Proceedings of the 1999 Winter Simulation Conference
P. A. Farrington, H. B. Nembhard, D. T. Sturrock, and G. W. Evans, eds.

COMPUTING THE DISTRIBUTION FUNCTION OF A CONDITIONAL EXPECTATION VIA MONTE CARLO: DISCRETE CONDITIONING SPACES

Shing-Hoi Lee

Fixed-Income Research Group
Morgan Stanley Dean Witter & Co.
1585 Broadway
Manhattan, NY 10036, U.S.A.

Peter W. Glynn

Department of Engineering-Economic Systems
and Operations Research
Stanford University
Stanford, CA 94305-4023, U.S.A.

ABSTRACT

We examine different ways of numerically computing the distribution function of conditional expectations where the conditioning element takes values in a finite or countably infinite outcome space. Both the conditional expectation and the distribution function itself are computed via Monte Carlo simulation. Given a limited (and fixed) computer budget, the quality of the estimator is gauged by the inverse of its mean square error. It is a function of the fraction of the budget allocated to estimating the conditional expectation versus the amount of sampling done relative to the "conditioning variable". We will present the asymptotically optimal rates of convergence for different estimators and resolve the trade-off between the bias and variance of the estimators. Moreover, central limit theorems are established for some of the estimators proposed. We will also provide algorithms for the practical implementation of the estimators and illustrate how confidence intervals can be formed in some cases. Major potential application areas include calculation of Value at Risk (VaR) in the field of mathematical finance and Bayesian performance analysis.

1 INTRODUCTION

Let X be a real-valued random variable (r.v.) and let Z be a random element taking values in a finite or countably infinite outcome space. For fixed $x \in \mathbb{R}$, our goal in this paper is to compute

$$\alpha \stackrel{\triangle}{=} \mathbb{P}(\mathbb{E}(X|Z) \leq x). \tag{1}$$

Thus, this paper is focused on computing the distribution function of a conditional expectation in the setting in which the conditioning random element Z is discrete.

There are several different applications contexts that have served to motivate our interest in this class of problems. The first such application concerns risk management portfolios that contain substantial numbers of financial derivative options. The theory of options pricing asserts that, under suitable conditions, an option's current value can be expressed as a conditional expectation, where the conditioning random element Z is the current price of the underlying asset(s) and the expectation is computed under the so-called "equivalent martingale measure";see, for example, Duffie (1996) for details.

Consequently, if a portfolio consists of a single option, (1) expresses the probability that the value of the portfolio is less than or equal to x. It is worth noting that in such an application, the inner conditional expectation is computed using the equivalent martingale measure, whereas the outer probability involves the postulated dynamics of the underlying asset. It is common to use diffusion processes to model the movement of the asset price. Hence, such applications typically give rise to a continuous conditioning element Z. The computational theory for (1) when Z is continuous is quite different both mathematically and algorithmically and can be found in a companion paper, Lee and Glynn (1999).

The second major class of applications that we have in mind concerns performance evaluation problems in which statistical uncertainty exists about the dynamics of the underlying mathematical model. Assuming that the model is known up to a finite-dimensional statistical parameter, it is often appropriate to model the residual uncertainty via a posterior distribution on the parameter space that incorporates both observational data and *a priori* knowledge. When the parameter space is discrete, this can lead to a problem of the form (1) in which the conditioning random element Z is discrete. To illustrate this point, an example is in order.

Consider a telecommunications service provider that needs to make a decision regarding capacity expansion in a certain neighborhood over the next year. The goal is to deliver requested web-pages to users in less than one second on average. Suppose that N is the number of subscribers in the neighborhood during the next year, and let X_i be the delivery time for the i-th web-page requested within the neighborhood (measured in seconds). It is reasonable to expect that $(X_i : i \geq 1)$ satisfies a law of large numbers (LLN) of the form

$$\frac{1}{n}\sum_{i=1}^{n} X_i \to \mathbb{E}(X_1|N) \quad \text{a.s..} \tag{2}$$

as $n \to \infty$, where a.s. denotes "almost surely". For example, (2) holds, under suitable conditions, if $(X_i : i \geq 1)$ is a stationary sequence which is conditionally ergodic, given N. One could attempt to design the system capacity so that $\mathbb{E}(X|Z) \leq 1$ a.s. However, such a system design would entail building capacity appropriate to dealing with the "worst case" performance scenario ("best case" in terms of revenue) associated with the neighborhood customer base N for the provider. Substantial potential savings can be realized by instead computing $\mathbb{P}(\mathbb{E}(X|N) \leq 1)$. If this probability is sufficiently close to one, then the design capacity is deemed adequate; otherwise, it needs to be increased.

We start, in Section2, by considering two estimators for α that used a fixed amount of sampling per outcome value of Z to compute the conditional expectation of X given that outcome value. Rates of convergence are studied, and a central limit theorem is obtained. Section 3 is concerned with studying the potential improvement in convergence rate that is achievable if one permits the amount of sampling done to compute $\mathbb{E}(X|Z)$ to depend on Z. Finally, Section 4 provides numerical results pertaining to the performance of the basic estimator.

2 ESTIMATION METHODOLOGY WITH SAMPLING RATE INDEPENDENT OF OUTCOME

Suppose that the range of the random element Z is $\{z_1, z_2, \ldots\}$. Our discussion in this section presumes the ability of the simulationist to:

1. draw samples from the distribution $\mathbb{P}(Z \in \cdot)$;
2. for each z_i $(i \geq 1)$, draw samples from the conditional distribution $\mathbb{P}(X \in \cdot|Z = z_i)$.

We consider here the "obvious estimator" for α. To precisely describe this estimator, let $(Z_i : 1 \leq i \leq n)$ be a sequence of independent identically distributed (i.i.d.) copies of the r.v. Z. Conditional on $(Z_i : 1 \leq i \leq n)$, the sample $\left(X_j(Z_i) : 1 \leq i \leq n, 1 \leq j \leq m\right)$ consists of independent r.v.'s in which $X_j(Z_i)$ follows the distribution $\mathbb{P}(X \in \cdot|Z_i)$. In other words,

$$\mathbb{P}(X_j(Z_i) \in A_{ij}, 1 \leq i \leq n, 1 \leq j \leq m|Z_i : i \geq 1) = \prod_{i=1}^{n}\prod_{j=1}^{m} \mathbb{P}(X \in A_{ij}|Z_i).$$

The obvious estimator is then

$$\alpha(m,n) = \frac{1}{n}\sum_{i=1}^{n} I\left(\bar{X}_m(Z_i) \leq x\right),$$

where $\bar{X}_m(Z_i) = m^{-1}\sum_{j=1}^{m} X_j(Z_i)$. Because the sample size m associated with $\bar{X}_m(Z_i)$ is independent of the outcome value Z_i, we call $\alpha(m,n)$ an estimator with outcome-independent sampling rate.

We wish to develop a central limit theorem (CLT) for this estimator that describes its rate of convergence. For a given computer budget c, let $m(c)$ and $n(c)$ be chosen so that the computational effort required to generate $\alpha(m(c), n(c))$ is approximately c. To this end, let δ_1 be the average amount of time required to generate Z_i and let δ_2 be the average amount of time required to generate $X_j(Z_i)$. Then, the aggregate effort required to compute $\alpha(m,n)$ is approximately $\delta_1 n + \delta_2 mn$. It follows that $\delta_1 n(c) + \delta_2 m(c)n(c) \approx c$. In order that $\bar{X}_{m(c)}(Z_i) \to \mathbb{E}(X|Z_i)$ a.s. as $c \to \infty$, we clearly need to impose the requirement that $m(c) \to \infty$ as $c \to \infty$. Consequently, $\delta_1 n(c) + \delta_2 m(c)n(c) \approx \delta_2 m(c)n(c)$ for c large. Finally, we may, without loss of generality, assume $\delta_2 = 1$ (for otherwise, we can simply re-define the units by which we choose to measure computer time). Given this analysis, it is evident that $(m(c), n(c))$ must be chosen to satisfy the asymptotic relation $m(c)n(c)/c \to 1$ as $c \to \infty$.

For a given sampling plan $((m(c), n(c)) : c \geq 0)$, let $\alpha_1(c) = \alpha(m(c), n(c))$ be the estimator available after expending c units of computational time. The key to understanding the behavior of $\alpha_1(c)$ is to develop an expression for the bias of $\alpha_1(c)$. This will permit us to perform the standard "bias-variance" trade-off necessary to compute the most efficient possible sampling plan.

Note that

$$\begin{aligned}\mathbb{E}\alpha(m,n) &= \sum_{z\in\Gamma_+} p(z)\mathbb{P}(\bar{X}_m(z) \leq x) \\ &\quad + \sum_{z\in\Gamma_-} p(z)(1 - \mathbb{P}(\bar{X}_m(z) > x)) \\ &= \alpha + \sum_{z\in\Gamma_+} p(z)\mathbb{P}(\bar{X}_m(z) \leq x) \\ &\quad - \sum_{z\in\Gamma_-} p(z)\mathbb{P}(\bar{X}_m(z) > x),\end{aligned} \tag{3}$$

where $\Gamma_+ = \{z_i : \mathbb{E}(X|Z = z_i) > x, i \geq 1\}$, $\Gamma_- = \{z_i : \mathbb{E}(X|Z = z_i) \leq x, i \geq 1\}$ and $p(z) = \mathbb{P}(Z = z)$. Thus, the rate at which the bias goes to zero is determined by the rate at which $\mathbb{P}(\bar{X}_m(z) \leq x) \to 0$ for $z \in \Gamma_+$ and the rate at which $\mathbb{P}(\bar{X}_m(z) > x) \to 0$ for $z \in \Gamma_-$. These rates involving the distribution of $\bar{X}_m(z)$ are of a type that have been extensively studied as part of the substantial literature on "large deviation".

We say that a r.v. X is lattice if X takes values only in the set $\{c + kd : k \in \mathbb{Z}\}$, where c is a fixed constant and $d > 0$ is the lattice spacing. One of the fundamental results in large deviations theory is the following; see p.121 of Bucklew (1990).

Theorem 2.1 *Let $(X_i : i \geq 1$ be an i.i.d. sequence of r.v.'s such that $\mathbb{E}X_1 < x$. Suppose $\varphi(\theta) = \mathbb{E}\exp(\theta X_1) < \infty$ for $\theta \in \mathbb{R}$ and that $\mathbb{P}(X_1 > x) > 0$. Then, if X_1 is not lattice,*

$$\lim_{m\to\infty} \mathbb{P}\left(\frac{1}{m}\sum_{i=1}^{m} X_i > x\right)\exp(m\eta)\sqrt{m} = \frac{1}{\sqrt{2\pi}\sigma},$$

where

$$\eta = \theta^* x - \log\varphi(\theta^*),$$
$$\frac{\varphi'(\theta^*)}{\varphi(\theta^*)} = x,$$
$$\sigma^2 = \frac{\varphi''(\theta^*)}{\varphi(\theta^*)} - x^2.$$

In view of Theorem 2.1, we now make the following assumptions:

A1. For $i \geq 1$ and $\theta \in \mathbb{R}$, $\mathbb{E}[\exp(\theta X)|Z = z_i] < \infty$;
A2. $\mathbb{P}(X > x|Z = z_i) > 0, z_i \in \Gamma_-$,
$\mathbb{P}(X \leq x|Z = z_i) > 0, z_i \in \Gamma_+$;
A3. $\mathbb{E}(X|Z = z_i) \neq x$ for $i \geq 1$;
A4. For $i \geq 1$, $X_1(z_i)$ is not lattice.

With A1-A4 in hand, Theorem 2.1 guarantees the existence, for each z_i, of finite constants $\gamma(z_i)$ and $\eta(z_i)$ such that

$$\begin{aligned} &\mathbb{P}(\bar{X}_m(Z_1) > x|Z_1 = z_i) \\ &\quad \sim \quad m^{-1/2}\gamma(z_i)\exp(-m\eta(z_i)), \quad z_i \in \Gamma_- \\ &\mathbb{P}(\bar{X}_m(Z_1) \leq x|Z_1 = z_i) \\ &\quad \sim \quad m^{-1/2}\gamma(z_i)\exp(-m\eta(z_i)), \quad z_i \in \Gamma_+ \end{aligned} \quad (4)$$

as $m \to \infty$, where $a_m \sim b_m$ as $m \to \infty$ means that $a_m/b_m \to 1$ as $m \to \infty$. With the aid of a couple of additional hypothesis, we can derive an asymptotic approximation for the bias of $\alpha_1(c)$. Put $\eta^* = \inf\{\eta(z_i) : i \geq 1\}$.

A5. $B^* = \{z_i : i \geq 1, \eta(z_i) = \eta^*\}$ is non-empty and finite.
A6. $\inf\{\eta(z_i) : i \geq 1, z_i \neq B^*\} > \eta^*$.

Note that both A5 and A6 are trivially satisfied when the range of Z is finite.

Proposition 2.1 *Assume A1-A6. If $m(c) \to \infty$ as $c \to \infty$, then*

$$m(c)^{1/2}\exp(\eta^* m(c))(\mathbb{E}\alpha_1(c) - \alpha) \to \gamma^*$$

as $m \to \infty$, where

$$\gamma^* \triangleq \sum_{z\in\Gamma_+\cap B^*} p(z)\gamma(z) - \sum_{z\in\Gamma_-\cap B^*} p(z)\gamma(z).$$

Proof. It follows from Markov's inequality that for $\theta \geq 0$,

$$\mathbb{P}(\bar{X}_m(Z_1) > x|Z_1 = z_i) \leq \exp(-\theta m x + m\Psi(\theta, z_i))$$

where $\Psi(\theta, z_i) = \log\mathbb{E}[\exp(\theta X)|Z = z_i]$. In particular, if $z_i \in \Gamma_-$, we may choose θ so that $\theta = \theta^*(z_i)$, where $\theta^*(z_i)$ satisfies $\Psi'(\theta^*(z_i), z_i) = x$. Then, we obtain the relation

$$\mathbb{P}(\bar{X}_m(Z_1) > x|Z_1 = z_i) \leq \exp(-m\eta(z_i)); \quad (5)$$

this inequality holds uniformly in $m \geq 1$ and $z_i \in \Gamma_-$. Similarly,

$$\mathbb{P}(\bar{X}_m(Z_1) \leq x|Z_1 = z_i) \leq \exp(-m\eta(z_i)) \quad (6)$$

for $m \geq 1$ and $z_i \in \Gamma_+$.

From (4), it is evident that

$$\begin{aligned} &m(c)^{1/2}\exp(\eta^* m(c))(\mathbb{E}\alpha(c) - \alpha) \quad (7)\\ &= \sum_{z\in\Gamma_+\cap B^*} p(z)m(c)^{1/2}e^{\eta^* m(c)}\mathbb{P}(\bar{X}_{m(c)}(z) \leq x) \\ &\quad - \sum_{z\in\Gamma_-\cap B^*} p(z)m(c)^{1/2}e^{\eta^* m(c)}\mathbb{P}(\bar{X}_{m(c)}(z) > x) \\ &\quad + \sum_{z\in\Gamma_+\cap B^{*c}} p(z)m(c)^{1/2}e^{\eta^* m(c)}\mathbb{P}(\bar{X}_{m(c)}(z) \leq x) \\ &\quad - \sum_{z\in\Gamma_-\cap B^{*c}} p(z)m(c)^{1/2}e^{\eta^* m(c)}\mathbb{P}(\bar{X}_{m(c)}(z) > x). \end{aligned}$$

Since B^* is finite, the difference of the first two sums on the right-hand side of (7) converges to γ^*; see (4). To handle the two final sums on the right-hand side of (7), observe that (5) and (6) yield the bound

$$\left| \sum_{z \in \Gamma_+ \cap B^{*c}} p(z) m(c)^{1/2} e^{\eta^* m(c)} \mathbb{P}(\bar{X}_{m(c)}(z) \le x) - \sum_{z \in \Gamma_- \cap B^{*c}} p(z) m(c)^{1/2} e^{\eta^* m(c)} \mathbb{P}(\bar{X}_{m(c)}(z) > x) \right| \le \sum_{z \notin B^*} p(z) m(c)^{1/2} e^{\eta^* m(c)} \exp(-\eta(z) m(c)). \quad (8)$$

But according to A6, for $z \notin B^*$, $\eta(z) - \eta^*$ is uniformly positive, so $m(c)^{1/2} \exp(\eta^* m(c)) \exp(-\eta(z) m(c)) \to 0$ uniformly in $z \notin B^*$ as $c \to \infty$. It follows from the Dominated Convergence theorem that the right-hand side of (8) goes to zero, completing the proof.

Hence, if $\gamma^* \neq 0$,

$$\mathbb{E}\alpha_1(c) - \alpha \sim \gamma^* m(c)^{-1/2} \exp(-\eta^* m(c)) \quad (9)$$

providing us with our desired bias asymptotic. We now turn to the variance of $\alpha_1(c)$. Note that

$$\begin{aligned} \mathrm{Var}\alpha_1(c) &= \frac{1}{n(c)} \mathrm{Var} I(\bar{X}_{m(c)}(Z_1) \le x) \\ &= \frac{1}{n(c)} \mathbb{E}\alpha_1(c)(1 - \mathbb{E}\alpha_1(c)) \quad (10) \\ &\sim \frac{1}{n(c)} \alpha(1-\alpha) \end{aligned}$$

as $c \to \infty$. If we now choose to optimize our choice of $(m(c), n(c))$ so as to minimize

$$\mathrm{MSE}(\alpha_1(c)) = \mathrm{Var}\alpha_1(c) + (\mathbb{E}\alpha_1(c) - \alpha)^2$$

subject to the constraint that $m(c)n(c) \approx c$, the approximations (9) and (10) suggest that the optimal choice of $m(c)$ satisfies the asymptotic

$$m^*(c) \sim (\log c)/2\eta^*$$

as $c \to \infty$. This asymptotic relation is supported by the following CLT for $\alpha_1(c)$; this is our main result in this section.

Theorem 2.2 *Assume Assumptions A1-A6. Suppose that $m(c) \to \infty$ and $n(c) \to \infty$ in such a way that $n(c)m(c)/c \to 1$ as $c \to \infty$. Then, if $m(c) = \lfloor a \log c \rfloor$ as $c \to \infty$ where $a \ge 1/2\eta^*$,*

$$\sqrt{\frac{c}{\log c}}\, (\alpha_1(c) - \alpha) \Rightarrow \sqrt{a\alpha(1-\alpha)}\, N(0,1)$$

as $c \to \infty$, where $\Rightarrow$ denotes weak convergence and $N(0,1)$ is a normally distributed r.v. with mean zero and unit variance. On the other hand, if $m(c) = \lfloor a \log c \rfloor$ with $0 < a < 1/2\eta^*$, then

$$c^{\eta^* a} \sqrt{\log c}\, (\alpha_1(c) - \alpha) \Rightarrow \frac{\gamma^*}{\sqrt{a}}$$

as $c \to \infty$.

Proof. Define $\chi_i(m) \triangleq I(\bar{X}_m(Z_i) \le x)$. Note that

$$\alpha_1(c) - \alpha = \frac{1}{n} \sum_{i=1}^{n} \hat{\chi}_i(m) + \mathbb{P}(\bar{X}_m(Z) \le x) - \alpha,$$

where $\hat{\chi}_i(m) = \chi_i(m) - \mathbb{P}(\bar{X}_m(Z) \le x)$ is the centered version of $\hat{X}_i(m)$. Then,

$$\alpha_1(c) - \alpha = \frac{1}{\sqrt{n}} \left(\sum_{i=1}^{n} \frac{\hat{\chi}_i(m)}{\sqrt{n}} \right) + \mathbb{P}(\bar{X}_m(Z) \le x) - \alpha.$$

Observe that for each i, $\hat{\chi}_i(m)$ is a bounded sequence of r.v.'s, it follows that the family $\{\hat{\chi}_i(m(c)) : i = 1, \ldots, n(c), c > 0\}$ is uniformly integrable. By Lemma A-1, the Lindeberg-Feller theorem (ref. Billingsley 1995) holds here. That is, as $c \to +\infty$,

$$\sum_{i=1}^{n(c)} \frac{\hat{\chi}_i(m(c))}{\sqrt{n(c)}} \Rightarrow \sigma \mathrm{N}(0,1),$$

where $\sigma = \sqrt{\alpha(1-\alpha)}$.

Since $n(c)m(c)/c \to 1$ and $m(c) = \lfloor a \log c \rfloor$, we have that $c/(n(c)\log c) \to a$. Hence, by the converging together theorem (ref. Billingsley 1995), we have that

$$\sqrt{\frac{c}{\log c}} \sum_{i=1}^{n(c)} \frac{\hat{\chi}_i(m(c))}{n(c)} \Rightarrow \sqrt{a\alpha(1-\alpha)}\, \mathrm{N}(0,1) \quad (11)$$

as $c \to \infty$. On the other hand,

$$\begin{aligned} &\sqrt{\frac{c}{\log c}}\, (\mathbb{P}(\bar{X}_{m(c)}(Z) \le x) - \alpha) \\ &= \sqrt{\frac{c}{\log c}}\, m(c)^{-1/2} e^{-\eta^* m(c)} \\ &\quad \cdot m(c)^{1/2} e^{\eta^* m(c)} (\mathbb{P}(\bar{X}_{m(c)}(Z) \le x) - \alpha). \end{aligned}$$

We know that the second term converges to γ^* by Proposition 2.1. For the first term, notice that

$$\sqrt{\frac{c}{m(c)\log c}}\, e^{-\eta^* m(c)} = c^{\frac{1}{2}-\frac{\eta^* m(c)}{\log c}} \cdot \sqrt{\frac{\log c}{m(c)}} \cdot \frac{1}{\log c}.$$

converges to 0 as $c \to \infty$ since $m(c)/\log c \to a$ and by assumption, $1/2 - a\eta^* \le 0$. By the converging theorem, we must have that

$$\sqrt{\frac{c}{\log c}}\, (\mathbb{P}(\bar{X}_{m(c)}(Z) \le x) - \alpha) \Rightarrow 0 \qquad (12)$$

as $c \to \infty$. Applying the converging together theorem once again, we thus obtain the first result by combining the converging results (11) and (12).

Similarly, if $m(c) = \lfloor a \log c \rfloor$ and $a\eta^* < 1/2$, then, we have

$$\frac{c^{a\eta^*}\sqrt{\log c}}{n(c)} = \sqrt{\frac{c}{n(c)m(c)}}\sqrt{\frac{m(c)}{\log c}}\; c^{a\eta^* - 1/2} \to 0.$$

as $c \to \infty$. By the converging together theorem, we have that

$$c^{a\eta^*}\sqrt{\log c} \sum_{i=1}^{n(c)} \frac{\hat{\chi}_i(m(c))}{n(c)} \Rightarrow 0 \qquad (13)$$

as $c \to \infty$. Also, by the converging together theorem,

$$\begin{aligned} & c^{a\eta^*}\sqrt{\log c}\, (\mathbb{P}(\bar{X}_{m(c)}(Z) \le x) - \alpha) \\ = \; & m^{1/2} e^{m(c)\eta^*} (\mathbb{P}(\bar{X}_{m(c)}(Z) \le x) - \alpha) \\ & \cdot \sqrt{\frac{\log c}{m(c)}}\, \exp(\eta^*(a \log c - m(c))) \\ \to \; & \frac{\gamma^*}{\sqrt{a}}. \end{aligned} \qquad (14)$$

as $c \to \infty$. Finally, we obtain the second result by combining the converging results (13) and (14).

The proof of Theorem 2.2 actually shows that if $n(c)m(c)/c \to 1$ as $c \to \infty$ with $m(c)/\log(c) \to +\infty$, then

$$\sqrt{n(c)}(\alpha_1(c) - \alpha) \Rightarrow \sqrt{\alpha(1-\alpha)}\mathrm{N}(0, 1)$$

as $c \to \infty$. It follows that if $m(c)/\log(c) \to \infty$, then

$$\left[\alpha_1(c) - z\sqrt{\frac{\alpha_1(c)(1-\alpha_1(c))}{n(c)}},\; \alpha_1(c) + z\sqrt{\frac{\alpha_1(c)(1-\alpha_1(c))}{n(c)}}\right]$$

is an approximate $100(1-\delta)\%$ confidence interval for α, provided c is large and z is selected so that $\mathbb{P}(-z \le \mathrm{N}(0,1) \le z) = 1 - \delta$. A natural choice for $m(c)$ here is to set $m(c) = \lfloor ac^r \rfloor$ for $a > 0$ with $r \in (0, 1)$. This suggests the following confidence interval procedure for computing α.

Algorithm 2.1

Step 0. *Initialization.* Input c, ν, and a.

Step 1. *Determine the sample sizes.* Set $(m, n) \triangleq (ac^\nu, a^{-1}c^{1-\nu})$.

Step 2. *Determine* $\hat{\alpha}$. Set

$$\hat{\alpha} \triangleq \frac{1}{n}\sum_{i=1}^{n} I\left(\frac{1}{m}\sum_{j=1}^{m} X_j(Z_i) \le x\right).$$

Step 3. *Form the* $(1-\xi) \times 100\%$ *confidence interval for* $\hat{\alpha}$. A consistent estimate of the standard error (s.e.), $s_{\hat{\alpha}}$, of $\hat{\alpha}$ is $\sqrt{\frac{\hat{\alpha}(1-\hat{\alpha})}{n}}$ since $\hat{\alpha} \sim \alpha$. The c.i. is then set to

$$\left[\hat{\alpha} - z_{\xi/2}s_{\hat{\alpha}}, \hat{\alpha} + z_{\xi/2}s_{\hat{\alpha}}\right],$$

where $z_{\xi/2}$ is the $\xi/2$-quantile of a N(0, 1) r.v.

In Section 4, we offer empirical data associated with the performance of this confidence interval procedure for α.

We conclude this section with a discussion of an alternative estimator that is applicable when the probability mass function of Z is known. For example, in our telecommunications service provider example, it may be that the distribution of the number of subscribers is modelled via a Poisson r.v. or binomial r.v., in which case the probability mass function is known explicitly. In particular, suppose that the simulationist:

1. has knowledge of the probability mass function $p(\cdot)$ corresponding to the random element Z;
2. has the ability to draw samples from the conditional distribution $\mathbb{P}(X \in \cdot \mid Z = z_i)$, for each z_i $(i \ge 1)$.

The estimator we have in mind here is

$$\alpha_2(m) = \sum_i p(z_i) I(\bar{X}_m(z_i) \leq x),$$

so that the sample size used to estimate $\mathbb{P}(\mathbb{E}(X|Z = z_i) \leq x)$ is again outcome-independent. The computer time required to generate $\alpha_2(m)$ is proportional to m multiplied by the number of outcome values for Z. Thus, the estimator can only be (exactly) computed when the number of outcome values for Z is finite. Throughout the remainder of this section, we will assume that this is the case. Then, m scales linearly in the computer budget c so that examining the rate of convergence as a function of m is equivalent to studying the rate of convergence as a function of c.

For $i \geq 1$, let

$$\kappa_m(z_i) = \mathbb{P}(\bar{X}_m(Z_1) \leq x | Z_1 = z_i)$$

if $z_i \in \Gamma_+$ and let

$$\kappa_m(z_i) = \mathbb{P}(\bar{X}_m(Z_1) \leq x | Z_1 = z_i)$$

if $z_i \in \Gamma_-$. Then,

$$\mathbb{E}\alpha_2(m) - \alpha = \sum_{z \in \Gamma_+} p(z)\kappa_m(z) - \sum_{z \in \Gamma_-} p(z)\kappa_m(z)$$

and

$$\text{Var}\alpha_2(m) = \sum_i p(z_i)^2 \kappa_m(z_i)(1 - \kappa_m(z_i)). \tag{15}$$

Assume that A1-A4 hold (and note that A5-A6) are automatic, in view of our finite outcome assumption). It follows that Proposition 2.1 asserts that if $\gamma^* \neq 0$, then

$$\mathbb{E}\alpha_2(m) - \alpha \sim \gamma^* m^{-1/2} \exp(-\eta^* m)$$

as $m \to \infty$. Furthermore, (4) and (15) together imply that

$$\text{Var}\ \alpha_2(m) \sim \beta^* m^{-1/2} \exp(-\eta^* m)$$

as $m \to \infty$, where

$$\beta^* = \sum_{z \in B^*} p(z)^2 \gamma(z).$$

As a consequence, the mean square error satisfies the asymptotic relation

$$\text{MSE}(\alpha_2(m)) \sim \beta^* m^{-1/2} \exp(-\eta^* m)$$

as $m \to \infty$, so that the mean square error converges to zero exponentially fast in this setting. Thus, in those settings where it applies, $\alpha_2(m)$ is to be preferred to $\alpha_1(c)$, at least asymptotically (for large computer budgets). This analysis also suggests that in the large-sample context, it is the sampling of the Z-values (the "outer sampling") that contributes primarily to the variability of $\alpha_1(c)$ (rather than the "inner sampling").

3 ESTIMATION METHODOLOGY WITH OUTCOME DEPENDENT SAMPLING RATE

The large deviations asymptotics expressed by (4) assert that the impact of m on the rate at which the individual bias terms in (3) go to zero is highly state-dependent. This suggests that the amount of sampling necessary to mitigate the effect of bias is highly outcome-dependent and that improved algorithms for estimating α can, at least in principle, be obtained by permitting the "inner sample size" m to be outcome- dependent. Our goal in this section is to explore the potential increases in efficiency that can be obtained via such an idea.

We start with the case in which the mass function of Z is known. In this case, a sampling plan is an assignment of sample sizes $\vec{m} = (m(z_i) : i \geq 1)$ to each possible outcome value of Z, leading to the estimator

$$\alpha_3(\vec{m}) = \sum_i p(z_i) I(\bar{X}_{m(z_i)}(z_i) \leq x).$$

The total computer time expended to calculate $\alpha_3(\vec{m})$ is approximately proportional to $\sum_i m(z_i)$. Thus, given a computer budget c, this effectively acts as a constraint on $\sum_i m(z_i)$. We wish to find a selection of the sample sizes $(m(z_i) : i \geq 1)$ which minimizes the mean square error of $\alpha_3(\vec{m})$, subject to the constraint that $\sum_i m(z_i) \leq c$. We will denote the corresponding estimator $\alpha_3(c)$.

Assume A1-A4. To simplify the (mathematical) technical issues involved, we will suppose, through the remainder of this section, that the number of different outcome values for Z is finite, so that A5 and A6 are also in force. Just as for the estimator $\alpha_2(m)$, the bias and variance of $\alpha_3(\vec{m})$ may easily be computed:

$$\mathbb{E}\alpha_3(\vec{m}) - \alpha = \sum_{z \in \Gamma_+} p(z)\kappa_{m(z)}(z) - \sum_{z \in \Gamma_-} p(z)\kappa_{m(z)}(z)$$

and

$$\text{Var}\ \alpha_3(\vec{m}) = \sum_i p(z)^2 \kappa_{m(z_i)}(z_i) \left(1 - \kappa_{m(z_i)}(z_i)\right).$$

The following result uses the above expressions to determine the optimal sampling plan $m^* = (m_c^*(z_i) : i \geq 1)$ (for large computer budgets c) and the associated rate of convergence.

Theorem 3.1 *Assume A1-A4. Then, for any choice of $\vec{m}_c = (m_c(z_i) : i \geq 1)$ satisfying $\sum_i m_c(z_i) \leq c$,*

$$\liminf_{c\to\infty} \frac{1}{c} \log MSE(\alpha_3(\vec{m}_c)) \geq -\tau^*$$

where

$$\tau^* = \left(\sum_i \eta(z_i)^{-1}\right)^{-1}.$$

Furthermore, if we choose $m_c^*(z_i) = \lfloor c\tau^*/\eta(z_i)\rfloor$,

$$\lim_{c\to\infty} \frac{1}{c} \log MSE(\alpha_3(m^*)) = -\tau^*.$$

Proof. We would like to minimize $\mathrm{MSE}(\alpha_3(\vec{m}))$ with respect to $(m(z_i) : 1 \leq i \leq K)$ where K is the cardinality of the outcome space of Z. In order to make the minimization problem a bit more tractable, we will determine some bounds on $\mathrm{MSE}(\alpha_3(\vec{m}))$. Specifically, we use the inequality $(\sum_{i=1}^n x_i)^2 \leq n\sum_{i=1}^n x_i^2$, to bound $\mathrm{MSE}(\alpha_3(\vec{m}))$ by

$$\begin{aligned}\mathrm{MSE}(\alpha_3) &\leq \mathrm{Var}(\hat{\alpha}) + K\sum_z p(z)^2 \frac{\gamma(z)^2}{m(z)} e^{-2m(z)\eta(z)} \\ &\leq (K+1)\sum_{z\in\Gamma} p(z)^2\gamma(z)e^{-m(z)\eta(z)} \qquad (16)\end{aligned}$$

by (4) since A1-A4 hold. Denote by V the term on the right hand side (RHS) of the last inequality.

Now, let's minimize (16) with respect to $\vec{m}$ subject to $\sum_z m(z) = c$ and $m(z) > 0$ for all z. Let λ be the Lagrange multiplier of the equality constraint. Taking partials with respect to $m(z)$, the optimizer $m^*(z)$ must satisfy

$$\frac{\partial V}{\partial m(z)} = p(z)^2\gamma(z)e^{-m^*(z)\eta(z)}(-\eta(z)) + \lambda = 0.$$

In other words,

$$m^*(z) = \frac{-1}{\eta(z)} \log\left(\frac{\lambda}{p(z)^2\gamma(z)\eta(z)}\right).$$

Substituting the above expression for $m^*(z)$ into the equality constraint, we deduce that λ satisfies

$$-\log\lambda = \frac{c - \sum_z \frac{1}{\eta(z)} \log(p(z)^2\gamma(z)\eta(z))}{\sum_z \eta(z)^{-1}}.$$

Hence, we have derived the expression for the optimal $m^*(z)$ that minimizes the upper bound:

$$\begin{aligned}m^*(z) = &\frac{\eta(z)^{-1}c}{\sum_w \eta(w)^{-1}} + \eta(z)^{-1}\log(p(z)^2\gamma(z)\eta(z)) \\ &- \frac{\eta(z)^{-1}}{\sum_w \eta(w)^{-1}} \sum_w \frac{\log(p(w)^2\gamma(w)\eta(w))}{\eta(w)}.\end{aligned}$$

The minimized upper bound on $\mathrm{MSE}(\alpha_3(\vec{m}))$ is thus equal to

$$V_{\min} = K\left(\prod_z (p(z)^2\gamma(z)\eta(z))^{\tau^*\eta(z)^{-1}}\right) e^{-\tau^* c},$$

where $\tau^* \triangleq (\sum_z \eta(z)^{-1})^{-1}$. In other words, the upper bound, V, on $\mathrm{MSE}(\alpha_3(\vec{m}))$ is converging to zero exponentially fast at the rate $O(e^{-\tau^* c})$. Notice that τ^* is largely determined by η^*.

Let's now examine a lower bound on the MSE. Notice that, for large $m(z)$,

$$\mathrm{MSE}(\alpha_3(\vec{m})) \geq \frac{1}{2}\sum_z p(z)^2 \frac{\gamma(z)}{\sqrt{m(z)}} \exp(-m(z)\eta(z))$$

since any one of two terms, $\mathbb{P}(\bar{X}_{m(z)}(z) \leq x)$ or $\mathbb{P}(\bar{X}_{m(z)}(z) > x)$ is greater than or equal to 1/2.

Note that in order to drive $\mathrm{MSE}(\alpha_3(\vec{m}))$ to 0, we must have that the $\vec{m}_c$ that minimizes $\mathrm{MSE}(\hat{\alpha})$ must satisfy $\vec{m}_c \nearrow +\infty$ for all z as $c \nearrow +\infty$. On the other hand, for each $\epsilon > 0$, there exists M_ϵ such that for all $m \geq M_\epsilon$, $m^{-1/2} \geq \exp(-\epsilon m)$. Combining these two remarks, we have that for all $\epsilon > 0$, there exists C_ϵ such that for all $c > C_\epsilon$

$$\begin{aligned}&\min \mathrm{MSE}(\alpha_3(\vec{m})) \\ &\geq \min \frac{1}{2}\sum_z p(z)^2\gamma(z)\exp(-(\eta(z)+\epsilon)m(z)) \\ &= \text{constant}\cdot\exp\left(\frac{-c}{\sum_z(\eta(z)+\epsilon)^{-1}}\right).\end{aligned}$$

In other words, for all $c > C_\epsilon$,

$$\frac{-1}{\sum_z \eta(z)^{-1}} \geq \frac{1}{c}\log(\min \mathrm{MSE}(\hat{\alpha})) \geq \frac{-1}{\sum_z(\eta(z)+\epsilon)^{-1}}.$$

This completes the proof.

Thus, the optimal choice m^* (rather than, for example, using constant sample sizes as for $\alpha_2(m)$) impacts the exponential rate at which the mean square error converges to zero. Of course, implementation of $\alpha_3(c)$ $(= \alpha_3(m^*))$ requires knowledge of $\eta(z_i)$ for $i \geq 1$. Note, however, that the formula for $m_c^*(z_i)$ asserts that the most critical outcome values are those for which $\eta(z_i)$ is close to zero. In such a setting, the corresponding "large deviations" involves looking at events that are relatively more likely. Such a regime is one in which the corresponding large deviations are relatively more Gaussian (since the deviation involves a tail event that is relatively closer to the mean of the distribution). In sampling Gaussian r.v.'s with mean $\mu < x$ and standard deviation σ, the likelihood of a deviation in the sample mean greater than x is approximately $\exp(-n(\mu-x)^2/2\sigma^2)$ (in "logarithmic scale"). This suggests the approximation $\eta(z_i) \approx (\mu(z_i) - x)^2/2\sigma^2(z_i)$ for $i \geq 1$, where $\mu(z_i)$ and $\sigma(z_i)$ are, respectively, the mean and standard deviation of the distribution $\mathbb{P}(X \in \cdot \,|Z = z_i)$. Of course, for outcome values z_i with large $\eta(z_i)$, $(\mu(z_i) - x)^2/2\sigma^2(z_i)$

will not give a good approximation to its η. As mentioned earlier, such $\eta(\cdot)$'s have a small contribution to τ^*. Hence, for each z_i, this heuristic would propose spending a small portion of the computational budget to estimate $(\mu(z_i)-x)^2/2\sigma^2(z_i)$, and then using this to estimate $\eta(z_i)$, followed by "production runs" to compute α.

The algorithm below gives a practical methodology for the implementation of $\alpha_3(c)$.

Algorithm 3.1

Step 0. *Initialization.* Input c, $0 < r < 1$, and $\{p(z_i) : i \geq 1\}$.

Step 1. *Estimate the* $\eta(\cdot)$. Let $m^* = c^r/K$. For each z, we sample m^* X's according to $\mathbb{P}(X \in \cdot \,|Z = z)$ and set

$$\hat{\eta}(z) \stackrel{\triangle}{=} \frac{1}{2}\frac{\bar{X}_{m^*}(z)^2}{\frac{1}{m^*-1}\sum_{j=1}^{m^*}(X_j(z) - \bar{X}_{m^*}(z))^2}.$$

Step 2. *Estimate the optimal* $m_c(\cdot)$. Let $\tau^* \stackrel{\triangle}{=} \sum_z \hat{\eta}(z)^{-1}$. Set $m(z) \stackrel{\triangle}{=} \eta(z)^{-1}c/\tau^*$.

Step 3. *Determine* $\hat{\alpha}$. Set

$$\alpha_3(c) = \sum_z p(z) I\left(\bar{X}_{m(z)}(z) \leq x\right);$$

i.e., we sample $m(z)$ X's under the d.f. $\mathbb{P}(X \in \cdot \,|Z = z)$ and take $\alpha_3(c)$ as the weighted sum of the indicator functions with the weights being equal to $p(z)$'s.

We conclude this section by discussing the use of outcome-dependent sampling in the context of random elements Z for which the probability mass function is unknown. In this setting, we must resort to sampling the Z_i's, as for the estimator $\alpha_1(c)$. Here, a sampling plan requires assigning, for a given computer budget c, an "outer sample size" $n = n(c)$. If outcome z_i is sampled, then the "inner sample size" $m = m_c(z_i)$ is utilized. For n large, the amount of "inner sampling" at outcome z_i will then be approximately $np(z_i)m_c(z_i)$ by the LLN. Consequently, the sampling plan $\vec{m} = (m_c(z_i) : i \geq 1)$ and $n = n(c)$ must be selected so that $\sum_i p(z_i)m_c(z_i) \cdot n(c) \approx c$. This leads to the estimator

$$\alpha_4(c) = \frac{1}{n(c)}\sum_{i=1}^{n(c)} I(\bar{X}_{m_c(Z_i)} \leq x).$$

Here,

$$\mathbb{E}\alpha_4(c) - \alpha = \sum_{z\in\Gamma_+} p(z)\kappa_{m(z)}(z) - \sum_{z\in\Gamma_-} p(z)\kappa_{m(z)}(z)$$

and

$$\text{Var}\ \alpha_4(c) = \frac{1}{n(c)}(\mathbb{E}\alpha_4(c))(1 - \mathbb{E}\alpha_4(c)).$$

An analysis very similar to that given in Section 2 for $\alpha_1(c)$ establishes the following CLT.

Theorem 3.2 *Assume A1-A4. Suppose that for* $i \geq 1$, $m_c(z_i) \to \infty$ *and* $n(c) \to \infty$ *in such a way that* $n(c) \cdot \sum_i p(z_i)m_c(z_i)/c \to 1$ *as* $c \to \infty$. *If* $m_c(z_i) = \lfloor a(\log c)\rfloor/\eta(z_i)$ *as* $c \to \infty$ *where* $a \geq 1/2$, *then*

$$\sqrt{\frac{c}{\log c}}(\alpha_4(c) - \alpha) \Rightarrow \sqrt{a\nu\alpha(1-\alpha)}\, N(0,1)$$

as $c \to \infty$, *where* $\nu = \sum_i p(z_i)/\eta(z_i)$. *On the other hand, if* $m_c(z_i) = \lfloor a(\log c)/\eta(z)\rfloor$ *as* $c \to \infty$ *where* $0 < a < 1/2$, *then*

$$c^a\sqrt{\log c}(\alpha_4(c) - \alpha) \Rightarrow (\sum_i p(z_i)\gamma(z_i)\eta(z_i)^{1/2})/a^{1/2}$$

as $c \to \infty$.

Comparing Theorem 2.2 to Theorem 3.2, we see that the qualitative form of the convergence rates and limit structure is identical. Furthermore, Theorem 2.2 identifies the optimal mean-square error achievable for a given (large) value of c as approximately $(1/2\eta^*) \cdot (\log c/c) \cdot \alpha(1-\alpha)$, whereas Theorem 3.2's optimal mean-square error looks asymptotically like $(\nu/2) \cdot (\log c/c) \cdot \alpha(1-\alpha)$. Hence, the improvement obtained by using outcome-dependent sampling rates is asymptotically in proportion to η^*/ν.

As for the implementation of $\alpha_3(c)$, heuristics need to be applied, in order to circumvent the difficulties inherent in $\eta(\cdot)$ being unknown. The Gaussian heuristic suggested earlier in this section is one alternative.

When compared with that of Section 2, the analysis of this section suggests that use of outcome-dependent sampling, while an improvement on outcome independent sampling, tends not to lead to "orders of magnitude" improvement in convergence rates. In particular, the $\sqrt{\log c/c}$ convergence rate is characteristic of both types of estimators when the mass function $p(\cdot)$ is unknown to the simulationist (and must be estimated computationally). In view of the ease of applicability of $\alpha_1(c)$, as well as its very general domain of applicability, we recommend the use of this estimator in lieu of additional problem structure that may shift the choice elsewhere.

4 NUMERICAL RESULTS

In this section, we will report the numerical results of the algorithms proposed in the Section 2. The example we have used is as follows: Assume that the conditioning random element $Z \stackrel{\mathcal{D}}{=}$ binomial(10, 0.4) and that conditioned on $Z = z$, $X \stackrel{\mathcal{D}}{=} \mathrm{N}(z/2 - 2.3, 1)$. The exact value of $\alpha \stackrel{\triangle}{=} \mathbb{P}(\mathbb{E}(X|Z) \leq 0)$ is given by

$$\sum_{z=0}^{4} \mathbb{P}(Z = z) = \sum_{z=0}^{4} \binom{10}{z} 0.4^z 0.6^{10-z} = 0.6331.$$

The proposed algorithm presented in this paper was programmed in ANSI-C.We replicate the estimator 200 times. Denote by $\{\hat{\alpha}_i(c) : i = 1, \ldots, 200\}$ the values of the 200 replicates of the estimator $\alpha_1(c)$, given that the computational budget is equal to c. We estimate the mean, standard error, bias, and MSE of the estimator as follows:

mean: set $\bar{\alpha}(c) \stackrel{\triangle}{=} (1/200) \sum_{i=1}^{200} \hat{\alpha}_i(c)$;

s.e.: set $s_{\hat{\alpha}}(c) \stackrel{\triangle}{=} \sqrt{(200-1)^{-1} \sum_{i=1}^{200} (\hat{\alpha}_i(c) - \bar{\alpha}(c))^2}$;

bias: set $b_{\hat{\alpha}}(c) \stackrel{\triangle}{=} \bar{\alpha}(c) - \alpha$, where α is the exact theoretical value;

MSE: set $\mathrm{MSE}_{\hat{\alpha}}(c) \stackrel{\triangle}{=} (200)^{-1} \sum_{i=1}^{200} (\hat{\alpha}_i(c) - \alpha)^2$.

We choose $\nu = 0.2$ in this example and apply Algorithm2.1. Table 1 summarizes the numerical results.

To deduce the rate of convergence of the estimator, we plot the $\log(\mathrm{MSE}(c))$ vs. $\log c$ and the plot (Figure 1) turns out to be linear.

This suggests that $\mathrm{MSE}(c) \sim V c^{\lambda}$ for some constants V and λ. We can estimate $\log V$ and λ by the y-intercept of the plot and its slope respectively. The theoretical slope and intercept are equal to $-(1-0.2) = -0.8$ and 0.7680 resp.; whereas the empirical slope and intercept are equal to -0.78 and 1.29 respectively. The slope estimate matches the theoretical value quite well.

Table 1: Numerical Results for Algorithm 2.1

c	mean	s.d.	bias	$\log(\mathrm{MSE})/c$
1024	0.6454	0.1226	0.0123	-0.0041
2048	0.6475	0.0963	0.0144	-0.0023
4096	0.6454	0.0784	0.0123	-0.0012
8192	0.6386	0.0589	0.0055	-0.0007
16384	0.6383	0.0468	0.0052	-0.0004
32768	0.6347	0.0352	0.0016	-0.0002
65536	0.6332	0.0251	0.0001	-0.0001
131072	0.6332	0.0189	0.0001	-0.0001

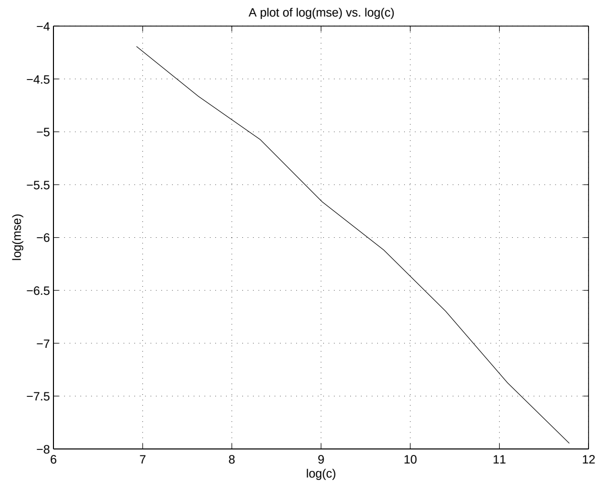

Figure 1: Distribution Function Estimator for the Discrete Case Example

Out of the 200 experiments, we tested the number of times, N, the confidence intervals covered the true value. The corresponding estimated coverage probability is then set to $\hat{p} \stackrel{\triangle}{=} N/200$. The standard error of the estimated coverage probability is given by $\sqrt{\hat{p}(1-\hat{p})/200}$ and is expressed inside the parenthesis beside the corresponding probability in the Table 2. All coverage probabilities converge to the correct values.

Table 2: Confidence Interval Coverage Probabilities of the Discrete Case Example

c	90% cov.	95% cov.	99% cov.
1024	0.89 (0.02)	0.91 (0.02)	0.97 (0.01)
2048	0.89 (0.02)	0.93 (0.02)	0.98 (0.01)
4096	0.89 (0.02)	0.91 (0.02)	0.99 (0.01)
8192	0.91 (0.02)	0.95 (0.02)	0.98 (0.01)
16384	0.88 (0.02)	0.94 (0.02)	0.98 (0.01)
32768	0.90 (0.02)	0.95 (0.02)	0.99 (0.01)
65536	0.90 (0.02)	0.97 (0.01)	0.99 (0.01)
131072	0.91 (0.02)	0.97 (0.01)	1.00 (0.00)

ACKNOWLEDGMENTS

Research partially supported by Army Research Office Grant No. DAAGSS-97-0377-P0001 and National Science Foundation Grant DMS-9704732-001.

APPENDIX

Lemma A-1 *Assume that the following conditions hold:*

1. *the r.v.'s* $(X_{c,j} : j = 1, 2, \ldots, n(c), c > 0)$ *is i.i.d. where* $n(c) \nearrow +\infty$ *as* $c \nearrow +\infty$*;*
2. $\mathbb{E}X_{c,1} = 0$, $\sigma_c^2 \stackrel{\triangle}{=} \mathbb{E}X_{c,1}^2$*;*
3. $\lim_{c\to\infty} \sigma_c^2 = \sigma^2 \in (0, \infty)$*;*
4. *the family* $\{X_{c,1}^2 : c > 0\}$ *is uniformly integrable.*

Then, $\{X_{c,i}^2 : i = 1, \ldots, n(c), c > 0\}$ *satisfies the Lindeberg-Feller condition; namely,*

$$\lim_{c\to\infty} \frac{1}{\sigma_c^2} \int_{|X_{c,1}|\geq\epsilon\sqrt{n(c)}\,\sigma_c} X_{c,1}^2 \, d\mathbb{P} = 0$$

for all $\epsilon > 0$.

Proof. We need to show that if Conditions 1–3 hold, then for all $\epsilon > 0$ and $\eta > 0$, there exists $C(\epsilon, \eta)$ such that for all $c \geq C(\epsilon, \eta)$,

$$\frac{1}{\sigma_c^2} \mathbb{E}\left[X_{c,1}^2; |X_{c,1}| \geq \epsilon\sqrt{n(c)}\,\sigma_c\right] < \eta.$$

By Condition 3, we know that there exists $\bar{C}$ such that for all $c > \bar{C}$, $\sigma_c^2 \in (\sigma^2/2, 3\sigma^2/2)$. Let $\xi = \sigma/\sqrt{2}$. Then, for all $c \geq \bar{C}$, we have that $\sigma_c^2 \geq \xi^2$. Now, by Condition 4 and the assumption that $n(c) \nearrow +\infty$ as $c \nearrow +\infty$, there exists $C(\epsilon, \eta) \geq \bar{C}$ such that

$$\mathbb{E}\left[X_{c,1}^2; |X_{c,1}| > \epsilon\sqrt{n(C(\epsilon,\eta))}\,\xi\right] < \eta\xi^2 \quad \forall c > 0.$$

Then, for all $c \geq C(\epsilon, \eta)$,

$$\begin{aligned}
& \frac{1}{\sigma_c^2} \mathbb{E}\left[X_{c,1}^2; |X_{c,1}| \geq \epsilon\sqrt{n(c)}\,\sigma_c\right] \\
\leq \quad & \frac{1}{\xi^2} \mathbb{E}\left[X_{c,1}^2; |X_{c,1}| \geq \epsilon\sqrt{n(c)}\,\sigma_c\right] \\
\leq \quad & \frac{1}{\xi^2} \mathbb{E}\left[X_{c,1}^2; |X_{c,1}| \geq \epsilon\sqrt{n(c)}\,\xi\right] \quad \text{since } c \geq \bar{C} \\
\leq \quad & \frac{1}{\xi^2} \mathbb{E}\left[X_{c,1}^2; |X_{c,1}| \geq \epsilon\sqrt{n(C(\epsilon,\eta))}\,\xi\right] \\
< \quad & \frac{1}{\xi^2} \cdot \eta\xi^2 = \eta.
\end{aligned}$$

Since $\epsilon > 0$ and $\eta > 0$ are arbitrary, we have proved the lemma.

REFERENCES

Billingsley, P. 1995. *Probability and Measure* 3rd-ed. New York: John Wiley and Sons.

Bucklew, J. 1990. *Large Deviation Techniques in Decision, Simulation and Estimation*. New York: John Wiley and Sons.

Duffie, D. 1996. *Dynamic Asset Pricing Theory*. 2nd ed. Princeton University Press, Princeton, New Jersey.

Lee, S.H. and Glynn, P.W. 1999. Computing the Distribution Function of a Conditional Expectation via Monte Carlo. Submitted for publication.

AUTHOR BIOGRAPHIES

SHINGHOI LEE is a research associate in the fixed-income research group of Morgan Stanley Dean Witter & Co. He received B.Sc. with Honors and M.Sc. degrees in Applied Mathematics from Auckland University; and he received a Ph.D. degree in Operations Research from Stanford University.

PETER W. GLYNN received his Ph.D. from Stanford University, after which he joined the faculty of the Department of Industrial Engineering at the University of Wisconsin-Madison. In 1987, he returned to Stanford, where he is the Thomas Ford Professor of Engineering in the Department of Engineering-Economic Systems and Operations Research. He was a co-winner of the 1993 Outstanding Simulation Publication Award sponsored by the TIMS college on Simulation and he is a Fellow of the Institute for Mathematical Statistics. His research interests include discrete-event simulation, computational probability, queueing, and general theory for stochastic systems.