \epsilon^2 t_\epsilon} \left| \frac{s}{\epsilon} (h(\bar{Y}(s/\epsilon^2)) - h(\mu)) \right| = \sup_{0 \leq s \leq \epsilon^2 t_\epsilon} |\epsilon s (h(\bar{Y}(s)) - h(\mu))|$$

Let $K_1$ denote the Lipschitz constant of $h$ at $\mu$, then we have, for any $\delta > 0$,

$$P \left( \sup_{0 \leq s \leq \epsilon^2 t_\epsilon} |\epsilon s (h(\bar{Y}(s)) - h(\mu))| > \delta \right)$$

$$\leq P \left( \sup_{0 \leq s \leq \epsilon^2 t_\epsilon} |\epsilon K_1 (Y(s) - \mu)| > \delta \right)$$

$$\leq P \left( \sup_{0 \leq k \leq \epsilon^2 t_\epsilon} |\epsilon K_1 (Y(k) - \mu)| > \delta \right)$$

$$\leq \frac{\epsilon^2 K_1^2 \epsilon^2 t_\epsilon E \{X(i) - \mu \}^2}{\delta^2} \to 0 \text{ as } \epsilon \to 0.$$

The last inequality holds by Doob’s Martingale inequality as $\{Y(k) - \mu k : k \geq 0 \}$ is a Martingale.

For the second part, as $\frac{1}{T} (\bar{Y}(t/\epsilon^2) - \mu) \Rightarrow \sigma_0 B(t)$ and $B$ is tight,

$$\sup_{0 \leq s \leq \epsilon^2 t_\epsilon} \left| h'(\mu) \frac{s}{\epsilon} (\bar{Y}(s/\epsilon^2) - \mu) \right| \to 0 \text{ in probability as } \epsilon \to 0.$$ 

For the third part, let $K_2$ denote the Lipschitz constant of $h'$ at $\mu$. Then, we have

$$\sup_{\epsilon^2 t_\epsilon \leq s \leq T} \left| \frac{s}{\epsilon} (h(\bar{Y}(s/\epsilon^2)) - h(\mu)) - h'(\mu) \frac{s}{\epsilon} (\bar{Y}(s/\epsilon^2) - \mu) \right|$$

$$= \sup_{\epsilon^2 t_\epsilon \leq s \leq T/\epsilon^2} \left| \epsilon s \{h(\bar{Y}(s)) - f(\mu) - h'(\mu)(\bar{Y}(s) - \mu)\} \right|$$

$$\leq \sup_{\epsilon^2 t_\epsilon \leq s \leq T/\epsilon^2} \left\{ K_2 \epsilon s \{\bar{Y}(s) - \mu\}^2 \right\}$$

$$\leq \sup_{\epsilon^2 t_\epsilon \leq s \leq T/\epsilon^2} \left\{ K_2 \frac{\sqrt{T}}{\sqrt{\epsilon}} s \{\bar{Y}(s) - \mu\}^2 \right\}$$
\[
\leq \sup_{k \geq \lfloor \epsilon t \rfloor} \left\{ \frac{K_2 \sqrt{T}}{k^{3/4}} \left( \frac{Y(k) - \mu k}{\frac{1}{k^{3/4}}} \right)^2 \right\}
\]

\[ \rightarrow 0 \text{ in probability as } \epsilon \rightarrow 0 \]

The convergence follows from the Law of Iterated Logarithm. Thus,

\[
\sup_{t \leq s \leq T} \left| 2 \left( f(Y(s/\epsilon^2)) - f(\mu) - \sigma f'(\mu)B(s) \right) \right| \rightarrow 0 \text{ in probability as } \epsilon \rightarrow 0.
\]

□

**Proof of Theorem 6**  
The proof follows from the proof of Theorem 6.2 in Sen (1972), we include it here for completeness. For simplicity of notation, we denote

\[ \Xi_\epsilon(t) = \frac{t}{\epsilon} \left( \hat{\xi}_{t/\epsilon^2, p} - \xi_p \right) \]

and

\[ \hat{\Xi}_\epsilon(t) = \frac{t}{\epsilon f(\xi_p)} \left( \hat{F}(t/\epsilon^2, \xi_p) - p \right) \]

We first notice that

\[ \hat{\Xi}_\epsilon(t) \Rightarrow \sigma B(t) \]

in \( D(0, \infty) \) as \( \epsilon \rightarrow 0 \) by Donsker’s theorem.

We next prove that for any \( T > 0, \) \( \sup_{0 \leq t \leq T} |\Xi_\epsilon(t) - \hat{\Xi}_\epsilon(t)| \rightarrow 0 \) in probability as \( \epsilon \rightarrow 0. \)

For every positive integer \( k, \) fix \( \delta > 0, \) we define

\[ k^* = |F^{-1}(k^{-1+\delta})| + |F^{-1}(1 - k^{-1+\delta})| + 1. \]

We then consider a sequence of positive integers \( k_\epsilon \) satisfying that

\[ k_\epsilon \rightarrow \infty \text{ and } \epsilon k_\epsilon k^*_\epsilon \rightarrow 0 \text{ as } \epsilon \rightarrow 0 \]

Then for any \( T < \infty \)

\[
\sup_{0 \leq t \leq T} |\Xi_\epsilon(t) - \hat{\Xi}_\epsilon(t)| \leq \sup_{0 \leq t \leq \epsilon^2 k_\epsilon} |\Xi_\epsilon(t)| + \sup_{0 \leq t \leq \epsilon^2 k_\epsilon} |\hat{\Xi}_\epsilon(t)| + \sup_{\epsilon^2 k_\epsilon \leq t \leq T} |\Xi_\epsilon(t) - \hat{\Xi}_\epsilon(t)|
\]

We shall show the three terms on the righthand converges to zero as \( \epsilon \rightarrow 0 \) one by one.

For \( \sup_{0 \leq t \leq \epsilon^2 k_\epsilon} |\Xi_\epsilon(t)|, \) we first notice that

\[
\sup_{0 \leq t \leq \epsilon^2 k_\epsilon} |\Xi_\epsilon(t)| = \max_{1 \leq n \leq k_\epsilon} \{ n \epsilon (\hat{\xi}_{n, p} - \xi_p) \} \leq \epsilon k_\epsilon (|X(\epsilon k_\epsilon) - \xi_p| + |X(\epsilon k_\epsilon) - \xi_p|).
\]
We then notice that as $k_{\epsilon} (1 - F(X_{(k_{\epsilon}, k_{\epsilon})})) \Rightarrow \text{Exp}(1)$ as $\epsilon \to 0$,

$$P(\hat{\xi}_p \leq X_{(k_{\epsilon}, k_{\epsilon})} \leq F^{-1}(1 - k_{\epsilon}^{1-\delta})) = P(k_{\epsilon}(1 - p) \geq k_{\epsilon}(1 - F(X_{(k_{\epsilon}, k_{\epsilon})})) > k_{\epsilon}^{1-\delta}) \to 1$$

as $\epsilon \to 0$. Similarly, as $k_{\epsilon} F(X_{(k_{\epsilon}, 1)}) \Rightarrow \text{Exp}(1)$ as $\epsilon \to 0$, we have $P(F^{-1}(k_{\epsilon}^{1-\delta}) \leq X_{(k_{\epsilon}, 1)} \leq \hat{\xi}_p) \to 1$ as $\epsilon \to 0$.

Thus,

$$\sup_{0 \leq t \leq 2k_{\epsilon}} |\Xi_{\epsilon}(t)| \to 0 \text{ in probability as } \epsilon \to 0.$$ 

For $\sup_{0 \leq t \leq 2k_{\epsilon}} |\widehat{\Xi}_{\epsilon}(t)|$, as $\widehat{\Xi}_{\epsilon}(t) \Rightarrow \sigma B(t)$ and $B$ is tight, $\sup_{0 \leq t \leq 2k_{\epsilon}} |\check{\Xi}_{\epsilon}(t)| \to 0$ in probability as $\epsilon \to 0$.

Lastly the Bahadur representation indicates that

$$\sup_{t^2 k_{\epsilon} \leq T} |\Xi_{\epsilon}(t) - \check{\Xi}_{\epsilon}(t)| = \sup_{k_{\epsilon} \leq n \leq T k_{\epsilon}^{-2}} \left\{ \epsilon n \left( \left( \hat{\xi}_{n,p} - \xi_p \right) - \frac{F_n(\xi_p) - p}{f(\xi_p)} \right) \right\} \to 0$$

in probability as $\epsilon \to 0$.

□

Acknowledgments

References


