Computing Bayesian Means Using Simulation

Sigrún Andradóttir, Georgia Institute of Technology
Peter W. Glynn, Stanford University

This paper is concerned with the estimation of $\alpha = E\{r(Z)\}$, where Z is a random vector and the function values $r(z)$ must be evaluated using simulation. Estimation problems of this form arise in the field of Bayesian simulation, where Z represents the uncertain (input) parameters of a system and $r(z)$ is the expected performance of the system when $Z = z$. Our approach involves obtaining (possibly biased) simulation estimates of the function values $r(z)$ for a number of different values of z, and then using a (possibly weighted) average of these estimates to estimate α. We start by considering the case where the chosen values of z are independent and identically distributed observations of the random vector Z (independent sampling). We analyze the resulting estimator as the total computational effort c grows and provide numerical results. Then we show that improved convergence rates can be obtained through the use of techniques other than independent sampling. Specifically, our results indicate that the use quasi-random sequences yields a better convergence rate than independent sampling, and that in the presence of a suitable special structure, it may be possible to use other numerical integration techniques (such as Simpson's rule) to achieve the best possible rate $c^{-1/2}$ as $c \to \infty$. Finally, we present and analyze a general framework of estimators for α that encompasses independent sampling, quasi-random sequences, and Simpson's rule as special cases.

Categories and Subject Descriptors: I.6.1 [**Simulation and Modeling**]: Simulation Theory; I.6.6 [**Simulation and Modeling**]: Simulation Output Analysis

General Terms: Measurement, Performance, Theory

Additional Key Words and Phrases: Bayesian simulation, nested simulation, independent sampling, quasi-random numbers, numerical integration, Simpson's rule, bias, consistency, convergence rate

ACM Reference Format:
Sigrún Andradóttir and Peter W. Glynn, 2013. Computing Bayesian Means Using Simulation. *ACM Trans. Model. Comput. Simul.* 9, 4, Article 39 (June 2013), 30 pages.
DOI:http://dx.doi.org/10.1145/0000000.0000000

1. INTRODUCTION

Consider the task of designing a manufacturing facility that can be modelled as a network of queues. Suppose that each of the interarrival and processing time distributions underlying the queueing network is assumed to be gamma. In this case, the network is characterized (statistically) by a parameter vector θ consisting of the scale and shape parameters for each of the underlying distributions.

The research of the first author was supported by the National Science Foundation under grants DMI–9523111, DMI–0000135, and DMI–0217860.
Author's addresses: S. Andradóttir, H. Milton Stewart School of Industrial and Systems Engineering, Georgia Institute of Technology, Atlanta, GA 30332-0205; Peter W. Glynn, Department of Management Science and Engineering, Stanford University, Stanford, CA 94305-4026; sa@gatech.edu and glynn@stanford.edu.

DOI:http://dx.doi.org/10.1145/0000000.0000000

Let $C_\theta(t)$ be the cost of running this facility over the time interval $[0, t]$ under θ. If the system is regenerative, then, in great generality, we have that

$$\frac{C_\theta(t)}{t} \Rightarrow \frac{E_\theta\{Y\}}{E_\theta\{\tau\}} \text{ as } t \to \infty,$$

where $\Rightarrow$ denotes weak convergence and $E_\theta\{Y\}$ and $E_\theta\{\tau\}$ are, respectively, the expected total cost (associated with running the facility) and the expected total time duration of a regenerative cycle under θ. The limit $E_\theta\{Y\}/E_\theta\{\tau\}$ is, of course, the long-run average cost per unit time associated with the facility having distributions determined by θ.

It is frequently the case that the exact value of θ is unknown prior to the operation of the facility. However, historical and subjective information often exists, permitting one to compute (using Bayesian methods) a (prior or posterior) distribution for θ. Given such a Bayesian framework, it is natural to wish to compute the mean steady-state cost, given by

$$\alpha = E\{r(\theta)\}, \tag{1}$$

where $r(\theta) = E_\theta\{Y\}/E_\theta\{\tau\}$ and the expectation appearing in equation (1) corresponds to an integration with respect to the distribution for θ.

This paper is concerned with the efficient computation of expectations like that appearing in equation (1). More generally, we will be concerned with the efficient numerical computation, via simulation, of expectations that can be expressed in the form:

$$\alpha = E\{r(Z)\}, \tag{2}$$

where Z is a random vector taking values in a set $\mathcal{Z}$ and having distribution μ, and the function $r(\cdot)$ is evaluated using simulation. Note that our motivating example (1) is precisely of this form, since the steady-state limit $r(\theta)$ is most naturally computed via a steady-state simulation of the network associated with the parameter vector θ. However, our problem formulation (2) is quite general, and is not restricted to regenerative systems. In fact, the use of Bayesian methods in a wide variety of stochastic modeling environments leads naturally to problems of the form given in equation (2), because it is typically the case that the performance measure $r(\cdot)$ of interest can not be expressed in closed form, and can only be computed via a simulation of the underlying system.

However, it should be pointed out that equation (2) also arises in other application settings. For example, suppose that it is of interest to compute $\alpha = E\{g(P(t))\}$, where $P(t)$ is the price at time t of a (derivative) option on some underlying security. The theory of option pricing asserts that, under quite general conditions, the price $P(t)$ can be expressed as a conditional expectation under an "equivalent martingale measure" in which the conditioning occurs with respect to the price $X(t)$ of the underlying asset at time t, see for example Duffie (1996). Thus, $g(P(t))$ may be re-expressed in the form $r(X(t))$. Of course, the function $r(\cdot)$ involves a conditional expectation that may be impossible to compute analytically. Repeated sampling under the equivalent martingale measure offers the opportunity to compute $r(\cdot)$ via simulation. Thus, this problem is a special case of our general framework (2).

Finally, note that the equality $E\{X\} = E\{E\{X|Z\}\}$ implies that any estimation problem of the form $\alpha = E\{X\}$ can be converted into an estimation problem of the form given in equation (2) by defining $r(Z) = E\{X|Z\}$. In the simulation literature, conditioning in this manner is generally used as a variance reduction technique, assuming that the conditional expected values $E\{X|Z\}$ can be computed exactly (or can be estimated more efficiently than through straightforward simulation). For an overview of the use of conditioning as a variance reduction technique, see for example Section 2.6

of Bratley, Fox, and Schrage (1987), Section 11.6 of Law and Kelton (2000), and Section V.4 of Asmussen and Glynn (2007).

For estimation problems of the form given in (2), it is natural to estimate α by generating observations $Z_1, \ldots, Z_m$ of Z from μ, using simulation to estimate $r(Z_i)$, where $i = 1, \ldots, m$, and finally averaging the m resulting estimates (possibly using different weights on the various estimates). For more discussion of this (nested, two-level, Bayesian) simulation approach, see for example Andradóttir and Bier (2000), Chick (2006), Lee and Glynn (2003), Steckley and Henderson (2003), Sun, Apley, and Staum (2011), Zouaoui and Wilson (2003), and the references therein. Recently, this simulation approach has been used for various finance applications, see for example Broadie, Du, and Moallemi (2011), Gordy and Juneja (2010), Lan, Nelson, and Staum (2010), and references therein.

In this paper, we determine the asymptotic behavior of the resulting estimator of α as the total available computational budget c grows, with focus on situations where the simulation estimates of the function values $r(Z_1), \ldots, r(Z_m)$ have some bias. Our motivation for allowing the estimates of the function values $r(Z_1), \ldots, r(Z_m)$ to be biased comes from quantile estimation problems and steady-state simulation problems (like the one described at the beginning of this section) with uncertain input parameters. We are particularly interested in determining the asymptotic convergence rate of the resulting estimator, and in investigating when it is possible to achieve the fastest possible convergence rate $c^{-1/2}$ in the expended computational effort c (this is the hoped for convergence rate because we estimate $r(Z_1), \ldots, r(Z_m)$ via simulation). Our proof approach will involve decomposing the error in our estimator into three components, namely:

(i) the noise associated with estimating $r(Z_1), \ldots, r(Z_m)$ via simulation;
(ii) the bias in the estimators of $r(Z_1), \ldots, r(Z_m)$; and
(iii) the error associated with the uncertainty about the value of Z, which is addressed by estimating α with a (possibly weighted) average of $r(Z_1), \ldots, r(Z_m)$.

We will identify the rate of convergence of each error component. The overall convergence rate will then be determined by the slowest of the convergence rates of the three error components.

The remainder of this paper is organized as follows. In Section 2, we present heuristic arguments that illustrate our main results. In Section 3, we consider independent sampling where the quantities $Z_1, \ldots, Z_m$ generated to estimate α are independent and identically distributed (i.i.d.) observations of the random variable Z. Both theoretical and numerical results about the asymptotic behavior of the resulting estimator are provided. In Section 4, we show that improved convergence rates can be achieved (relative to independent sampling) by using other approaches (specifically quasi-random numbers and Simpson's rule) to generate $Z_1, \ldots, Z_m$. In Section 5, we present and analyze a broad framework for estimating α via simulation that contains independent sampling, quasi-random numbers, and Simpson's rule as special cases. Finally, Section 6 contains some concluding remarks. For related research on confidence interval estimation, see for example Lan, Nelson, and Staum (2010) and references therein. An earlier version of this paper can be found in Andradóttir and Glynn (2002).

Although the focus of this paper is on situations where the estimates of the values of the function r are obtained using simulation, see Sections 3 and 4, it is also possible to use other numerical integration techniques (besides simulation) to estimate the values of the function r. The techniques used in this paper can be used to consider such approaches, but this is outside the scope of the present paper.

2. HEURISTIC ARGUMENTS

We want to estimate $\alpha = E\{r(Z)\}$, where $r(\cdot)$ is smooth and must be estimated by simulation for any given realization of Z, and Z has distribution μ. In the great majority of the applications we have in mind, the (outer) "integration" over Z is low-dimensional, whereas the (inner) "integration" (i.e., number of random variables needed to estimate $r(\cdot)$) is high dimensional. Therefore, although we only consider doing the inner integration by (Monte Carlo) simulation, the outer integration can be done by Monte-Carlo or non-Monte Carlo methods. A key methodological contribution of our paper is that our theory shows that the low-dimensional integration over Z should be done by non-Monte Carlo methods when the problem at hand is sufficiently smooth.

We start by considering the case where we use Monte Carlo to sample Z values from μ and $r(\cdot)$ can be evaluated without error. Then we can compute α using the unweighted estimator

$$\frac{1}{m}\sum_{i=1}^{m} r(Z_i) \simeq \alpha + \frac{\sigma W}{m^{1/2}},$$

where σ is the standard deviation of $r(Z)$ and W is a zero mean, unit variance normal random variable.

Turning next to the case in which $r(z)$ is to be estimated via a Monte Carlo estimator $\hat{r}_t(z)$, our estimator of α takes the form

$$\hat{\alpha}(c) = \frac{1}{m}\sum_{i=1}^{m} \hat{r}_t(Z_i),$$

where c denotes the available computer budget and t denotes the computational effort used to estimate $r(Z_i)$ for each i. We assume throughout that $c = m \times t$, so that the computational effort involved in obtaining $Z_1, \ldots, Z_m$ and other overhead is negligible relative to the effort associated with generating the estimates $\hat{r}_t(Z_i)$ for $i = 1, \ldots, m$. The resulting dependence of the parameters m and t on c is implicit in this "heuristic arguments" section, but will be made explicit in the rigorous derivations in subsequent sections. Suppose now that

$$\hat{r}_t(z) \simeq r(z) + \frac{\sigma(z)G(z)}{t^{\eta}} + \frac{b(z)}{t^{\gamma}} \tag{3}$$

for all $z \in \mathcal{Z}$, where $G(\cdot)$ is a mean zero, unit variance random field, independent of W, with $G(z)$ being independent of $G(z')$ for $z \neq z'$, $\sigma(\cdot)$ and $b(\cdot)$ are functions, and $\eta \leq 1/2$. (Typically, $\eta = 1/2$. The possibility that $\eta < 1/2$ would arise, for example, when the inner integration involves a stochastic differential equation, as in the finance setting; see Duffie and Glynn (1995) for the appropriate convergence rates.) Then

$$\begin{aligned}
\hat{\alpha}(c) &\simeq \frac{1}{m}\sum_{i=1}^{m} r(Z_i) + \frac{1}{mt^{\eta}}\sum_{i=1}^{m}\sigma(Z_i)G(Z_i) + \frac{1}{mt^{\gamma}}\sum_{i=1}^{m} b(Z_i) \\
&\simeq \alpha + \frac{\sigma W}{m^{1/2}} + \frac{1}{mt^{\eta}}\sqrt{\sum_{i=1}^{m}\sigma^2(Z_i)}\,G + \frac{b}{t^{\gamma}} \\
&= \alpha + \frac{\sigma W}{m^{1/2}} + \sqrt{\frac{1}{m}\sum_{i=1}^{m}\sigma^2(Z_i)}\,\frac{G}{m^{1/2}t^{\eta}} + \frac{b}{t^{\gamma}} \\
&\simeq \alpha + \frac{\sigma W}{m^{1/2}} + \frac{\sigma G}{m^{1/2}t^{\eta}} + \frac{b}{t^{\gamma}},
\end{aligned}$$

where $b = E\{b(Z)\}$, $\sigma^2(\cdot) = [\sigma(\cdot)]^2$, and G is a zero mean, unit variance normal random variable independent of W. If m is of the order c^w and t is of the order c^{1-w}, then the rate of convergence of $\hat{\alpha}(c)$ is $c^{-\rho}$, where

$$\rho = \min(w/2, w/2 + (1-w)\eta, (1-w)\gamma). \tag{4}$$

In the most common case, $\gamma = 1$ and $\eta = 1/2$, so the minimum rate is $\rho^* = 1/3$, which is attained when $w = 2/3$.

We now consider the case where $\alpha = \int r(z)f(z)dz$, f is the density of Z and is assumed to be known, the functions r, σ, b, and f are smooth, and the outer integral is to be evaluated using non-Monte Carlo methods. If we could evaluate $r(\cdot)$ without error, we could compute the outer integral for α via a sum of the form

$$\sum_{i=1}^{m} w(z_i)r(z_i)f(z_i)$$

for some sequence of points $(z_1, \ldots, z_m)$. The particular points $z_1, \ldots, z_m$ could depend on m (as in a quadrature rule) or $z_1, \ldots, z_m$ could be the first m points in an infinite sequence (which would be more natural when one is sequentially refining the estimator to achieve a given accuracy). The points $z_1, \ldots, z_m$ could be selected randomly (in which case we would denote them typically by $Z_1, \ldots, Z_m$) or they could be selected non-randomly. The "weights" $w(z_i)$ could be identically equal to $1/m$ or they could be non-constant; when the weights are non-constant, they would typically depend on m.

Suppose that we know that for the particular non-Monte Carlo integration rule (characterized by the weights and points) and for suitably "smooth" (or "regular") integrands $r(\cdot)f(\cdot)$, the estimate of α satisfies

$$\sum_{i=1}^{m} w(z_i)r(z_i)f(z_i) = \alpha + E_m,$$

where E_m is the error. For quasi-random methods, the weights are $1/m$ and E_m is generally of order $[\log m]^{\iota}/m^{\beta}$, where $\iota, \beta \in R^+$. For quadrature methods with r sufficiently smooth, E_m would typically be of order $1/m^p$, where $p \in \mathbb{R}^+$; the weights would be of order $1/m$.

Turning next to the case in which $r(z)$ is to be estimated via a Monte Carlo estimator $\hat{r}_t(z)$, our estimator takes the form

$$\hat{\alpha}(c) = \sum_{i=1}^{m} w(z_i)\hat{r}_t(z_i)f(z_i).$$

Then,

$$\begin{aligned}
\hat{\alpha}(c) &\simeq \sum_{i=1}^{m} w(z_i)r(z_i)f(z_i) + \frac{1}{t^\eta}\sum_{i=1}^{m} w(z_i)\sigma(z_i)f(z_i)G(z_i) + \frac{1}{t^\gamma}\sum_{i=1}^{m} w(z_i)b(z_i)f(z_i) \\
&\simeq \alpha + E_m + \frac{1}{t^\eta}\sqrt{\sum_{i=1}^{m} w(z_i)^2\sigma^2(z_i)f(z_i)^2}\,G + \frac{b}{t^\gamma} + \frac{E'_m}{t^\gamma} \\
&\simeq \alpha + E_m + \sqrt{m\sum_{i=1}^{m} w(z_i)^2\sigma^2(z_i)f(z_i)^2}\,\frac{G}{m^{1/2}t^\eta} + \frac{b}{t^\gamma} \\
&= \alpha + E_m + I_m + \frac{b}{t^\gamma},
\end{aligned} \tag{5}$$

where E'_m is the error in the integration rule for b. (The approximation uses the fact E'_m is always smaller than at least one of the terms preceding, so will never determine the rate of convergence.)

So if we use quasi-random methods with m of the order c^w and t of the order c^{1-w}, the weights are all equal to $1/m$ and we obtain that E_m is of order $[\log c]^\iota / c^{w\beta}$, I_m is of order $1/c^{w/2+(1-w)\eta}$, and the final term in (5) is of order $1/c^{(1-w)\gamma}$. Up to logarithmic terms, the rate of convergence of $\hat{\alpha}(c)$ equals $c^{-\rho}$, where

$$\rho = \min(w\beta, w/2 + (1-w)\eta, (1-w)\gamma). \tag{6}$$

In the classical case where $\beta = 1$, $\gamma = 1$, and $\eta = 1/2$, the minimum rate is $\rho^* = 1/2$ (up to logarithmic terms) if $w = 1/2$, so we get the canonical Monte Carlo rate of convergence. Hence, we definitely do not want to sample the outer integral.

For a quadrature method with weights of order $1/m$ and rate of convergence $1/m^p$, if m is of order c^w and t is of order c^{1-w}, we have that E_m is of order $1/c^{wp}$, I_m is of order $1/c^{w/2+(1-w)\eta}$, and the final term in (5) is of order $1/c^{(1-w)\gamma}$. The rate of convergence of $\hat{\alpha}(c)$ is clearly given by $c^{-\rho}$, where

$$\rho = \min(wp, w/2 + (1-w)\eta, (1-w)\gamma). \tag{7}$$

(The previous case corresponds to $p = \beta - \epsilon$ for all $\epsilon > 0$.) If $p = 4$ (for example), $\gamma = 1$, and $\eta = 1/2$, the minimum rate is $\rho^* = 1/2$ (with w between 1/8 and 1/2). Here, the rate is faster by a logarithmic factor, and the range of good w values does not contract to a single value (unlike the classical case considered in the previous paragraph, where only $w = 1/2$ yielded the rate $1/2$ up to logarithmic terms).

The above discussion of quasi-random and quadrature methods assumed that Z has a known density f. If f is unknown, then quadrature rules cannot be used, but quasi-random methods can be used to sample the outer integral. Under appropriate smoothness assumptions, the rate of convergence for the known f case still applies, so again it is better to use quasi-random numbers on the outer integral.

Note that the first, second, and third terms in equations (4), (6), and (7) correspond to the error associated with the uncertainty about the value of Z, the noise associated with estimating $r(Z_1), \ldots, r(Z_m)$ via simulation, and the bias in the estimators of $r(Z_1), \ldots, r(Z_m)$, respectively, see items (iii), (i), and (ii) in Section 1. Clearly, if there is no uncertainty about the value of Z (as would be the case in a typical steady-state simulation), then the first term vanishes. Similarly, if the values of $r(Z_1), \ldots, r(Z_m)$ can be estimated without noise or without bias, then the second or third terms vanish.

We would like to point out that our results hold when $\alpha = E\{g(r(Z))\}$, where g is a known, smooth function; just put $r'(z) = g(r(z))$ and apply our theory. This will, for example, allow us to estimate expected values of functions of steady-state performance under parameter uncertainty. Moreover, when smoothness in g is violated, some of our assumptions can break down. Such an example (of practical interest) is when g is an indicator function, say the indicator of the interval $(-\infty, x]$. One key element that breaks down is the bias expansion for $E\{g(r(Z))\}$ (see (3)); note that for z with $r(z) = x$, the bias is of order 1. So, our paper does not cover such examples; a different theory is needed (see, e.g., Lee and Glynn, 2003).

In the remainder of this paper, we will fill in the rigorous details of the above heuristic arguments and provide more thorough analysis and discussion of our results. We will start by considering some special cases in Sections 3 and 4, and then analyze a general framework in Section 5.

3. INDEPENDENT SAMPLING

In this section, we analyze the estimator of the unknown quantity α obtained by generating i.i.d. samples of the random vector Z and averaging simulation estimates of the values of the function r at the sampled values of Z. Given a total available computational budget $c \in \mathbb{R}^+$, let $m(c) \in \mathbb{N}$ be the number of different values of the random vector Z used in the estimation of α, and let $t(c) \in \mathbb{R}^+$ be the (constant) computational effort expended to obtain the estimate $\hat{r}_{t(c)}(Z_i)$ of $r(Z_i)$ for each $i = 1, \ldots, m(c)$. As in Section 2, we assume that $c = m(c) \times t(c)$, and the resulting dependence of $m(c)$ and $t(c)$ on c is now indicated explicitly. (Note that the processes $\hat{r}_t(\cdot)$, where $t \geq 0$, may also depend on i. However, we believe that the fact that our notation does not explicitly show this dependence will not confuse the reader.) Our estimator of α obtained with the computational budget c is then given by

$$\hat{\alpha}(c) = \frac{1}{m(c)} \sum_{i=1}^{m(c)} \hat{r}_{t(c)}(Z_i), \tag{8}$$

where $Z_1, \ldots, Z_{m(c)}$ are independent observations of the random variable Z.

This section is organized as follows. We first analyze the asymptotic behavior of the estimator (8) as the computational budget c grows in Section 3.1. We then study the behavior of the estimator (8) for finite computational budgets c in Section 3.2.

3.1. Theoretical results

In this section, we study the asymptotic behavior of the estimator $\hat{\alpha}(c)$ defined in equation (8) as $c \to \infty$ when $Z_1, \ldots, Z_{m(c)}$ are independent observations of the random vector Z. We first show that in order for $\hat{\alpha}(c)$ to be an asymptotically unbiased and consistent estimator of α as $c \to \infty$, we generally need both $t(c) \to \infty$ and $m(c) \to \infty$ as $c \to \infty$ (Propositions 3.1 and 3.2 and Theorem 3.3). Then we present the main result in this section (Theorem 3.4), which identifies the convergence rate of $\hat{\alpha}(c)$ to α as $c \to \infty$ as a function of the respective growth rates of $t(c)$ and $m(c)$ with c. The following assumption describes more precisely the situation considered in this section.

ASSUMPTION 3.1. *Assume that:*

(i). The random variable $r(Z)$ is integrable (implying that $|\alpha| < \infty$).
(ii). For all $c \in \mathbb{R}^+$, the parameters $m(c)$ and $t(c)$ satisfy $c = m(c) \times t(c)$.
(iii). The random variables Z_i, where $i \in \mathbb{N}$, are independent observations of the random variable Z.
(iv). For all $t \in \mathbb{R}^+$, the random variables $\hat{r}_t(Z_i)$, where $i \in \mathbb{N}$, are independent observations of the random variable $\hat{r}_t(Z)$.
(v). For all $t \in \mathbb{R}^+$ and $m \in \mathbb{N}$, the random numbers used to generate the estimators $\hat{r}_t(Z_1), \ldots, \hat{r}_t(Z_m)$ are independent of the values of $Z_1, \ldots, Z_m$.

For all $x \in \mathbb{R}$, let $\lfloor x \rfloor$ denote the integer part of x. The following two propositions are concerned with the bias and consistency of $\hat{\alpha}(c)$ as $c \to \infty$ and either $t(c)$ or $m(c)$ remains constant. The proofs of these propositions are straightforward, and are omitted.

PROPOSITION 3.1. *Suppose that Assumption 3.1 holds, that $t(c) = t > 0$ for all $c \geq 0$ and that $|E\{\hat{r}_t(Z)\}| < \infty$. Then $m(c) = \lfloor c/t \rfloor$ for all $c \geq 0$, $E\{\hat{\alpha}(c)\} = E\{\hat{r}_t(Z)\}$ for all $c \geq 0$, and $\hat{\alpha}(c) \to E\{\hat{r}_t(Z)\}$ almost surely as $c \to \infty$.*

PROPOSITION 3.2. *Suppose that Assumption 3.1 holds, that $m(c) = m > 0$ for all $c \geq 0$, and that $\hat{r}_t(z) \Rightarrow r(z)$ as $t \to \infty$ for all $z \in \mathcal{Z}$. Then $t(c) = c/m$ for all $c \geq 0$, $E\{\hat{\alpha}(c)\} = E\{\hat{r}_{c/m}(Z)\}$ for all $c \geq 0$, and $\hat{\alpha}(c) \Rightarrow \sum_{i=1}^{m} r(Z_i)/m$ as $c \to \infty$, where $Z_1, \ldots, Z_m$ are independent observations of the random variable Z. If also the set of random variables $\{\hat{r}_{c/m}(Z) : c \geq 0\}$ is uniformly integrable, then $E\{\hat{\alpha}(c)\} \to E\{r(Z)\} = \alpha$ as $c \to \infty$.*

Propositions 3.1 and 3.2 show that in order for $\hat{\alpha}(c)$ to be a consistent estimator for α as $c \to \infty$, it is generally necessary to have that as $c \to \infty$, both $t(c) \to \infty$ and $m(c) \to \infty$. We now turn our attention to this case. For all $z \in \mathcal{Z}$ and $t \geq 0$, let $r_t(z) = E\{\hat{r}_t(z)\} = E\{\hat{r}_t(Z)|Z = z\}$ (see part (v) of Assumption 3.1). We will need the following assumption:

ASSUMPTION 3.2. *Assume that:*

(i). The set of random variables $\{[\hat{r}_t(Z) - r_t(Z)]^2 : t \geq 0\}$ *is uniformly integrable.*
(ii). The random variables $\hat{r}_t(z)$ *satisfy* $\hat{r}_t(z) \Rightarrow r(z)$ *as* $t \to \infty$ *for all* $z \in \mathcal{Z}$.
(iii). The function r_t *satisfies* $r_t(z) = r(z) + b(z)/t^\gamma + e(z)o(1/t^\gamma)$ *as* $t \to \infty$ *for all* $z \in \mathcal{Z}$*, where* $b(z), e(z) \in \mathbb{R}$ *for all* $z \in \mathcal{Z}$*,* $\gamma > 0$*, and the* $o(1/t^\gamma)$ *term is uniform in* $z \in \mathcal{Z}$.
(iv). The random variables $b(Z)$ *and* $e(Z)$ *are integrable.*

The following result is concerned with the bias and consistency of the estimator $\hat{\alpha}(c)$ when both $t(c) \to \infty$ and $m(c) \to \infty$ as $c \to \infty$.

THEOREM 3.3. *Suppose that Assumption 3.1 and parts (ii), (iii), and (iv) of Assumption 3.2 hold, that the set of random variables* $\{\hat{r}_t(Z) - r_t(Z) : t \geq 0\}$ *is uniformly integrable, and that as* $c \to \infty$*, both* $m(c) \to \infty$ *and* $t(c) \to \infty$*. Then* $E\{\hat{\alpha}(c)\} = E\{\hat{r}_{t(c)}(Z)\} \to E\{r(Z)\} = \alpha$ *and* $\hat{\alpha}(c) \to \alpha$ *in probability as* $c \to \infty$.

Proof: It is clear that by part (v) of Assumption 3.1, parts (ii) and (iii) of Assumption 3.2, and the fact that $t(c) \to \infty$ as $c \to \infty$, we have that $\hat{r}_{t(c)}(Z) - r_{t(c)}(Z) \Rightarrow 0$ as $c \to \infty$. Therefore, part (iv) of Assumption 3.1, parts (iii) and (iv) of Assumption 3.2, the uniform integrability of the random variables $\hat{r}_t(Z) - r_t(Z)$, where $t \geq 0$, and the fact that $t(c) \to \infty$ as $c \to \infty$ give that

$$E\{\hat{\alpha}(c)\} - \alpha = E\{\hat{r}_{t(c)}(Z) - r_{t(c)}(Z)\} + E\{r_{t(c)}(Z) - r(Z)\} \to 0$$

as $c \to \infty$. In the remainder of the proof, we show that $\hat{\alpha}(c) \to \alpha$ in probability as $c \to \infty$.

For all $c \geq 0$, we clearly have that

$$\hat{\alpha}(c) - \alpha = \hat{\alpha}_1(c) + \hat{\alpha}_2(c) + \hat{\alpha}_3(c), \tag{9}$$

where

$$\begin{aligned}
\hat{\alpha}_1(c) &= \frac{1}{m(c)} \sum_{i=1}^{m(c)} [\hat{r}_{t(c)}(Z_i) - r_{t(c)}(Z_i)], \\
\hat{\alpha}_2(c) &= \frac{1}{m(c)} \sum_{i=1}^{m(c)} [r_{t(c)}(Z_i) - r(Z_i)], \quad \text{and} \\
\hat{\alpha}_3(c) &= \frac{1}{m(c)} \sum_{i=1}^{m(c)} [r(Z_i) - \alpha].
\end{aligned}$$

By parts (i) and (iii) of Assumption 3.1, the strong law of large numbers, and the fact that $m(c) \to \infty$ as $c \to \infty$, it is clear that $\hat{\alpha}_3(c) \to 0$ almost surely as $c \to \infty$. Moreover, part (iii) of Assumption 3.2 implies that

$$\hat{\alpha}_2(c) = \frac{1}{m(c)t(c)^\gamma} \sum_{i=1}^{m(c)} [b(Z_i) + e(Z_i)o(1)]. \tag{10}$$

Part (iii) of Assumption 3.1, parts (iii) and (iv) of Assumption 3.2, the strong law of large numbers, and the facts that $\gamma > 0$ and as $c \to \infty$, both $m(c) \to \infty$ and $t(c) \to \infty$, now imply that $\hat{\alpha}_2(c) \to 0$ almost surely as $c \to \infty$. Finally, note that Markov's inequality and parts (iii) and (v) of Assumption 3.1 imply that for all $\epsilon > 0$,

$$P\{|\hat{\alpha}_1(c)| > \epsilon\} \leq \frac{1}{\epsilon} \times E\{|\hat{\alpha}_1(c)|\} \leq \frac{1}{\epsilon} \times E\{|\hat{r}_{t(c)}(Z) - r_{t(c)}(Z)|\}.$$

The facts that $t(c) \to \infty$ as $c \to \infty$, that the random variables $\hat{r}_t(Z) - r_t(Z)$ are uniformly integrable, and that $\hat{r}_{t(c)}(Z) - r_{t(c)}(Z) \Rightarrow 0$ as $c \to \infty$ yield that $\hat{\alpha}_1(c) \to 0$ in probability as $c \to \infty$. The result now follows from (9). □

In the proof of Theorem 3.3, note that the terms $\hat{\alpha}_1(c)$, $\hat{\alpha}_2(c)$, and $\hat{\alpha}_3(c)$ in the decomposition (9) of the error in the estimator $\hat{\alpha}(c)$ correspond to the noise associated with estimating $r(Z_1), \ldots, r(Z_m)$ via simulation, the bias in the estimators of $r(Z_1), \ldots, r(Z_m)$, and the error associated with the uncertainty about the value of Z, respectively, see items (i), (ii), and (iii) in Section 1.

We are now ready to present the main result in this section. Let $b = E\{b(Z)\}$, where the function $b(\cdot)$ is defined in part (iii) of Assumption 3.2 (see also equation (3)). Moreover, for all $x, y \in \mathbb{R}$, let $N(x, y^2)$ denote the normal distribution with mean x and variance y^2 (if $y = 0$ then $N(x, y^2)$ equals x). Theorem 3.4 establishes that as the total available computational effort c grows, the estimator $\hat{\alpha}(c)$ is asymptotically normal. Theorem 3.4 also provides the rate at which the estimator $\hat{\alpha}(c)$ converges to α as $c \to \infty$ for different growth rates of $m(c)$ with c.

THEOREM 3.4. *Suppose that Assumptions 3.1 and 3.2 hold and that the random variable Z satisfies $\sigma^2 = \mathrm{Var}\{r(Z)\} < \infty$. Then, the following statements hold:*

(a). Assume that when $c \to \infty$, we have that $m(c)/c^{2\gamma/(2\gamma+1)} \to \infty$. Then

$$t(c)^{\gamma}(\hat{\alpha}(c) - \alpha) \Rightarrow b \quad \textit{as} \quad c \to \infty.$$

(b). Assume that when $c \to \infty$, we have that $m(c)/c^{2\gamma/(2\gamma+1)} \to 0$. Then

$$m(c)^{1/2}(\hat{\alpha}(c) - \alpha) \Rightarrow N(0, \sigma^2) \quad \textit{as} \quad c \to \infty.$$

(c). Assume that when $c \to \infty$, we have that $m(c)/c^{2\gamma/(2\gamma+1)} \to \ell$, where $0 < \ell < \infty$. Then

$$c^{\gamma/(2\gamma+1)}(\hat{\alpha}(c) - \alpha) \Rightarrow N(\ell^{\gamma} b, \sigma^2/\ell) \quad \textit{as} \quad c \to \infty.$$

Proof: For all $c \geq 0$, let $\hat{\alpha}_1(c)$, $\hat{\alpha}_2(c)$, and $\hat{\alpha}_3(c)$ be defined as in the proof of Theorem 3.3. Let $\epsilon > 0$. Observe that parts (iii) and (iv) of Assumption 3.1 give that

$$m(c)E\{(\hat{\alpha}_1(c))^2\} = E\{[\hat{r}_{t(c)}(Z) - r_{t(c)}(Z)]^2\}. \tag{11}$$

We start by considering part *(a)*. By Markov's inequality, we have

$$P\{|t(c)^{\gamma}\hat{\alpha}_1(c)| > \epsilon\} \leq \frac{t(c)^{2\gamma}}{\epsilon^2 m(c)} \times m(c)E\{(\hat{\alpha}_1(c))^2\}.$$

Equation (11), part (i) of Assumption 3.2, and the facts that $c = m(c) \times t(c)$ and $m(c)/c^{2\gamma/(2\gamma+1)} \to \infty$ as $c \to \infty$ now show that $t(c)^{\gamma}\hat{\alpha}_1(c) \to 0$ in probability as $c \to \infty$. Moreover, from equation (10), the strong law of large numbers, part (iii) of Assumption 3.1, parts (iii) and (iv) of Assumption 3.2, and the fact that $m(c) \to \infty$ as $c \to \infty$, it is clear that $t(c)^{\gamma}\hat{\alpha}_2(c) \to b$ almost surely as $c \to \infty$. Finally, part (iii) of Assumption 3.1 and the facts that $\sigma^2 < \infty$ and $m(c) \to \infty$ as $c \to \infty$ clearly imply that

$$m(c)^{1/2}\hat{\alpha}_3(c) \Rightarrow N(0, \sigma^2) \quad \text{as} \quad c \to \infty. \tag{12}$$

Together with the facts that $c = m(c) \times t(c)$ and $m(c)/c^{2\gamma/(2\gamma+1)} \to \infty$ as $c \to \infty$, this shows that $t(c)^{\gamma}\hat{\alpha}_3(c) \Rightarrow 0$ as $c \to \infty$. Equation (9) now gives the result of part *(a)*.

Since $t(c) \to \infty$ as $c \to \infty$ in parts *(b)* and *(c)*, it follows from Assumptions 3.1 and 3.2 that $[\hat{r}_{t(c)}(Z) - r_{t(c)}(Z)]^2 \Rightarrow 0$ as $c \to \infty$ (see the proof of Theorem 3.3 for a similar argument). Hence, part (i) of Assumption 3.2 and equation (11) yield that

$$m(c)E\{(\hat{\alpha}_1(c))^2\} \to 0 \quad \text{as} \quad c \to \infty. \tag{13}$$

We now consider part *(b)*. Note that Markov's inequality gives

$$P\{|m(c)^{1/2}\hat{\alpha}_1(c)| > \epsilon\} \leq \frac{1}{\epsilon^2} \times m(c)E\{(\hat{\alpha}_1(c))^2\}.$$

Equation (13) now shows that $m(c)^{1/2}\hat{\alpha}_1(c) \to 0$ in probability as $c \to \infty$. Moreover, from equation (10), the strong law of large numbers, parts (ii) and (iii) of Assumption 3.1, parts (iii) and (iv) of Assumption 3.2, and the fact that $m(c)/c^{2\gamma/(2\gamma+1)} \to 0$ as $c \to \infty$, it is clear that $m(c)^{1/2}\hat{\alpha}_2(c) \to 0$ almost surely as $c \to \infty$. Putting the above together with equations (9) and (12) gives the result of part *(b)*.

Finally, for part *(c)*, note that Markov's inequality gives

$$P\{|c^{\gamma/(2\gamma+1)}\hat{\alpha}_1(c)| > \epsilon\} \leq \frac{c^{2\gamma/(2\gamma+1)}}{\epsilon^2 m(c)} \times m(c)E\{(\hat{\alpha}_1(c))^2\}.$$

Equation (13) and the fact that $m(c)/c^{2\gamma/(2\gamma+1)} \to \ell > 0$ as $c \to \infty$ now show that $c^{\gamma/(2\gamma+1)}\hat{\alpha}_1(c) \to 0$ in probability as $c \to \infty$. Moreover, from equation (10), the strong law of large numbers, part (ii) and (iii) of Assumption 3.1, parts (iii) and (iv) of Assumption 3.2, and the fact that $m(c)/c^{2\gamma/(2\gamma+1)} \to \ell > 0$ as $c \to \infty$, it is clear that $c^{\gamma/(2\gamma+1)}\hat{\alpha}_2(c) \to \ell^{\gamma}b$ almost surely as $c \to \infty$. Finally, equation (12) and the facts that $c = m(c) \times t(c)$ and $m(c)/c^{2\gamma/(2\gamma+1)} \to \ell > 0$ as $c \to \infty$ also show that $c^{\gamma/(2\gamma+1)}\hat{\alpha}_3(c) \Rightarrow N(0, \sigma^2/\ell)$ as $c \to \infty$. Putting the above together with equation (9) gives the result of part *(c)*. □

Remark 3.5. *If in part (a) of Theorem 3.4 we let $t(c) = \kappa c^{\delta}$ for all $c \geq 0$, where $\kappa, \delta > 0$, then the facts that $c = m(c) \times t(c)$ and $m(c)/c^{2\gamma/(2\gamma+1)} \to \infty$ as $c \to \infty$ imply that $\delta < 1/(2\gamma+1)$. Similarly, if in part (b) of Theorem 3.4 we let $m(c) = \lfloor \kappa c^{\delta} \rfloor$ for all $c \geq 0$, where $\kappa, \delta > 0$, then the fact that $m(c)/c^{2\gamma/(2\gamma+1)} \to 0$ as $c \to \infty$ implies that $\delta < 2\gamma/(2\gamma+1)$. Therefore, Theorem 3.4 shows that the maximum convergence rate for the estimator $\hat{\alpha}(c)$ is of the order of $1/c^{\gamma/(2\gamma+1)}$, with $m(c)$ growing at the rate $c^{2\gamma/(2\gamma+1)}$ and $t(c)$ growing at the rate $c^{1/(2\gamma+1)}$ as $c \to \infty$. Note that $\gamma/(2\gamma+1) < 1/2$ for all $\gamma > 0$, that $\gamma/(2\gamma+1)$ increases with γ, and that $\gamma/(2\gamma+1) \to 1/2$ as $\gamma \to \infty$ (i.e., as the bias in the estimators $\hat{r}_t(z)$, where $t \geq 0$ and $z \in \mathcal{Z}$, is reduced, see part (iii) of Assumption 3.2). Finally, note that as long as $\gamma > 1/2$ (as would typically be the case in practice), the maximum convergence rate $1/c^{\gamma/(2\gamma+1)}$ is obtained by letting $m(c)$ grow at a faster rate than $t(c)$ as $c \to \infty$.*

Remark 3.6. *It is frequently the case that simulation estimators obtained from a sample path of length t have a principal bias term of the order $1/t$; see for example Glynn and Heidelberger (1992) and Awad and Glynn (2007) for conditions that guarantee this. This suggests that the special case when $\gamma = 1$ is of particular interest. We have shown that when $\gamma = 1$, then the best possible convergence rate of $\hat{\alpha}(c)$ to α is $1/c^{1/3}$, which is a considerably slower convergence rate than $1/c^{1/2}$, the best possible convergence rate expected in a simulation environment. However, better convergence rates can be obtained through the use of bias reduction techniques such as jackknifing that remove the highest order bias term; see for example Section 2.7 of Bratley, Fox, and Schrage*

(1987), Appendix 9A of Law and Kelton (2000), and Glynn and Heidelberger (1992) for an introduction to the jackknifing bias reduction technique and Awad and Glynn (2007) for a discussion of low-bias steady-state estimators. The use of these techniques with a sample path of length t would typically yield simulation estimates with a principal bias term of the order $1/t^2$, corresponding to $\gamma = 2$. Our results then show that the best possible convergence rate of $\hat{\alpha}(c)$ to α is $1/c^{2/5}$, which is a considerable improvement over the convergence rate $1/c^{1/3}$ obtained previously, but nevertheless substantially worse than the desired convergence rate $1/c^{1/2}$. In Section 4 below, we discuss other estimation techniques than can achieve the desired convergence rate $1/c^{1/2}$ even when $\gamma = 1$.

3.2. Numerical results

In this section, we provide insights into the behavior of the independent-sampling estimator $\hat{\alpha}(c)$ given in equation (8) for finite computational budgets c. The specific example that we consider involves an autoregressive process $\{X_n\}$ of order one with an unknown multiplier Z that is believed to be uniformly distributed on the interval $[0.1, 0.5]$. More specifically, suppose that

$$X_{n+1} = Z \times X_n + \epsilon_n$$

for all $n \geq 0$, where $X_0 \in \{0.1, 1, 10, 100\}$ is a scalar, Z is uniformly distributed on the interval $[0.1, 0.5]$, and $\epsilon_1, \epsilon_2, \ldots$ are $N(0, 1)$ random variables that are independent of each other and of Z. We are interested in estimating the steady-state mean $\alpha = E\{r(Z)\}$, where for all $z \in [0.1, 0.5]$, we have $r(z) = z + r'(z)$ and $r'(z)$ is the steady-state mean of the autoregressive process $\{X_n\}$ given that $Z = z$. It is clear that both the functions $r(\cdot)$ and $r'(\cdot)$, and hence also the scalar α, can be computed analytically. In particular, $r'(z) = 0$ and $r(z) = z$ for all $z \in [0.1, 0.5]$, and hence $\alpha = 0.3$. This facilitates using this example to illustrate the approach and results discussed in Section 3.1.

For all $z \in [0.1, 0.5]$, let $\{X_n(z)\}$ represent the autoregressive process $\{X_n\}$ given that $Z = z$, and let the total computational budget c be measured in terms of the maximum number of normal random variables that can be generated in the numerical experiment. Consider estimators $\hat{\alpha}(c) = \hat{\alpha}_{v,\ell}(c)$ of α of the form given in equation (8), where $Z_1, \ldots, Z_{m(c)}$ are sampled at random from the uniform distribution with range $[0.1, 0.5]$, $\hat{r}_t(z) = z + \sum_{n=0}^{t} X_n(z)/t$ for all $z \in [0.1, 0.5]$ and $t \in \mathbb{N}$, c satisfies $c = c' \times 1{,}000{,}000$ with $c' \in \{10, 20, 50, 100, 200, 300, \ldots, 1000\}$, and $m(c) = \ell c^{v}$ and $t(c) = c^{1-v}/\ell$, where $0 < v < 1$ and $\ell > 0$. Then it is clear that parts (i) through (iii) of Assumption 3.1 are satisfied, and we conduct the simulation in such a way that parts (iv) and (v) of Assumption 3.1 hold. Moreover, it is not difficult to show that

$$r_t(z) = E\{\hat{r}_t(z)\} = z + \frac{X_0}{t} \times \frac{1 - z^{t+1}}{1 - z} \tag{14}$$

for all $z \in [0.1, 0.5]$ and $t \in \mathbb{N}$, so that parts (iii) and (iv) of Assumption 3.2 are satisfied with $\gamma = 1$ and $b(z) = -e(z) = X_0/(1 - z)$ for all $z \in [0.1, 0.5]$. Finally, it is also not difficult to show that for all $z \in [0.1, 0.5]$ and $t \in \mathbb{N}$,

$$\hat{r}_t(z) = r_t(z) + Y_t(z),$$

where $Y_t(z)$ has a $N(0, \sigma_t^2(z))$ distribution with

$$\sigma_t^2(z) \leq \frac{1}{t} \times \left(\frac{1 - z^t}{1 - z} \right)^2 .$$

This implies that parts (i) and (ii) of Assumption 3.2 hold. Since $\sigma^2 = \text{Var}\{r(Z)\} = \text{Var}\{Z\} = 1/75 < \infty$, Theorem 3.4 now shows that the best asymptotic convergence

Table I. Performance of the independent-sampling estimator $\hat{\alpha}_{\upsilon,\ell^*}(c)$ for different X_0 and c with the asymptotically optimal multiplier ℓ^*

c	$X_0 = 0.1$			$X_0 = 1$			$X_0 = 10$			$X_0 = 100$		
(millions)	OR	AD	RD	OR	AD	RD	OR	AD	RD	OR	AD	RD
10	21/30	2.41×10^{-7}	0.61	21/30	6.21×10^{-7}	0.38	22/30	9.36×10^{-6}	1.41	22/30	2.69×10^{-5}	1.11
20	21/30	6.86×10^{-8}	0.28	20/30	0	0	21/30	1.19×10^{-6}	0.25	21/30	4.52×10^{-6}	0.21
50	22/30	3.25×10^{-8}	0.29	21/30	1.07×10^{-6}	2.30	22/30	1.48×10^{-6}	0.64	21/30	7.00×10^{-6}	0.78
100	21/30	8.13×10^{-8}	1.24	20/30	0	0	21/30	1.34×10^{-6}	1.55	21/30	5.54×10^{-6}	1.00
200	20/30	0	0	21/30	4.30×10^{-7}	2.77	21/30	9.30×10^{-7}	1.23	22/30	1.90×10^{-6}	0.41
300	21/30	1.18×10^{-8}	0.93	20/30	0	0	21/30	1.84×10^{-6}	6.08	21/30	2.37×10^{-6}	0.78
400	20/30	0	0	21/30	4.35×10^{-8}	0.71	21/30	1.27×10^{-6}	6.92	21/30	2.41×10^{-6}	1.19
500	19/30	1.08×10^{-8}	0.67	21/30	1.61×10^{-8}	0.53	21/30	8.32×10^{-7}	3.32	21/30	1.58×10^{-5}	8.15
600	20/30	0	0	21/30	1.57×10^{-7}	3.54	20/30	0	0	21/30	3.75×10^{-6}	3.57
700	20/30	0	0	21/30	2.21×10^{-7}	6.21	21/30	1.10×10^{-7}	0.43	21/30	2.66×10^{-7}	0.15
800	19/30	2.18×10^{-9}	0.81	20/30	0	0	20/30	0	0	21/30	1.12×10^{-7}	0.13
900	19/30	2.26×10^{-10}	0.09	20/30	0	0	22/30	3.22×10^{-7}	0.67	21/30	1.00×10^{-6}	0.63
1000	20/30	0	0	21/30	8.55×10^{-9}	0.38	21/30	1.24×10^{-7}	0.48	21/30	2.37×10^{-6}	2.22

rate (as $c \to \infty$) is obtained with $\upsilon = \upsilon^* = 2\gamma/(2\gamma + 1) = 2/3$. Therefore, we consider $\upsilon \in \Upsilon$, where $\Upsilon = \{15/30, 16/30, \ldots, 24/30\}$. Similarly, the value of ℓ that minimizes the mean-squared error of the asymptotic distribution $N(\ell^\gamma b, \sigma^2/\ell)$ when $\gamma = 1$ is

$$\ell^* = \left(\frac{\sigma^2}{2b^2}\right)^{1/3} = \left(\frac{0.4^2}{150 \times X_0^2 \times [\ln(1.8)]^2}\right)^{1/3} = \left(937.5 \times X_0^2 \times [\ln(1.8)]^2\right)^{-1/3}.$$

Therefore, we consider ℓ satisfying $\ell/\ell^* \in \mathcal{L}$, where $\mathcal{L} = \{0.01, 0.1, 0.2, 0.5, 1, 2, 5, 10, 100\}$. (The expressions for $m(c)$ and $t(c)$ are not necessarily integer-valued. Therefore, in our numerical experiments, we let $m(c) = \lfloor \ell c^\upsilon \rfloor$ and $t(c) = \lfloor c/m(c) \rfloor$, unless this leads to the number $c - m(c) \times t(c)$ of unused normal random variables being greater than or equal to $t(c)$, in which case we let $t(c) = \lfloor c^{1-\upsilon}/\ell \rfloor$ and $m(c) = \lfloor c/t(c) \rfloor$.)

We conducted two sets of numerical experiments. In the first set of experiments, we used the asymptotically optimal multiplier ℓ^* and identified the (empirically) optimal rate (OR) $\upsilon \in \Upsilon$ for each initial state X_0 and computational budget c. In the second set of experiments, we used the asymptotically optimal rate υ^* and identified the (empirically) optimal multiplier (OM) $\ell/\ell^* \in \mathcal{L}$ for each X_0 and c. In both cases, the objective was to minimize the mean-squared error (MSE) of the estimator of α. In all cases, our results were obtained by replicating the estimation process 100 times using common random numbers for different X_0, c, υ, and ℓ values.

The results of our first set of numerical experiments are shown in Table I. For each choice of X_0 and c, we show the (empirically) optimal rate (OR) (i.e., the $\upsilon \in \Upsilon$ with the smallest average $[\hat{\alpha}_{\upsilon,\ell^*}(c) - \alpha]^2$), and also the absolute and relative differences (AD and RD) between the MSE obtained with the observed optimal rate and with the best asymptotic rate $\upsilon^* = 2/3$. The results of our second set of numerical experiments are shown in Table II. Similar to Table I, for each X_0 and c, we show the (empirically) optimal multiplier (OM) (i.e., the ℓ/ℓ^* with the smallest average $[\hat{\alpha}_{\upsilon^*,\ell}(c) - \alpha]^2$), and also the absolute and relative differences (AD and RD) between the MSE obtained with the observed optimal multiplier and with the best asymptotic multiplier ℓ^*.

Tables I and II show that for finite c, the number $m(c)$ of values of Z that yields the smallest MSE is usually strictly larger than that predicted by the asymptotic theory. In particular, the best choice of $\upsilon \in \Upsilon$ in Table I (OR) is in general larger than the asymptotically optimal $\upsilon^* = 20/30$, and the best choice of $\ell/\ell^* \in \mathcal{L}$ in Table II is mostly larger than the asymptotically optimal $\ell/\ell^* = 1$. However, the difference between the empirically optimal and asymptotically optimal parameter choices are not large (i.e., no larger than 2/30 in Table I and no larger than 4 in Table II).

Table II. Performance of the independent-sampling estimator $\hat{\alpha}_{\upsilon^*,\ell}(c)$ for different X_0 and c with the asymptotically optimal rate υ^*

c	$X_0 = 0.1$			$X_0 = 1$			$X_0 = 10$			$X_0 = 100$		
(millions)	OM	AD	RD	OM	AD	RD	OM	AD	RD	OM	AD	RD
10	2	1.12×10^{-8}	1.81	2	7.21×10^{-7}	0.47	2	1.02×10^{-5}	1.75	2	2.60×10^{-5}	1.04
20	2	4.55×10^{-8}	0.17	2	3.07×10^{-7}	0.37	2	2.21×10^{-6}	0.59	2	4.04×10^{-6}	0.19
50	1	0	0	2	1.19×10^{-6}	3.40	2	2.17×10^{-6}	1.36	2	4.65×10^{-6}	0.41
100	0.5	2.95×10^{-8}	0.25	2	3.64×10^{-8}	0.11	2	1.05×10^{-6}	0.91	2	4.15×10^{-6}	0.60
200	1	0	0	0.5	4.42×10^{-7}	3.07	2	6.99×10^{-7}	0.71	2	2.29×10^{-6}	0.54
300	1	0	0	2	5.16×10^{-8}	0.63	2	1.75×10^{-6}	4.44	2	1.81×10^{-6}	0.50
400	1	0	0	1	0	0	2	1.09×10^{-6}	2.99	2	1.82×10^{-6}	0.69
500	2	1.85×10^{-8}	2.18	1	0	0	2	5.89×10^{-7}	1.19	2	1.57×10^{-5}	7.77
600	1	0	0	2	1.45×10^{-7}	2.59	1	0	0	5	6.87×10^{-7}	0.17
700	1	0	0	2	2.11×10^{-7}	4.63	0.5	1.59×10^{-7}	0.78	1	0	0
800	0.5	3.44×10^{-10}	0.08	1	0	0	1	0	0	1	0	0
900	1	0	0	1	0	0	2	7.66×10^{-7}	19.53	2	1.55×10^{-6}	1.46
1000	1	0	0	1	0	0	0.5	2.76×10^{-7}	2.62	2	7.39×10^{-7}	0.27

Tables I and II also show that the behavior of the independent-sampling estimator $\hat{\alpha}_{\upsilon,\ell}(c)$ depends heavily on the choice of the initial state X_0. When X_0 is small, then the observed optimal rate and multiplier are close to the optimal asymptotic rate and multiplier, sometimes coincide with the optimal values, and occasionally are smaller than optimal. On the other hand, for larger X_0, the observed optimal rate and multiplier are usually larger than optimal, occasionally optimal, but not smaller than optimal. These results are reasonable because we do not perform any truncation to remove initialization bias while estimating the steady-state mean $r'(z)$ of the autoregressive process $\{X_n(z)\}$, where $z \in [0.1, 0.5]$. Therefore, X_0 has a heavy influence on the bias in the estimator of α (see also equation (14)), and longer sample path lengths are required to reduce the bias for large X_0 than for small X_0 (in other words, the asymptotical results derived in Section 3.1 come into play for larger values of c when X_0 is large than when X_0 is small). Tables I and II also show that the rate at which the observed optimal rate (OR) and multiplier (OM) approach the asymptotically optimal rate and multiplier are slow, reflecting the slow growth rate of the best asymptotic sample path length $t(c) \simeq c^{1-\upsilon^*} = c^{1/3}$ with respect to c.

We conclude this section by showing the behavior of the MSE of the estimator $\hat{\alpha}_{\upsilon,\ell^*}(c)$ as a function of the growth rate $\upsilon \in \Upsilon$, and the behavior of the MSE of the estimator $\hat{\alpha}_{\upsilon^*,\ell}(c)$ as a function of the multiplier $\ell/\ell^* \in \mathcal{L}$. The results are provided for $X_0 = 1$ and $c = c' \times$ 1,000,000, where $c' \in \{100, 500, 900\}$, and are shown on a logarithmic scale in Figures 1 and 2, respectively.

Figures 1 and 2 show that the MSE of the estimator depends heavily on the choice of the parameters υ and ℓ, and that this sensitivity is larger when the computational budget c is large. Specifically, in Figure 1, the difference in MSE from using a suboptimal rate υ can be almost two orders of magnitude for $c =$ 100,000,000 and almost three orders of magnitude for $c =$ 900,000,000. Similarly, in Figure 2, the difference in MSE from using a suboptimal multiplier ℓ can be almost three orders of magnitude for $c =$ 100,000,000 and almost four orders of magnitude for $c =$ 900,000,000.

4. EXAMPLE PROCEDURES WITH BETTER CONVERGENCE RATES

In this section, we show that improved convergence rates can be achieved (relative to the independent sampling approach considered in Section 3) by using other methods to generate the sampled values of the random variable Z, see equation (2). More specifically, in Section 4.1, we analyze the case where $Z_i = z_i = h(u_i)$, for $i = 1, \ldots, m(c)$, and $\{u_n\}$ is a quasi-random sequence defined on $[0, 1]^d$. In Section 4.2, we discuss how modest improvements in the convergence rate (over the rates given in Section 4.1) some-

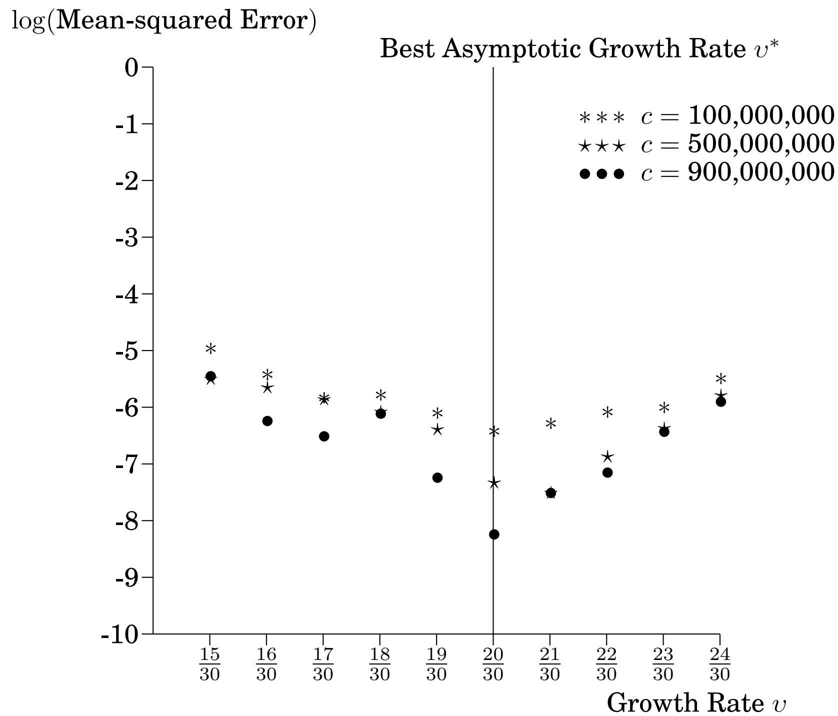

Fig. 1. Performance of the independent-sampling estimator $\hat{\alpha}_{\upsilon,\ell^*}(c)$ for $X_0 = 1$, $\upsilon \in \Upsilon$

times can be obtained by using numerical integration techniques that exploit special structure, and illustrate this idea using Simpson's rule.

4.1. Quasi-random numbers

In this section, we determine the asymptotic behavior of the estimator $\hat{\alpha}(c)$ defined in equation (8) as $c \to \infty$ when $Z_1 = z_1, \ldots, Z_{m(c)} = z_{m(c)}$ are generated using a quasi-random sequence. As in Section 3.1, we first show that in order for $\hat{\alpha}(c)$ to be asymptotically unbiased and consistent, we generally need both $t(c) \to \infty$ and $m(c) \to \infty$ as $c \to \infty$ (Propositions 4.2 and 4.3 and Theorem 4.5). Then we present the main result in this section (Theorem 4.9), which identifies the convergence rate of $\hat{\alpha}(c)$ to α as $c \to \infty$. We will be using the following assumption throughout this section.

ASSUMPTION 4.1. *Assume that:*

(i). The random variable $r(Z)$ is integrable.
(ii). For all $c \in \mathbb{R}^+$, the parameters $m(c)$ and $t(c)$ satisfy $c = m(c) \times t(c)$.
(iii). The random vector Z can be expressed as $Z = h(U)$, where U is a uniformly distributed random vector on the set $[0,1]^d$, where $d \in \mathbb{N}$, and $h : [0,1]^d \to \mathcal{Z}$ is a known function.
(iv). The sequence $\{u_n\}$ is a deterministic (quasi-random) sequence taking values in $[0,1]^d$ with the (star) discrepancy of $u_1, \ldots, u_N$ being of order $O([\log N]^\iota / N^\beta)$ for all $N \in \mathbb{N}$, where $\iota, \beta \in \mathbb{R}^+$ (see for example Niederreiter, 1992, page 14, for the definition of the star discrepancy of a sequence of points).
(v). For all $t \in \mathbb{R}^+$, the random variables $\hat{r}_t(z_i)$, where $z_i = h(u_i)$ and $i \in \mathbb{N}$, are independent.

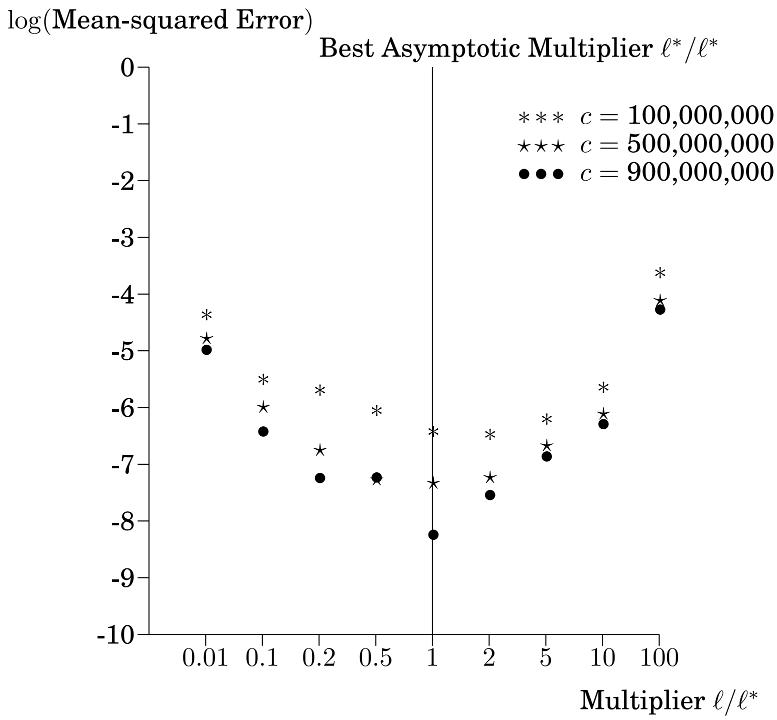

Fig. 2. Performance of the independent-sampling estimator $\hat{\alpha}_{v^*,\ell}(c)$ for $X_0 = 1$, $\ell/\ell^* \in \mathcal{L}$

Remark 4.1. *For results that can be used to show that part (iv) of Assumption 4.1 holds with $\beta = 1$, see for example Theorems 3.6 and 3.8 of Niederreiter (1992). Note that the sequences $\{u_n\}$ and $\{z_n\}$ defined in parts (iv) and (v) of Assumption 4.1, respectively, may depend on the value of $c \in \mathbb{R}^+$ (this would for example be the case when Theorem 3.8 of Niederreiter (1992) is used to generate the sequence $\{u_n\}$ and $m(c)$ is not constant in $c \in \mathbb{R}^+$), although we suppress this in our notation. There is an extensive literature on the development and analysis of quasi- and randomized quasi-Monte Carlo sequences, including measures of discrepancy other than the star discrepancy we consider here (see part (iv) of Assumption 4.1), see for example L'Ecuyer (2009) for a recent review.*

As in Section 3.1, we start by analyzing the bias and consistency of the estimator $\hat{\alpha}(c)$ defined in equation (8). Let $\hat{\alpha}_1(c)$, $\hat{\alpha}_2(c)$, and $\hat{\alpha}_3(c)$ be defined as in equation (9) for all $c \in \mathbb{R}^+$. We first consider the case where $c \to \infty$ and $t(c)$ remains constant.

PROPOSITION 4.2. *Suppose that Assumption 4.1 holds, that $t(c) = t > 0$ for all $c \geq 0$, that the function r_t defined by $r'_t(u) = r_t(h(u))$ for all $u \in [0,1]^d$ is Riemann integrable, and that $\sup_{i \in \mathbb{N}} E\{[\hat{r}_t(z_i) - r_t(z_i)]^2\} < \infty$, where the sequence $\{z_n\}$ is defined in part (v) of Assumption 4.1. Then $m(c) = \lfloor c/t \rfloor$ for all $c \geq 0$, $E\{\hat{\alpha}(c)\} = \sum_{i=1}^{m(c)} r_t(z_i)/m(c) \to E\{\hat{r}_t(Z)\}$ as $c \to \infty$, and $\hat{\alpha}(c) \to E\{\hat{r}_t(Z)\}$ in probability as $c \to \infty$, where Z is defined in part (iii) of Assumption 4.1 and the random numbers used to generate $\hat{r}_t(Z)$ are independent of the value of Z.*

Proof: From equation (8), parts (ii) and (iv) of Assumption 4.1, the fact that the function r'_t is Riemann integrable, and pages 14 and 17 of Niederreiter (1992), it is clear that

$$E\{\hat{\alpha}(c)\} - E\{\hat{r}_t(Z)\} = \frac{1}{m(c)} \sum_{i=1}^{m(c)} r_t(z_i) - E\{\hat{r}_t(Z)\} \to 0 \quad \text{as} \quad c \to \infty. \tag{15}$$

Moreover,

$$\hat{\alpha}(c) = \hat{\alpha}_1(c) + E\{\hat{\alpha}(c)\}. \tag{16}$$

Let $\epsilon > 0$. By Markov's inequality and part (v) of Assumption 4.1, we have that

$$P\{|\hat{\alpha}_1(c)| > \epsilon\} \le \frac{E\{[\hat{\alpha}_1(c)]^2\}}{\epsilon^2} = \frac{1}{\epsilon^2 m(c)^2} \sum_{i=1}^{m(c)} E\{[\hat{r}_t(z_i) - r_t(z_i)]^2\}.$$

Therefore, part (ii) of Assumption 4.1 and the fact that $\sup_{i \in \mathbb{N}} E\{[\hat{r}_t(z_i) - r_t(z_i)]^2\} < \infty$ imply that $\hat{\alpha}_1(c) \to 0$ in probability as $c \to \infty$. The convergence in probability of $\hat{\alpha}(c)$ to $E\{\hat{r}_t(Z)\}$ now follows from equations (15) and (16). □

We now consider the case when $c \to \infty$ and $m(c)$ remains constant. The proof of the following proposition is straightforward, and is omitted.

PROPOSITION 4.3. *Suppose that Assumption 4.1 holds, that $m(c) = m > 0$ for all $c \ge 0$, that the values of the sequence $\{u_n\}$ defined in part (iv) of Assumption 4.1 do not depend on $c \in \mathbb{R}^+$, and that $\hat{r}_t(z_i) \Rightarrow r(z_i)$ as $t \to \infty$ for $i = 1, \ldots, m$. Then $t(c) = c/m$ for all $c \ge 0$, $E\{\hat{\alpha}(c)\} = \sum_{i=1}^{m} r_{c/m}(z_i)/m$ for all $c \ge 0$, and $\hat{\alpha}(c) \Rightarrow \sum_{i=1}^{m} r(z_i)/m$ as $c \to \infty$. If also the sets of random variables $\{\hat{r}_{c/m}(z_i) : c \ge 0\}$, where $i = 1, \ldots, m$, are uniformly integrable, then $E\{\hat{\alpha}(c)\} \to \sum_{i=1}^{m} r(z_i)/m$ as $c \to \infty$.*

Propositions 4.2 and 4.3 show that when $\{z_n\}$ is generated using a quasi-random sequence, we generally need to have that as $c \to \infty$, both $t(c) \to \infty$ and $m(c) \to \infty$ in order for $\hat{\alpha}(c)$ to be a consistent estimator for α as $c \to \infty$ (this is consistent with the results obtained earlier for independent sampling, see Section 3.1). We now turn our attention to this case. For all $c \in \mathbb{R}^+$ and $i \in \mathbb{N}$, let $Y_i(c) = \sqrt{t(c)}(\hat{r}_{t(c)}(z_i) - r_{t(c)}(z_i))$, where the quantities $z_1, \ldots, z_{m(c)}$ are defined in part (v) of Assumption 4.1. Furthermore, let the symbol $\circ$ denote the composition of two functions and let I_A denote the indicator of the set A for all A. The results given in the remainder of this section will require some (or all) the parts of the following assumption.

ASSUMPTION 4.2. *Assume that:*

(i). For all $z \in \mathcal{Z}$, the random variables $\hat{r}_t(z)$, where $t \ge 0$, satisfy $E\{t[\hat{r}_t(z) - r_t(z)]^2\} \to \sigma^2(z)$ uniformly in $z \in \mathcal{Z}$ as $t \to \infty$, where $\sigma^2(z) \in \mathbb{R}^+$.
(ii). For all $\epsilon > 0$, the random variables $Y_i(c)$, where $i \in \mathbb{N}$ and $c \in \mathbb{R}^+$, satisfy

$$\lim_{c \to \infty} \sum_{i=1}^{m(c)} \frac{1}{s_c^2} E\{[Y_i(c)]^2 I_{\{|Y_i(c)| \ge \epsilon s_c\}}\} = 0,$$

where $s_c^2 = \sum_{i=1}^{m(c)} E\{[Y_i(c)]^2\}$ for all $c \in \mathbb{R}^+$.
(iii). The function $\sigma^2 \circ h : [0,1]^d \to \mathbb{R}^+$ is Riemann integrable, where the function σ^2 is defined in part (i) of this assumption and the function h is defined in part (iii) of Assumption 4.1.
(iv). The function r_t satisfies $r_t(z) = r(z) + b(z)/t^\gamma + e(z)o(1/t^\gamma)$ as $t \to \infty$ for all $z \in \mathcal{Z}$, where $\gamma > 0$, $b(z), e(z) \in \mathbb{R}$ for all $z \in \mathcal{Z}$, and the $o(1/t^\gamma)$ term is uniform in $z \in \mathcal{Z}$.

(v). The function $b \circ h : [0,1]^d \to \mathbb{R}$ is Riemann integrable, where the function b is defined in part (iv) of this assumption.
(vi). The function $e \circ h : [0,1]^d \to \mathbb{R}$ is Riemann integrable, where the function e is defined in part (iv) of this assumption.
(vii). The function $r \circ h : [0,1]^d \to \mathbb{R}$ has bounded variation in the sense of Hardy and Krause (see for example Niederreiter, 1992, page 19, for the definition of bounded variation in the sense of Hardy and Krause), where the function r is defined in Section 1.

Remark 4.4. *Note that part (ii) of Assumption 4.2 holds if for each $c \in \mathbb{R}^+$, the random variables $Y_i(c)$, where $i \in \mathbb{N}$, are identically distributed with finite and positive variance. Part (ii) of Assumption 4.2 also holds if there exists $\epsilon > 0$ such that $\sup_{i,c} E\{[Y_i(c)]^{2+\epsilon}\} < \infty$ and parts (ii) and (iv) of Assumption 4.1, parts (i) and (iii) of Assumption 4.2, and Assumption 4.3 hold (note that $s_c^2/m(c) \to \sigma^2$ as $c \to \infty$ under these conditions, see the proof of Lemma 4.6 in the online appendix).*

We have:

THEOREM 4.5. *Suppose that Assumption 4.1 and parts (iv), (v), (vi), and (vii) of Assumption 4.2 hold, that*

$$\lim_{t\to\infty} \sup_{i\in\mathbb{N}} E\{|\hat{r}_t(z_i) - r_t(z_i)|\} = 0, \tag{17}$$

and that as $c \to \infty$, both $m(c) \to \infty$ and $t(c) \to \infty$. Then $E\{\hat{\alpha}(c)\} = \sum_{i=1}^{m(c)} r_{t(c)}(z_i)/m(c) \to E\{r(Z)\} = \alpha$ and $\hat{\alpha}(c) \to \alpha$ in probability as $c \to \infty$.

Proof: From part (iv) of Assumption 4.2, we have

$$E\{\hat{\alpha}(c)\} = \frac{1}{m(c)} \sum_{i=1}^{m(c)} r(z_i) + \frac{1}{t(c)^\gamma} \left(\frac{1}{m(c)} \sum_{i=1}^{m(c)} [b(z_i) + e(z_i)o(1)] \right).$$

By part (iv) of Assumption 4.1, parts (v), (vi), and (vii) of Assumption 4.2, the Koksma-Hlawka inequality (see for example Theorem 2.11 in Niederreiter, 1992), pages 14 and 17 of Niederreiter (1992), and the fact that as $c \to \infty$, both $m(c) \to \infty$ and $t(c) \to \infty$, the first term in the above expression converges to $\alpha = E\{r(Z)\}$ and the second term converges to zero. This shows that $E\{\hat{\alpha}(c)\} \to \alpha$ as $c \to \infty$.

Let $\epsilon > 0$. Markov's inequality gives that

$$P\{|\hat{\alpha}_1(c)| > \epsilon\} \le \frac{1}{\epsilon m(c)} \sum_{i=1}^{m(c)} E\{|\hat{r}_{t(c)}(z_i) - r_{t(c)}(z_i)|\}.$$

Therefore, equation (17) and the fact that $t(c) \to \infty$ as $c \to \infty$ imply that $\hat{\alpha}_1(c) \to 0$ in probability as $c \to \infty$. Equation (16) now gives that $\hat{\alpha}(c) \to \alpha$ in probability as $c \to \infty$. □

In the remainder of this section, we will use the following assumption:

ASSUMPTION 4.3. *For all $c \ge C$, where $C \in \mathbb{R}^+$, assume that the parameter $m(c)$ satisfies $m(c) = \lfloor \kappa c^\delta \rfloor$, where $\kappa > 0$ and $0 < \delta < 1$.*

We are now ready to study the rate of convergence of $\hat{\alpha}(c)$ to α as $c \to \infty$. Let $\sigma^2 = E\{\sigma^2(Z)\}$ and $e = E\{e(Z)\}$, and recall that $b = E\{b(Z)\}$ (note that part (iii) of Assumption 4.1 and parts (iii), (v), and (vi) of Assumption 4.2 imply that the random variables $b(Z)$, $e(Z)$, and $\sigma^2(Z)$ are integrable). Recall that the functions $\hat{\alpha}_1(c)$,

$\hat{\alpha}_2(c)$, and $\hat{\alpha}_3(c)$ are defined in the proof of Theorem 3.3. We shall need the following three lemmas whose proofs are provided in the online appendix to this paper.

LEMMA 4.6. *Suppose that Assumptions 4.1, 4.3, and parts (i), (ii), and (iii) of Assumption 4.2 hold. Then*

$$\sqrt{c}\hat{\alpha}_1(c) \Rightarrow N(0,\sigma^2) \text{ as } c \to \infty.$$

LEMMA 4.7. *Suppose that Assumption 4.1, 4.3, and parts (iv), (v), and (vi) of Assumption 4.2 hold. Then*

$$\hat{\alpha}_2(c) = \frac{b\kappa^\gamma}{c^{\gamma(1-\delta)}} + o\left(\frac{1}{c^{\gamma(1-\delta)}}\right) \text{ as } c \to \infty.$$

LEMMA 4.8. *Suppose that Assumption 4.1, 4.3, and part (vii) of Assumption 4.2 hold. Then*

$$|\hat{\alpha}_3(c)| \le O\left(\frac{[\log c]^\iota}{c^{\delta\beta}}\right) \text{ as } c \to \infty.$$

We now present the main result in this section. The following theorem specifies the rate at which the estimator $\hat{\alpha}(c)$ converges to α as $c \to \infty$ as a function of the choice of the parameter δ (see Assumption 4.3).

THEOREM 4.9. *Suppose that Assumptions 4.1, 4.2, and 4.3 hold. Then, the following statements hold:*

(a). If $\delta > \gamma/(\gamma+\beta)$ and $\gamma(1-\delta) > 1/2$, then

$$\sqrt{c}(\hat{\alpha}(c) - \alpha) \Rightarrow N(0,\sigma^2) \quad \text{as} \quad c \to \infty.$$

(b). If $\delta > \gamma/(\gamma+\beta)$ and $\gamma(1-\delta) = 1/2$, then

$$\sqrt{c}(\hat{\alpha}(c) - \alpha) \Rightarrow N(b\kappa^\gamma,\sigma^2) \quad \text{as} \quad c \to \infty.$$

(c). If $\delta > \gamma/(\gamma+\beta)$ and $\gamma(1-\delta) < 1/2$, then

$$c^{\gamma(1-\delta)}(\hat{\alpha}(c) - \alpha) \Rightarrow b\kappa^\gamma \quad \text{as} \quad c \to \infty.$$

(d). If $\delta \le \gamma/(\gamma+\beta)$ and $\delta > 1/(2\beta)$, then

$$\sqrt{c}(\hat{\alpha}(c) - \alpha) \Rightarrow N(0,\sigma^2) \quad \text{as} \quad c \to \infty.$$

(e). If $\delta \le \gamma/(\gamma+\beta)$ and $\delta \le 1/(2\beta)$, then

$$c^{\delta\beta}|\hat{\alpha}(c) - \alpha|/[\log c]^\iota \le X(c) + O(1) \quad \text{as} \quad c \to \infty,$$

where $X(c) \Rightarrow 0$ as $c \to \infty$.

Proof: Note that $\delta > \gamma/(\gamma+\beta)$ if and only if $\delta\beta > \gamma(1-\delta)$. The result now follows from equation (9), Lemmas 4.6, 4.7, and 4.8, and the continuous mapping theorem (see for example Theorems 4.4 and 5.1 of Billingsley, 1968). □

Remark 4.10. *The specific convergence rates provided by part (iv) of Assumption 4.1 and by the Koksma-Hlawka inequality are needed only for Lemma 4.8 and part (e) of Theorem 4.9. Also, in Theorem 4.5, part (vii) of Assumption 4.2 can be replaced by the assumption that the function $r \circ h : [0,1]^d \to \mathbb{R}$ is Riemann integrable.*

Remark 4.11. *Consider the classical case where $\beta = 1$. From parts (a), (b), and (d) of Theorem 4.9, it is clear that when $1/2 < \delta \le (2\gamma-1)/(2\gamma)$ and Assumptions 4.1, 4.2, and 4.3 hold, then the estimator $\hat{\alpha}(c)$ converges to α at the rate $c^{-1/2}$ as $c \to \infty$. This is an improvement over the convergence rate obtained using independent sampling, see*

Remark 3.5. However, it is only possible to select δ in this range when $\gamma > 1$. When $\gamma = 1$ and $\delta > 1/2$, then part (c) of Theorem 4.9 gives that the estimator $\hat{\alpha}(c)$ converges to α at the rate $c^{\delta-1}$ as $c \to \infty$, and when $\gamma = 1$ and $\delta \leq 1/2$, then part (e) of Theorem 4.9 gives that the estimator $\hat{\alpha}(c)$ converges to α at the rate $[\log(c)]^{\iota}/c^{\delta}$ as $c \to \infty$. By choosing $\delta = 1/2$ when $\gamma \geq 1$, it is again clear that we get a computational improvement by generating $\{z_n\}$ using a quasi-random sequence $\{u_n\}$, relative to the situation considered in Section 3. However, this improvement is achieved using additional assumptions, including the smoothness assumptions in parts (iii), (v), (vi), and (vii) of Assumption 4.2, which may be difficult to verify in practice. Moreover, this improvement is asymptotic, and need not be observed in practice for realistic computational budgets c, especially if the dimension d is large, see for example L'Ecuyer (2009) for additional details. Note that when $\delta = 1/2$, then $m(c)$ and $t(c)$ will grow at the same rate as c increases. Thus, when quasi-random numbers are used, it is not necessary to let $m(c)$ grow as rapidly as when independent sampling is used, see Theorem 3.5 and Remark 3.5.

4.2. Other numerical integration techniques

As is clear from Theorem 4.9 and Remark 4.11, and given that the estimates $\hat{r}_{t(c)}(z_i)$ of $r(z_i)$, where $i = 1, \dots, m(c)$, are generated using simulation (so that one would expect $c^{-1/2}$ to be the best possible convergence rate), there is not much room for improving the rate at which the estimator $\hat{\alpha}(c)$ defined in equation (8) converges to the quantity of interest α as the total computational budget c grows, relative to the convergence rate obtained when the sequence $\{z_n\}$ is generated using a quasi-random sequence $\{u_n\}$. However, in the presence of some special structure, it is sometimes possible to use numerical integration techniques other than the ones considered in Sections 3 and 4.1 (i.e., other than independent sampling and quasi-random sequences) to obtain (slightly) better rate of convergence results than Theorem 4.9. We illustrate this approach by analyzing a single other numerical integration technique, namely Simpson's rule (see, e.g., Davis and Rabinowitz, 1984, Section 2.2).

More specifically, in this section we assume that the underlying integration problem is one-dimensional (i.e., $Z = h(U)$, where U is uniformly distributed on $[0, 1]$) and smooth (in a sense that is specified later). In this case, we can use Simpson's rule to improve upon Theorem 4.9. This involves using an estimator of the form

$$\tilde{\alpha}(c) = \sum_{i=1}^{m(c)} w_i(c)\hat{r}_{t(c)}(z_i(c)), \tag{18}$$

where $c \in \mathbb{R}^+$, to estimate α. Details on how $z_1(c), \dots, z_{m(c)}(c)$ and the weights $w_1(c), \dots, w_{m(c)}(c)$ are selected are provided in the online appendix to this paper, together with our analysis of the estimator (18). The main conclusion is that the estimator (18) converges to α at the rate $c^{-1/2}$ as $c \to \infty$ when $1/8 \leq \delta \leq (2\gamma - 1)/(2\gamma)$ under the assumptions stated in the online appendix. Moreover, the interval $[1/8, (2\gamma - 1)/(2\gamma)]$ is non-empty for all $\gamma \geq 4/7$ and includes the value $\delta = 1/2$ for all $\gamma \geq 1$ (as would typically be the case in practice). This is an improvement over the rate of convergence results obtained in Sections 3.1 and 4.1, see Remarks 3.5 and 4.11.

5. GENERAL FRAMEWORK

In Sections 3 and 4, we studied three specific methods for estimating the quantity α defined in equation (2). In all cases, we provided theoretical results specifying the rate of convergence of the estimator under consideration to α as the total available computational budget c grows. In this section, we present a unified framework for proving such rate of convergence results for a broad class of estimators that includes the estimators (8) and (18) considered in Sections 3, 4.1 and in Section 4.2, respectively, as special cases. However, as the form of the general estimator and the associated analysis are relatively abstract, we believe it is of value to include the analysis of the specific estimators of Sections 3 and 4 as well.

More specifically, in this section the estimator of α obtained with the computational budget c is given by

$$\bar{\alpha}(c) = \sum_{i=1}^{m(c)} W_i(c)\hat{r}_{t(c)}(Z_i(c)),$$

where $Z_1(c), \ldots, Z_{m(c)}(c)$ are the different (and possibly random) values (locations) of the random vector Z used in the estimation of α, $\hat{r}_{t(c)}(Z_1(c)), \ldots, \hat{r}_{t(c)}(Z_{m(c)}(c))$ are the estimates of $r(Z_1(c)), \ldots, r(Z_{m(c)}(c))$ obtained using the computational effort $t(c)$, and $W_1(c), \ldots, W_{m(c)}(c)$ are the (possibly random) weights given to the estimates $\hat{r}_{t(c)}(Z_1(c)), \ldots, \hat{r}_{t(c)}(Z_{m(c)}(c))$, respectively.

Let $\mathcal{F}$ denote the σ-algebra generated by the locations $Z_1(c), \ldots, Z_{m(c)}(c)$ and weights $W_1(c), \ldots, W_{m(c)}(c)$, for all $c \in \mathbb{R}^+$, and let $F_t(z, x) = P\{\hat{r}_t(z) \leq x\}$ for all $z \in \mathcal{Z}$, $t \geq 0$, and $x \in \mathbb{R}$. The following assumption describes more precisely the framework for estimating α considered in this section.

ASSUMPTION 5.1. *Assume that:*

(i). For all $c \in \mathbb{R}^+$, the parameters $m(c)$ and $t(c)$ satisfy $c = m(c) \times t(c)$.
(ii). The parameter $m(c)$ satisfies $m(c)/c^\delta \to d$ as $c \to \infty$, where $d \in \mathbb{R}^+$ and $0 < \delta < 1$.
(iii). For all $c \in \mathbb{R}^+$, the random variables $\hat{r}_{t(c)}(Z_1(c)), \ldots, \hat{r}_{t(c)}(Z_{m(c)}(c))$ satisfy

$$P\{\hat{r}_{t(c)}(Z_i(c)) \leq x_i, \forall i = 1, \ldots, m(c)|\mathcal{F}\} = \prod_{i=1}^{m(c)} F_{t(c)}(Z_i(c), x_i)$$

for all $x_1, \ldots, x_{m(c)} \in \mathbb{R}$.
(iv). The weights $W_1(c), \ldots, W_{m(c)}(c)$ satisfy $\sum_{i=1}^{m(c)} W_i(c) \Rightarrow 1$ as $c \to \infty$ and $l \leq m(c)W_i(c) \leq u$ for all $i = 1, \ldots, m(c)$ and $c \in \mathbb{R}^+$, where l and u are positive constants.

For all $z \in \mathcal{Z}$ and $t \geq 0$, let $r_t(z) = E\{\hat{r}_t(z)\}$. Moreover, for all $z \in \mathcal{Z}$ and $t \geq 0$, let $Y_t(z) = \sqrt{t}[\hat{r}_t(z) - r_t(z)]$. The results given in the remainder of this section will require some (or all) the parts of the following technical assumption.

ASSUMPTION 5.2. *Assume that:*

(i). There exists a function $\sigma^2 : \mathcal{Z} \to \mathbb{R}^+$ such that for all $z \in \mathcal{Z}$, the random variables $Y_t(z)$, where $t \geq 0$, satisfy $E\{[Y_t(z)]^2\} \to \sigma^2(z)$ uniformly in $z \in \mathcal{Z}$ as $t \to \infty$, where $\inf_{z \in \mathcal{Z}} \sigma^2(z) > 0$.
(ii). There exists $\epsilon > 0$ such that the random variables $Y_t(z)$, where $z \in \mathcal{Z}$ and $t \geq 0$, satisfy $\sup_{z \in \mathcal{Z}, t \geq 0} E\{[Y_t(z)]^{2+\epsilon}\} < \infty$.
(iii). There exist functions $b : \mathcal{Z} \to \mathbb{R}$ and $e : \mathcal{Z} \to \mathbb{R}$ such that the function r_t satisfies $r_t(z) = r(z) + b(z)/t^\gamma + e(z)o(1/t^\gamma)$ as $t \to \infty$ for all $z \in \mathcal{Z}$, where $\gamma > 0$ and the $o(1/t^\gamma)$ term is uniform in $z \in \mathcal{Z}$.
(iv). The locations $Z_1(c), \ldots, Z_{m(c)}(c)$, weights $W_1(c), \ldots, W_{m(c)}(c)$, and function r satisfy

$$\sum_{i=1}^{m(c)} W_i(c)r(Z_i(c)) = \alpha + \frac{X_3(c)}{g(c)},$$

where $g : \mathbb{R}^+ \to \mathbb{R}^+$ is a deterministic function, $X_3(c) \Rightarrow X_3$ as $c \to \infty$, and X_3 is a (proper) random variable.
(v). The locations $Z_1(c), \ldots, Z_{m(c)}(c)$, weights $W_1(c), \ldots, W_{m(c)}(c)$, and functions σ^2, b, and e defined in parts (i) and (iii) of this assumption satisfy $\sum_{i=1}^{m(c)} W_i(c)\sigma^2(Z_i(c)) \Rightarrow E\{\sigma^2(Z)\}$, $\sum_{i=1}^{m(c)} W_i(c)b(Z_i(c)) \Rightarrow E\{b(Z)\}$, and $\sum_{i=1}^{m(c)} W_i(c)e(Z_i(c)) \Rightarrow E\{e(Z)\}$ as $c \to \infty$.

Note that the function g in part (iv) of Assumption 5.1 is not uniquely defined. As will become clear later (see Theorem 5.4 below), it is best to define g such that $g(c)$ grows as rapidly as possible with c, so that X_3 is different from zero with positive probability.

As in equation (9), it will be useful to express the estimate $\bar{\alpha}(c)$ as follows:

$$\bar{\alpha}(c) - \alpha = \bar{\alpha}_1(c) + \bar{\alpha}_2(c) + \bar{\alpha}_3(c), \tag{19}$$

where

$$\begin{aligned}
\bar{\alpha}_1(c) &= \sum_{i=1}^{m(c)} W_i(c)[\hat{r}_{t(c)}(Z_i(c)) - r_{t(c)}(Z_i(c))], \\
\bar{\alpha}_2(c) &= \sum_{i=1}^{m(c)} W_i(c)[r_{t(c)}(Z_i(c)) - r(Z_i(c))], \quad \text{and} \\
\bar{\alpha}_3(c) &= \sum_{i=1}^{m(c)} W_i(c)r(Z_i(c)) - \alpha
\end{aligned}$$

for all $c \geq 0$. Define $\sigma^2 = E\{\sigma^2(Z)\}$ and $b = E\{b(Z)\}$. We shall need the following two lemmas.

LEMMA 5.1. *Suppose that Assumption 5.1 and parts (i), (ii), and (v) of Assumption 5.2 hold. Then, for all $c \in R^+$, there exist random variables $X_1(c)$ and $\bar{\alpha}_1'(c)$ such that*

$$\sqrt{c}\bar{\alpha}_1(c) = X_1(c) \times \bar{\alpha}_1'(c), \quad \sqrt{l} \leq X_1(c) \leq \sqrt{u}, \quad \textit{and} \quad \bar{\alpha}_1'(c) \Rightarrow Y \sim N(0, \sigma^2) \textit{ as } c \to \infty.$$

Moreover, $(\bar{\alpha}_1'(c), X_3(c)) \Rightarrow (Y, X_3)$ as $c \to \infty$, where Y and X_3 are independent.

Proof: For all $c \geq 0$, define

$$\begin{aligned}
\hat{S}_c^2 &= \sum_{i=1}^{m(c)} [W_i(c)]^2 E\{[Y_{t(c)}(Z_i(c))]^2 | \mathcal{F}\}, \\
\tilde{S}_c^2 &= \sum_{i=1}^{m(c)} W_i(c) E\{[Y_{t(c)}(Z_i(c))]^2 | \mathcal{F}\},
\end{aligned}$$

and $X_1(c) = \sqrt{m(c)}\hat{S}_c/\tilde{S}_c$. We have

$$\tilde{S}_c^2 = \sigma^2 + \left(\sum_{i=1}^{m(c)} W_i(c)\sigma^2(Z_i(c)) - \sigma^2 \right) + \sum_{i=1}^{m(c)} W_i(c) \left[E\{[Y_{t(c)}(Z_i(c))]^2 | \mathcal{F}\} - \sigma^2(Z_i(c)) \right].$$

It now follows from parts (iii) and (iv) of Assumption 5.1, parts (i) and (v) of Assumption 5.2, and the converging together lemma (see, e.g., Theorem 25.4 of Billingsley, 1995) that $\tilde{S}_c^2 \Rightarrow \sigma^2$ as $c \to \infty$. Moreover, part (iv) of Assumption 5.1 implies that $l \leq [X_1(c)]^2 \leq u$.

From part (i) of Assumption 5.1, we have that for all $c \geq 0$, $\sqrt{c}\bar{\alpha}_1(c) = X_1(c) \times \bar{\alpha}_1'(c)$, where

$$\bar{\alpha}_1'(c) = \tilde{S}_c \times \frac{\sum_{i=1}^{m(c)} W_i(c) Y_{t(c)}(Z_i(c))}{\hat{S}_c}.$$

It follows from parts (i), (ii), (iii), and (iv) of Assumption 5.1 and part (i) of Assumption 5.2, that there exists $C > 0$ such that

$$\hat{S}_c^2 \geq \frac{l^2}{m(c)} \times \frac{\inf_{z \in \mathcal{Z}} \sigma^2(z)}{2} > 0$$

for all $c \geq C$. Therefore, parts (iii) and (iv) of Assumption 5.1 and part (ii) of Assumption 5.2 imply that

$$\begin{aligned} 0 \;&\leq\; \sum_{i=1}^{m(c)} \frac{[W_i(c)]^{2+\epsilon} E\{[Y_{t(c)}(Z_i(c))]^{2+\epsilon} | \mathcal{F}\}}{\hat{S}_c^{2+\epsilon}} \\ &\leq \left(\frac{2u^2}{l^2 \inf_{z \in \mathcal{Z}} \sigma^2(z)} \right)^{1+\epsilon/2} \times \frac{\sup_{z \in \mathcal{Z}, t \geq 0} E\{[Y_t(z)]^{2+\epsilon}\}}{[m(c)]^{\epsilon/2}}, \end{aligned}$$

for all $c \geq C$, and hence that

$$\lim_{c \to \infty} \sum_{i=1}^{m(c)} \frac{[W_i(c)]^{2+\epsilon} E\{[Y_{t(c)}(Z_i(c))]^{2+\epsilon} | \mathcal{F}\}}{\hat{S}_c^{2+\epsilon}} = 0$$

by part (ii) of Assumption 5.1. By part (iii) of Assumption 5.1 and Lindeberg's theorem (see for example Theorems 27.2 and 27.3 of Billingsley, 1995), we now have that the conditional distribution of the term $\sum_{i=1}^{m(c)} W_i(c) Y_{t(c)}(Z_i(c))/\hat{S}_c$ given $\mathcal{F}$ always converges to the standard normal distribution $N(0,1)$ as $c \to \infty$ (for all possible locations and weights). The weak convergence of $\bar{\alpha}_1'(c)$ to $Y \sim N(0, \sigma^2)$ as $c \to \infty$ now follows from the continuous mapping theorem, the bounded convergence theorem (see for example Theorem 16.5 of Billingsley, 1995), and the fact that $\tilde{S}_c^2 \Rightarrow \sigma^2$ as $c \to \infty$. Similarly, the asymptotic independence of $\bar{\alpha}_1'(c)$ and $X_3(c)$ now follows from the bounded convergence theorem. □

Remark 5.2. *If $(X_1(c), X_3(c)) \Rightarrow (X_1, X_3)$ as $c \to \infty$, then the bounded convergence theorem can be used to show that $(\bar{\alpha}_1'(c), X_1(c), X_3(c)) \Rightarrow (Y, X_1, X_3)$ as $c \to \infty$, where Y and (X_1, X_3) are independent (see the proof of Lemma 5.1 for a similar argument).*

LEMMA 5.3. *Suppose that parts (i) and (ii) of Assumption 5.1 and parts (iii) and (v) of Assumption 5.2 hold. Then*

$$\bar{\alpha}_2(c) = X_2(c)/[t(c)]^{\gamma}, \text{ where } X_2(c) \Rightarrow b \text{ as } c \to \infty.$$

Proof: By part (iii) of Assumption 5.2, we clearly have that

$$\bar{\alpha}_2(c) = \frac{b}{[t(c)]^{\gamma}} + \frac{1}{[t(c)]^{\gamma}} \left[\sum_{i=1}^{m(c)} W_i(c) b(Z_i(c)) - b \right] + o\left(\frac{1}{[t(c)]^{\gamma}} \right) \sum_{i=1}^{m(c)} W_i(c) e(Z_i(c))$$

as $c \to \infty$. The result now follows from parts (i) and (ii) of Assumption 5.1 and part (v) of Assumption 5.2. □

The following theorem specifies the rate at which $\bar{\alpha}(c)$ converges to α as $c \to \infty$ as a function of the choice of the parameter δ (see part (ii) of Assumption 5.1).

THEOREM 5.4. *Suppose that Assumptions 5.1 and 5.2 hold. Then the following statements hold:*

(a). When $\delta < (2\gamma - 1)/(2\gamma)$*, there are two cases:*
(i). If $\sqrt{c}/g(c) \to g$ *as* $c \to \infty$*, where* $g \in \mathbb{R}$*, then*

$$\sqrt{c}(\bar{\alpha}(c) - \alpha) = X_1(c)\bar{\alpha}_1'(c) + gX_3(c) + \epsilon(c),$$

where $\sqrt{l} \le X_1(c) \le \sqrt{u}$ *for all* $c \in \mathbb{R}^+$*,* $(\bar{\alpha}_1'(c), X_3(c)) \Rightarrow (Y, X_3)$ *as* $c \to \infty$ *with* Y *and* X_3 *being independent, and* $\epsilon(c) \Rightarrow 0$ *as* $c \to \infty$*.*
(ii). If $\sqrt{c}/g(c) \to \infty$ *as* $c \to \infty$*, then*

$$g(c)(\bar{\alpha}(c) - \alpha) \Rightarrow X_3 \quad \textit{as} \quad c \to \infty.$$

(b). When $\delta = (2\gamma - 1)/(2\gamma)$*, there are two cases:*
(i). If $\sqrt{c}/g(c) \to g$ *as* $c \to \infty$*, where* $g \in \mathbb{R}$*, then*

$$\sqrt{c}(\bar{\alpha}(c) - \alpha) = X_1(c)\bar{\alpha}_1'(c) + bd^\gamma + gX_3(c) + \epsilon(c),$$

where $\sqrt{l} \le X_1(c) \le \sqrt{u}$ *for all* $c \in \mathbb{R}^+$*,* $(\bar{\alpha}_1'(c), X_3(c)) \Rightarrow (Y, X_3)$ *as* $c \to \infty$ *with* Y *and* X_3 *being independent, and* $\epsilon(c) \Rightarrow 0$ *as* $c \to \infty$*.*
(ii). If $\sqrt{c}/g(c) \to \infty$ *as* $c \to \infty$*, then*

$$g(c)(\bar{\alpha}(c) - \alpha) \Rightarrow X_3 \quad \textit{as} \quad c \to \infty.$$

(c). When $\delta > (2\gamma - 1)/(2\gamma)$*, there are two cases:*
(i). If $c^{\gamma(1-\delta)}/g(c) \to g$ *as* $c \to \infty$*, where* $g \in \mathbb{R}$*, then*

$$c^{\gamma(1-\delta)}(\bar{\alpha}(c) - \alpha) \Rightarrow bd^\gamma + gX_3 \quad \textit{as} \quad c \to \infty.$$

(ii). If $c^{\gamma(1-\delta)}/g(c) \to \infty$ *as* $c \to \infty$*, then*

$$g(c)(\bar{\alpha}(c) - \alpha) \Rightarrow X_3 \quad \textit{as} \quad c \to \infty.$$

Proof: Note that equation (19), Lemmas 5.1 and 5.3, part (i) of Assumption 5.1, and part (iv) of Assumption 5.2 imply that

$$\bar{\alpha}(c) - \alpha = \frac{X_1(c)\bar{\alpha}_1'(c)}{\sqrt{c}} + \frac{X_2(c)}{c^{\gamma(1-\delta)}} \times \left(\frac{m(c)}{c^\delta}\right)^\gamma + \frac{X_3(c)}{g(c)}$$

for all $c \in \mathbb{R}^+$. Moreover, it is clear that $\gamma(1 - \delta) > 1/2$ if and only if $\delta < (2\gamma - 1)/(2\gamma)$. The result now follows from Lemmas 5.1 and 5.3, part (ii) of Assumption 5.1, part (iv) of Assumption 5.2, and the continuous mapping theorem. □

Remark 5.5. *It is clear from Theorem 5.4 and its proof that when Assumptions 5.1 and 5.2 hold with* $b \neq 0$*, then* $\delta \le (2\gamma-1)/(2\gamma)$ *is a necessary condition for the estimator* $\bar{\alpha}(c)$ *to converge to* α *at the best possible rate* $c^{-1/2}$*, and* $\delta \le (2\gamma-1)/(2\gamma)$ *and* $\sqrt{c}/g(c) \to g$ *as* $c \to \infty$*, where* $g \in \mathbb{R}$*, is a sufficient condition for this result. Also, it is frequently the case that simulation estimators obtained from a sample path of length* t *have a principal bias term of the order* $1/t$*, see Remark 3.6, and hence the special case when* $\gamma = 1$ *is of particular interest. In this case, the necessary condition for obtaining the best possible convergence rate* $c^{-1/2}$ *is given by* $\delta \le 0.5$*, implying that* $\bar{\alpha}(c)$ *will only converge to* α *at the best possible rate when the number of locations* $m(c)$ *grows no faster than the computational effort* $t(c)$ *used to estimate the value of the function* r *at each location. This is consistent with Theorems 4.9 and B.4, see Remarks 4.11 and B.5 (Theorem B.4 and Remark B.5 are provided in the online appendix to this paper).*

6. CONCLUSION

The use of Bayesian methods to determine the expected performance of a stochastic system often requires the computation of the quantity $\alpha = E\{r(Z)\}$, where the vector Z represents the uncertain (input) parameters of the system and the function values $r(z)$ represent the expected performance of the system when $Z = z$. We have studied the bias, consistency, and rate of convergence of three classes of simulation estimators for α as the total computational effort c grows. We have also provided a general framework for estimating α, and have characterized the convergence rate of the resulting estimator.

The three specific classes of estimators we consider all involve using simulation to estimate the function values $r(z)$ for a number of different values z of the random vector Z. The primary difference between the three approaches lies in the choice of the values z of Z for which the function values $r(z)$ are estimated. The first approach generates these values using independent sampling, the second approach uses a quasi-random sequence, and the third approach is based on Simpson's numerical integration rule. We show that the estimators based on Simpson's rule have the best possible convergence rate $c^{-1/2}$ and that the use of a quasi-random sequence leads to a better convergence rate than the use of independent sampling. Other specific methods could of course be used to choose the values of Z (e.g., stratification, Latin hypercube sampling, etc.). The study of these methods is a valuable direction for future work, but is outside the scope of the current paper.

ACKNOWLEDGMENTS

We thank the Area Editor, Associate Editor, and two anonymous reviewers for comments that led to substantial improvements in the paper.

REFERENCES

S. Andradóttir and V. M. Bier. Applying Bayesian Ideas in Simulation. *Simulation Practice and Theory*, 2000, 253–280.

S. Andradóttir and P. W. Glynn. Computing Bayesian Means Using Simulation. Technical Report, School of Industrial and Systems Engineering, Georgia Institute of Technology, 2002.

S. Asmussen and P. W. Glynn. *Stochastic Simulation: Algorithms and Analysis*. Springer, New York, NY, 2007.

H. P. Awad and P. W. Glynn. On the Theoretical Comparison of Low-Bias Steady-state Estimators. *ACM Transactions on Modeling and Computer Simulation*, 2007.

P. Billingsley. *Convergence of Probability Measures*. Wiley, New York, NY, 1968.

P. Billingsley. *Probability and Measure*, Third Edition. Wiley, New York, NY, 1995.

P. Bratley, B. L. Fox, and L. E. Schrage. *A Guide to Simulation*, Second Edition. Springer-Verlag, New York, NY, 1987.

M. Broadie, Y. Du, and C. C. Moallemi. Efficient Risk Estimation via Nested Sequential Simulation. *Management Science*, 2011, 1172–1194.

S. E. Chick. Subjective Probability and Bayesian Methodology. Chapter 9 in *Handbooks in Operations Research and Management Science: Simulation*, edited by Shane G. Henderson and Barry L. Nelson. North-Holland, Amsterdam, The Netherlands, 2006.

P. J. Davis and P. Rabinowitz. *Methods of Numerical Integration*, Second Edition. Academic Press, Orlando, FL, 1984.

D. Duffie. *Dynamic Asset Pricing Theory*, Second Edition. Princeton University Press, Princeton, NJ, 1996.

D. Duffie and P. Glynn. Efficient Monte Carlo Simulation of Security Prices. *The Annals of Applied Probability*, 1995, 897–905.

P. W. Glynn and P. Heidelberger. Jackknifing under a Budget Constraint. *ORSA Journal on Computing*, 1992, 226–234.

M. B. Gordy and S. Juneja. Nested Simulation in Portfolio Risk Measurement. *Management Science*, 2010, 1833–1848

H. Lan, B. L. Nelson, and J. Staum. A Confidence Interval Procedure for Expected Shortfall Risk Measurement via Two-Level Simulation. *Operations Research*, 2010, 1481–1490.

A. M. Law and W. D. Kelton. *Simulation Modeling and Analysis*, Third Edition. McGraw-Hill, Boston, MA, 2000.

P. L'Ecuyer. Quasi-Monte Carlo Methods with Applications in Finance. *Finance and Stochastics*, 2009, 307–349.

S.-H. Lee and P. W. Glynn. Computing the Distribution Function of a Conditional Expectation via Monte Carlo: Discrete Conditioning Spaces. *ACM Transactions on Modeling and Computer Simulation*, 2003, 238–258.

H. Niederreiter. *Random Number Generation and Quasi-Monte Carlo Methods*. SIAM, Philadelphia, PA, 1992.

S. G. Steckley and S. G. Henderson. A Kernel Approach to Estimating the Density of a Conditional Expectation. *Proceedings of the 2003 Winter Simulation Conference*, edited by S. Chick, P. J. Sánchez, D. Ferrin, and D. J. Morrice, 383–391. IEEE Press.

Y. Sun, D. W. Apley, and J. Staum. Efficient Nested Simulation for Estimating the Variance of a Conditional Expectation. *Operations Research*, 2011, 998–1007.

F. Zouaoui and J. R. Wilson. Accounting for Parameter Uncertainty in Simulation Input Modeling. *IIE Transactions*, 2003, 781–792

Online Appendix
Computing Bayesian Means Using Simulation
Sigrún Andradóttir, Georgia Institute of Technology
Peter W. Glynn, Stanford University

A. SUPPORTING MATERIAL FOR SECTION 4.1 IN THE MAIN PAPER

Proof of Lemma 4.6: For all $c \geq 0$, we have

$$\frac{s_c^2}{m(c)} = \sigma^2 + \left(\frac{1}{m(c)} \sum_{i=1}^{m(c)} \sigma^2(z_i) - \sigma^2 \right) + \left(\frac{1}{m(c)} \sum_{i=1}^{m(c)} \left[E\{[Y_i(c)]^2\} - \sigma^2(z_i) \right] \right).$$

By part (iv) of Assumption 4.1, part (iii) of Assumption 4.2, Assumption 4.3, and pages 14 and 17 of Niederreiter (1992), it is clear that the second term in the above expression converges to zero as $c \to \infty$. Similarly, it follows from part (ii) of Assumption 4.1, part (i) of Assumption 4.2, and Assumption 4.3 that the third term in the above expression converges to zero as $c \to \infty$. This shows that $s_c^2/m(c) \to \sigma^2$ as $c \to \infty$. From part (ii) of Assumption 4.1, we have that for all $c \geq 0$,

$$\sqrt{c}\hat{\alpha}_1(c) = \frac{s_c}{\sqrt{m(c)}} \times \frac{\sum_{i=1}^{m(c)} Y_i(c)}{s_c}.$$

By part (v) of Assumption 4.1, part (ii) of Assumption 4.2, and Lindeberg's theorem we have that $\sum_{i=1}^{m(c)} Y_i(c)/s_c \Rightarrow N(0,1)$ as $c \to \infty$. The result now follows from the continuous mapping theorem. □

Proof of Lemma 4.7: By part (iv) of Assumption 4.2, we clearly have that

$$\hat{\alpha}_2(c) = \frac{b\kappa^\gamma}{c^{\gamma(1-\delta)}} + \frac{b}{c^{\gamma(1-\delta)}} \left(\frac{c^{\gamma(1-\delta)}}{t(c)^\gamma} - \kappa^\gamma \right) + \frac{1}{t(c)^\gamma} \left[\frac{1}{m(c)} \sum_{i=1}^{m(c)} b(z_i) - b \right]$$
$$+ o\left(\frac{1}{t(c)^\gamma} \right) \frac{1}{m(c)} \sum_{i=1}^{m(c)} e(z_i)$$

as $c \to \infty$. From part (ii) of Assumption 4.1 and Assumption 4.3, we have that $c^{\gamma(1-\delta)}/t(c)^\gamma \to \kappa^\gamma$ as $c \to \infty$. The result now follows from the fact that part (iv) of Assumption 4.1, parts (v) and (vi) of Assumption 4.2, Assumption 4.3, and pages 14 and 17 of Niederreiter (1992) imply that $\frac{1}{m(c)} \sum_{i=1}^{m(c)} b(z_i)$ converges to b and $\frac{1}{m(c)} \sum_{i=1}^{m(c)} e(z_i)$ to e as $c \to \infty$. □

Proof of Lemma 4.8: By the Koksma-Hlawka inequality and part (vii) of Assumption 4.2, we have that

$$|\hat{\alpha}_3(c)| = \left| \frac{1}{m(c)} \sum_{i=1}^{m(c)} r(z_i) - \alpha \right| \leq O\left(\frac{[\log m(c)]^\iota}{[m(c)]^\beta} \right) \text{ as } c \to \infty.$$

The result now follows from Assumption 4.3. □

B. SUPPORTING MATERIAL FOR SECTION 4.2 IN THE MAIN PAPER

The following assumption gives additional details on how Simpson's rule would be used to estimate α (see the estimator (18)).

ASSUMPTION B.1. *Assume that:*

(i). The random variable $r(Z)$ is integrable.
(ii). For all $c \in \mathbb{R}^+$, the parameters $m(c)$ and $t(c)$ satisfy $c = m(c) \times t(c)$.
(iii). The random vector Z can be expressed as $Z = h(U)$, where U is a uniformly distributed random variable on the set $[0,1]$ and $h : [0,1] \to \mathcal{Z}$ is a known function.
(iv). For all $c \in \mathbb{R}^+$ and $i = 1, \ldots, m(c)$, let $u_i(c) = (i-1)/(m(c)-1)$.
(v). For all $c \in \mathbb{R}^+$, the random variables $\hat{r}_{t(c)}(z_i(c))$, where $z_i(c) = h(u_i(c))$ and $i \in \mathbb{N}$, are independent.
(vi). For all $c \in \mathbb{R}^+$, the weights $w_i(c)$, where $i = 1, \ldots, m(c)$, satisfy

$$w_i(c) = \begin{cases} 1/[3(m(c)-1)] & \text{if } i \in \{1, m(c)\}, \\ 4/[3(m(c)-1)] & \text{if } 1 < i < m(c) \text{ and } i \text{ is even}, \\ 2/[3(m(c)-1)] & \text{if } 1 < i < m(c) \text{ and } i \text{ is odd}. \end{cases}$$

(vii). For all $c \geq C$, where $C \in \mathbb{R}^+$, the parameter $m(c)$ satisfies $m(c) = 1 + 2\lfloor \kappa c^{\delta} \rfloor$, where $\kappa > 0$ and $0 < \delta < 1$.

For all $c \in \mathbb{R}^+$, let

$$Y_i'(c) = \begin{cases} \sqrt{t(c)}(\hat{r}_{t(c)}(z_i(c)) - r_{t(c)}(z_i(c)))/2 & \text{if } i \in \{1, m(c)\}, \\ \sqrt{t(c)}(\hat{r}_{t(c)}(z_i(c)) - r_{t(c)}(z_i(c))) & \text{if } 1 < i < m(c), \end{cases}$$

where $z_1(c), \ldots, z_{m(c)}(c)$ are defined in part (v) of Assumption B.1. The results given in the remainder of this section will require some (or all) the parts of the following technical assumption.

ASSUMPTION B.2. *Assume that:*

(i). For all $z \in \mathcal{Z}$, the random variables $\hat{r}_t(z)$, where $t \geq 0$, satisfy $E\{t[\hat{r}_t(z) - r_t(z)]^2\} \to \sigma^2(z)$ uniformly in $z \in \mathcal{Z}$ as $t \to \infty$, where $\sigma^2(z) \in \mathbb{R}^+$.
(ii). For all $\epsilon > 0$, the random variables $Y_i'(c)$, where $i \in \mathbb{N}$ and $c \in \mathbb{R}^+$, satisfy

$$\lim_{c\to\infty} \sum_{j=1}^{(m(c)-1)/2} \frac{1}{(s_c^e)^2} E\{[Y_{2j}'(c)]^2 I_{\{|Y_{2j}'(c)| \geq \epsilon s_c^e\}}\} = 0 \quad \textit{and}$$

$$\lim_{c\to\infty} \sum_{j=1}^{(m(c)+1)/2} \frac{1}{(s_c^o)^2} E\{[Y_{2j-1}'(c)]^2 I_{\{|Y_{2j-1}'(c)| \geq \epsilon s_c^o\}}\} = 0,$$

where $(s_c^e)^2 = \sum_{j=1}^{(m(c)-1)/2} E\{[Y_{2j}'(c)]^2\}$ and $(s_c^o)^2 = \sum_{j=1}^{(m(c)+1)/2} E\{[Y_{2j-1}'(c)]^2\}$ for all $c \in \mathbb{R}^+$.
(iii). The function $\sigma^2 \circ h : [0,1] \to \mathbb{R}^+$ is two times continuously differentiable.
(iv). The function r_t satisfies $r_t(z) = r(z) + b(z)/t^{\gamma} + e(z)o(1/t^{\gamma})$ as $t \to \infty$ for all $z \in \mathcal{Z}$, where $\gamma > 0$, $b(z), e(z) \in \mathbb{R}$ for all $z \in \mathcal{Z}$, and the $o(1/t^{\gamma})$ term is uniform in $z \in \mathcal{Z}$.
(v). The function $b \circ h : [0,1] \to \mathbb{R}$ is two times continuously differentiable.
(vi). The function $e \circ h : [0,1] \to \mathbb{R}$ is two times continuously differentiable.
(vii). The function $r \circ h : [0,1] \to \mathbb{R}$ is four times continuously differentiable.

To conserve space, we do not present results on the bias and consistency of the estimator $\tilde{\alpha}(c)$ as $c \to \infty$, but focus on studying the rate of convergence of $\tilde{\alpha}(c)$ to α as $c \to \infty$. As in equation (9), let

$$\tilde{\alpha}(c) - \alpha = \tilde{\alpha}_1(c) + \tilde{\alpha}_2(c) + \tilde{\alpha}_3(c), \tag{20}$$

where

$$\tilde{\alpha}_1(c) = \sum_{i=1}^{m(c)} w_i(c)[\hat{r}_{t(c)}(z_i(c)) - r_{t(c)}(z_i(c))],$$
$$\tilde{\alpha}_2(c) = \sum_{i=1}^{m(c)} w_i(c)[r_{t(c)}(z_i(c)) - r(z_i(c))], \quad \text{and}$$
$$\tilde{\alpha}_3(c) = \sum_{i=1}^{m(c)} w_i(c)[r(z_i(c)) - \alpha]$$

for all $c \geq 0$. As before, define $b = E\{b(Z)\}$, $e = E\{e(Z)\}$, and $\sigma^2 = E\{\sigma^2(Z)\}$, and note that part (iii) of Assumption B.1 and parts (iii), (v), and (vi) of Assumption B.2 imply that the random variables $b(Z)$, $e(Z)$, and $\sigma^2(Z)$ are integrable. We shall need the following three lemmas.

LEMMA B.1. *Suppose that Assumption B.1 and parts (i), (ii), and (iii) of Assumption B.2 hold. Then*

$$\sqrt{c}\tilde{\alpha}_1(c) \Rightarrow N(0, 10\sigma^2/9) \text{ as } c \to \infty.$$

Proof: For all $c \geq 0$, we have

$$\frac{(s_c^e)^2}{m(c)} = \frac{\sigma^2}{2} + \left(\frac{1}{m(c)} \sum_{j=1}^{(m(c)-1)/2} \sigma^2(z_{2j}(c)) - \frac{\sigma^2}{2} \right) + \left(\frac{1}{m(c)} \sum_{j=1}^{(m(c)-1)/2} \left[E\{[Y'_{2j}(c)]^2\} - \sigma^2(z_{2j}(c)) \right] \right), \tag{21}$$

$$\frac{(s_c^o)^2}{m(c)} = \frac{\sigma^2}{2} + \left(\frac{1}{m(c)} \sum_{j=2}^{(m(c)-1)/2} \sigma^2(z_{2j-1}(c)) - \frac{\sigma^2}{2} \right) + \frac{1}{m(c)} \left[E\{[Y'_1(c)]^2\} + E\{[Y'_{m(c)}(c)]^2\} \right] + \left(\frac{1}{m(c)} \sum_{j=2}^{(m(c)-1)/2} \left[E\{[Y'_{2j-1}(c)]^2\} - \sigma^2(z_{2j-1}(c)) \right] \right). \tag{22}$$

By part (vii) of Assumption B.1, part (iii) of Assumption B.2, and equation (2.1.12) in Davis and Rabinowitz (1984), we have that the second term in equation (21) converges to zero as $c \to \infty$. Similarly, by part (vii) of Assumption B.1, part (iii) of Assumption B.2, and equation (2.1.11) in Davis and Rabinowitz (1984), we have that the second term in equation (22) converges to zero as $c \to \infty$. Moreover, from part (vii) of Assumption B.1 and part (i) of Assumption B.2, we have that the third term in equation (21) and the third and fourth terms in equation (22) converge to zero as $c \to \infty$. This shows that $(s_c^e)^2/m(c) \to \sigma^2/2$ and $(s_c^o)^2/m(c) \to \sigma^2/2$ as $c \to \infty$.

From parts (ii) and (vi) of Assumption B.1, we have that for all $c \geq 0$,

$$\sqrt{c}\tilde{\alpha}_1(c) = \frac{4s_c^e\sqrt{m(c)}}{3(m(c)-1)} \times \frac{\sum_{j=1}^{(m(c)-1)/2} Y'_{2j}(c)}{s_c^e} + \frac{2s_c^o\sqrt{m(c)}}{3(m(c)-1)} \times \frac{\sum_{j=1}^{(m(c)+1)/2} Y'_{2j-1}(c)}{s_c^o}.$$

By part (v) of Assumption B.1, part (ii) of Assumption B.2, and Lindeberg's theorem, we have that the terms $\sum_{j=1}^{(m(c)-1)/2} Y'_{2j}(c)/s_c^e$ and $\sum_{j=1}^{(m(c)+1)/2} Y'_{2j-1}(c)/s_c^o$ both converge weakly to the standard normal distribution $N(0,1)$ as $c \to \infty$. Since part (v) of Assumption B.1 implies that the two terms in the above expression are independent, the result now follows from the continuous mapping theorem and the additive property of the normal distribution. □

LEMMA B.2. *Suppose that Assumption B.1 and parts (iv), (v), and (vi) of Assumption B.2 hold. Then*

$$\tilde{\alpha}_2(c) = \frac{b2^\gamma \kappa^\gamma}{c^{\gamma(1-\delta)}} + o\left(\frac{1}{c^{\gamma(1-\delta)}}\right) \text{ as } c \to \infty.$$

Proof: By part (iv) of Assumption B.2, we clearly have that

$$\tilde{\alpha}_2(c) = \frac{b2^\gamma \kappa^\gamma}{c^{\gamma(1-\delta)}} + \frac{b}{c^{\gamma(1-\delta)}}\left(\frac{c^{\gamma(1-\delta)}}{t(c)^\gamma} - 2^\gamma\kappa^\gamma\right) + \frac{1}{t(c)^\gamma}\left[\sum_{i=1}^{m(c)} w_i(c)b(z_i(c)) - b\right] + o\left(\frac{1}{t(c)^\gamma}\right)\sum_{i=1}^{m(c)} w_i(c)e(z_i(c))$$

as $c \to \infty$. From parts (ii) and (vii) of Assumption B.1, we have that $c^{\gamma(1-\delta)}/t(c)^\gamma \to 2^\gamma\kappa^\gamma$ as $c \to \infty$. The result now follows from the fact that part (vii) of Assumption B.1, parts (v), and (vi) of Assumption B.2, and equations (2.1.11) and (2.1.12) in Davis and Rabinowitz (1984) imply that

$$\left|\sum_{i=1}^{m(c)} w_i(c)b(z_i(c)) - b\right| \le \left|\sum_{j=1}^{(m(c)-1)/2} w_{2j}(c)b(z_{2j}(c)) - \frac{2b}{3}\right| + \left|\sum_{j=1}^{(m(c)+1)/2} w_{2j-1}(c)b(z_{2j-1}(c)) - \frac{b}{3}\right| \le O\left(\frac{1}{[m(c)]^2}\right) = O\left(\frac{1}{c^{2\delta}}\right)$$

and $\left|\sum_{i=1}^{m(c)} w_i(c)e(z_i(c)) - e\right| \le O\left(\frac{1}{c^{2\delta}}\right)$ as $c \to \infty$. □

LEMMA B.3. *Suppose that Assumption B.1 and part* (vii) *of Assumption B.2 hold. Then*

$$|\tilde{\alpha}_3(c)| \le O\left(\frac{1}{c^{4\delta}}\right) \text{ as } c \to \infty.$$

Proof: By part (vii) of Assumption B.2 and equation (2.2.6) in Davis and Rabinowitz (1984), we have that

$$|\tilde{\alpha}_3(c)| = \left|\sum_{i=1}^{m(c)} w_i(c)r(z_i(c)) - \alpha\right| \le O\left(\frac{1}{[m(c)]^4}\right) \text{ as } c \to \infty.$$

The result now follows from part (vii) of Assumption B.1. □

The following theorem specifies the rate at which $\tilde{\alpha}(c)$ converges to α as $c \to \infty$ as a function of the choice of the parameter δ (see part (vii) of Assumption B.1).

THEOREM B.4. *Suppose that Assumptions B.1 and B.2 hold. Then, the following statements hold:*

(a). If $\delta > \gamma/(\gamma+4)$ and $\gamma(1-\delta) > 1/2$, then

$$\sqrt{c}(\tilde{\alpha}(c) - \alpha) \Rightarrow N(0, 10\sigma^2/9) \quad \text{as} \quad c \to \infty.$$

(b). If $\delta > \gamma/(\gamma+4)$ and $\gamma(1-\delta) = 1/2$, then

$$\sqrt{c}(\tilde{\alpha}(c) - \alpha) \Rightarrow N(b2^\gamma\kappa^\gamma, 10\sigma^2/9) \quad \textit{as} \quad c \to \infty.$$

(c). If $\delta > \gamma/(\gamma+4)$ and $\gamma(1-\delta) < 1/2$, then

$$c^{\gamma(1-\delta)}(\tilde{\alpha}(c) - \alpha) \Rightarrow b2^\gamma\kappa^\gamma \quad \textit{as} \quad c \to \infty.$$

(d). If $\delta \leq \gamma/(\gamma+4)$ and $\delta > 1/8$, then

$$\sqrt{c}(\tilde{\alpha}(c) - \alpha) \Rightarrow N(0, 10\sigma^2/9) \quad \textit{as} \quad c \to \infty.$$

(e). If $\delta \leq \gamma/(\gamma+4)$ and $\delta = 1/8$, then

$$\sqrt{c}|\tilde{\alpha}(c) - \alpha| \leq X(c) + O(1) \quad \textit{as} \quad c \to \infty,$$

where $X(c) \Rightarrow N(0, 10\sigma^2/9)$ as $c \to \infty$.
(f). If $\delta \leq \gamma/(\gamma+4)$ and $\delta < 1/8$, then

$$c^{4\delta}|\tilde{\alpha}(c) - \alpha| \leq X(c) + O(1) \quad \textit{as} \quad c \to \infty,$$

where $X(c) \Rightarrow 0$ as $c \to \infty$.

Proof: Note that $\delta > \gamma/(\gamma+4)$ if and only if $4\delta > \gamma(1-\delta)$. The result now follows from equation (20), Lemmas B.1, B.2, and B.3, and the continuous mapping theorem. □

Remark B.5. *From parts (a), (b), (d), and (e) of Theorem B.4, it is clear that when $1/8 \leq \delta \leq (2\gamma-1)/(2\gamma)$ and Assumptions B.1 and B.2 hold, then the estimator $\hat{\alpha}(c)$ converges to α at the rate $c^{-1/2}$ as $c \to \infty$. Moreover, the interval $[1/8, (2\gamma-1)/(2\gamma)]$ is non-empty for all $\gamma \geq 4/7$ and includes the value $\delta = 1/2$ for all $\gamma \geq 1$ (as would typically be the case in practice). Thus, we can always obtain the best possible convergence rate $c^{-1/2}$ in this situation (as long as Assumptions B.1 and B.2 hold and $\gamma \geq 4/7$), and when $\gamma \geq 1$, we can always obtain the best possible convergence rate with $\delta = 1/2$. This is an improvement over the rate of convergence results obtained in Sections 3.1 and 4.1, see Remarks 3.5 and 4.11. However, as in Section 4.1, this improvement is achieved using assumptions that may be difficult to verify in practice, including the smoothness assumptions in parts (iii), (v), (vi), and (vii) of Assumption B.2. Note that we have more flexibility here in the choice of the growth rates of $m(c)$ and $t(c)$ with respect to c, and that we can even let $t(c)$ grow at a faster rate than $m(c)$ and still obtain the best possible convergence rate. This is again an improvement over the independent sampling and quasi-random approaches, see Remarks 3.5 and 4.11.*

Remark B.6. *If part (vii) of Assumption B.2 is replaced by the assumption that the function $r \circ h : [0,1] \to \mathbb{R}$ is two times continuously differentiable, then one can use equations (2.1.11) and (2.1.12) of Davis and Rabinowitz (1984) to show that $|\tilde{\alpha}_3(c)| \leq O(1/c^{2\delta})$ as $c \to \infty$. This fact can then be used together with Lemmas B.1 and B.2 to obtain a revised version of Theorem B.4.*