Optimal Budget Allocation in the Evaluation of Simulation-Optimization Algorithms

ERIC ANJIE GUO and SHANE G. HENDERSON, Cornell University

To efficiently evaluate simulation-optimization algorithms, we propose three different performance measures and their respective estimators. Only one estimator achieves the canonical Monte Carlo convergence rate $O(T^{-1/2})$, while the other two converge at the sub-canonical rate of $O(T^{-1/3})$. For each estimator, we study how the computational budget should be allocated between the execution of the optimization algorithm and the assessment of the output, so that the mean squared error of the estimator is minimized.

General Terms: Simulation, Optimization, Performance

ACM Reference Format:
DOI = 10.1145/0000000.0000000 http://doi.acm.org/10.1145/0000000.0000000

1. INTRODUCTION

Simulation Optimization captures the special case of mathematical programming problems where the objective function and constraints are expressed implicitly through a stochastic simulation. The stochastic simulation, when provided with an input $x$, is “executed at the design $x$” to provide an estimate of the objective function at $x$. The objective function $g(x)$ is defined as the expectation of the random variable $G(x)$:

$$g(x) := E G(x).$$

The random variable $G(x)$ is the estimate provided by the stochastic simulation, and the replications are assumed to be independent and identically distributed. In practice, the overhead of any simulation-optimization algorithm is usually dominated by the computational effort that has to be expended on the evaluation of the estimates $G(x)$. In light of that, we measure the computational budget for simulation-optimization algorithms by the number of times the stochastic simulation is executed, and denote the budget by $t$. We view $t$ as a fixed parameter specified by the user.

Let $V(t)$ denote the true objective function value associated with the estimated (and therefore random) optimal solution $X(t)$ after the budget $t$ has been expended, for a given algorithm, namely

$$V(t) := g(X(t)). \quad (1)$$

In this paper we develop estimators of the mean, distribution function, and quantiles of $V(t)$. These estimators can be used to characterize the performance of a given algorithm, or to rank different algorithms applied to a class of simulation-optimization problems. The variance of the estimators converges to 0 as we run multiple replications of the simulation-optimization algorithm and get estimated optimal solutions $X(t)$. Our estimators of the distribution function and quantiles of $V(t)$ are biased, and the bias is reduced as we compute more precise estimates of $g(\cdot)$ at each $X(t)$. Within a given computational effort $T$, not to be confused with the budget $t$ for the simulation-optimization algorithm on one replication, we trade off variance and bias to minimize the mean squared error (MSE) of the estimators. Our results indicate that the estimator of the mean converges at the canonical Monte Carlo convergence rate $T^{-1/2}$, while sub-canonical rates are achieved for the distribution function and quantiles.

We further generalize our results to the scenario where the budget $t$ is allowed to assume multiple values, and we might then plot the true objective function value...
$g(X(t))$ associated with the estimated optimal solution $X(t)$ as a function of $t$. This plot illustrates how the quality of the optimal solution as estimated by the algorithm evolves over time. Such performance measures plots are present in various commercial simulation-optimization packages. Figure 1 shows one example of the Performance Measures Plot as in SimRunner [ProModel 2011].

When multiple estimators with different budgets $t_k, k = 1, \ldots, l$ are considered, the optimal budget allocation scheme is also solved with the goal being to minimize the sum or maximum of MSEs across all budgets. Whether the performance measure is the mean, distribution function or quantile, the same convergence rate as in the single-budget case is attained. Furthermore, different simulation approaches give rise to different computational cost structure, due to the possibility of reusing sample paths for some classes of simulation-optimization algorithms.

For each scenario discussed above, we provide an intuitive interpretation for the optimal budget allocation scheme and the resulting convergence rate of the estimator with respect to the total computational budget.

The estimators discussed in this paper are all constructed via nested simulation, which has been adopted to estimate the probability of loss, Value at Risk (VaR), and expected shortfall (ES) of portfolios in financial applications. The VaR was first analyzed in the context of nested simulation by Lee [1998]; Lee and Glynn [2003], where central limit theorems were established for estimators of the distribution function and quantiles. We exploit these results and generalize the central limit theorem to a multidimensional scenario in order to demonstrate the asymptotic properties of our estimators in the multiple-budget case. Nested-simulation problems were subsequently considered by Gordy and Juneja [2010], who suggested an optimal budget allocation that minimizes the MSE of each estimator. Despite the different context, the optimization problems in Gordy and Juneja [2010] bear a strong resemblance to the optimal budget allocation problems of a single-budget in this paper, and the resulting solutions show a lot of similarity as well. We also refer the reader to the work of Broadie et al. [2011], which proposed a nonuniform nested simulation algorithm. That paper is different from previous work and this paper in the sense that the computational effort is not determined prior to the simulation via deterministic optimization, but allocated in a sequential fashion. It may be possible to improve the convergence rate of some of the estimators discussed in this paper by adopting an adaptive inner simulation.
length and applying the methodology of Broadie et al. [2011]. That would require a sequential allocation scheme and analysis that is substantially different from the static allocations considered herein.

Our work should not be confused with Optimal Computing Budget Allocation (OCBA), another line of research that primarily focuses on attaining the highest simulation decision quality using a fixed computing budget $t$. An introduction to OCBA ideas can be found in Chen et al. [2008]; Fu et al. [2008]; Chen and Lee [2011].

In our view the primary contributions of this paper are as follows.

— We draw attention to the finite-time properties of simulation-optimization algorithms, as a complement to convergence results. Finite-time properties for practically relevant simulation-optimization budgets $(t)$ are of great interest, but appear to be difficult to obtain in any kind of analytic fashion. The results in this paper are relevant for estimating these properties on test problems, as in a research agenda advanced in Fu [2002]; Glynn [2002]; Pasupathy and Henderson [2006, 2011].

— We highlight $V(t)$ as an important performance measure in these settings, and show how to estimate its distributional properties through nested simulation.

— We obtain the asymptotically optimal (non-sequential) design of nested simulations in the sense of minimizing mean-squared error. The resulting expressions provide insight on how to design simulation experiments to estimate properties of $V(t)$, both in the single budget case (one value of $t$) and for multiple budgets.

— We develop multivariate central limit theorems for the distribution function and quantiles in the multiple-budget case.

The rest of this paper is organized as follows. Section 2 lays out the framework and basic notation for optimal budget allocation in the evaluation of simulation-optimization algorithms. Estimators of the mean, distribution function, and quantiles of $V(t)$ are analyzed in Section 3, Section 4, and Section 5, respectively. Each section treats the single-budget and multiple-budget cases separately, and budget allocation schemes are established for both. In particular, the optimal convergence rates can be different depending on whether the feasible region is discrete or continuous, as will be seen in Sections 4 and 5. In Section 6, we focus on the scenario where sample paths are reused, and thus correlation is introduced between the estimators for different budgets $t_k, k = 1, \ldots, l$. Multivariate central limit theorems are established for each type of estimator, which can be proved by generalizing the results of Lee [1998]; Lee and Glynn [2003]. A summary of the results is given in Section 7. We provide proof sketches of some of the results herein, and complete proofs in an online appendix.

2. FRAMEWORK

Suppose that the budget $t$ is fixed, and we wish to estimate one of the aforementioned performance measures. A simulation approach to this problem is as follows. Run the simulation-optimization algorithm with a budget $t$, $n$ independent times. Let $X_1(t), \ldots, X_n(t)$ be the resulting estimates of optimal solutions. Now estimate $g(X_i(t))$ at each point $X_i(t)$ with a simulation run of length $m$, for each $i = 1, \ldots, n$, yielding estimates $Z_1(t), \ldots, Z_n(t)$ of $g(X_1(t)), \ldots, g(X_n(t))$ respectively, each of which is a sample average of (conditionally) i.i.d. observations:

$$Z_i(t) := \frac{1}{m} \sum_{j=1}^{m} G_j(X_i(t)). \quad (2)$$

Then the performance measure is estimated using $Z_1(t), \ldots, Z_n(t)$. For example, for estimating the $p$th quantile of $g(X(t))$, we use the $p$th sample quantile of $Z_1(t), \ldots, Z_n(t)$. 

ACM Transactions on Modeling and Computer Simulation, Vol. 9, No. 4, Article 39, Publication date: March 2010.
We can also vary the budget \( t \) to obtain further information on the performance of a single algorithm on a single problem. Let \( t_1, \ldots, t_l \) be the budgets to be considered. Without loss of generality, they are arranged in ascending order, namely \( t_1 < \cdots < t_l \). Let \( V(t_k) \) denote the true objective function value associated with the estimated optimal solution \( X(t_k) \) after the budget \( t_k \) has been expended, i.e., \( V(t_k) := g(X(t_k)) \).

The number of times the simulation-optimization algorithm is run is allowed to vary with the budget \( t \). For each budget \( t_k, k = 1, \ldots, l \), run the simulation-optimization algorithm \( n_k \) independent times. Let \( X_i(t_k), i = 1, \ldots, n_k \) be the resulting estimates of optimal solutions. The estimates of \( V(t_k) \) are of the same form as (2):

\[
Z_i(t_k) := \frac{1}{m_k} \sum_{j=1}^{m_k} G_j(X_i(t_k)).
\]

The estimators of the performance measures with respect to each budget \( t_k \) are formulated in the same way as in the single-budget context.

Let \( T \) denote the total computational budget available for evaluating the simulation-optimization algorithm, not to be confused with the budget \( t \) for running the simulation-optimization algorithm itself. We measure \( T \) in the same units as \( t \), namely the number of times the stochastic simulation is executed. For a given \( t \), let \( n, m \) be chosen so that the computational effort required to estimate the performance measure is approximately \( T \). Since \( n \) instances of \( X(t) \) need to be simulated, and each requires \( m \) estimates of \( G(X(t)) \), the aggregate effort is \( tn + mn \), and thus

\[
n(t + m) = T. \tag{3}
\]

In practice, \( n \) and \( m + t \) must be positive integers and lie in the interval \([1, T]\). We relax these restrictions, because our goal is insight into the form of the optimal allocations as the computational budget \( T \to \infty \) and imposing these restrictions with the ensuing notational complication does not lead to further insight or clarity. For some problem parameters, this relaxation could lead to "optimal" values of \( t + m \) or \( n \) that are less than 1 or greater than \( T \). In such cases the optimal value is instead the projection onto the interval \([1, T]\), because our recommendations arise from solving convex minimization problems. In any case, for sufficiently large \( T \), both \( n \) and \( m \) will be smaller than \( T \), so the upper bound of \( T \) is unnecessary from an asymptotic standpoint.

In principle, in order to generate a performance measures plot as in Figure 1, we need to estimate the value of \( V(t) \) for all \( t > 0 \). From a practical standpoint, we restrict our attention to a finite set of budgets \( t_1 < t_2 < \cdots < t_l \). We assume that these budgets are given exogenously, rather than attempting to select them as part of our estimation procedures. Indeed, these budgets will likely be related to computational budgets that are practically feasible, and therefore should be chosen by the user of a simulation optimization algorithm. Consequently, we focus on delivering high-quality estimators of \( V(t) \) for any specified budget \( t \), but we do not view \( t \) as a decision variable.

There are two possible cost structures associated with the multiple-budget scenario. First, if the simulation has to be done independently for each budget \( t_k \), then the total computational budget is simply

\[
T = \sum_{k=1}^{l} T_k, \tag{4}
\]

where

\[
T_k := n_k(t_k + m_k) \tag{5}
\]

is the aggregate effort spent to compute the performance measure for budget \( t_k \). Second, for many simulation-optimization algorithms, the partial sample paths of esti-
Optimal Budget Allocation in the Evaluation of SO Algorithms

3. EXPECTATIONS

One way to examine the quality of an estimated optimal solution $X(t)$ produced by a simulation-optimization algorithm is to take the expectation of the true objective function value, i.e., $E V(t) = E g(X(t))$; see (1). In this section, we propose an estimator $\hat{\mu}$ for this expectation, and present the optimal budget allocation for both single and multiple budgets. It turns out that the estimator achieves the canonical Monte Carlo convergence rate $O(T^{-1/2})$.

3.1. The Single-Budget Case

Suppose that the budget $t$ is fixed. In order for the performance measure to be well-defined we assume that $V(t)$ has finite expectation, which holds, for example, when $g(\cdot)$ is a continuous function and the feasible region of $x$ is compact.

Fig. 2. Performance Measures Plot of Scaled and Shifted Kiefer-Wolfowitz Stochastic Approximation Algorithm applied to $G(x) = -x^T x + \mathcal{N}(0, 1)$, $x \in \mathbb{R}^5$: RMSEs are displayed at each computational budget.
The obvious estimator of \( E V(t) \) is

\[
\hat{\mu} := \frac{1}{n} \sum_{i=1}^{n} Z_i(t),
\]

which is unbiased.

**Definition 3.1.** Let \( \rho(t) \) be the correlation between the estimated objective function values at a common estimated optimal solution \( X(t) \) after the budget \( t \) has been expended, so that

\[
\rho(t) := \text{corr}(G_1(X(t)), G_2(X(t))).
\]

To compute \( \rho(t) \), first we compute

\[
\text{Cov}(G_1(X(t)), G_2(X(t))) = E[(G_1(X(t)) - E V(t))(G_2(X(t)) - E V(t))]
\]

\[
= E[E[(G_1(X(t)) - E V(t))(G_2(X(t)) - E V(t))|X(t)]]
\]

\[
= E[E[G_1(X(t)) - E V(t)|X(t)] E[G_2(X(t)) - E V(t)|X(t)]]
\]

\[
= E[(g(X(t)) - E V(t))(g(X(t)) - E V(t))]
\]

\[
= \text{Var}(g(X(t))).
\]

It follows that the correlation \( \rho(t) \) can also be written as

\[
\rho(t) = \frac{\text{Var}(g(X(t)) / \text{Var} G(X(t)))}{\text{Var} G(X(t))}.\]

In other words, it is the proportion of variability of \( G(X(t)) \) caused by the common factor \( X(t) \).

Consequently, the correlation \( \rho(t) \) is well-defined if and only if \( \text{Var} G(X(t)) \) is. Thus, we also assume finite variance of the estimate at the random estimated optimal solution, or

\[
\text{Var} G(X(t)) \leq \infty.
\]

If \( g(\cdot) \) is a continuous function on a compact space, then the condition holds, e.g., when

\[
\text{Var} G(x) \text{ is bounded uniformly on its domain.}
\]

By conditioning on the estimated optimal solution \( X(t) \), we obtain the following result.

**Proposition 3.2.** The correlation \( \rho(t) \) is nonnegative, and the MSE of \( \hat{\mu} \) with respect to \( E V(t) \) is

\[
\text{MSE} \hat{\mu} := \text{MSE}(\hat{\mu}, E V(t)) = \frac{1 + (m - 1)\rho(t)}{nnm} \text{Var} G(X(t)).
\]

We now turn to identifying the values of \( n \) and \( m \) that minimize this MSE, subject to the constraint that \( n(t + m) = T \) as in (3), where we relax the constraint that \( n \) and \( t + m \) should be integers in \([1, T]\).

**Theorem 3.3.** The values of \( n \) and \( m \) that minimize \( \text{MSE} \hat{\mu} \) are

\[
n = \sqrt{T} \left[ \sqrt{1/\rho(t)} - 1 + \sqrt{T} \right], \quad m = \sqrt{1/\rho(t) - 1} t,
\]

and the resulting minimum MSE is

\[
\text{MSE} \hat{\mu} = \left[ \sqrt{1 - \rho(t)} + \sqrt{\rho(t)} t \right] \frac{\text{Var} G(X(t))}{T}.
\]

A proof is given in the Electronic Appendix. This result can also be viewed as a special case of splitting; cf. Asmussen and Glynn [2007, p. 147–149]. To see why, recall that in splitting we wish to estimate \( E \varphi(X, Y) \) for some (known) function \( \varphi \) of two
random variables \(X\) and \(Y\). In our setting \(\phi(x, y) = y, X = X(t)\) and \(Y = G(X(t))\). The method of splitting involves using an estimator of the form

\[
\hat{\mu} = \frac{1}{n} \sum_{i=1}^{n} \frac{1}{m} \sum_{j=1}^{m} G_j(X_i(t)) = \frac{1}{nm} \sum_{i=1}^{n} \sum_{j=1}^{m} \varphi(X_i(t), G_j(X_i(t))),
\]

(7)

where we use multiple samples of \(Y\) for each sample of \(X\). Since \(\hat{\mu}\) is unbiased, to minimize its MSE it is sufficient to find the values of \(n\) and \(m\) that minimize its variance, which follow directly from the general theory in Asmussen and Glynn [2007] for splitting.

It can be seen from (6) that \(m\) is increasing in \(t\) but decreasing in \(\rho(t)\), which matches intuition. Indeed, when \(t\) is large, the optimal solutions \(X_1(t), \ldots, X_n(t)\) are costly to estimate, and so it is optimal to limit \(n\) and use long “internal” simulation run lengths \(m\). Meanwhile, if the correlation \(\rho(t)\) is large, then information is shared across \(X_i(t)\)'s and incurs bigger covariance, and therefore the optimal \(m\) decreases.

Notice that \(m\) stays constant and does not increase with \(T\). This is a direct consequence of the unbiasedness of the estimator. In this case the optimal value of \(m\) is determined purely by the variance, which depends on \(T\) only through the number of “macro” replications \(n\).

The expression for the optimal MSE is of order \(T^{-1}\), which shows that the estimator of \(EV(t)\) converges at the so-called canonical Monte Carlo convergence rate \(T^{-1/2}\); see Asmussen and Glynn [2007, p. 70] for a discussion of this nomenclature.

3.2. The Multiple-Budget Case

Now we vary the budget \(t\). Let \(t_1 < t_2 < \cdots < t_l\) be the budgets under consideration, and \(EV(t_k), k = 1,\ldots, l\) the performance measures that need to be estimated. The obvious estimator for each \(EV(t_k)\) is

\[
\hat{\mu}_k := \frac{1}{n_k} \sum_{i=1}^{n_k} Z_i(t_k).
\]

Definition 3.4. For each \(k = 1,\ldots, l\), let \(\rho(t_k)\) be the correlation between the estimated objective function values at a common estimated optimal solution \(X(t_k)\) after the budget \(t_k\) has been expended:

\[
\rho(t_k) := \text{corr}(G_1(X(t_k)), G_2(X(t_k))) = \frac{\text{Var} g(X(t_k))}{\text{Var} G(X(t_k))},
\]

where the latter expression ensures that \(\rho(t_k)\) is nonnegative; see Proposition 3.2.

We first consider the scenario where sample paths cannot be reused, and simulations for each budget \(t_k\) have to be performed independently. The case where sample paths are reused is discussed in Section 3.3.

In Section 2, we have established the constraints (4) and (5), namely \(n_k(t_k + m_k) = T_k\) and \(\sum_{k=1}^{l} T_k = T\). Theorem 3.3 guarantees that the optimal values of \(n_k\) and \(m_k\) must satisfy

\[
n_k = \frac{T_k}{\sqrt{t_k} \left[ \sqrt{1/\rho(t_k)} - 1 + \sqrt{t_k} \right]}, \quad m_k = \sqrt{(1/\rho(t_k) - 1) t_k},
\]

(8)

given that \(T\) is sufficiently large so that \(m_k \leq T_k\) for all \(k\), and the minimum MSE of each \(\hat{\mu}_k\) with respect to \(EV(t_k)\) is

\[
\text{MSE}\hat{\mu}_k := \text{MSE}(\hat{\mu}_k, EV(t_k)) = \left[ \sqrt{1 - \rho(t_k)} + \sqrt{\rho(t_k) t_k} \right]^2 \frac{\text{Var} G(X(t_k))}{T_k}.
\]

(9)
The minimum sum of MSE $\hat{\mu}_k$, $k = 1, \ldots, l$ is

$$\sum_{k=1}^l \text{MSE } \hat{\mu}_k = \left[ \sum_{k=1}^l \left( \sqrt{1 - \rho(t_k)} + \sqrt{\rho(t_k) t_k} \right) \sqrt{\text{Var} \left( X(t_k) \right)} \right]^2 \frac{1}{T}, \quad (10)$$

and is attained when

$$T_k = \frac{\left( \sqrt{1 - \rho(t_k)} + \sqrt{\rho(t_k) t_k} \right) \sqrt{\text{Var} \left( X(t_k) \right)}}{\sum_{h=1}^l \left( \sqrt{1 - \rho(t_h)} + \sqrt{\rho(t_h) t_h} \right) \sqrt{\text{Var} \left( X(t_h) \right)}} T, \quad (11)$$

for each $k = 1, \ldots, l$.

It can be seen from (11) that each $T_k$ is proportional to $T$, which implies $T_k^{-1/2} = O(T^{-1/2})$. Combining with Theorem 3.3, it follows that each estimator $\hat{\mu}_k$ achieves the canonical Monte Carlo convergence rate $O(T^{-1/2})$, which is reflected in (10).

The intuition of (8) is identical to the single-budget case. Also, due to the constraint (4), each $T_k$ decreases as $T_h, h \neq k$ increases, and is therefore adversely affected by $t_h$ and $\rho(t_h)$.

Proof Sketch. Apply the Cauchy-Schwarz inequality to the product of (4) and (9) to get a lower bound of the sum of MSEs. The optimal values of $T_k$ are derived from the condition for the equality to hold.

Corollary 3.6. The optimal values of $n_k$ and $m_k$ that minimize the sum of MSE $\hat{\mu}_k$, $k = 1, \ldots, l$ are

$$n_k = \frac{\sqrt{\text{Var} \left( X(t_k) \right)} \rho(t_k)/t_k}{\sum_{h=1}^l \sqrt{\text{Var} \left( X(t_h) \right)} \left( \sqrt{1 - \rho(t_h)} + \sqrt{\rho(t_h) t_h} \right)} T, \quad m_k = \sqrt{1/\rho(t_k) - 1} t_k.$$  

By now it should not be surprising that the optimal value of $m_k$ remains constant (for sufficiently large $T$), which is the same as in the single-budget case. This is a direct consequence of splitting. Indeed, in Asmussen and Glynn [2007, p. 147–149], it was argued that the optimal length of the inner-level simulation can be determined without knowledge of the total computational budget. In the context of multiple-budget evaluation of simulation-optimization algorithms, this observation means that $m_k$ is independent of $T_k$ for each $k = 1, \ldots, l$, and hence independent of $T$ as well.

An alternative to minimizing the sum of the MSEs is to minimize the maximum of the MSEs, which is also equivalent to minimizing the width of the longest error bar in Figure 2.

Theorem 3.7. The smallest possible maximum of MSE $\hat{\mu}_k$, $k = 1, \ldots, l$ is

$$\max_{k=1, \ldots, l} \text{MSE } \hat{\mu}_k = \frac{1}{T} \sum_{k=1}^l \text{Var} \left( X(t_k) \right) \left[ \sqrt{1 - \rho(t_k)} + \sqrt{\rho(t_k) t_k} \right]^2, \quad (12)$$

and it is attained when

$$T_k = \frac{\text{Var} \left( X(t_k) \right) \left[ \sqrt{1 - \rho(t_k)} + \sqrt{\rho(t_k) t_k} \right]^2}{\sum_{h=1}^l \text{Var} \left( X(t_h) \right) \left[ \sqrt{1 - \rho(t_h)} + \sqrt{\rho(t_h) t_h} \right]^2} T, \quad (13)$$

for each $k = 1, \ldots, l$.  

ACM Transactions on Modeling and Computer Simulation, Vol. 9, No. 4, Article 39, Publication date: March 2010.
The estimated optimal solution \( t \) is obtained when \( \text{MSE} \hat{\mu}_k, k = 1, \ldots, l \) is least when they are all equal. Imposing such a condition in (9), and the optimal values of \( T_k \) follow immediately. □

The ratio \( T_k/T \) is independent of the total computational budget \( T \), whether we minimize the sum or maximum of the MSE. Moreover, the convergence rate is unaffected by the choice of sum or maximum. In fact, as long as each ratio \( T_k/T \) is independent of the total computational budget \( T \), the optimal convergence rate will be attained in both cases, and the performance will differ only by a constant multiplier. Therefore, the choice of performance measure in the multiple-budget case is not vital in terms of convergence rate, and we only consider the “maximum” performance measure, which is mathematically more tractable, in the remainder of the paper.

3.3. Reusing Partial Sample Paths

Suppose the budgets \( t_k, k = 1, \ldots, l \) are increasing in \( k \), namely \( t_1 < \cdots < t_l \). For many simulation-optimization algorithms terminated at a computational budget \( t_k \) with an estimated optimal solution \( X(t_k) \), the partial sample path \( X(t_1), \ldots, X(t_{k-1}) \) can be reused in \( \hat{\mu}_1, \ldots, \hat{\mu}_{k-1} \), respectively.

For example, consider the class of stochastic approximation algorithms which includes Kiefer-Wolfowitz [Kiefer and Wolfowitz 1952] and Robbins-Monro [Robbins and Monro 1951] algorithms. A typical iteration of these algorithms is of the form

\[
X(j + 1) \leftarrow X(j) + \epsilon(j) Y(j + 1),
\]

where \( K \) is a given square matrix and, conditional on \( X(j) \), \( Y(j + 1) \) is an estimator of \( \nabla g(X(j)) \).

Typically, \( \{\epsilon(j)\}_{j \geq 0} \) is a sequence of pre-specified positive deterministic constants that does not depend on the simulation-optimization budget \( t \). If the simulation-optimization algorithm is used with budget \( t_k \), then when it terminates with the estimated optimal solution \( X(t_k) \), we would have iterated through the sequence of solutions \( \{X(t)\}_{t \geq 0} \), and hence the intermediate optimal solutions \( X(t_1), \ldots, X(t_{k-1}) \) for budgets \( t_1, \ldots, t_{k-1} \) are available with no additional computational cost. Hence the budget constraint takes the form of (14).

Sometimes, however, the sequence \( \{\epsilon(j)\}_{j \geq 0} \) depends on the computational budget \( t \). For example, in Nemirovski et al. [2008], \( \epsilon_j = \epsilon \) does not depend on \( j \), but does depend on the (prespecified) budget \( t \). In this case, the sample paths with each budget \( t_k \) cannot be reused for other budgets, since different values of \( \epsilon \) are used for each budget.

Assuming that the estimated optimal solutions \( X(t_1), \ldots, X(t_{k-1}) \) are reused, their effective computational costs are all \( 0 \). Alternatively, one could think of \( X(t_1) \) having a computational cost of \( t_1 \), but the cost of generating each \( X(t_k) \) with \( k > 1 \) to be \( t_k - t_{k-1} \), as the first \( t_{k-1} \) units of time have been expended to compute \( X(t_1), \ldots, X(t_{k-1}) \). Hence, for given \( n_k, m_k, k = 1, \ldots, l \), the total computational cost for all the estimators \( \hat{\mu}_1, \ldots, \hat{\mu}_l \) is

\[
T = \sum_{k=1}^{l} n_k (t_k - t_{k-1} + m_k), \tag{14}
\]

where \( t_0 = 0 \).

The new budget constraint (14) takes the same form as the budget constraint (4) and (5) without sample-path reuse, only with each \( t_k \) replaced by \( t_k - t_{k-1} \). Therefore, if no other constraints are imposed on \( n_k \) and \( m_k \), then their optimal values take the same
form as in Corollary 3.6, only with each \( t_k \) replaced by \( t_k - t_{k-1} \), i.e.,

\[
\begin{align*}
    n_k &= \frac{\sqrt{\rho(t_k) \text{Var} G(X(t_k))/ (t_k - t_{k-1})}}{\sum_{h=1}^{l} \sqrt{\text{Var} G(X(t_h))} \left[ \sqrt{1 - \rho(t_h)} + \sqrt{\rho(t_k) (t_k - t_{h-1})} \right]} T, \\
    m_k &= \frac{\sqrt{(1/\rho(t_k) - 1)(t_k - t_{k-1})}}{\sum_{h=1}^{l} \sqrt{\text{Var} G(X(t_h))} \left[ \sqrt{1 - \rho(t_h)} + \sqrt{\rho(t_k) (t_k - t_{h-1})} \right]} T,
\end{align*}
\]

and the minimum sum of MSE \( \hat{\mu}_k \) is

\[
\sum_{k=1}^{l} \text{MSE} \hat{\mu}_k = \left[ \sum_{k=1}^{l} \left( \sqrt{\text{Var} G(X(t_k))} \left[ 1 - \frac{1}{\rho(t_k)} \right] \right) \right]^{2} \frac{1}{T}.
\]

This derivation relies on the assumption that all the sample paths are reused. Under this assumption, the sample sizes \( n_k, k = 1, 2, \ldots, l \) are nested, so there is an additional implicit constraint that \( n_1 \geq \cdots \geq n_l \). This constraint will always be satisfied by the optimal solution. Indeed, whether the performance measure is the sum or the minimum of the MSEs, it is certainly monotone increasing in each of the MSEs. A non-nested solution would imply that we would not use some of the partial sample paths, in effect reducing the sample size \( n_k \) used in computing \( \hat{\mu}_k \) for some \( k \). But this would increase the MSE \( \hat{\mu}_k \) relative to reusing the sample paths. Therefore, we do not need to impose a constraint that the sample sizes \( n_k \) are decreasing in \( k \).

4. DISTRIBUTION FUNCTION

The mean is sensitive to extreme values. In particular, if a simulation-optimization algorithm outputs a good quality solution most of the time, our intuition would suggest that this is a good algorithm. However, if the algorithm occasionally goes awry, giving an extremely poor solution with a small probability, then the estimator \( \hat{\mu} \) could be significantly affected. We might then draw the conclusion that the algorithm is unacceptable, which is not the case.

An alternative performance measure is the distribution function value \( F(r) := P(V(t) \leq r) \) for some fixed \( r \). Essentially, \( P(V(t) \leq r) \) measures the probability that the estimated optimal solution \( X(t) \) is of low quality in terms of the true objective function value \( g(X(t)) \) (for maximization problems).

In this section, we propose an estimator \( \hat{F}(r) \) for the performance measure \( F(r) \) and show that it converges at a sub-canonical rate. In particular, if the simulation-optimization problem in question has a discrete feasible region, then the convergence rate is \( O(\sqrt{(\log T)/T}) \), while the rate is \( O(T^{-1/3}) \) if the feasible region is continuous. We also solve for the optimal simulation run length parameters \( n, m \) that minimize the asymptotic (as the budget available for evaluating the simulation-optimization algorithm increases) MSE (AMSE). Our argument involves expressing \( F(r) \) as the distribution function of a conditional expectation, and invoking results from Lee [1998]; Lee and Glynn [2003]. This is also the approach we follow in Section 5 and Section 6.

4.1. Monte Carlo Computation of Conditional Expectation Quantiles

Using (1), \( F(r) \) can be expressed as

\[
F(r) := P(V(t) \leq r) = P(g(X(t)) \leq r) = P(E[G(X(t)|X(t))] \leq r),
\]

which is the distribution function of a conditional expectation. The obvious estimator is then

\[
\hat{F}(r) := \frac{1}{n} \sum_{i=1}^{n} \mathbb{I}(Z_i(t) \leq r).
\]
The convergence rate of (16) is presented in Lee [1998]; Lee and Glynn [2003], where central limit theorems are also established. As the main result of this paper concerns the optimal values of $n$ and $m$, only the cases with the optimal rate of convergence are stated here with an adaptation to the simulation-optimization algorithm context. For central limit theorems with a suboptimal rate of convergence, see Lee [1998]; Lee and Glynn [2003].

For central limit theorems in both the discrete and continuous feasible region cases, we require a large number of technical conditions, as stated in Appendix A and Appendix B, respectively. These appear to be unavoidable, as they arise in the proof of the central limit theorems derived in Lee [1998]; Lee and Glynn [2003]. A different set of conditions is provided in Gordy and Juneja [2010].

4.1.1. Discrete Feasible Region. Suppose that the range of the estimated optimal solution $X(t)$ is the discrete set $S = \{x_1, x_2, \ldots\}$.

Theorem 4.1 is a central limit theorem for $\hat{F}(r)$ in the case of a discrete feasible region, and is established in Lee and Glynn [2003, p. 12]. Here, $\eta^*$ is a positive constant relating to the large-deviations behavior of $G(x)$ that does not depend on $t$, and its definition can be found in Appendix A.

In the sequel we use the notation $g(T) \sim f(T)$ to mean that $\lim_{T \to \infty} g(T)/f(T) = 1$. For example, when we write $m \sim a \log T$, we mean that $m$ depends on $T$, and $m = m(T)$ is such that $m(T)/(a \log T) \to 1$ as $T \to \infty$.

**Theorem 4.1.** Assume the conditions A1-A6 as stated in Appendix A, and that $m \sim a \log T, n \sim T/m, as T \to \infty$, where $\alpha \geq 1/2\eta^*$. Then

$$\sqrt{\frac{T}{\log T}} \left[ \hat{F}(r) - F(r) \right] \Rightarrow \sqrt{aF(r)(1 - F(r))}N(0, 1)$$

as $T \to \infty$.

We will use the following specialized definition of the asymptotic mean squared error (AMSE). Let $W_n$ be a sequence of estimators of the parameter $\theta$, and $\{a_n\}$ a sequence of positive numbers satisfying $a_n \to \infty$ or $a_n \to a > 0$. Suppose that the central limit theorem holds in the sense of $a_n(W_n - \theta) \Rightarrow Z$ with $0 < \mathbb{E}Z^2 < \infty$, then the asymptotic mean squared error of $W_n$ with respect to $\theta$ is defined as

$$AMSE_{W_n} := \frac{\mathbb{E}Z^2}{a_n^2}.$$

Under a uniform integrability assumption, we obtain the following result for the AMSE of $\hat{F}(r)$.

**Corollary 4.2.** In addition to the conditions of Theorem 4.1, assume that the square of the left-hand side in Theorem 4.1 is uniformly integrable. The asymptotic mean squared error (AMSE) of $\hat{F}(r)$ with respect to $F(r)$ is

$$AMSE_{\hat{F}}(r) := AMSE(\hat{F}(r), F(r)) = aF(r)(1 - F(r)) \frac{\log T}{T}.$$

Like in Section 3, for any total computational budget $T$, the MSE can be minimized by choosing appropriate parameters in the expression of $n, m$. 

ACM Transactions on Modeling and Computer Simulation, Vol. 9, No. 4, Article 39, Publication date: March 2010.
THEOREM 4.3. Under the assumptions of Theorem 4.1 and Corollary 4.2, the optimal value of \( a \) that minimizes AMSE \( \hat{F}(r) \) is
\[
a = \frac{1}{2\eta^*},
\]
and the minimum AMSE of \( \hat{F}(r) \) is
\[
\text{AMSE} \hat{F}(r) = \frac{F(r)(1 - F(r)) \log T}{2\eta^* T}.
\]

Here we see that \( \hat{F}(r) \) has a sub-canonical convergence rate of \( O\left(\sqrt{\log T}/T\right) \) when the feasible region is discrete. In fact, none of the estimators discussed in this paper other than \( \hat{\mu} \) attains the canonical convergence rate \( O(T^{-1/2}) \).

4.1.2. Continuous Feasible Region. We now establish an analogous result when the estimated optimal solutions \( X(t) \) have a continuous support \( \Gamma \).

Define
\[
Y(x) := g(x) - r \sigma(x),
\]
for each possible \( x \in \Gamma \), where
\[
g(x) := \mathbb{E} G(x) \quad \text{and} \quad \sigma^2(x) := \text{Var} G(x).
\]
Let \( Y \) denote \( Y(X(t)) \), and assume that it has a density \( f_Y(\cdot) \). Denote by \( f_Y^{(k)}(\cdot) \) the \( k \)-th derivative of \( f_Y(\cdot) \).

The central limit theorem for \( \hat{F}(r) \) for continuous feasible region is established in Lee [1998, p. 56]. Here we only state the particular case where the optimal convergence rate is attained.

THEOREM 4.4. Assume the conditions B1-B7 in Appendix B hold. If \( T^{-1/3}m \to a > 0 \) and \( T^{-2/3}n \to b > 0 \), where \( 0 < a, b < \infty \), then as \( T \to \infty \),
\[
T^{1/3} \left[ \hat{F}(r) - F(r) \right] \Rightarrow \sqrt{\frac{F(r)(1 - F(r))}{b}} N(0, 1) + \frac{f_Y^{(1)}(0)}{2a}.
\]

Since we balance the squared bias and the variance with the given choice of scalings, they converge at the same rate of \( O(T^{-1/3}) \), and hence both show up in the asymptotic distribution.

Similar to the discrete feasible region case, the approximate formula for the asymptotic mean squared error follows immediately from the central limit theorem.

COROLLARY 4.5. In addition to the conditions of Theorem 4.4, assume that the square of the left-hand side in Theorem 4.4 is uniformly integrable. Then the AMSE of \( \hat{F}(r) \) with respect to \( F(r) \) is
\[
\text{AMSE} \hat{F}(r) = \left[ \frac{f_Y^{(1)}(0)^2}{4a^2} + \frac{F(r)(1 - F(r))}{b} \right] T^{-2/3}.
\]

The values of \( a, b \) can then be picked so that the AMSE is minimized.

COROLLARY 4.6. Under the conditions of Corollary 4.5, the optimal values of \( a \) and \( b \) that minimize AMSE \( \hat{F}(r) \) are
\[
a = 2^{-1/3} f_Y^{(1)}(0)^{2/3} F(r)^{-1/3}(1 - F(r))^{-1/3},
\]
\[
b = 2^{1/3} f_Y^{(1)}(0)^{-2/3} F(r)^{1/3}(1 - F(r))^{1/3},
\]
and the minimum AMSE of $\hat{F}(r)$ is

$$\text{AMSE} \hat{F}(r) = \frac{3}{2\sqrt{2}} f_Y^{(1)}(0)^{2/3} F(r)^{2/3}(1 - F(r))^{2/3} T^{-2/3}. $$

This result indicates how the optimal values of $n$ and $m$ depend on the real value of the performance measure $F(r)$. Specifically, if $F(r)$ is extreme, i.e., close to 0 or 1, then the variance of its estimator $\hat{F}(r)$ can be very small. Therefore, most of the computational budget should be allocated to reduce the bias. Thus $m$, the length of the inner-level simulation, increases as $F(r)$ approaches 0 or 1.

In the continuous feasible region case, the estimator $\hat{F}(r)$ converges at the rate of $O(T^{-1/3})$. It does not achieve the canonical Monte Carlo convergence rate $O(T^{-1/2})$, and is also slower than the convergence rate in the discrete feasible region case. As discussed in the introduction, Broadie et al. [2011] have developed an estimator with a faster convergence rate by sequentially allocating computational effort in the inner simulation with different values of $m$ for different $X(t)$. We do not pursue a sequential approach here.

4.2. The Multiple-Budget Case

Let $t_1 < t_2 < \cdots < t_l$ be the budgets under consideration, and $F_k(r) := P(V(t_k) \leq r), k = 1, \ldots, l$ be the performance measures that need to be estimated. The obvious estimator for each $F_k(r)$ is

$$\hat{F}_k(r) := \frac{1}{n_k} \sum_{i=1}^{n_k} 1(Z_i(t_k) \leq r). \quad (17)$$

Suppose that no partial sample paths are reused. It follows that the optimal values of $n_k$ and $m_k$ can be chosen independently for each $k = 1, \ldots, l$. Hence the minimum AMSE $\text{AMSE} \hat{F}_k(r)$ with respect to $F_k(r)$ for each $k = 1, \ldots, l$ is given by Corollary 4.2 or Corollary 4.5, namely

$$\text{AMSE} \hat{F}_k(r) := \text{AMSE}(\hat{F}_k(r), F_k(r)) = \frac{F_k(r)(1 - F_k(r)) \log T_k}{2n_k T_k} $$

if the feasible region of the simulation-optimization problem is discrete, or

$$\text{AMSE} \hat{F}_k(r) = \frac{3}{2\sqrt{2}} f_Y^{(1)}(0)^{2/3} F_k(r)^{2/3}(1 - F_k(r))^{2/3} T^{-2/3} $$

if the feasible region is continuous.

**Theorem 4.7.** Suppose the simulation-optimization problem has discrete feasible region and the conditions in Corollary 4.2 are satisfied for each computational budget $t_k$. Then a budget allocation achieves the optimal convergence rate if and only if

$$\frac{T_k}{T} \rightarrow \tau_k^*,$$

where

$$\tau_k^* = \frac{\sqrt{F_k(r)(1 - F_k(r))/2\eta_k^2}}{\sum_{h=1}^{l} \sqrt{F_h(r)(1 - F_h(r))/2\eta_h^2}}.$$
and the minimum sum of the AMSE $\hat{F}_k(r)$ is
\[
\sum_{k=1}^{l} \text{AMSE} \hat{F}_k(r) = \left[ \sum_{k=1}^{l} \sqrt{\frac{F_k(r)(1 - F_k(r))}{2\eta_k^*}} \right]^2 \frac{\log T}{T}.
\]

**Proof Sketch.** We first show that the sum of AMSE $\hat{F}_k(r)$ can be bounded to the order of $O((\log T)/T)$. It is then left to argue that no better rate of convergence is possible.

It is a basic fact in real analysis that $T_k/T$ has a subsequence converging to some $\tau_k$. The optimal values $\tau_k^*$ can be determined via the inequality of arithmetic and geometric means, and then it follows that the entire sequence $T_k/T$ has to converge to $\tau_k$. \qed

The form of $T_k$ here is very similar to that in Theorem 3.5. Indeed, both are proportional to the total computational budget $T$, and the proportionality constant depends only on the parameters at the budget $t_k$, scaled by the total. This happens because the MSE at each budget $t_k$ depends on the others only through the total budget constraint.

As mentioned in Section 3, the optimal values of $n$ and $m$ do not have closed-form expressions if the sum of MSEP is used as the performance measure in the continuous feasible region case, but the maximum of the MSEP is more tractable. We now give the result when minimizing the maximum of the MSEP. The resulting optimal values have the same convergence rate as would be obtained if we were minimizing the sum of the MSEP, and the optimal values of $n$ and $m$ for the two performance measures are equivalent in the sense that they differ only by a constant factor.

**Theorem 4.8.** Suppose the simulation-optimization problem has continuous feasible region and the conditions in Corollary 4.5 are satisfied for each computational budget $t_k$. Then the smallest possible maximum of AMSE $F_k(r)$ is
\[
\max_{k=1,\ldots,l} \text{AMSE} \hat{F}_k(r) = \frac{3}{2\sqrt{2}} \left[ \sum_{h=1}^{l} f_{Y_h}(0)F_h(r)(1 - F_h(r)) \right]^{2/3} \frac{T^{-2/3}}{T},
\]
and is achieved when
\[
\frac{T_k}{T} \to \tau_k^*;
\]
where
\[
\tau_k^* = \frac{f_{Y_h}(0)F_h(r)(1 - F_h(r))}{\sum_{h=1}^{l} f_{Y_h}(0)F_h(r)(1 - F_h(r))}
\]
for each $k = 1,\ldots,l$.

**Proof Sketch.** The argument in the proof of Theorem 4.7 applies here as well, so that $T_k/T$ converges to some $\tau_k$ for each $k$. The optimal values of $\tau_k$ are obtained using the same idea as in Theorem 3.7. \qed

As discussed in Section 3, here we also see that the optimal convergence rate is attained if and only if
\[
\frac{T_k}{T} \to \tau_k
\]
for each $k = 1,\ldots,l$, where each $\tau_k$ is a nonzero constant. The convergence rate is unaffected whether we minimize the sum or maximum of AMSEs. We chose to minimize the sum of AMSEs in the discrete feasible region case, and the maximum of the AMSE's
in the continuous feasible region case, simply because the optimization problems are mathematically tractable and explicit expressions can be given for the optimal solutions.

5. QUANTILES

In this section, we consider the \( p \)-th quantile of \( g(X(t)) \), the true objective function value of the estimated optimal solution \( X(t) \). We adopt the definition of the \( q \)-th quantile as in Lee [1998]:

\[
q = q(p) := \sup\{r : F(r) < p\}.
\]

The obvious estimator of \( q \), and the one we adopt, is

\[
Q = Q(p) := \sup\{r : \hat{F}(r) < p\}.
\]

5.1. The Single-Budget Case

Suppose the range of the estimated optimal solution \( X(t) \) is continuous. Then the asymptotic behavior of \( Q(p) \) has been discussed in Lee [1998]. The result and a set of sufficient conditions are quoted below, with adaptation to the simulation-optimization algorithm performance measure context.

Given the probability \( p \), the corresponding benchmark \( r \) is simply \( q(p) \). Thus we redefine the quantity \( Y(x) \) defined in Section 4 to

\[
Y(x) := \frac{g(x) - q}{\sigma(x)}.
\]

Recall that \( f_Y(\cdot) \) is the density of \( Y := Y(X(t)) \).

**Theorem 5.1** (Lee [1998]). Let \( 0 < p < 1 \). Assume that the conditions C1-C7 in Appendix C hold, that \( F(\cdot) \) is continuously differentiable at \( q \) with \( F'(q) > 0 \), and \( f_Y^{(1)}(\cdot) \) is continuous at the point 0. If \( T^{-1/3}m \to a \) and \( T^{-2/3}n \to b \), where \( 0 < a, b < \infty \), then as \( T \to \infty \),

\[
T^{1/3} [Q(p) - q(p)] \to \sqrt{\frac{p(1-p)}{b}} \frac{1}{F'(q)} \mathcal{N}(0, 1) - \frac{1}{2a} f_Y^{(1)}(0) \frac{1}{F'(q)}.
\]

The asymptotic distribution of \( Q(p) \) is the same as that of \( \hat{F}(r) \), except for the factor of \( F'(q) \) and the flipping of the sign of the bias term. As a result, its AMSE is similar to that of \( \hat{F}(r) \) too.

**Corollary 5.2.** In addition to the conditions of Theorem 5.1, if the square of the left-hand side in Theorem 5.1 is uniformly integrable, then the AMSE of \( Q(p) \) with respect to \( q(p) \) is

\[
\text{AMSE}\{Q(p)\} := \text{AMSE}(Q(p), q(p)) = \left[ \frac{f_Y^{(1)}(0)^2}{4a^2} + \frac{p(1-p)}{b} \right] T^{-2/3} F'(q)^2.
\]

The AMSE here is a scaled version of that for the distribution function in Section 4, so we immediately have the following analog of Corollary 4.6.

**Corollary 5.3.** Assume the conditions in Theorem 5.1, as well as the uniform integrability in Corollary 5.2. The optimal values of \( a \) and \( b \) that minimize \( \text{AMSE}\{Q(p)\} \) are

\[
a = 2^{-1/3} f_Y^{(1)}(0)^{2/3} p^{-1/3} (1-p)^{-1/3},
\]

\[
b = 2^{1/3} f_Y^{(1)}(0)^{-2/3} p^{1/3} (1-p)^{1/3}.
\]
and the minimum AMSE of $Q(p)$ is
\[
\text{AMSE}_Q(p) = \frac{3}{2\sqrt{2}} f_Y(0)^{2/3} p^{2/3} (1 - p)^{2/3} F'(q)^{-2} T^{-2/3}.
\]

The estimator $Q(p)$ has a convergence rate of $O(T^{-1/3})$, which is also sub-canonical. In particular, it converges at the same speed as its counterpart $\hat{F}(r)$.

### 5.2. The Multiple-Budget Case

Let $t_1 < t_2 < \cdots < t_l$ be the budgets under consideration, and $q_k(p) := \sup\{r : F_k(r) < p\}, k = 1, \ldots, l$ be the performance measures that need to be estimated. The obvious estimator for each $q_k(p)$ is
\[
Q_k(p) := \sup\{r : \hat{F}_k(r) < p\},
\]
and the minimum AMSE of each $Q_k(p)$ with respect to $q_k(p)$ is given by Corollary 5.3:
\[
\text{AMSE}_{Q_k(p)} := \text{AMSE}(Q_k(p), q_k(p)) = \frac{3}{2\sqrt{2}} f_Y(0)^{2/3} p^{2/3} (1 - p)^{2/3} F'(q_k)^{-2} T^{-2/3}.
\]

**Theorem 5.4.** Suppose the simulation-optimization problem has continuous feasible region and the conditions in Corollary 5.3 are satisfied for each computational budget $t_k$. Then the smallest possible maximum of $\text{AMSE}_{Q_k(p)}$ is
\[
\max_{k=1,\ldots,l} \text{AMSE}_{Q_k(p)} = \frac{3}{2\sqrt{2}} \left[ \sum_{h=1}^l f_Y(0)^{2/3} p^{2/3} (1 - p)^{2/3} F'(q_h)^{-3} \right] T^{-2/3},
\]
and is achieved when
\[
\frac{T_k}{T} \to \tau^*_k,
\]
where
\[
\tau^*_k = \frac{f_Y(0)^{2} F'(q_k)^{-3}}{\sum_{h=1}^l f_Y(0)^{2} F'(q_h)^{-3}}.
\]

The form of $\tau^*_k$ follows from the same reasoning as in Theorem 4.7, because of the similarity of the normal distribution in Theorem 5.1 to that in Theorem 4.4.

**Proof Sketch.** Simply replace the result of Corollary 4.5 by that of Corollary 5.2, and apply the same argument used in the proof of Theorem 4.8.

### 6. Multidimensional Central Limit Theorems

In the scenario where the sample paths are reused, the optimal solutions $X(t_k)$ with different computational budgets are obtained from shared sample paths, and exhibit some correlation. Therefore, besides the AMSE of the estimators with each budget $t_k$, we are also interested in the joint distribution of the estimators for the various budgets, since they are not independent.

The central limit theorems in Lee [1998]; Lee and Glynn [2003] provide conditions under which the estimators in the single-budget case are asymptotically normally distributed. In this section, we generalize this result, and show that the vector of all estimators with budgets $t_k, k = 1, \ldots, l$ is asymptotically jointly normally distributed.

These multidimensional central limit theorems not only give the asymptotic distribution of each estimator with budget $t_k$ as its single-budget counterparts do, but more
importantly, they also reveal the dependence across different budgets when \( T \) is large. When the computational budget \( T \) is a fixed large quantity, the univariate central limit theorems can be used to compute the marginal accuracy of each estimator, but the multivariate central limit theorems help by quantifying the jointly distributed error of the estimators.

6.1. Preliminaries
In this section we list some definitions and theorems that will be used in the proof of the central limit theorems.

Recall that the Hadamard product of two \( n \times m \) matrices \( A = (a_{ij}) \) and \( B = (b_{ij}) \) is the element-wise product, i.e., the \( n \times m \) matrix with \((i, j)\)-th entry \( a_{ij}b_{ij} \).

**Lemma 6.1** (Slutsky’s Theorem for Random Vectors). Let \( \{X_n\}, \{Y_n\} \) be sequences of random vectors in \( \mathbb{R}^m \). If
\[
X_n \Rightarrow X, \quad Y_n \Rightarrow c,
\]
then
\[
X_n + Y_n \Rightarrow X + c, \\
X_n \circ Y_n \Rightarrow c \circ X.
\]

The proof of Slutsky’s Theorem can be found in various references, e.g., [Billingsley 1986], and Lemma 6.1 is a straightforward generalization to multiple dimension.

Before stating the multivariate central limit theorems, we need to study the covariance between the estimators associated with the estimated optimal solutions at each budget \( t_k \). The following result shows that one can replace \( Z_i(t_k) \) in the covariance calculations by \( V(t_k) \) asymptotically, i.e., the effect on the variance of inner-stage sampling vanishes.

**Lemma 6.2.** Let \( q_k, k = 1, 2, \ldots, l \) be given and suppose that \( P(V(t_k) = q_k) = 0 \) for all \( k = 1, 2, \ldots, l \). Suppose further that, for each \( k = 1, \ldots, l \), \( m_k = m_k(T) \to \infty \) as \( T_k \to \infty \). Then for any pair \( k, h = 1, \ldots, l \), and for any sequence of constants \( \{s_k(T)\} \) and \( \{s_h(T)\} \) with \( s_k(T) \to 0 \) and \( s_h(T) \to 0 \) as \( T \to \infty \), we have
\[
\operatorname{Cov} (\mathbb{I}(Z_i(t_k) \leq q_k + s_k(T)), \mathbb{I}(Z_i(t_h) \leq q_h + s_h(T))) 
\xrightarrow{T \to \infty} \operatorname{Cov} (\mathbb{I}(V(t_k) \leq q_k), \mathbb{I}(V(t_h) \leq q_h)).
\]

Lemma 6.2 is used in the proof of the central limit theorem for the quantiles in the multiple-budget case. The corresponding result for the distribution function is simpler.

**Corollary 6.3.** Let \( r_k, k = 1, 2, \ldots, l \) be given and suppose that \( P(V(t_k) = r_k) = 0 \) for all \( k = 1, 2, \ldots, l \). Suppose that, for each \( k = 1, \ldots, l \), \( m_k \to \infty \) as \( T \to \infty \). For any pair \( k, h = 1, 2, \ldots, l \), we have
\[
\operatorname{Cov} (\mathbb{I}(Z_i(t_k) \leq r_k), \mathbb{I}(Z_i(t_h) \leq r_h)) \xrightarrow{T \to \infty} \operatorname{Cov} (\mathbb{I}(V(t_k) \leq r_k), \mathbb{I}(V(t_h) \leq r_h)).
\]

**Proof.** Set \( q_k = r_k \) and \( s_k(T) = 0 \) in Lemma 6.2, and the result follows. \( \square \)

To simplify the notation for the covariance matrix, we adopt the following definition.

**Definition 6.4.** Let \( \Lambda_T \) be the covariance matrix of \( \mathbb{I}(Z_i(t_k) \leq r_k), k = 1, \ldots, l \), and \( \Lambda \) the covariance matrix of \( \mathbb{I}(V(t_k) \leq r_k), k = 1, \ldots, l \), namely
\[
\Lambda_{Tkh} := \operatorname{Cov} (\mathbb{I}(Z_i(t_k) \leq r_k), \mathbb{I}(Z_i(t_h) \leq r_h)),
\Lambda_{kh} := \operatorname{Cov} (\mathbb{I}(V(t_k) \leq r_k), \mathbb{I}(V(t_h) \leq r_h)).
\]
Corollary 6.3 above basically states that $\Lambda_T \rightarrow \Lambda$ entry-wise as the inner sample size increases.

**Definition 6.5.** For simplicity of notation, we use $v^r$ to denote the vector whose entries are those of $v$ raised to the power $r$, i.e., $(v^r)_i = v_i^r$. This is equivalent to defining $v^r$ so that $\text{Diag}(v^r) = (\text{Diag}(v))^r$, where $\text{Diag}(v)$ is the diagonal matrix with the elements of $v$ on the diagonal.

### 6.2. Distribution Function

In this section, we extend Theorem 4.1 and Theorem 4.4 to multivariate central limit theorems. Since the true objective function values $g(X(t_k))$ at the estimated optimal solutions $X(t_k)$ with different computational budgets $t_k$ are estimated independently, the simulation run lengths $n_k, k = 1, \ldots, l$ may be different from each other.

As before, we discuss separately the case with discrete feasible region and the case with continuous feasible region. In each case, the central limit theorem is stated with the assumption that $n_k$ is the same for each $k = 1, \ldots, l$, which is relaxed later to obtain a more general result.

**Theorem 6.6.** Assume A1-A6 for each $t_k, k = 1, \ldots, l$. Suppose $n_k = n$ for each $k = 1, \ldots, l$. Then, if $m_k \sim a_k \log T$ for each $k = 1, \ldots, l$ and $n \sim b T/ \log T > 0$ as $T \rightarrow \infty$ where $a_k \geq 1/2 \eta_k^*$ for each $k = 1, \ldots, l$, then

$$\sqrt{\frac{T}{\log T}} \left[ \hat{F}_k(r_k) - F_k(r_k) \right]_{k} \Rightarrow b^{-1/2} N(0, \Lambda),$$

where $[v(k)]_k$ represents a vector with $k$-th entry being $v(k)$.

**Proof Sketch.** Using the Cramér-Wold Theorem [Billingsley 1986], we reduce the multiple-budget version to a univariate central limit theorem concerning an arbitrary linear combination of the estimators $\hat{F}(r_k), k = 1, \ldots, l$. The single-budget central limit theorem, namely Theorem 4.1, has been proved in Lee and Glynn [2003]. The proof centralizes the random variable and calculates separately the limit of the standardized random variable and its bias. By adopting the same technique, we isolate the bias of the linear combination, and show that the standardized linear combination also converges to a normal random variable using Lemma D.1, the same lemma used in the proof of the single-budget central limit theorem. □

It is natural to further extend the above theorem and allow $n_k, k = 1, \ldots, l$ to differ, as the next theorem states. In fact, Theorem 6.6 is completely captured by Theorem 6.7, yet they are both fully stated here, since Theorem 6.6 serves as an important intermediate step in the proof of Theorem 6.7.

**Theorem 6.7.** Assume A1-A6 for each $t_k, k = 1, \ldots, l$. Suppose $m_k \rightarrow \infty$ and $n_k \rightarrow \infty$ for each $k = 1, \ldots, l$. Also assume that the $n_k$’s are in decreasing order. Then, if $m_k \sim a_k \log T$ and $n_k \sim b_k T/ \log T$ for each $k = 1, \ldots, l$ as $T \rightarrow \infty$ where $a_k \geq 1/2 \eta_k^*$ for each $k = 1, \ldots, l$, then

$$\sqrt{\frac{T}{\log T}} \left[ \hat{F}_k(r_k) - F_k(r_k) \right]_{k} \Rightarrow N(0, B_{\text{min}} \circ \Lambda),$$

where $B_{\text{min}}$ is an $l$-dimensional square matrix with entries $B_{\text{min}}(k, h) := b_{k \wedge h}^{-1}$, and $k \wedge h = \min\{k, h\}$.

**Proof Sketch.** In contrast to the assumptions in Theorem 6.6, the $n_k$’s here are no longer assumed to be equal to each other, and so grouping the summands into i.i.d.
linear combinations does not work directly. However, in each estimator \( \hat{F}_k(r_k) \), the \( n_k \) summands are independent, and can be divided into \( k \) groups, each of which consists of identically distributed random variables. Hence the central limit theorem can be shown by breaking down the random vector into \( l \) independent parts, and proving the theorem for each part using the same argument in Theorem 6.6. □

For the case with continuous feasible region, we proceed in the same way as above, and obtain the following results.

**Theorem 6.8.** Assume B1-B7 for each \( t_k, k = 1, \ldots, l \). If \( T^{-1/3}m_k \to a_k > 0 \) and \( n_k = n \) for each \( k = 1, \ldots, l \) with \( T^{-2/3}n \to b \), where \( 0 < a_k, b < \infty \), then as \( T \to \infty \),

\[
T^{1/3} \left[ \hat{F}_k(r_k) - F_k(r_k) \right]_k \Rightarrow \mathcal{N}(0, b^{-1}A) + \left[ \frac{f_{Y_k}(0)}{2a_k} \right]_k.
\]

**Proof Sketch.** The proof is almost the same as that of Theorem 6.6, except that the bias is no longer zero. Since the bias of each entry is unrelated to the rest, their values are directly given by the single-budget central limit theorem, Theorem 4.4. □

**Theorem 6.9.** Assume B1-B7 for each \( t_k, k = 1, \ldots, l \). Assume further that \( T^{-1/3}m_k \to a_k > 0 \) and \( T^{-2/3}n_k \to b_k > 0 \) for each \( k = 1, \ldots, l \). If the \( n_k \)'s are in decreasing order, then as \( T \to \infty \),

\[
T^{1/3} \left[ \hat{F}_k(r_k) - F_k(r_k) \right]_k \Rightarrow \mathcal{N}(0, B_{\text{min}} \circ \Lambda) + \left[ \frac{f_{Y_k}(0)}{2a_k} \right]_k.
\]

**Proof Sketch.** Apply the same technique in Theorem 6.7 to generalize Theorem 6.8. □

### 6.3. Quantiles

The central limit theorem for quantiles as stated in Section 5 can also be generalized in the same fashion. Again, the case with \( n_k \) being the same for all \( k = 1, \ldots, l \) is shown first, and serves as an intermediate result for the more general central limit theorem.

**Theorem 6.10.** Assume that the conditions in Theorem 5.1 hold for each \( t_k, k = 1, \ldots, l \). If \( T^{-1/3}m_k \to a_k \) and \( n_k = n \) for each \( k = 1, \ldots, l \) with \( T^{-2/3}n \to b \), where \( 0 < a_k, b < \infty \), then as \( T \to \infty \),

\[
T^{1/3} \left[ Q_k(p_k) - q_k(p_k) \right]_k \Rightarrow \left[ \frac{1}{F'_k(q_k(p_k))} \right]_k \circ \mathcal{N}(0, b^{-1}A) - \left[ \frac{f_{Y_k}(0)}{2a_k} \right]_k.
\]

**Proof Sketch.** The Cramér-Wold Theorem [Billingsley 1986] cannot be applied directly. We study the joint distribution function of \( Q_k(p_k) \), \( k = 1, \ldots, l \), and reduce it to a weighted sum of Bernoulli random variables \( \chi_{ki}, k = 1, \ldots, l, i = 1, \ldots, n \).

The random variables \( \chi_{ki}, k = 1, \ldots, l, i = 1, \ldots, n \) are additive, so we can then apply the Cramér-Wold Theorem. The proof then follows that of the central limit theorem for quantiles in Lee [1998], and the resulting weak convergence in terms of \( \chi_{ki}, k = 1, \ldots, l, i = 1, \ldots, n \) can be "lifted" back to that of \( Q_k(p_k), k = 1, \ldots, l \). □

**Theorem 6.11.** Assume that the conditions in Theorem 5.1 hold for each \( t_k, k = 1, \ldots, l \). Suppose further that \( T^{-1/3}m_k \to a_k \) and \( T^{-2/3}n_k \to b_k \) for each \( k = 1, \ldots, l \), where \( 0 < a_k, b_k < \infty \). If the \( n_k \)'s are in decreasing order, then as \( T \to \infty \),

\[
T^{1/3} \left[ Q_k(p_k) - q_k(p_k) \right]_k \Rightarrow \left[ \frac{1}{F'_k(q_k(p_k))} \right]_k \circ \mathcal{N}(0, B_{\text{min}} \circ \Lambda) - \left[ \frac{f_{Y_k}(0)}{2a_k} \right]_k.
\]
7. SUMMARY
We have proposed performance measures for simulation-optimization algorithms, including the mean, distribution function, and quantiles of $V(t)$, which denotes the true objective function value associated with the estimated optimal solution. Estimators for each of the performance measures have been analyzed, and we have shown that the estimator of the mean converges at the canonical Monte Carlo convergence rate. Estimators of the other performance measures converge at sub-canonical Monte Carlo convergence rates.

When the computational budget $t$ is allowed to vary and multiple budgets are considered, we have obtained the optimal budget allocation schemes for each estimator. In particular, we have shown that the optimal convergence rate is achieved as long as the computational effort spent on each $t_k$ is maintained as a constant fraction of the total computational budget $T$, and the convergence rate is unaffected whether one minimizes the sum or maximum of the AMSEs.

The expressions we have derived for the optimal simulation run lengths $n$ and $m$ are important because they confirm our intuition in some cases and show the tradeoffs involved between the variance and bias of the estimators, exactly as in, for example, the expressions for the optimal step length in finite-difference estimators of sensitivities; see Glasserman [2003]. Some of the constants appearing in these expressions are unlikely to be estimable in practice, but they provide insight into what features of the problem influence $n$ and $m$.

Finally, multivariate central limit theorems for each performance measure are established by generalizing results in Lee [1998], and indicate that a vector of estimators for different budgets $t_k, k = 1, \ldots$, converges in distribution to a multivariate normal random variable.

A. CONDITIONS FOR THEOREM 4.1
Define
\[
\Gamma_+ = \{ x_i : E G(x_i) > r, i \geq 1 \},
\]
\[
\Gamma_- = \{ x_i : E G(x_i) \leq r, i \geq 1 \}.
\]

One of the fundamental results in large deviations theory [Bucklew 1990] implies that for $X(t) \in \Gamma_-$, $P(Z_i(t) > r)$ typically converges to 0 exponentially fast; whereas for $X(t) \in \Gamma_+$, $P(Z_i(t) \leq r)$ typically converges to 0 exponentially fast.

We also use stronger assumptions than are necessary for the central limit theorem, for simplicity. For all $i \geq 1$ and all $\theta \in \mathbb{R}$, suppose
\[
\varphi_i(\theta) := E[exp(\theta G(x_i))] < \infty.
\]

The constant $\theta^*_i$ is defined as the root of the equation
\[
\varphi_i'(\theta^*_i)/\varphi_i(\theta^*_i) = r,
\]
and $\eta(x_i)$ is given by
\[
\eta(x_i) := \theta^*_i r - \log \varphi_i(\theta^*_i) = \sup_{\theta} \{ \theta r - \log \varphi_i(\theta) \}.
\]

Let $\eta^*$ be the slowest decay rate:
\[
\eta^* := \inf \{ \eta(x_i) : i \geq 1 \}.
\]

Assume the following conditions for Theorem 4.1:

A1. For $i \geq 1$ and $\theta \in \mathbb{R}$, $\varphi_i(\theta) < \infty$;
A2. $P(G(x_i) > r) > 0, x_i \in \Gamma_-, P(G(x_i) \leq r) > 0, x_i \in \Gamma_+;
A3. \( E(G(x_i) \neq r \) for all \( i \geq 1 \);
A4. For all \( i \geq 1 \), \( G(x_i) \) is a continuous random variable;
A5. \( B^* = \{ x_i : i \geq 1, \eta(x_i) = \eta^* \} \) is nonempty and finite;
A6. \( \inf \{ \eta(x_i) : i \geq 1, x_i \notin B^* \} > \eta^* \).

**B. CONDITIONS FOR THEOREM 4.4**

Let \( \mathcal{C}\beta \) be the class of \( k \)-times continuously differentiable functions whose derivatives of order less than or equal to \( k \) are bounded.

Define

\[
\beta_k(x) := E \frac{|G(x) - \mu(x)|^k}{\sigma(x)^k}
\]

for each \( k = 1, 2, \ldots, \) and

\[
\bar{\beta}_k(y) := E [\beta_k(X(t))] Y(X(t)) = y .
\]

Define \( \gamma_p(x) \) as the \( p \)-th cumulant of \( G(x) \). In particular, if

\[
c(s) := \log E \left[ e^{sG(x)} \right]
\]

is the cumulant-generating function, then \( \gamma_p(x) \) is the series that satisfies

\[
c(s) = \sum_{p=1}^{\infty} \gamma_p(x) \frac{s^p}{p!}
\]

Given \( \nu \geq 1 \), define

\( -\Sigma(\nu) := \{ s : k_1, \ldots, k_\nu \in \mathbb{N}_0 \text{ solving } k_1 + 2k_2 + \cdots + \nu k_\nu = \nu, s = k_1 + \cdots + k_\nu \} \)

\( -\kappa(\nu, s) := \{ k_1, \ldots, k_\nu \in \mathbb{N}_0 : k_1 + 2k_2 + \cdots + \nu k_\nu, s = k_1 + \cdots + k_\nu \} \text{ for each } s \in \Sigma(\nu) \)

\( -\chi_{\nu, s}(x) := \sum_{(k_1, \ldots, k_\nu) \in \kappa(\nu, s)} \prod_{p=1}^{\nu} \frac{1}{k_p!} \left[ \frac{\gamma_{p+2}(x)}{(p+2)\sigma^{p+2}(x)} \right]^{k_p} \text{ for all } z \in \Gamma \) and \( s \in \Sigma(\nu) \).

For notational purpose, for any two real-valued function \( f(\cdot) \) and \( g(\cdot) \), denote \( (f \cdot g)(x) := f(x) \cdot g(x) \). Assume for Theorem 4.4 that for some \( k \geq 1 \):

**B1.** \( \|G(X(t))\|^{2k+2} < \infty ; \)

**B2.** there exists a density \( \tilde{g}(\cdot) \) and \( \epsilon > 0 \) such that \( P(G(X(t)) \in \cdot | X(t)) \geq \epsilon \tilde{g}(\cdot) \) almost surely;

**B3.** \( \beta_{2k+2}(\cdot) \in \mathcal{C}\beta_1 ; \)

**B4.** \( f_Y(\cdot) \in \mathcal{C}\beta_k ; \)

**B5.** \( (f_Y \cdot \chi_{1,1})(\cdot) \in \mathcal{C}\beta_{2k-1} ; \)

Assume further that for all \( s = k_1 + \cdots + k_\nu \), where \( (k_1, \ldots, k_\nu) \) is any non-negative integer solution of the equation \( k_1 + 2k_2 + \cdots + \nu k_\nu = \nu , \)

**B6.** \( (f_Y \cdot \chi_{2\nu, s})(\cdot) \in \mathcal{C}\beta_{2(k-\nu)} \) for \( s \in \Sigma(2\nu) , 1 \leq \nu \leq k-1 ; \)

**B7.** \( (f_Y \cdot \chi_{2\nu+1, s})(\cdot) \in \mathcal{C}\beta_{2(k-\nu)-1} \) for \( s \in \Sigma(2\nu+1) , 1 \leq \nu \leq k-1 . \)

**C. CONDITIONS FOR THEOREM 5.1**

Given the probability \( p \), the corresponding value of \( r \) is \( q(p) \). Redefine

\[
Y(x) := \frac{g(x) - q}{\sigma(x)} .
\]

Assume the following conditions for Theorem 5.1:

ACM Transactions on Modeling and Computer Simulation, Vol. 9, No. 4, Article 39, Publication date: March 2010.
C1. $E |G(X(t))|^4 < \infty$;
C2. there exists a density function $\tilde{g}(\cdot)$ such that $P(G(X(t)) \in \cdot | X(t)) \geq c \tilde{g}(\cdot) \ a.s.$;
C3. $E|\beta_\epsilon(X(t))|Y(X(t)) = \cdot \in C_0^\epsilon$ for $|\cdot - q| < \epsilon$;
C4. $f_\epsilon(\cdot) \in C_0^\epsilon$ for $|\cdot - 1| < \epsilon$;
C5. for $|\cdot - q| < \epsilon$, $(f_\epsilon \cdot \chi_{1,1})(\cdot) \in C_0^\epsilon$.

Let $h(\cdot, \cdot)$ denote the joint density function of $(\mu(X(t)), \sigma(X(t)))$, which is assumed to exist. Assume further that, for a fixed $\epsilon > 0$,

C6. $\sup_{|x-q|<\epsilon} \sup_{\xi \in \mathbb{R}} \int_0^\infty z_2^2 h^2(x + \xi z_2, z_2) \, dz_2 < \infty$;
C7. $\sup_{|x-q|<\epsilon} \sup_{\xi \in \mathbb{R}} \int_0^\infty z_2^4 \left[ \frac{\partial}{\partial z_1} h(x + \xi z_2, z_2) \right]^2 \, dz_2 < \infty$.

**D. LEMMA FOR THE PROOF OF MULTIVARIATE CENTRAL LIMIT THEOREMS**

**LEMMA D.1.** Assume that the following conditions hold:

(a) For each $c > 0$, the sequence $(X_{c,j} : j \geq 1)$ consists of independent and identically distributed random variables;
(b) $E X_{c,1} = 0$, Var $X_{c,1} := \sigma_{c,j}^2$;
(c) $\lim_{c \to \infty} \sigma_{c,j}^2 = \sigma^2 \in (0, \infty)$;
(d) the family $\{X_{c,j}^2 : c > 0\}$ is uniformly integrable.

If $n(c) \to \infty$ as $c \to \infty$, then $\{X_{c,j} : 1 \leq j \leq n(c), c > 0\}$ satisfies the Lindeberg-Feller condition, namely for each $\epsilon > 0$,

$$\lim_{c \to \infty} \frac{1}{\text{Var} S_c} \sum_{j=1}^{n(c)} E \left[ X_{c,j}^2 | (X_{c,j}^2 > \epsilon^2 \text{Var} S_c) \right] = 0$$

where $S_c = \sum_{j=1}^{n(c)} X_{c,j}$.

**ELECTRONIC APPENDIX**

The electronic appendix for this article can be accessed in the ACM Digital Library.

**ACKNOWLEDGMENTS**

This work was partially supported by National Science Foundation grants CMMI-0800688 and CMMI-1200315. We thank the editorial team for comments that improved the paper.

**REFERENCES**


PROMODEL. 2011. SimRunner.


Received February 2007; revised March 2009; accepted June 2009
A. PROOFS

PROOF OF PROPOSITION 3.2. Recall that \( \rho(t) = \frac{\text{Var} g(X(t))}{\text{Var} G(X(t))} \geq 0 \). Let \( \mu = \mathbb{E} V(t) = \mathbb{E} G(X(t)) \). Since \( \hat{\mu} \) is an unbiased estimator for \( \mu \), the MSE of \( \hat{\mu} \) is its variance, which is

\[
\text{Var} \hat{\mu} = \text{Var} \left( \frac{1}{n} \sum_{i=1}^{n} Z_i(t) \right) = \frac{1}{n} \text{Var} \left( \frac{1}{m} \sum_{j=1}^{m} G_j(X(t)) \right)
\]

\[
= \frac{1}{nm^2} \left[ m \text{Var} G(X(t)) + m(m-1) \text{Cov}(G_1(X(t)), G_2(X(t))) \right]
\]

\[
= \frac{1 + (m-1)\rho(t)}{nm} \text{Var} G(X(t)).
\]

PROOF OF THEOREM 3.3. By Proposition 3.2 and the budget constraint (3), the optimization problem can be formulated as

\[
\begin{align*}
\min & \quad \frac{1 + (m-1)\rho(t)}{nm} \text{Var} G(X(t)) \\
\text{s.t.} & \quad n(t + m) = T,
\end{align*}
\]

where the integer constraints and bounds on \( n, m \) are relaxed. Its Lagrangian function is

\[
\mathcal{L}(n, m; \lambda) = \frac{1 + (m-1)\rho(t)}{nm} \text{Var} G(X(t)) - \lambda [n(t + m) - T].
\]

By the Karush-Kuhn-Tucker conditions, e.g., [Nocedal and Wright 2006], at an optimal point the partial derivatives satisfy

\[
\frac{\partial}{\partial n} \mathcal{L}(n, m; \lambda) = -\frac{1 + (m-1)\rho(t)}{n^2m} \text{Var} G(X(t)) - \lambda = 0,
\]

\[
\frac{\partial}{\partial m} \mathcal{L}(n, m; \lambda) = -\frac{1 - \rho(t)}{nm^2} \text{Var} G(X(t)) - \lambda n = 0.
\]

The above system of equations leads to the solution

\[
m = \sqrt{\left( \frac{1}{\rho(t)} - 1 \right)t},
\]

and the budget constraint (3) gives

\[
n = \frac{T}{t + m} = \frac{T}{\sqrt{t \left[ \sqrt{\frac{1}{\rho(t)} - 1} + \sqrt{t} \right]}}.
\]

Direct substitution of \( n, m \) into (3.2) gives the value of the minimum MSE. \( \square \)
Proof of Theorem 3.5. Conditioning on the value of each $T_k$, the minimum MSE of $\hat{\mu}_k$ is given by (9), and therefore

$$\sum_{k=1}^{l} \text{MSE} \hat{\mu}_k = \sum_{k=1}^{l} \left[ \sqrt{1-\rho(t_k)} + \sqrt{\rho(t_k)} t_k \right]^2 \frac{\text{Var} G(X(t_k))}{T_k}.$$

Multiply both sides by the total budget constraint (4), and then apply the Cauchy-Schwarz inequality,

$$T \sum_{k=1}^{l} \text{MSE} \hat{\mu}_k = \sum_{k=1}^{l} T_k \sum_{k=1}^{l} \left[ \sqrt{1-\rho(t_k)} + \sqrt{\rho(t_k)} t_k \right]^2 \frac{\text{Var} G(X(t_k))}{T_k} \geq \left( \sum_{k=1}^{l} \left[ \sqrt{1-\rho(t_k)} + \sqrt{\rho(t_k)} t_k \right] \sqrt{\text{Var} G(X(t_k))} \right)^2,$$

with equality if and only if

$$T_k = C \left[ \sqrt{1-\rho(t_k)} + \sqrt{\rho(t_k)} t_k \right]^2 \frac{\text{Var} G(X(t_k))}{T_k},$$

for each $k = 1, \ldots, l$, for a constant $C$ independent of $k$, which leads to

$$T_k = \frac{\left[ \sqrt{1-\rho(t_k)} + \sqrt{\rho(t_k)} t_k \right] \sqrt{\text{Var} G(X(t_k))}}{\sum_{h=1}^{l} \left[ \sqrt{1-\rho(t_h)} + \sqrt{\rho(t_h)} t_h \right] \sqrt{\text{Var} G(X(t_h))}} T.$$

\[\square\]

Proof of Corollary 3.6. The optimal value of each $m_k$ is already given in (8). By substituting the value of each $T_k$ in (11) into (8), we further obtain the expression for the optimal $n_k$. \[\square\]

Proof of Theorem 3.7. Conditioning on each $T_k$, the minimum MSE of each $\hat{\mu}_k$ is given by (9), which is a decreasing function of $T_k$. Hence with the total budget constraint (4), the maximum of MSE $\hat{\mu}_k, k = 1, \ldots, l$ is least when they are equal to each other.

It follows that

$$T_k \propto \left[ \sqrt{1-\rho(t_k)} + \sqrt{\rho(t_k)} t_k \right]^2 \text{Var} G(X(t_k)),$$

which leads to (13). A direct substitution into (9) gives the value of (12). \[\square\]

Proof of Corollary 4.2. Recall that $\text{MSE} \hat{F}(r) = \text{E}[\hat{F}(r) - F(r)]^2$ by definition. By applying Theorem 25.12 in [Billingsley 1986], Theorem 4.1 and the assumption that the square of the left-hand side is uniformly integrable give us the convergence in expectation:

$$\frac{T}{\log T} \text{MSE} \hat{F}(r) = \frac{T}{\log T} \text{E} \left[ \hat{F}(r) - F(r) \right]^2 = \text{E} \left[ \sqrt{\frac{T}{\log T}} \left[ \hat{F}(r) - F(r) \right] \right]^2 \rightarrow \text{E} \left[ \sqrt{aF(r)(1-F(r))} \mathcal{N}(0,1) \right]^2 = aF(r)(1-F(r)),$$

as $T \rightarrow \infty$, which gives the expression for the AMSE. \[\square\]
Optimal Budget Allocation in the Evaluation of SO Algorithms

Proof of Theorem 4.3. It is clear from Corollary 4.2 that AMSE $\hat{F}(r)$ is strictly increasing in $a$. Since the condition $a \geq 1/2\eta^*$ is needed for the optimal convergence rate, the optimal value of $a$ is the lower bound $1/2\eta^*$. A direct substitution gives the value of the AMSE. □

Proof of Corollary 4.5. By following the same argument as in the proof of Corollary 4.2, we have

$$T^{2/3} \text{MSE } \hat{F}(r) = T^{2/3} \mathbb{E} \left[ (\hat{F}(r) - F(r))^2 \right] = \mathbb{E} \left[ T^{1/3} \left( \hat{F}(r) - F(r) \right)^2 \right]$$

$$= \mathbb{E} \left[ \frac{F(r)(1 - F(r))}{b} N(0, 1) + \frac{f_Y^{(1)}(0)^2}{2a} \right]^2 = \frac{f_Y^{(1)}(0)^2}{4a^2} + \frac{F(r)(1 - F(r))}{b},$$

which gives the expression for the AMSE:

$$\text{AMSE } \hat{F}(r) = \left[ \frac{f_Y^{(1)}(0)^2}{4a^2} + \frac{F(r)(1 - F(r))}{b} \right] T^{-2/3}.$$

□

Proof of Corollary 4.6. The conditions of Theorem 4.4 together with the budget constraint (3) implies that

$$ab = 1.$$

Thus, $b = 1/a$, and from Corollary 4.5 the AMSE is

$$\text{AMSE } \hat{F}(r) = \left[ \frac{f_Y^{(1)}(0)^2}{4a^2} + aF(r)(1 - F(r)) \right] T^{-2/3}.$$

The minimizing value of $a$ can be found by one-dimensional optimization, and then $b$ is obtained from $b = 1/a$.

$$a = 2^{-1/3} f_Y^{(1)}(0)^{2/3} F(r)^{-1/3} (1 - F(r))^{-1/3},$$

$$b = 2^{1/3} f_Y^{(1)}(0)^{-2/3} F(r)^{1/3} (1 - F(r))^{1/3}.$$

□

Proof of Theorem 4.7. We first show that the sum of AMSE $\hat{F}_k(r)$ can indeed be bounded to the order of $O((\log T)/T)$. To see this, choose constants $0 < \tau_k < 1$ such that $\sum_{k=1}^l \tau_k = 1$ and let

$$T_k = \tau_k T \quad (21)$$

for each $k = 1, \ldots, l$. Then it follows from (18) that, for sufficiently large $T$, since $\tau_k$ is fixed and $\log \tau_k < 0$, we have

$$\text{AMSE } \hat{F}_k(r) = \frac{F_k(r)(1 - F_k(r))(\log \tau_k + \log T)}{2\eta^*_k \tau_k^2 T} \leq \frac{F_k(r)(1 - F_k(r)) \log T}{2\eta^*_k T} \leq \frac{\log T}{T},$$

ACM Transactions on Modeling and Computer Simulation, Vol. 9, No. 4, Article 39, Publication date: March 2010.
and consequently
\[ \sum_{k=1}^{l} \text{AMSE} \hat{F}_k(r) = O\left(\frac{\log T}{T}\right). \]

Next we show that the linear growth rate of \( T_k \) in (21) has to be satisfied asymptotically by the optimal \( T_k \). Starting with \( T_1 \), consider the ratio \( \frac{T_1}{T} \). Since \( 0 \leq \frac{T_1}{T} \leq 1 \) and \([0, 1]\) is a compact set in \( \mathbb{R} \), there must exist a convergent subsequence, i.e.,
\[ \frac{T_1'}{T} \to \tau_1 \]
for some \( 0 \leq \tau_1 \leq 1 \). By repeating this argument for \( T_2, \ldots, T_l \), a subsequence can be identified that satisfies
\[ \bar{T}_k \bar{T}_k \to \tau_k \in [0, 1] \quad (22) \]
for each \( k = 1, \ldots, l \). Without loss of generality, it can be assumed that (22) also holds for the full sequences of \( T \) and \( T_1, \ldots, T_l \). A justification of the assumption is given at the end of the proof.

Assume that \( \tau_k = 0 \) for some \( k \), i.e., \( T_k/T \to 0 \) as \( T \to \infty \). Then for any constant \( M > 0 \), \( T_k/T < 1/M \) for sufficiently large \( T \), i.e., \( MT_k < T \). Since \( (\log T)/T \) is a decreasing function of \( T \) for \( T > 1 \), it follows that for \( T \) and \( T_k \) sufficiently large so that \( M \leq T_k \),
\[ \frac{(\log T_k)/T_k}{(\log T)/T} \geq \frac{(\log T_k)/T_k}{\log(MT_k)/(MT_k)} \geq \frac{(\log T_k)/T_k}{2\log T_k/(MT_k)} = \frac{M}{2}, \]
which implies that
\[ \frac{(\log T_k)/T_k}{(\log T)/T} \to \infty. \]

This shows that
\[ \text{AMSE} \hat{F}_k(r) \in O\left(\frac{\log T_k}{T_k}\right) \]
grows strictly faster than the optimal rate \( O((\log T)/T) \), and hence \( \tau_k = 0 \) is not optimal.

Now we determine the optimal values of \( \tau_k, k = 1, \ldots, l \). Thus far we know that \( T_k \sim \tau_k T \) with \( \tau_k \neq 0 \) for each \( k \), so that
\[ \text{AMSE} \hat{F}_k(r) = \frac{F_k(r)(1 - F_k(r))}{2\eta_k^* \tau_k} \log T \frac{T}{T}. \]
Hence the sum of the AMSE's satisfies
\[ \sum_{k=1}^{l} \text{AMSE} \hat{F}_k(r) \sim \sum_{k=1}^{l} \frac{F_k(r)(1 - F_k(r))}{2\eta_k^* \tau_k} \log T \frac{T}{T} \]
\[ = \sum_{k=1}^{l} \frac{\tau_k}{\eta_k^*} \sum_{k=1}^{l} \frac{F_k(r)(1 - F_k(r))}{2\eta_k^* \tau_k} \log T \frac{T}{T} \]
\[ \geq \left[ \sum_{k=1}^{l} \frac{F_k(r)(1 - F_k(r))}{2\eta_k^*} \right]^2 \log T \frac{T}{T}. \]
by the Cauchy-Schwarz inequality, with equality if and only if
\[ \tau_k = C \frac{F_k(r)(1 - F_k(r))}{2\eta_k^2 \tau_k}, \]
for some \( C \), which leads to the optimal solution
\[ \tau_k^* = \frac{\sqrt{F_k(r)(1 - F_k(r))/2\eta_k^2}}{\sum_{h=1}^{l} \sqrt{F_h(r)(1 - F_h(r))}/2\eta_h^2}. \]

It is only left to argue that (22) must hold for the entire sequence with \( \tau_k = \tau_k^* \) for each \( k = 1, \ldots, l \). If this is not true for some \( k' \), then there must exist a second subsequence such that
\[ \frac{T_{k'}}{T} \to \tau_{k'} \]
for some \( \tau_{k'} \neq \tau_{k'}^* \). Recursively restricting to a further subsequence, it can be assumed without loss of generality that \( \bar{T}_k/\bar{T} \to \tau_k \) for each \( k = 1, \ldots, l \). But \( \tau \neq \tau^* \) and \( \tau^* \) is the unique optimal solution, so the new subsequence must be suboptimal, contradicting the optimality of the full original sequence. \( \square \)

**Proof of Theorem 4.8.** Using the same argument as in the proof of Theorem 4.7, (22) must hold for an optimal sequence of \( T_k \) values, \( k = 1, 2, \ldots, l \). It is then left to find the optimal values of \( \tau_k \).

Recall the AMSE \( \hat{F}_k(r) \) given by (19):
\[ \text{AMSE} \hat{F}_k(r) = \frac{3}{2\sqrt{2}} f_{Y_k}^{(1)}(0) (2/3) F_k(r)^{2/3}(1 - F_k(r))^{2/3} T_k^{-2/3}. \]

Under the constraint
\[ \sum_{k=1}^{l} \tau_k = 1, \]
the optimal values of \( \tau_k \) must satisfy
\[ \tau_k^{2/3} \propto \frac{3}{2\sqrt{2}} f_{Y_k}^{(1)}(0) (2/3) F_k(r)^{2/3}(1 - F_k(r))^{2/3}, \]
which leads to the solution
\[ \tau_k = \frac{f_{Y_k}^{(1)}(0) F_k(r)(1 - F_k(r))}{\sum_{h=1}^{l} f_{Y_h}^{(1)}(0) F_h(r)(1 - F_h(r))}, \]
and a direct substitution gives the minimum value of the AMSEs. \( \square \)

**Proof of Corollary 5.2.** Notice that the asymptotic distribution of \( Q(p) \) in Theorem 5.1 is exactly the same as that of \( \hat{F}(r) \) in Theorem 4.4, up to a constant factor of \( 1/F'(p) \) and with \( F(r) \) substituted by \( p \). Therefore the AMSE of \( Q(p) \) is
\[ \text{AMSE} Q(p) = \frac{\text{AMSE} \hat{F}(q)}{F'(q)^2} = \left[ \frac{f_Y^{(1)}(0)^2}{4a^2} + \frac{p(1-p)}{b} \right] T^{-2/3}/F'(q)^2. \]
\( \square \)

**Proof of Corollary 5.3.** Let \( F(r) \) be substituted by \( p \). Corollary 5.2 has shown that \( \text{AMSE} Q(p) \) takes the same form as \( \text{AMSE} \hat{F}(r) \), only scaled by a constant factor.
It immediately follows that their optimal values also differ by the same factor, and the optimal values of \( a \) and \( b \) do not change. \( \square \)

**Proof of Theorem 5.4.** Since the optimal value of each AMSE \( Q_k(p) \) in Corollary 5.3 is of the same order as AMSE \( \hat{F}_k(r) \) in Corollary 4.6, the theorem follows by simply substituting in Theorem 4.8 with the new constant factor. \( \square \)

**Proof of Lemma 6.2.** By the definition of covariance,
\[
\text{Cov}(\mathbb{I}(Z_i(t_k) \leq q_k + s_k(T)), \mathbb{I}(Z_i(t_h) \leq q_h + s_h(T))) = E[\mathbb{I}(Z_i(t_k) \leq q_k + s_k(T)) \cdot \mathbb{I}(Z_i(t_h) \leq q_h + s_h(T))] - E[\mathbb{I}(Z_i(t_k) \leq q_k + s_k(T))] \cdot E[\mathbb{I}(Z_i(t_h) \leq q_h + s_h(T))].
\]

We compute each of the expectations separately. First,
\[
E[\mathbb{I}(Z_i(t_k) \leq q_k + s_k(T))] = E[E[\mathbb{I}(Z_i(t_k) \leq q_k + s_k(T)) | X(t_k)]],
\]
\[
= E[P[Z_i(t_k) \leq q_k + s_k(T) | X(t_k)]].
\]
Conditional on \( X(t_k) \), the weak law of large numbers [Billingsley 1986] ensures that
\[
Z_i(t_k) = \frac{1}{m_k} \sum_{j=1}^{m_k} G_j(X(t_k)) \Rightarrow g(X(t_k)) = V(t_k),
\]
as \( m_k \to \infty \) and Slutsky’s Theorem [Billingsley 1986] gives
\[
Z_i(t_k) - s_k(T) \Rightarrow V(t_k).
\]
We assumed that \( P(V(t_k) = q_k) = 0 \), and hence
\[
P[Z_i(t_k) \leq q_k + s_k(T) | X(t_k)] = P[Z_i(t_k) - s_k(T) \leq q_k | X(t_k)] \to P[V(t_k) \leq q_k | X(t_k)].
\]
Finally, by the above convergence and the dominated convergence theorem,
\[
E[\mathbb{I}(Z_i(t_k) \leq q_k + s_k(T))] = E[E[P[Z_i(t_k) \leq q_k + s_k(T) | X(t_k)]],
\]
\[
= E[P[V(t_k) \leq q_k | X(t_k)]] = E[\mathbb{I}(V(t_k) \leq q_k)].
\]
This also shows that
\[
E[\mathbb{I}(Z_i(t_h) \leq q_h + s_h(T))] \to E[\mathbb{I}(V(t_h) \leq q_h)].
\]
Applying almost exactly the same argument establishes that
\[
E[\mathbb{I}(Z_i(t_k) \leq q_k + s_k(T)) \cdot \mathbb{I}(Z_i(t_h) \leq q_h + s_h(T))] \to E[\mathbb{I}(V(t_k) \leq q_k) \cdot \mathbb{I}(V(t_h) \leq q_h)]
\]
as \( T \to \infty \) and this completes the proof. \( \square \)

**Proof of Theorem 6.6.** From the condition \( n \sim bT / \log T \) it follows immediately that
\[
\sqrt{\frac{T}{n \log T}} \to b^{-1/2}
\]
as \( T \to \infty \). Therefore, by Slutsky’s theorem, it suffices to show that
\[
\sqrt{n} \left[ \hat{F}_k(r_k) - F_k(r_k) \right]_k \Rightarrow \mathcal{N}(0, \Lambda)
\]
(23)
as $T \to \infty$. The Cramér-Wold device [Billingsley 1986, p. 383] states that a necessary and sufficient condition for (23) to hold is that
\[ \sqrt{n} \sum_{k=1}^{l} \zeta_k \left[ \hat{F}_k(r_k) - F_k(r_k) \right] \Rightarrow \zeta^T N(0, \Lambda) \]
for any $\zeta \in \mathbb{R}^l$.

Adopt the notation
\[ \hat{F}_k^\zeta(r) := \sum_{k=1}^{l} \zeta_k \left[ \hat{F}_k(r_k) - F_k(r_k) \right]. \]

For each $k = 1, \ldots, l$ and $i = 1, \ldots, n_k$, define
\[ \chi_{ki} := I \left( \frac{1}{m_k} \sum_{j=1}^{m_k} G_j(X_i(t_k)) \leq r_k \right), \]
and
\[ \hat{\chi}_{ki} := \chi_{ki} - E \chi_{ki} = \chi_{ki} - P \left( \frac{1}{m_k} \sum_{j=1}^{m_k} G_j(X_{ki}) \leq r_k \right) \]
to be the centered version of $\chi_{ki}$. Then
\[ \sqrt{n} \hat{F}_k^\zeta(r) = \sqrt{n} \sum_{k=1}^{l} \zeta_k \left[ \hat{F}_k(r_k) - F_k(r_k) \right] \]
\[ = \sum_{k=1}^{l} \zeta_k n^{1/2} \left[ \frac{1}{n} \sum_{i=1}^{n} \hat{\chi}_{ki} + P \left( \frac{1}{m_k} \sum_{j=1}^{m_k} G_j(X_i(t_k)) \leq r_k \right) - F_k(r_k) \right] \]
\[ = \sum_{k=1}^{l} \zeta_k n^{-1/2} \sum_{i=1}^{n} \hat{\chi}_{ki} \]
\[ + \sum_{k=1}^{l} \zeta_k n^{1/2} \left[ P \left( \frac{1}{m_k} \sum_{j=1}^{m_k} G_j(X_i(t_k)) \leq r_k \right) - F_k(r_k) \right] \]

(24)

For (25), for each $k = 1, \ldots, l$, the proof of Theorem 4.1 in Lee and Glynn [2003] argues that each of the summands satisfies
\[ n^{1/2} \left[ P \left( \frac{1}{m_k} \sum_{j=1}^{m_k} G_j(X_i(t_k)) \leq r_k \right) - F_k(r_k) \right] \Rightarrow 0. \]

By Slutsky’s Theorem, it then suffices to show that for (24),
\[ \sum_{k=1}^{l} \zeta_k n^{-1/2} \sum_{i=1}^{n} \hat{\chi}_{ki} \Rightarrow \zeta^T N(0, \Lambda). \]

Define $S_i := \sum_{k=1}^{l} \zeta_k \hat{\chi}_{ki}$ for each $i = 1, \ldots, n$, and then (24) becomes
\[ \sum_{k=1}^{l} \zeta_k n^{-1/2} \sum_{i=1}^{n} \hat{\chi}_{ki} = n^{-1/2} \sum_{i=1}^{n} S_i \]

ACM Transactions on Modeling and Computer Simulation, Vol. 9, No. 4, Article 39, Publication date: March 2010.
If the sequence \{S_i : i = 1, \ldots, n\} satisfies the conditions in Lemma D.1 [Lee and Glynn 2003] with \(\sigma^2 = \zeta^T \Lambda \zeta\), then the Lindeberg-Feller Theorem holds here; cf. [Billingsley 1986]. That is, as \(T \to \infty\), we have the convergence
\[
\sum_{k=1}^{l} \zeta_k n^{-1/2} \sum_{i=1}^{n} \hat{\chi}_{ki} \Rightarrow \zeta^T N(0, \Lambda).
\]

Now we verify that \{S_i : i = 1, \ldots, n\} indeed satisfies the appropriate conditions in Lemma D.1:

(a) The sample path quantities \(X_i(t_k), k = 1, \ldots, l\) are independent for different \(i = 1, \ldots, n\). Each of the sample paths is estimated from the same simulation-optimization algorithm, and hence has the same distribution. Since each \(S_i\) is defined as the same function of \(X_i(t_k), k = 1, \ldots, l\), the values of \(S_i, i = 1, \ldots, n\) are independent and identically distributed.

(b) Since \(\hat{\chi}_{ki}\) is the centered version of \(\chi_{ki}\), we have \(\mathbb{E} \hat{\chi}_{ki} = 0\) for each \(k = 1, \ldots, l\). It then follows that \(\mathbb{E} S_i = 0\). Meanwhile, for each \(i = 1, \ldots, n\),
\[
\text{Var} S_i = \sum_{i=1}^{l} \zeta_k^2 \text{Var} \hat{\chi}_{ki} + 2 \sum_{1 \leq k < h \leq l} \zeta_k \zeta_h \text{Cov}(\hat{\chi}_{ki}, \hat{\chi}_{hi})
\]
\[
= \sum_{i=1}^{l} \zeta_k^2 \text{Var} \chi_{ki} + 2 \sum_{1 \leq k < h \leq l} \zeta_k \zeta_h \text{Cov}(\chi_{ki}, \chi_{hi})
\]
\[
= \sum_{i=1}^{l} \zeta_k^2 \Lambda_{Tkk} + 2 \sum_{1 \leq k < h \leq l} \zeta_k \zeta_h \Lambda_{Tkhh} = \zeta^T \Lambda \zeta.
\]

(c) Corollary 6.3 established that \(\Lambda_T \to \Lambda\) entry-wise. As a result,
\[
\lim_{T \to \infty} \Lambda_T = \Lambda.
\]

(d) Notice that, for each \(k = 1, \ldots, l\),
\[
|\hat{\chi}_{ki}| = |\chi_{ki} - \mathbb{E} \chi_{ki}| = \left| \left( \frac{1}{m_k} \sum_{j=1}^{m_k} G_j(X_i(t_k)) \leq r_k \right) - \mathbb{P} \left( \frac{1}{m_k} \sum_{j=1}^{m_k} G_j(X_i(t_k)) \leq r_k \right) \right| \leq 1.
\]
Therefore, \(|S_i| \leq \sum_{k=1}^{l} |\zeta_k| \cdot |\hat{\chi}_{ki}| = \sum_{k=1}^{l} |\zeta_k|\). Hence \(S_i, i = 1, \ldots, n\) are uniformly bounded, and therefore uniformly integrable.

\[\Box\]

**Proof of Theorem 6.7.** First consider the context of Theorem 6.6, and recall that \(\hat{F}_k(r_k) := \frac{1}{n} \sum_{i=1}^{n} I(Z_i(t_k) \leq r_k)\) as defined in (17). In this context, \(b_k\) does not depend on \(k\), so \(b_k\) can be viewed as a scalar. Scale (20) by \(n(\log T)/T \sim b\) to get
\[
\sqrt{\frac{\log T}{T}} \left[ \sum_{i=1}^{n} I(Z_i(t_k) \leq r_k) - nF(r_k) \right]_k \Rightarrow \mathcal{N}(0, b\Lambda).
\]
Recall that the sample paths \((X_i(t_k), k = 1, \ldots, l)\) are independent in \(i\). To generalize to the case with distinct \(n_k\), decompose the vector into independent summands:

\[
\sum_{i=1}^{n_k} \begin{bmatrix} \mathbb{I}(Z_i(t_k) \leq r_k) - n_k F(r_k) \\
\vdots \\
\end{bmatrix} = \sum_{k=1}^{l} \begin{bmatrix} \sum_{i=n_{k+1}}^{n_k} \mathbb{I}(Z_i(t_1) \leq r_1) - (n_k - n_{k+1}) F(r_1) \\
\vdots \\
\end{bmatrix},
\]

where we define \(n_{l+1} = 0\). For each summand, notice that \(n_k - n_{k+1} \sim (b_k - b_{k+1}) T / \log T\) as \(T \to \infty\). Therefore (26) implies that

\[
\sqrt{\frac{\log T}{T}} \left[ \sum_{i=1}^{n_k} \begin{bmatrix} \mathbb{I}(Z_i(t_k) \leq r_k) - n_k F(r_k) \\
\vdots \\
\end{bmatrix} \right] \Rightarrow \mathcal{N}(0, b_k - b_{k+1}) \mathcal{I}_k \circ \Lambda,
\]

where \(\mathcal{I}_k\) is an \(l\)-dimensional square matrix with the top-left \(k \times k\) block having all entries equal to 1 and the other entries equaling 0. The effect of \(\mathcal{I}_k\) is to preserve the top-left \(k \times k\) block and to truncate the rest. By summing over \(k\), we get

\[
\sqrt{\frac{\log T}{T}} \left[ \sum_{i=1}^{n_k} \begin{bmatrix} \mathbb{I}(Z_i(t_k) \leq r_k) - n_k F(r_k) \\
\vdots \\
\end{bmatrix} \right] \Rightarrow \sum_{k=1}^{l} \mathcal{N}_k(0, b_k - b_{k+1}) \mathcal{I}_k \circ \Lambda = \mathcal{N}(0, B_{\text{max}} \circ \Lambda),
\]

where each \(\mathcal{N}_k\) represents an independent normal random variable, and \(B_{\text{max}}\) is an \(l\)-dimensional square matrix with entries \(B_{\text{max}}(k, h) = b_{k \wedge h}\). Divide the \(k\)th entry by \((\log T) / T \cdot n_k \sim b_k\) for each \(k\) to get

\[
\sqrt{\frac{T}{\log T}} \left[ \tilde{F}_k(r_k) - F(r_k) \right] \Rightarrow b^{-1} \circ \mathcal{N}(0, B_{\text{max}} \circ \Lambda) = \mathcal{N}(0, B_{\text{min}} \circ \Lambda).
\]

\[
\square
\]

**Proof of Theorem 6.8.** From the condition \(T^{-2/3 n} \to b\) it follows immediately that

\[
\frac{T^{1/3}}{\sqrt{n}} \to b^{-1/2}.
\]

Therefore, by Slutsky’s theorem, it suffices to show that

\[
\sqrt{n} \left[ \tilde{F}_k(r_k) - F_k(r_k) \right] \Rightarrow \mathcal{N}(0, \Lambda) + \left[ \frac{f^{(1)}(0) \sqrt{b}}{2 \sigma_k} \right]_k
\]

Apply the Cramér-Wold theorem in the same way as in the proof of Theorem 6.6, and it is then sufficient to consider the asymptotic property of \(\tilde{F}_k(r) := \sum_{k=1}^{l} \zeta_k \left[ \tilde{F}_k(r_k) - F_k(r_k) \right]\). Adopting the same notation as in the proof of Theorem 6.6,
we have the following identity:

\[
\sqrt{n} F_\zeta(r) = \sum_{k=1}^l \frac{1}{\sqrt{m_k}} \sum_{i=1}^n \sum_{j=1}^{m_k} \chi_{ki} G_j(X_i(t_k)) \leq r_k \]  

(27)

\[
+ \sum_{k=1}^l \zeta_k n^{1/2} \left[ P \left( \frac{1}{m_k} \sum_{j=1}^{m_k} G_j(X_i(t_k)) \leq r_k \right) - F_k(r_k) \right] 
\]

(28)

As in the proof of Theorem 6.6, (27) converges in distribution to \( \zeta^T N(0, \Lambda) \). Hence it remains to discuss the asymptotic property of (28). To this end, for each \( k = 1, \ldots, l \), the proof of Theorem 4.4 [Lee 1998, p. 56] ensures that

\[
n^{1/2} \left[ P \left( \frac{1}{m_k} \sum_{j=1}^{m_k} G_j(X_i(t_k)) \leq r_k \right) - F_k(r_k) \right] \Rightarrow \frac{f_{\chi_k}^{(1)}}{2a_k} (0) \sqrt{b},
\]

completing the proof. \( \Box \)

**Proof of Theorem 6.9.** Apply the technique used in the proof of Theorem 6.7 to decompose the random vector in question, and the theorem follows directly from the result of Theorem 6.8. \( \Box \)

**Proof of Theorem 6.10.** We follow the argument in Lee [1998, p. 69]. For each \( k = 1, \ldots, l \), let \( A_k \) be a constant whose value is determined later. For \( s \in \mathbb{R}^l \), define

\[
G(s, m, n) := P \left( \frac{n^{1/2}(Q_k(p_k) - q_k(p_k))}{A_k} \leq s_k, \forall k \right) = P \left( Q_k(p_k) \leq q_k(p_k) + A_k s_k n^{-1/2}, \forall k \right) 
\]

\[
= P \left( \frac{p_k \leq \bar{F}_k(q_k(p_k) + A_k s_k n^{-1/2}), \forall k \right) = P \left( n p_k \leq \sum_{i=1}^n \chi_{ki}, \forall k \right) 
\]

\[
= P \left( 0 \leq n^{-1/2} \sum_{i=1}^n (\chi_{ki} - p_k), \forall k \right),
\]

(29)

where \( \chi_{ki} := \mathbb{I}(Z_i(t_k) \leq q_k(p_k) + A_k s_k n^{-1/2}) \) for each \( k = 1, \ldots, l \) and \( i = 1, \ldots, n \).

To compute the value of (29), consider the joint distribution of \( n^{-1/2} \sum_{i=1}^n (\chi_{ki} - p_k) \) for all \( k \). We will apply the Cramér-Wold device. For some \( A_0 = A_0(T) \) that we specify later, a vector of arbitrary multipliers \( \zeta \), and any \( s_0 \in \mathbb{R} \), define

\[
G(s_0) := P \left( s_0 \leq \frac{\sum_{k=1}^l \zeta_k n^{-1/2} \sum_{i=1}^n (\chi_{ki} - p_k)}{A_0} \right) 
\]

\[
= P \left( \sum_{i=1}^n \sum_{k=1}^l \zeta_k (\chi_{ki} - \Delta_k) \geq s_0 - \frac{n^{1/2} \sum_{k=1}^l \zeta_k (\Delta_k - p_k)}{A_0} \right),
\]

where \( \Delta_k := E \chi_{ki} = P \left( Z_i(t_k) \leq q_k(p_k) + A_k s_k n^{-1/2} \right) \). Let

\[
A_0 := \text{Var} \frac{1}{2} \left( \sum_{k=1}^l \zeta_k \chi_{ki} \right) = \sqrt{\zeta^T \Lambda \zeta},
\]

ACM Transactions on Modeling and Computer Simulation, Vol. 9, No. 4, Article 39, Publication date: March 2010.
where $\Lambda_T$ is the covariance matrix in Definition 6.4 (with an appropriate value for $r_k$). Define

$$Z_n := \sum_{i=1}^{n} \sum_{k=1}^{l} \zeta_k \chi_{ki}$$

and its standardized form

$$Z_n^* := \frac{\sum_{i=1}^{n} \sum_{k=1}^{l} \zeta_k (\chi_{ki} - \Delta_k)}{\sqrt{n^{1/2} \zeta^T \Lambda_T \zeta}}.$$ 

We have shown that

$$G(s_0) = P(Z_n^* \geq -c_{m,n}) = 1 - P(Z_n^* < -c_{m,n}) = \Phi(c_{m,n}) + \Phi(-c_{m,n}) - P(Z_n^* < -c_{m,n}),$$

where

$$c_{m,n} := -s_0 + \frac{n^{1/2} \sum_{k=1}^{l} \zeta_k (\Delta_k - p_k)}{\sqrt{\zeta^T \Lambda_T \zeta}}.$$ 

Now utilizing the Berry-Esséen inequality, we write

$$\sup_{-\infty < x < \infty} |P(Z_n^* < x) - \Phi(x)| \leq C \frac{\rho}{\sigma^3 n^{1/2}},$$

where $C$ is a universal constant, $\sigma^2 := \text{Var} Z_1 = \zeta^T \Lambda_T \zeta$, and

$$\rho := E \left| Z_1 - \sum_{k=1}^{l} \zeta_k \Delta_k \right|^3 \leq \left( \sum_{k=1}^{l} |\zeta_k| \right)^3.$$ 

Hence we have

$$|P(Z_n^* < -c_{m,n}) - \Phi(-c_{m,n})| \leq \sup_{-\infty < x < \infty} |P(Z_n^* < x) - \Phi(x)| \leq C \frac{\rho}{\sigma^3 n^{1/2}} \to 0$$

as $T \to \infty$.

It remains to investigate $c_{m,n}$ as $m,n \to \infty$. But $c_{m,n}$ is a linear combination of analogous quantities studied in Lee [1998, p. 71], and so it satisfies an asymptotic property derived there, namely that

$$c_{m,n} = -s_0 + \frac{n^{1/2} \sum_{k=1}^{l} \zeta_k (\Delta_k - p_k)}{\sqrt{\zeta^T \Lambda_T \zeta}} = -s_0 + \frac{\sum_{k=1}^{l} \zeta_k n^{1/2} (\Delta_k - p_k)}{\sqrt{\zeta^T \Lambda_T \zeta}} 
\to -s_0 + \sum_{k=1}^{l} \zeta_k \left[ \frac{\mathcal{F}'_k(q_k(p_k))}{\sqrt{\zeta^T \Lambda_\zeta}} s_k A_k + \frac{\sqrt{b}}{2 a_k \sqrt{\zeta^T \Lambda_\zeta}} f_{Y_k}^{(1)}(0) \right].$$

Set $A_k := \sqrt{b} / \mathcal{F}'_k(q_k(p_k))$ for each $k = 1, \ldots, l$. Then, as $T \to \infty$, for any $s_0 \in \mathbb{R}$,

$$\lim_{T \to \infty} G(s_0) = \lim_{T \to \infty} \Phi(c_{m,n}) = \Phi \left( -s_0 + \frac{1}{\sqrt{\zeta^T \Lambda_\zeta}} \frac{\sum_{k=1}^{l} \zeta_k \sqrt{b}}{2 a_k} \left[ s_k + \frac{f_{Y_k}^{(1)}(0)}{2 a_k} \right] \right) 
= 1 - \Phi \left( -s_0 + \frac{1}{\sqrt{\zeta^T \Lambda_\zeta}} \frac{\sum_{k=1}^{l} \zeta_k \sqrt{b}}{2 a_k} \left[ s_k + \frac{f_{Y_k}^{(1)}(0)}{2 a_k} \right] \right),$$

which allows us to conclude that

$$\sum_{k=1}^{l} \zeta_k n^{-1/2} \sum_{i=1}^{n} (\chi_{ki} - p_k) \to \mathcal{N}(0, \zeta^T \Lambda_\zeta) + \sum_{k=1}^{l} \zeta_k \sqrt{b} \left[ s_k + \frac{f_{Y_k}^{(1)}(0)}{2 a_k} \right].$$
Since \( \zeta \in \mathbb{R}^l \) was arbitrary, the Cramér-Wold device ensures that
\[
\left[ n^{-1/2} \sum_{i=1}^{n} (\chi_{ki} - p_k) \right] \to \mathcal{N}(0, \Lambda) + \left[ \sqrt{b} \left( s_k + \frac{f_{Y_k}'(0)}{2a_k} \right) \right]_k. \tag{30}
\]
Therefore, as \( T \to \infty \),
\[
G(s, m, n) \to P \left( 0 \leq \mathcal{N}(0, \Lambda) + \left[ \sqrt{b} \left( s_k + \frac{f_{Y_k}'(0)}{2a_k} \right) \right]_k \right) = P \left( \mathcal{N}(0, b^{-1}\Lambda) - \left[ \frac{f_{Y_k}'(0)}{2a_k} \right]_k \leq s \right),
\]
which means that
\[
\left[ T^{1/3} (Q_k(p_k) - q_k(p_k)) F_k'(q_k(p_k)) \right]_k \Rightarrow \mathcal{N}(0, b\Lambda) - \left[ \frac{f_{Y_k}'(0)}{2a_k} \right]_k.
\]
The theorem then follows by dividing each entry by \( F_k'(q_k(p_k)) \). \( \Box \)

**Proof of Theorem 6.11.** In the context of Theorem 6.10, \( b_k \) does not depend on \( k \), so \( b \) can be viewed as a scalar. Recall the intermediate result (30) that appeared in the proof of Theorem 6.10, which can be re-written as
\[
T^{-1/3} \left[ \sum_{i=1}^{n} (\chi_{ki} - p_k) \right]_k \Rightarrow \mathcal{N}(0, b\Lambda) + b \left[ s_k + \frac{f_{Y_k}'(0)}{2a_k} \right]_k.
\]
If the assumption \( n_k = n \) is relaxed, then the left-hand side can be decomposed in the same way as in the proof of Theorem 6.7 and Theorem 6.9:
\[
T^{-1/3} \left[ \sum_{i=1}^{n_k} (\chi_{ki} - p_k) \right]_k \Rightarrow \mathcal{N}(0, (b_k-b_{k+1})I_k \circ \Lambda) + \left[ s_k + \frac{f_{Y_k}'(0)}{2a_k} \right]_k.
\]
Each summand then has the same asymptotic property:
\[
T^{-1/3} \left[ \sum_{i=n_{k+1}+1}^{n_k} (\chi_{ki} - p_k) \right] \Rightarrow \mathcal{N}(0, (b_k-b_{k+1})I_k \circ \Lambda) + \left[ s_k + \frac{f_{Y_k}'(0)}{2a_k} \right]_k.
\]
(Take \( b_{l+1} = 0 \).) Summing over \( k = 1, \ldots, l \), we get
\[
T^{-1/3} \left[ \sum_{i=1}^{n_k} (\chi_{ki} - p_k) \right]_k \Rightarrow \mathcal{N}(0, B_{\max} \circ \Lambda) + \left[ b_k \left( s_k + \frac{f_{Y_k}'(0)}{2a_k} \right) \right]_k. \tag{31}
\]
Also, modify the definition of $G(t, m, n)$:

$$G(t, m, n) := P\left(\frac{n_k^{1/2}(Q_k(p_k) - q_k(p_k))}{A_k} \leq s_k, \forall k\right)$$

with $A_k := \sqrt{b_k/F_k'(q_k(p_k))}$. Follow the analysis in the proof of Theorem 6.10, and apply (31), to get

$$G(t, m, n) = P\left(0 \leq T^{-1/3} \sum_{i=1}^{n_k}(\chi_{ki} - p_k), \forall k\right)$$

$$\rightarrow P\left(0 \leq \mathcal{N}(0, B_{\text{max}} \circ \Lambda) + \left[b_k \left(s_k + \frac{f_Y^{(1)}(0)}{2a_k}\right)\right]_k\right)$$

$$= P\left(\mathcal{N}(0, B_{\text{min}} \circ \Lambda) - \left[b_k \left(s_k + \frac{f_Y^{(1)}(0)}{2a_k}\right)\right]_k\right) \leq s_k, \forall k.$$ 

which means that

$$T^{1/3} \left[(Q_k(p_k) - q_k(p_k)) F_k'(q_k(p_k))\right]_k \Rightarrow \mathcal{N}(0, B_{\text{min}} \circ \Lambda) - \left[b_k \left(s_k + \frac{f_Y^{(1)}(0)}{2a_k}\right)\right]_k.$$ 

Hence the theorem follows by dividing each entry by $F_k'(q_k(p_k))$.  \qed