Green Simulation: Reusing the Output of Repeated Experiments

MINGBIN FENG, University of Waterloo
JEREMY STAUM, Northwestern University

We introduce a new paradigm in simulation experiment design and analysis, called "green simulation," for the setting in which experiments are performed repeatedly with the same simulation model. Green simulation means reusing outputs from previous experiments to answer the question currently being asked of the simulation model. As one method for green simulation, we propose estimators that reuse outputs from previous experiments by weighting them with likelihood ratios, when parameters of distributions in the simulation model differ across experiments. We analyze convergence of these estimators as more experiments are repeated, while a stochastic process changes the parameters used in each experiment. As another method for green simulation, we propose an estimator based on stochastic kriging. We find that green simulation can reduce mean squared error by more than an order of magnitude in examples involving catastrophe bond pricing and credit risk evaluation.

CCS Concepts: • **Computing methodologies** → **Modeling and simulation**; **Model development and analysis**; **Simulation types and techniques**;

Additional Key Words and Phrases: Likelihood ratio method, multiple importance sampling, score function method, simulation metamodeling

ACM Reference format:
Mingbin Feng and Jeremy Staum. 2017. Green Simulation: Reusing the Output of Repeated Experiments. *ACM Trans. Model. Comput. Simul.* 27, 4, Article 23 (October 2017), 28 pages.
https://doi.org/10.1145/3129130

1 INTRODUCTION

Consider a setting in which simulation experiments are performed repeatedly, using the same simulation model with different values of its inputs. As we discuss in detail below, such settings occur when a simulation model is used routinely to support a business process, and over the lifecycle of a simulation model as it goes from development to application in repeated simulation studies. In these settings, the standard practice is that each new simulation experiment is designed to answer a particular question without using the output of previous simulation experiments. We advocate a paradigm of *green simulation* for repeated experiments, meaning that one should reuse output from previous experiments to answer new questions. The benefit of green simulation is greater computational efficiency. In this article, we show that when old simulation output is reused well,

Authors' addresses: M. Feng, University of Waterloo, 200 University Ave W, M3 3141, Waterloo, Ontario, Canada, N2L 3G1; email: ben.feng@uwaterloo.ca; J. Staum, Northwestern University, 2145 Sheridan Road, Tech Institute C210, Evanston, IL, 60208; email: j-staum@northwestern.edu.
Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. Copyrights for components of this work owned by others than ACM must be honored. Abstracting with credit is permitted. To copy otherwise, or republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee. Request permissions from permissions@acm.org.
© 2017 ACM 1049-3301/2017/10-ART23 $15.00
https://doi.org/10.1145/3129130

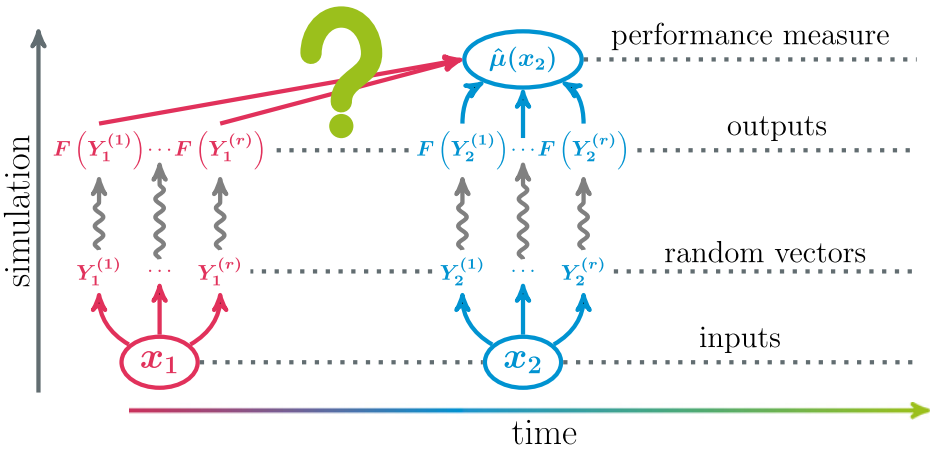

Fig. 1. Setting of repeated simulation experiments.

it provides greater accuracy when combined with a new simulation experiment than would be achieved by the same new simulation experiment alone.

Green simulation entails a new perspective on management of simulation experiments. The standard practice is to discard or ignore the output of a simulation experiment after it has delivered the desired answer. When a new question arises, a new simulation experiment is designed to answer it without using the output of previous simulation experiments. From this perspective, running the simulation model is a computational cost or expense. From the green simulation perspective, running the simulation model is a computational investment that provides future benefits, because simulation output is a valuable resource to be used in answering questions that will be asked of the simulation model in the future.

1.1 Settings

One setting of repeated simulation experiments occurs when simulation supports a business process. For example, in finance and insurance, simulation models support pricing and risk management decisions that are made periodically. At each period, current information, such as prices and forecasts, is used to update the inputs to the model, and a simulation experiment is performed to answer a question about price or risk. Similarly, in manufacturing, service, and logistics systems, simulation models can be used routinely to provide information about expected completion times and to support decisions about such matters as dispatching and staffing. A simulation experiment is run whenever information is required, using inputs that describe the current state of the system.

Another setting of repeated simulation experiments occurs in the lifecycle of a simulation model as it goes from development to application in simulation studies. First, experiments are performed for purposes of verification and validation of the model. They may also be performed for model calibration: to choose realistic values of unknown inputs. For these purposes, the simulation model is run many times with different values of its inputs, to see how its outputs change with its input, and where this behavior is reasonable and realistic. Once model development is complete, experiments are performed for purposes such as making predictions for particular values of the inputs, metamodeling, sensitivity analysis, and optimization. Moreover, some simulation models are used in many simulation studies. Consequently, each model is used in several or many experiments.

Figure 1 illustrates a sequence of two repeated experiments, each with a single run. To clarify our terminology, by "a run," we mean one or more replications of simulation output generated

with the inputs to the model held fixed. By a simulation "experiment," we mean a collection of one or more runs of a simulation model, designed for the purpose of answering a specific question. In Figure 1, the first experiment has a single run with input x_1, and the second experiment has a single run with input x_2. Each run has r replications. The purpose of the nth experiment is to estimate $\mu(x_n)$, the mean output of the simulation model when the input is x_n. The standard practice is to estimate $\mu(x_2)$ in the second experiment using only the single run with input x_2. The green question mark in Figure 1 indicates the question answered in this article: "How can we reuse the simulation output from the first experiment to improve our answer in the second experiment?"

1.2 Contributions

The main contribution of this article is to introduce the paradigm of green simulation for repeated experiments and to demonstrate theoretically and experimentally that it yields significant benefits in computational efficiency. Specifically, we investigate the case in which a simulation experiment is run routinely (e.g., to support a business process) with updated values of inputs to the simulation model. For this case, we propose, analyze, and test green simulation estimators based on the likelihood ratio (LR) method and on stochastic kriging (SK). For the LR estimators to be applicable, the changing inputs to the model must serve as parameters of distributions of random variables generated in the simulation. In particular, we prove a novel theorem about how the LR estimators converge as the number of repeated experiments increases, while the number of simulation replications per experiment remains constant. In this setting, the estimator based on standard practice does not converge at all, but our LR estimators converge at the canonical rate for Monte Carlo: variance inversely proportional to total computational budget. They converge at this rate even though the total computational budget includes all previous experiments, which are not obviously relevant to answering the question currently being asked of the simulation model. A secondary contribution of this article is to import methods from multiple importance sampling (where importance sampling is used for variance reduction) into the LR method (which reuses simulation output), and to provide evidence that there can be great practical value to using one particular weighting scheme in the LR method.

1.3 Literature Review

The core idea of green simulation, reusing simulation output, has been applied to isolated experiments that contain multiple runs. Many types of simulation experiments use multiple runs to learn about the model's *response surface*, the function that maps the model's inputs to a performance measure. In stochastic simulation, this performance measure is often the expected output of the simulation model. There are metamodeling and sensitivity analysis experiments that run the model at different input values to learn about how the response surface varies, globally or locally. In nested simulation experiments, an outer-level simulation generates random values of the inputs at which it is desired to learn the value of the response surface of an inner-level simulation. For example, in assessing the impact of uncertainty about a simulation model's input on the conclusions of a simulation study, this model is the inner-level simulation model, and the outer-level simulation samples values of the inputs from an appropriate distribution. In optimization via simulation experiments, the model is run at different input values in a search for optimal input values. If simulation output from runs at some values of the inputs can be reused in estimating the value of the response surface at another value of the inputs, then experiment designs that involve multiple runs can be modified to be cheaper. It is unnecessary to run many replications at every value of the inputs for which it is desired to estimate the value of the response surface if estimates of these values can reuse output from runs at other values of the inputs. For example, the nested simulation methods of Barton et al. (2014) and Xie et al. (2014) use stochastic

kriging (Ankenman et al. 2010) to reuse output from a moderate number of runs to estimate the value of the response surface for many values of the inputs, thus reducing the required number of runs in the simulation experiment. The LR method, also known as the score function method, has also been applied to reuse output within isolated experiments for metamodeling, sensitivity analysis, and optimization, see, for example, L'Ecuyer (1990, 1993), Rubinstein and Shapiro (1993), Kleijnen and Rubinstein (1996), Glasserman and Xu (2014), and Fu et al. (2015). We adopt the LR method and SK as tools for reusing output across experiments.

In applications and surveys of the LR method that we have seen, such as Beckman and McKay (1987), Rubinstein and Shapiro (1993), Kleijnen and Rubinstein (1996), and Glasserman and Xu (2014), there is an isolated experiment within which each estimator reuses the output from a single run. The exception is Maggiar et al. (2015): they construct an estimator by reusing output from multiple runs in an isolated experiment. We reuse output from multiple runs that come from multiple experiments. More specifically, we propose to exploit the temporal view of inputs as experiments are repeated over time. Even in the simple setting of a single run per experiment, which we analyze in this article, reusing output from multiple experiments entails reusing output from multiple runs. Estimation using multiple simulation runs, weighted based on likelihood ratios, was given the name *multiple importance sampling* by Veach (1997). On multiple importance sampling, see, for example, Hesterberg (1988, 1995), Owen and Zhou (2000), Veach and Guibas (1995), and Veach (1997). Drawing on this literature on multiple importance sampling, Maggiar et al. (2015) and Feng and Staum (2015) use two of the LR estimators we use, the individual likelihood ratio (ILR) estimator (Section 2.1) and the mixture likelihood ratio (MLR) estimator (Section 2.3); we also treat two new green estimators: the weighted likelihood ratio (WLR) estimator (Section 2.2) and the green stochastic kriging (GSK) estimator (Section 3).

The work of Maggiar et al. (2015) and our work differ in setting and findings. The key differences are that their work focuses on an isolated experiment with a deterministic simulation model, whereas ours focuses on repeated experiments with a stochastic simulation model. Their goal is optimization of the response surface after smoothing by convolution with a Gaussian kernel, to reduce the influence of numerical noise in the deterministic simulation. They apply the likelihood ratio method to the corresponding Gaussian random variable. Their convergence theorem describes convergence to an optimal solution within an isolated optimization experiment with an increasing number of iterations. Our green simulation paradigm applies broadly to stochastic simulation, and we emphasize the setting of repeated experiments. Our convergence theorems describe convergence of estimators of values of the response surface at some or all points as the number of repeated experiments increases. Another difference between the work of Maggiar et al. (2015) and our work is in the findings about the ILR and MLR estimators. They report that the choice between ILR and MLR estimators makes "only a small difference on the performance" of their optimization procedure when applied to a test bed of optimization problems. In Section 6, we find that the difference between ILR and MLR estimators could be large, depending on the simulation model and sequence of repeated experiments. We recommend the MLR estimator, which is theoretically superior, because we find that it can work well in practice, even in cases where the ILR estimator yields poor results.

2 GREEN SIMULATION VIA THE LIKELIHOOD RATIO METHOD

We develop green simulation estimators via the LR method in a setting of repeated experiments with the same simulation model and some changing parameters that affect the likelihood of simulated random variables. Let X_n represent the parameters in the nth experiment, for example, prices observed in the market on day n or forecasted arrival rates for period n. We treat $\{X_n : n = 1, 2, \ldots\}$ as a discrete-time stochastic process taking values in a Polish space $\mathcal{X}$. We use "state" to refer to

an element of $\mathcal{X}$ or a random variable taking values in $\mathcal{X}$, and "current" to refer to quantities associated with the nth experiment. Thus, X_n is the current state. We suppose that the current state is observable at the current time step n, when the current experiment needs to be run, but was not observable earlier.

Given the current state X_n, the current simulation experiment samples a random vector Y_n according to the conditional likelihood $h(\cdot; X_n)$. For example, $h(y; x)$ could be the conditional probability density for a stock's price to be y in one year given that the stock's current price is x. As another example, $h(y; x)$ could be the conditional probability density for the vector y of interarrival times and service times given the arrival rate x, regarding service rate as fixed and not included in x. We will write expectations in the form of integrals, which implicitly assumes that Y_n has a conditional probability density $h(\cdot; X_n)$, but this is not essential; the setting allows for continuous, discrete, and mixed conditional distributions.

The simulation output or simulated performance of the stochastic system is $F(Y_n)$, where the function $F : \mathcal{Y} \mapsto \mathbb{R}$ represents the logic of the simulation model. For example, $F(y)$ could be the discounted payoff of a stock option if the stock's price in one year is y, or $F(y)$ could be the average customer waiting time in a queue given the vector y of interarrival times and service times.

In the current experiment, we wish to estimate the conditional expected performance $\mu(X_n)$ of the stochastic system given the current state X_n; $\mu(X_n)$ can also be described as the current expected performance. The expected performance for state x is

$$\mu(x) = \mathbb{E}\left[F(Y_n) \,|X_n = x\right] = \int_{\mathcal{Y}} F(y)\, h(y; x)\, dy, \tag{1}$$

which is the same for all n. The current expected performance $\mu(X_n)$ is a random variable, because the current state X_n was not observable at time step 0. Figures of merit for an estimator, such as bias and variance, should be evaluated conditional on the current state X_n.

The standard practice in the setting of repeated experiments is to estimate $\mu(X_n)$ by the *Standard Monte Carlo (SMC)* estimator,

$$\widehat{\mu}_r^{SMC}(X_n) = \frac{1}{r}\sum_{j=1}^{r} F\left(Y_n^{(j)}\right), \tag{2}$$

based on running r replications of the simulation model with the parameters set according to the current state X_n. For simplicity in notation, we have assumed that the number r of replications is fixed, but this is not essential. We refer to $\{F(Y_n^{(j)}) : j = 1, \ldots, r\}$ as the output of the current experiment. Clearly, $\widehat{\mu}_r^{SMC}(X_n)$ is conditionally unbiased for $\mu(X_n)$, given X_n. The variance for state x is

$$\sigma^2(x) = \mathbb{V}\mathrm{ar}\left[F(Y_n) \,|X_n = x\right] = \int_{\mathcal{Y}} \left(F(y) - \mu(x)\right)^2 h(y; x)\, dy, \tag{3}$$

which is the same for all n. The conditional variance of the SMC estimator, given the current state X_n, is $\sigma^2(X_n)/r$. Thus, to reduce the conditional variance of the SMC estimator, one must increase the number r of replications in the current experiment.

We propose three LR estimators that reuse the output of the $n-1$ previous experiments and combine it with the current (nth) experiment's output. In the current experiment, the *target distribution* appears in the conditional expectation $\mu(X_n) = \mathbb{E}\left[F(Y_n) \,|X_n\right]$ that we are estimating. The target distribution has the likelihood $h(\cdot; X_n)$. In a previous experiment at time step $k < n$, the *sampling distribution* had likelihood $h(\cdot; X_k)$. To use the output of a previous experiment in estimating $\mu(X_n)$ in a way that is conditionally unbiased given the state history $X_1, \ldots, X_n$, we can adjust the old output by using the likelihood ratio between the target distribution and the sampling

distribution:

$$\mathbb{E}\left[\frac{h\left(Y_k;X_n\right)}{h\left(Y_k;X_k\right)}F\left(Y_k\right)|X_1,\ldots,X_n\right]=\int_{\mathcal{Y}}\frac{h\left(y;X_n\right)}{h\left(y;X_k\right)}h\left(y;X_k\right)F\left(y\right)dy=\mathbb{E}\left[F\left(Y_n\right)|X_n\right]. \tag{4}$$

In Section 5, we show that, under some conditions, the conditional variance of the LR estimators goes to zero as the number n of experiments goes to infinity, even if the number r of replications per experiment is fixed.

We make the following assumptions to support the LR estimators. Although not all of the assumptions in the theorems are transparent, we show in the Appendix A that they can be verified in a realistic example.

- (A1) The nth simulation experiment is affected by n and the stochastic process $\{X_n : n = 1, 2, \ldots\}$ only through the conditional likelihood $h\left(\cdot;X_n\right)$ of the random vector Y_n. Specifically, the simulation logic F does not depend on X_n.
- (A2) For any n, $n' \neq n$, j, and j', given the state X_n, $Y_n^{(j)}|X_n$ is conditionally independent of the state $X_{n'}$ and the simulated random vector $Y_{n'}^{(j')}|X_{n'}$.
- (A3) For any n, j, and $j' \neq j$, given the state X_n, the random vector $Y_n^{(j)}$ simulated in the jth replication of the nth experiment is conditionally independent of $Y_n^{(j')}$.
- (A4) For all $x \in \mathcal{X}$, the likelihoods $h\left(\cdot;x\right)$ have the same support $\mathcal{Y}$.
- (A5) For any state $x \in \mathcal{X}$ and any $y \in \mathcal{Y}$, the likelihood $h\left(y;x\right)$ can be evaluated.
- (A6) For any states $x, x' \in \mathcal{X}$, the target-x-sample-x' variance defined as

$$\sigma_x^2\left(x'\right)=\int_{\mathcal{Y}}\left(F\left(y\right)\frac{h\left(y;x\right)}{h\left(y;x'\right)}-\mu(x)\right)^2h\left(y;x'\right)dy, \tag{5}$$

  and the expected performance $\mu(x)$ defined in Equation (1) are finite.

In Assumption 1, it is essential for the LR method that the parameters that change affect only likelihoods, not the simulation logic F. Assumptions (A1) and (A2) imply that the current simulation experiment is not affected directly by its time index n or by any past state $X_{n'}$ for $n' < n$ or by any of the output of a past experiment. We use these assumptions on the sequence of repeated experiments to support our analysis of the LR estimators. Assumption (A3), which asserts that each simulation experiment has independent replications, is made for the sake of simplicity in the analysis. It is not essential to the LR method or its analysis: for example, it would be acceptable to use stratified sampling within each experiment, and this would introduce the usual complications in analyzing and estimating variance. Assumptions (A4) and (A5) ensure that the likelihood ratio $h\left(y;x\right)/h\left(y;x'\right)$ is finite and can be computed; it is not enough merely to be able to sample according to the likelihood $h\left(\cdot;x\right)$. Assumption (A6) is needed to give the LR estimators finite conditional variance.

The need to satisfy these assumptions limits the applicability of the LR method. Furthermore, the LR estimators could work poorly if the variances associated with Equation (5) are too large. This can happen, for example, if the input x affects the likelihood of many independent terms. These are known limitations of the LR method in general, not specific to green simulation. Despite them, interesting examples may fit well into the LR framework. One such example, in Section 6.1, involves simulating a random number of independent random variables and summing them. It might appear to pose a difficulty for the LR method if Y_n is a random vector including the number of terms in the sum and every term in the sum. We avoid this difficulty by defining Y_n to be the sum and working with its likelihood, not with the joint likelihood of the number of terms in the sum and every term in the sum. By using such techniques if necessary, the LR method can be made to apply

even to non-trivial examples. It is not always possible to make the LR method work well; in such situations, one can use other green simulation methods, such as stochastic kriging (SK) (Section 3).

2.1 Individual Likelihood Ratio Estimator

Based on the fundamental idea of LR estimation expressed in Equation (4), we get the following unbiased estimator of $\mu(x)$ from the kth experiment:

$$\widehat{\mu}_{k,r}^{LR}(x) = \frac{1}{r}\sum_{j=1}^{r}\frac{h\left(Y_k^{(j)};x\right)}{h\left(Y_k^{(j)};X_k\right)}F\left(Y_k^{(j)}\right).$$

Averaging these estimators from all previous experiments and the current (nth) experiment yields the *Individual Likelihood Ratio (ILR)* estimator of $\mu(x)$:

$$\widehat{\mu}_{n,r}^{ILR}(x) = \frac{1}{n}\sum_{k=1}^{n}\widehat{\mu}_{k,r}^{LR}(x) = \sum_{k=1}^{n}\frac{1}{n}\left(\frac{1}{r}\sum_{j=1}^{r}\frac{h\left(Y_k^{(j)};x\right)}{h\left(Y_k^{(j)};X_k\right)}F\left(Y_k^{(j)}\right)\right) \tag{6a}$$

$$= \sum_{j=1}^{r}\sum_{k=1}^{n}\frac{1}{nr}\frac{h\left(Y_k^{(j)};x\right)}{h\left(Y_k^{(j)};X_k\right)}F\left(Y_k^{(j)}\right). \tag{6b}$$

The ILR estimator is so named because it contains likelihood ratios that each involve one individual sampling distribution; this distinguishes it from the mixture likelihood ratio (MLR) estimator proposed in Section 2.3. In particular, $\widehat{\mu}_{n,r}^{ILR}(X_n)$ is our estimator of the current expected performance $\mu(X_n)$. However, the LR method enables us to estimate expected performance given a state x that we never used in a sampling distribution.

The ILR estimator in Equation (6a) can be seen as the average of the individual LR estimators $\widehat{\mu}_{k,r}^{LR}(x)$ for $k = 1,\ldots,n$. If the target state is the current state, that is, $x = X_n$, then $\widehat{\mu}_{n,r}^{LR}(X_n)$ is the SMC estimator. It is worth emphasizing the difference between the present application of likelihood ratios and their typical application of importance sampling. In importance sampling, the designer of the simulation experiment chooses the sampling distribution with the aim of reducing variance. In this article, we do not address the choice of sampling distribution. We assume that the sampling distribution in the current experiment is determined by the current state X_n, as is standard practice in the setting of repeated experiments. We use likelihood ratios not to reduce variance, but to enable reuse of simulation output based on different sampling distributions.

Of course, the effect of likelihood ratios on variance needs to be considered. The target-x-sample-x' variance defined in Equation (5) could be more or less than the target-x-sample-x variance $\sigma_x^2(x) = \sigma^2(x)$ associated with standard Monte Carlo. The target-x-sample-X_k variance $\sigma_x^2(X_k)$ can be estimated by

$$\widehat{\sigma}_x^2(X_k) = \frac{1}{r}\sum_{j=1}^{r}\left(\frac{h\left(Y_k^{(j)};x\right)}{h\left(Y_k^{(j)};X_k\right)}F\left(Y_k^{(j)}\right) - \widehat{\mu}_{k,r}^{LR}(x)\right)^2.$$

Based on Equation (6a), we have

$$\mathbb{V}\mathrm{ar}\left[\widehat{\mu}_{n,r}^{ILR}(x)\middle| X_1,\ldots,X_n\right] = \frac{1}{n^2 r}\sum_{k=1}^{n}\sigma_x^2(X_k), \tag{7}$$

which can be estimated by $\sum_{k=1}^{n}\widehat{\sigma}_x^2(X_k)/(n^2 r)$. If none of $\sigma_{X_n}^2(X_1),\ldots,\sigma_{X_n}^2(X_n)$ is too large, then the ILR estimator has lower conditional variance $\sum_{k=1}^{n}\sigma_{X_n}^2(X_k)/(n^2 r)$ than $\sigma^2(X_n)/r$, which is the conditional variance of the SMC estimator. Considering that the variances $\sigma_{X_n}^2(X_1),\ldots,\sigma_{X_n}^2(X_n)$

may be unequal, one might try to construct a lower-variance estimator by using unequal weights instead of the equal weights $1/n$ in Equation (6a). This leads to the estimator proposed in Section 2.2. Even better results are possible if we replace the equal weights $1/nr$ in Equation (6b) with unequal weights that depend on the random vector $Y_k^{(j)}$. This topic is addressed in the research literature on multiple importance sampling, which leads to the estimator proposed in Section 2.3.

2.2 Weighted Likelihood Ratio Estimator

Instead of the ILR estimator Equation (6a) with its equal weights $1/n$, consider a weighted average estimator of the form $\sum_{k=1}^n w_k \widehat{\mu}_{k,r}^{LR}(x)$ where the weights $w_1, \ldots, w_n$ sum to one but may be unequal. Its conditional variance given the state history $X_1, \ldots, X_n$ is $\sum_{k=1}^n w_k^2 \sigma_x^2(X_k)/r$. Choosing the weights

$$w_k^{WLR} = \frac{\sigma_x^{-2}(X_k)}{\sum_{i=1}^n \sigma_x^{-2}(X_i)} \tag{8}$$

that minimize the conditional variance yields the *Weighted Likelihood Ratio (WLR)* estimator

$$\widehat{\mu}_{n,r}^{WLR}(x) = \sum_{k=1}^n w_k^{WLR} \widehat{\mu}_{k,r}^{LR}(x) = \sum_{k=1}^n w_k^{WLR} \left[\frac{1}{r} \sum_{j=1}^r F\left(Y_k^{(j)}\right) \frac{h\left(Y_k^{(j)}; x\right)}{h\left(Y_k^{(j)}; X_k\right)} \right]. \tag{9}$$

Proposition 2.1. *Under Assumptions (A1)–(A6), for any $n, r \in \mathbb{N}_+$ and $x, X_1, \ldots, X_n \in \mathcal{X}$, the weights given in Equation (8) minimize $\mathbb{V}ar[\sum_{k=1}^n w_k \widehat{\mu}_{k,r}^{LR}(x) | X_1, \ldots, X_n]$ subject to the constraint $\sum_{k=1}^n w_k = 1$. The resulting conditional variance is*

$$\mathbb{V}ar\left[\widehat{\mu}_{n,r}^{WLR}(x) \middle| X_1, \ldots, X_n\right] = \frac{1}{r \sum_{k=1}^n \sigma_x^{-2}(X_k)}. \tag{10}$$

Proof. The Lagrangian for this minimization problem is $L(w_1, \ldots, w_n, \lambda) = \sum_{k=1}^n w_k^2 \sigma_x^2(X_k) + \lambda(1 - \sum_{k=1}^n w_k)$. From the first-order condition $\partial L(w_1, \ldots, w_n, \lambda)/\partial w_k^{WLR} = 2 w_k^{WLR} \sigma_x^2(X_k) - \lambda = 0$ we get $w_k^{WLR} = \lambda / 2\sigma_x^2(X_k)$. Therefore, w_k^{WLR} is proportional to $\sigma_x^{-2}(X_k)$, and the denominator in Equation (8) is what is required to satisfy the constraint $\sum_{k=1}^n w_k^{WLR} = 1$. Substituting w_k^{WLR} into the conditional variance formula $\sum_{k=1}^n w_k^2 \sigma_x^2(X_k)/r$ yields $1/r \sum_{k=1}^n \sigma_x^{-2}(X_k)$. □

Proposition 2.3 and its proof show how much better the WLR estimator is than Equation (11), the ILR estimator, and Equation (13), the SMC estimator, in terms of conditional variance. It also shows in Equation (12) that the WLR estimator has a defensive property that the ILR estimator lacks: the conditional variance of the WLR estimator is bounded above by that of the best LR estimator generated from any single sampling distribution. The WLR estimator can only be improved by including samples from another sampling distribution, even if the associated variance is large; if so, the WLR estimator compensates by giving this sampling distribution a small weight. Proposition 2.3 relies on the following lemma, which holds due to convexity of the function that maps a to $1/a$, or equivalently, due to the well-known inequality between arithmetic and harmonic mean.

Lemma 2.2. *For any n and $a_1, \ldots, a_n \geq 0$, $\sum_{k=1}^n a_k^{-1}/n \geq 1/(\sum_{k=1}^n a_k/n)$, with strict inequality if there exist k, k' such that $a_k \neq a_{k'}$.*

Proposition 2.3. *Under Assumptions (A1)–(A6), for any $n, r \in \mathbb{N}_+$ and $x, X_1, \ldots, X_n \in \mathcal{X}$,*

$$\frac{\mathbb{V}ar\left[\widehat{\mu}_{n,r}^{ILR}(x) \middle| X_1, \ldots, X_n\right]}{\mathbb{V}ar\left[\widehat{\mu}_{n,r}^{WLR}(x) \middle| X_1, \ldots, X_n\right]} = \frac{1}{n^2} \left(\sum_{k=1}^n \sigma_x^2(X_k) \right) \left(\sum_{k=1}^n \sigma_x^{-2}(X_k) \right) \geq 1, \tag{11}$$

with strict inequality if $\sigma_x^2(X_k) \neq \sigma_x^2(X_{k'})$ *for some* $k, k' \in \{1, \ldots, n\}$*; also*

$$\frac{1}{nr} \min_{k\in\{1,\ldots,n\}} \sigma_x^2(X_k) \leq \mathbb{V}ar\left[\widehat{\mu}_{n,r}^{WLR}(x)\middle| X_1, \ldots, X_n\right] \leq \frac{1}{r} \min_{k\in\{1,\ldots,n\}} \sigma_x^2(X_k), \tag{12}$$

and

$$\frac{\mathbb{V}ar\left[\widehat{\mu}_r^{SMC}(X_n)\middle| X_1, \ldots, X_n\right]}{\mathbb{V}ar\left[\widehat{\mu}_{n,r}^{WLR}(X_n)\middle| X_1, \ldots, X_n\right]} = \sigma_{X_n}^2(X_n) \sum_{k=1}^{n} \sigma_{X_n}^{-2}(X_k) \geq 1, \tag{13}$$

with strict inequality if $\sigma_{X_n}^2(X_k) < \infty$ *for some* $k < n$.

Proof. By Lemma 2.2, $\sum_{k=1}^n \sigma_x^{-2}(X_k)/n \geq 1/(\sum_{k=1}^n \sigma_x^2(X_k)/n)$. Using this with Equations (7) and (10), the inequality in Equation (11) follows. Equation (12) follows from Equation (10) and taking the reciprocal of

$$\max_{k\in\{1,\ldots,n\}} \sigma_x^{-2}(X_k) \leq \sum_{k=1}^{n} \sigma_x^{-2}(X_k) \leq n \max_{k\in\{1,\ldots,n\}} \sigma_x^{-2}(X_k).$$

Equation (13) follows from Equation (10) and $\mathbb{V}\text{ar}[\widehat{\mu}_r^{SMC}(X_n)|X_1, \ldots, X_n] = \sigma_{X_n}^2(X_n)/r$. □

Proposition 2.3 showcases the theoretical advantages of the WLR estimator using the weights in Equation (8), which involve target-x-sample-X_k variances $\sigma_x^2(X_k)$. Unfortunately, these variances are usually unknown. Therefore, in a practical implementation, one would replace these variances by their estimates $\widehat{\sigma}_x^2(X_k)$. The resulting weights would be suboptimal, and Proposition 2.3 would not apply to this *empirical WLR (EWLR)* estimator. As our numerical experiments show, the EWLR estimator could perform worse than the ILR and SMC estimators. We do not recommend using EWLR unless the variances $\widehat{\sigma}_x^2(X_k)$ can be accurately estimated. In the next section, we propose an estimator that enjoys good theoretical properties, like the WLR estimator, and can be implemented without losing its theoretical properties.

2.3 Mixture Likelihood Ratio Estimator

In our setting of repeated experiments, we have r replications sampled from each of n distributions. The collection of nr observations can be viewed as a stratified sample from an equally-weighted mixture of these n distributions. The likelihood of the mixture is denoted by $\bar{h}(\cdot; X_1, \ldots, X_n)$, where

$$\bar{h}(y; x_1, \ldots, x_n) = \frac{1}{n} \sum_{k=1}^{n} h(y; x_k). \tag{14}$$

Hesterberg (1988) and Veach and Guibas (1995) advocated replacing the equal weights $1/nr$ in Equation (6b) with "balance heuristic" weights that, in our setting, are $h(Y_k^{(j)}; X_k)/nr\bar{h}(Y_k^{(j)}; X_1, \ldots, X_n)$. This leads to the *mixture likelihood ratio (MLR)* estimator

$$\widehat{\mu}_{n,r}^{MLR}(x) = \sum_{k=1}^{n} \sum_{j=1}^{r} \frac{1}{nr} \frac{h\left(Y_k^{(j)}; x\right)}{\bar{h}\left(Y_k^{(j)}; X_1, \ldots, X_n\right)} F\left(Y_k^{(j)}\right). \tag{15}$$

The MLR estimator is the likelihood ratio estimator that arises when we consider the pooled outputs of all simulation experiments performed so far, $\{Y_k^{(j)} : k = 1, \ldots, n, j = 1, \ldots, r\}$, as stratified sampling from the mixture distribution that has likelihood $\bar{h}$ with r independent samples allocated to each of n strata (Hesterberg 1995). It follows from this interpretation, or immediately from the results of Veach and Guibas (1995, Section 3.2), that the MLR estimator is conditionally unbiased for $\mu(x)$ given $X_1, \ldots, X_n$.

Proposition 2.4. *Under Assumptions (A1)–(A6), for any* $n, r \in \mathbb{N}_+$ *and* $x, X_1, \ldots, X_n \in \mathcal{X}$, $\mathbb{E}[\widehat{\mu}_{n,r}^{MLR}(x)|X_1, \ldots, X_n] = \mu(x)$.

The conditional variance of the MLR estimator is better than that of the ILR estimator (Proposition 2.5). The MLR estimator has a defensive property similar to that of the WLR estimator: an upper bound on its conditional variance related to the conditional variance of the best LR estimator generated from any single sampling distribution (Proposition 2.6). The conditional variance of the MLR estimator can be estimated by

$$\widehat{\mathbb{V}\mathrm{ar}}\left[\widehat{\mu}_{n,r}^{MLR}(x)\middle| X_1, \ldots, X_n\right] = \frac{1}{r}\sum_{j=1}^{r}\left(\sum_{k=1}^{n}\frac{h\left(Y_k^{(j)};x\right)F\left(Y_k^{(j)}\right)}{n\bar{h}\left(Y_k^{(j)};X_1,\ldots,X_n\right)} - \widehat{\mu}_{n,r}^{MLR}(x)\right)^2.$$

The following proposition and proof are adopted from Martino et al. (2015, Theorem A.2), expanded to provide sufficient conditions for a strict inequality. They involve the conditional expectation of the kth term in Equation (15),

$$m_k(x; X_1, \ldots, X_n) = \mathbb{E}\left[\frac{F(Y_k)\,h(Y_k;x)}{\bar{h}(Y_k; X_1, \ldots, X_n)}\middle| X_1, \ldots, X_n\right] = \int_{\mathcal{Y}} \frac{F(y)\,h(y;x)}{\bar{h}(y; X_1, \ldots, X_n)} h(y; X_k)\, dy. \tag{16}$$

Proposition 2.5. *Under Assumptions (A1)–(A6), for any* $n, r \in \mathbb{N}_+$ *and* $x, X_1, \ldots, X_n \in \mathcal{X}$,

$$\mathbb{V}ar\left[\widehat{\mu}_{n,r}^{MLR}(x)\middle| X_1, \ldots, X_n\right] \le \mathbb{V}ar\left[\widehat{\mu}_{n,r}^{ILR}(x)\middle| X_1, \ldots, X_n\right]. \tag{17}$$

This inequality is strict if there exist k, k' *such that* $m_k(x; X_1, \ldots, X_n) \neq m_{k'}(x; X_1, \ldots, X_n)$.

Proof. Let $\{\widetilde{Y}_k^{(j)} : k = 1, \ldots, n, j = 1, \ldots, r\}$ be an i.i.d. sample with the likelihood $\bar{h}(\cdot; X_1, \ldots, X_n)$ and define

$$\widehat{\mu}_{n,r}^{Mix}(x) = \sum_{k=1}^{n}\sum_{j=1}^{r}\frac{1}{nr}\frac{h\left(\widetilde{Y}_k^{(j)};x\right)}{\bar{h}\left(\widetilde{Y}_k^{(j)}; X_1, \ldots, X_n\right)} F\left(\widetilde{Y}_k^{(j)}\right).$$

The estimators $\widehat{\mu}_{n,r}^{Mix}$ and $\widehat{\mu}_{n,r}^{MLR}$ are similar: the latter is a stratified-sampling version of the former. Therefore, $\mathbb{E}[\widehat{\mu}_{n,r}^{Mix}(x)|X_1, \ldots, X_n] = \mathbb{E}[\widehat{\mu}_{n,r}^{MLR}(x)|X_1, \ldots, X_n]$, which equals $\mu(x)$ by Proposition 2.4. We have

$$\begin{aligned}&\mathbb{V}\mathrm{ar}\left[\widehat{\mu}_{n,r}^{ILR}(x)\middle| X_1, \ldots, X_n\right] - \mathbb{V}\mathrm{ar}\left[\widehat{\mu}_{n,r}^{Mix}(x)\middle| X_1, \ldots, X_n\right]\\ &\quad= \frac{1}{nr}\int_{\mathcal{Y}} (F(y)h(y;x))^2\left(\frac{1}{n}\sum_{k=1}^{n}\frac{1}{h(y;X_k)} - \frac{1}{\bar{h}(y; X_1, \ldots, X_n)}\right) dy \ge 0,\end{aligned}$$

because, by Lemma 2.2,

$$\frac{1}{n}\sum_{k=1}^{n}\frac{1}{h(y;X_k)} \ge \frac{1}{\frac{1}{n}\sum_{k=1}^{n} h(y;X_k)} = \frac{1}{\bar{h}(y; X_1, \ldots, X_n)}.$$

Next, we observe that $\widehat{\mu}_{n,r}^{MLR}$ is a stratified-sampling version of $\widehat{\mu}_{n,r}^{Mix}$, with equal number of samples allocated to the n equally weighted strata: the sampling likelihood of the kth stratum is $h(\cdot; X_k)$. Therefore, $\mathbb{V}\mathrm{ar}[\widehat{\mu}_{n,r}^{MLR}(x)|X_1, \ldots, X_n] \le \mathbb{V}\mathrm{ar}[\widehat{\mu}_{n,r}^{Mix}(x)|X_1, \ldots, X_n]$. The inequality is strict if there exist strata k, k' with different means $m_k(x; X_1, \ldots, X_n) \neq m_{k'}(x; X_1, \ldots, X_n)$. □

The next proposition follows from Veach (1997, Theorem 9.2) and Owen and Zhou (2000, Equation (8)).

Proposition 2.6. *Under Assumptions (A1)–(A6), for any $n, r \in \mathbb{N}_+$ and $x, X_1, \ldots, X_n \in \mathcal{X}$, if F is a non-negative function or there exists $k \in \{1, \ldots, n\}$ such that $x = X_k$, then*

$$\mathbb{V}ar\left[\widehat{\mu}_{n,r}^{MLR}(x)\middle| X_1, \ldots, X_n\right] \leq \frac{1}{r} \min_{k\in\{1,\ldots,n\}} \sigma_x^2(X_k) + \left(\frac{1}{r} - \frac{1}{nr}\right)(\mu(x))^2. \tag{18}$$

In the context of green simulation Equation (18) provides two kinds of protections against potential adversarial cases for ILR and (E)WLR: First, it suggests that MLR has finite variance if *at least one* of the ILR component's variance is finite, whereas finite ILR variance requires *all* components to have finite variance. Secondly, it also suggests that MLR's variance cannot be much worse than the smallest variance among all variances of the ILR components'. Proposition 2.6 is a worst-case bound on the performance of the MLR estimator. As shown in our numerical experiments, in practical applications, the MLR estimator could have better performance than this worst-case bound.

3 GREEN SIMULATION VIA STOCHASTIC KRIGING

The motivation to propose another kind of green simulation estimator, other than the LR estimators in Section 2, is that the latter may be inapplicable or ineffective in some situations. The LR estimators are intended to be used when Assumptions (A1)–(A6) hold. However, there are many situations in which Assumption (A1) does not hold and the LR method is inapplicable, because the current state affects the simulation model not only by affecting a parameter of a likelihood. It may also be that the LR method is applicable, but the target-X_n-sample-$X_{n'}$ variances $\sigma^2_{X_n}(X_{n'})$ are large, which makes the LR estimators ineffective. Therefore, we propose the *green stochastic kriging (GSK)* estimator based on stochastic kriging (SK), a metamodeling technique for stochastic simulation (Ankenman et al. 2010). The following assumption supports the GSK estimator.

(B1) Given the current state X_n, the SMC estimator $\widehat{\mu}_r^{SMC}(X_n)$ in the current experiment is normally distributed with finite mean $\mu(X_n)$ and finite variance $\sigma^2_{X_n}(X_n)$.

The reason to make Assumption (B1) is that Ankenman et al. (2010) showed that the SK estimator is an MSE-optimal linear predictor if the simulation output is normally distributed.

The GSK estimator $\widehat{\mu}_{n,r}^{GSK}(x)$ is the estimate of $\mu(x)$ generated by SK when it is supplied with the output of the first n experiments of r replications each. Let $\widehat{\mu}_{n,r}^{SMC} = [\widehat{\mu}_r^{SMC}(X_1), \ldots, \widehat{\mu}_r^{SMC}(X_n)]^\top$ represent the vector of SMC estimators for the first n experiments. SK involves fitting a model $f(x)^\top\beta$ to the relationship between $\widehat{\mu}_r^{SMC}(X_k)$ and X_k by choosing the coefficient vector β. In our experiments, we used a linear model, $f(x) = x$; we found that the results were markedly superior to using a constant model, $f(x) = 1$. SK also uses a Gaussian random field (GRF) model, treating the response surface μ as though it were a realization of a GRF with a so-called *extrinsic* covariance function whose parameters are also chosen in a process of fitting. In our experiments, we chose the GRF to be a generalized integrated Brownian field (Salemi et al. 2013); we found that the results were markedly superior to the more common choice of GRF with a Gaussian correlation function. Based on the GRF, it is possible to compute the $1 \times n$ vector $\Sigma(x)$ of extrinsic covariances between $\mu(x)$ and $\mu(X_1), \ldots, \mu(X_n)$, and the $n \times n$ matrix Σ of extrinsic covariances among $\mu(X_1), \ldots, \mu(X_n)$. Finally, SK also uses estimates of the so-called *intrinsic* covariances between simulation outputs. The intrinsic covariance matrix is denoted $C = [C_{ij}]$, where $C_{ij} = \mathrm{Cov}[\widehat{\mu}(X_i), \widehat{\mu}(X_j)]$, where $\widehat{\mu}(X_i)$ and $\widehat{\mu}(X_j)$ are the estimators used to calibrate the SK model. In particular, if the SMC estimators are used and the simulation experiments launched at times $k = 1, \ldots, n$ are independent, then C is an $n \times n$ diagonal matrix whose kth diagonal element is $\frac{1}{r}\sigma^2(X_k)$.

With the above notations, the GSK estimator is given by

$$\widehat{\mu}_{n,r}^{GSK}(x) = f(x)^\top \beta + \Sigma(x)(\Sigma + C)^{-1}\left(\widehat{\mu}_{n,r}^{SMC} - [f(X_1)^\top \beta, \ldots, f(X_n)^\top \beta]^\top\right). \tag{19}$$

In our implementation, we performed the SK fitting process after every experiment, using all available information, including all outputs in previous and the current experiment.

Despite their mathematical similarities, the proposed GSK differs from the classical SK in the calibration of the metamodel: In classical SK metamodeling, a fixed set of simulation outputs is used to calibrate the model then *the same model* is used for possibly multiple predictions. The proposed GSK estimator, however, changes over time as more simulation experiments are done. In particular, it recalibrates after each simulation experiment to incorporate new simulation outputs, in the hope to improve its prediction quality overtime. Numerically, the recalibration could be done efficiently with a *warm-starting* algorithm discussed in Section 4.

By substituting Equation (2) into Equation (19), the GSK estimator can be seen as including a weighted sum of the simulation outputs $\{F(Y_k^{(j)}) : k = 1, \ldots, n; j = 1, \ldots, r\}$. The LR estimators are also weighted sums of the simulation outputs. However, the GSK weights behave quite differently from the LR weights. The LR weights depend on the likelihoods associated with the states $X_1, \ldots, X_n$, and they cannot be negative. Like weights in kriging in general, the GSK weights depend primarily on the locations of the states $X_1, \ldots, X_n$, as related by the extrinsic covariance function, and they can be negative.

4 GREEN ALGORITHMS FOR GREEN SIMULATION ESTIMATORS

This section addresses computationally efficient implementation of the green simulation estimators we proposed.

First, we propose and analyze algorithms for the LR estimators in the setting of repeated experiments. The algorithms are also green, in the sense that they store and reuse likelihood evaluations as well as simulation output. Suppose that one simulation replication has computational cost C_F, one evaluation of a likelihood has computational cost C_h, and the computational cost of basic arithmetic operations such as addition, multiplication, and division is negligible in comparison to these. We envision a situation in which C_F is large, C_h is smaller but need not be negligible, storage space is abundant, and memory access is fast. We consider a sequence of experiments, indexed $n = 1, 2, \ldots,$ of r replications each. For each experiment in the sequence, the SMC, ILR, EWLR, or MLR estimators of the current expected performance $\mu(X_n)$ are computed, and the green simulation procedures store some information to be reused in the next experiment. We analyze the storage requirement and computation cost of the nth experiment.

For benchmarking purposes, we first consider the SMC estimator. It has zero storage requirement in the sense that no information is stored from one experiment to the next. Its computation cost is rC_F.

Algorithm 1 applies to the ILR and EWLR estimators. Consider the ILR estimator $\widehat{\mu}_{n,r}^{ILR}(X_n)$ in Equation (6). The likelihood $h(Y_k^{(j)}; X_k)$ in the denominator does not change as n increases. Therefore, we store and reuse likelihoods from one experiment to the next in Algorithm 1. The storage requirements and non-negligible computation costs for the nth experiment are shown on the right in Algorithm 1. Its storage requirement is $2nr$ and its computation cost is $rC_F + nrC_h$ for the nth experiment. The linear growth rate in n is reassuring; it suggests that it is affordable to reuse the outputs of many experiments in the ILR and EWLR estimators. Compared to the ILR estimator, the WLR estimator requires more basic arithmetic operations to estimate the WLR weights.

For the MLR estimator $\widehat{\mu}_{n,r}^{MLR}(X_n)$, a green algorithm is especially valuable. Inspection of Equations (14) and (15) suggests that the MLR estimator requires $n^2 r$ likelihood evaluations: $h(Y_k^{(j)}; x_\ell)$, for all $j = 1, \ldots, r$ and $k, \ell = 1, \ldots, n$. A quadratic growth rate of computation cost in n could be

ALGORITHM 1: Green Implementation of ILR or EWLR Estimator in the nth Experiment

Initialization: Observe X_n and initialize $\widehat{\mu}^{ILR}_{n,r}(X_n) \leftarrow 0$ or $\widehat{\mu}^{EWLR}_{n,r}(X_n) \leftarrow 0$;
for $j = 1, \ldots, r$ **do**
  Sample $Y_n^{(j)}$ and evaluate $F(Y_n^{(j)})$; /* rC_F computation */
  Append to output storage $F(Y_n^{(j)})$; /* nr storage */
  Calculate likelihood $h(Y_n^{(j)}; X_n)$; /* rC_h computation */
  Append to likelihood storage $h(Y_k^{(n)}; X_n)$; /* nr storage */
end
Set $\widehat{\mu}^{LR}_{r,X_k}(X_n) \leftarrow \frac{1}{r}\sum_{j=1}^{r} F(Y_n^{(j)})$;
for $k = 1, \ldots, n-1$ **do**
  for $j = 1, \ldots, r$ **do**
    Retrieve $F(Y_k^{(j)})$ and $h(Y_k^{(j)}; X_k)$ from storage;
    Calculate likelihood $h(Y_k^{(j)}; X_n)$; /* $(n-1)rC_h$ computation */
  end
  Set $\widehat{\mu}^{LR}_{r,X_k}(X_n) \leftarrow \frac{1}{r}\sum_{j=1}^{r} \frac{h(Y_k^{(j)};X_n)}{h(Y_k^{(j)};X_k)} F(Y_k^{(j)})$
end
if *ILR estimator* **then**
  Set $\widehat{\mu}^{ILR}_{n,r}(X_n) \leftarrow \frac{1}{n}\sum_{k=1}^{n} \widehat{\mu}^{LR}_{r,X_k}(X_n)$ and output ; /* End of algorithm for ILR */
end
if *EWLR estimator* **then**
  Set $\widehat{\sigma}^2_{X_n}(X_k) \leftarrow \frac{1}{r}\sum_{j=1}^{r}\left(\frac{h(Y_k^{(j)};X_n)}{h(Y_k^{(j)};X_k)}F(Y_k^{(j)}) - \widehat{\mu}^{LR}_{r,X_k}(X_n)\right)^2$, for $k = 1, \ldots, n$;
  Set $w_k^{EWLR} \leftarrow \widehat{\sigma}^{-2}_{X_n}(X_k) / \sum_{i=1}^{n} \widehat{\sigma}^{-2}_{X_n}(X_i)$, for $k = 1, \ldots, n$;
  Set $\widehat{\mu}^{EWLR}_{n,r}(X_n) \leftarrow \frac{1}{n}\sum_{k=1}^{n} w_k^{EWLR}\widehat{\mu}^{LR}_{r,X_k}(X_n)$ and output; /* End of algorithm for EWLR */
end

an obstacle for using the MLR estimator when reusing the output of many experiments. By storing and reusing likelihoods from one experiment to the next in Algorithm 2, we avoid this quadratic growth and achieve linear growth of the computation cost in n, as was the case for the ILR estimator. Algorithm 2 has storage requirement $2nr$ and computation cost $rC_F + (2n-1)rC_h$ for the nth experiment. This result for MLR is similar to the result for ILR, but MLR requires almost twice as many likelihood evaluations.

For the GSK estimator, we propose a green algorithm that re-uses parameters fitted after one experiment to warm-start the fitting of parameters after the next experiment. After the nth experiment, SK performs an optimization to find a vector of parameters that make the GRF fit best to the inputs and outputs of the first n experiments. We set the initial value in this optimization equal to the parameter vector chosen after the $(n-1)$st experiment. In our numerical experiments, we found that this warm-starting significantly improved the speed and numerical stability of SK parameter-fitting.

5 CONVERGENCE OF GREEN SIMULATION ESTIMATORS

In this section, we analyze the convergence of the LR estimators as the number n of experiments grows while the number r of replications per experiment is fixed. To this end, we make an assumption on the stochastic process $\{X_n : n = 1, 2, \ldots\}$, which determines the sampling distributions.

ALGORITHM 2: Green Implementation of MLR Estimator in the nth Experiment

Initialization: Observe X_n and initialize $\widehat{\mu}_{n,r}^{MLR}(X_n) \leftarrow 0$;
for $j = 1, \ldots, r$ **do**
  for $k = 1, \ldots, n$ **do**
    if $k < n$ **then**
      Retrieve $F(Y_k^{(j)})$ and $\bar{h}(Y_k^{(j)})$ from storage
    else
      Sample $Y_n^{(j)}$ and evaluate $F(Y_n^{(j)})$; /* rC_F computation */
      Append to output storage $F(Y_n^{(j)})$; /* nr storage */
      Set $\bar{h}(Y_n^{(j)}) \leftarrow 0$;
      for $\ell = 1, \ldots, n-1$ **do**
        Calculate likelihood $h(Y_n^{(j)}; X_\ell)$; /* $(n-1)rC_h$ computation */
        Set $\bar{h}(Y_n^{(j)}) \leftarrow \bar{h}(Y_n^{(j)}) + \frac{1}{n-1} h(Y_n^{(j)}; X_\ell)$
      end
    end
    Calculate likelihood $h(Y_k^{(j)}; X_n)$; /* nrC_h computation */
    Set $\bar{h}(Y_k^{(j)}) \leftarrow \frac{n-1}{n} \bar{h}(Y_k^{(j)}) + \frac{1}{n} h(Y_k^{(j)}; X_n)$;
    Update/Append to likelihood storage $\bar{h}(Y_n^{(j)})$; /* nr storage */
  end
end
Set $\widehat{\mu}_{n,r}^{MLR}(X_n) \leftarrow \frac{1}{nr} \sum_{k=1}^{n} \sum_{j=1}^{r} h(Y_k^{(j)}; X_n) F(Y_k^{(j)}) / \bar{h}(Y_k^{(j)})$ and output;

(C1) The stochastic process $\{X_n : n = 1, 2, \ldots\}$ is ergodic.

We adopt the following definition of ergodicity, which is consistent with the well-known Birkhoff ergodic theorem (Birkhoff 1931). For the class of Markov chains, this definition of ergodicity is standard; for a Markov chain to be ergodic, it is sufficient for it to be positive Harris recurrent and aperiodic (Meyn and Tweedie 2009; Nummelin 2004).

Definition 5.1. A stochastic process $\{X_n : n = 1, 2, \ldots\}$ taking values in a Polish state space $\mathcal{X}$ is *ergodic* if

(i) it has a stationary probability measure π on the Borel σ-algebra of $\mathcal{X}$, and
(ii) for any random variable $f(X)$ that has a finite expectation under π, that is, $f \in L_1(\mathcal{X}, \mathcal{B}(\mathcal{X}), \pi)$,

$$\lim_{n\to\infty} \frac{1}{n} \sum_{k=1}^{n} f(X_k) = \int_{\mathcal{X}} f(x) d\pi(x), \quad \text{a.s. and in } L_1. \tag{20}$$

The reason to make Assumption (C1) is as follows. For any target state $x \in \mathcal{X}$, we envision that it has a neighborhood such that, for every x' in this neighborhood, the sampling distribution associated with x' is a good sampling distribution for the target distribution associated with x. A good sampling distribution would be one for which the target-x-sample-x' variance $\sigma_x^2(x')$ defined in Equation (5) is sufficiently small. An ergodic process returns to this neighborhood infinitely often. The consequence is that, as $n \to \infty$, the number of good samples to be used in estimating $\mu(x)$ also grows without bound.

Theorem 5.2 shows that, under our assumptions, the conditional variance of the ILR estimator evaluated at a fixed target state x, given the state history $X_1, \ldots, X_n$, goes to zero at the rate $O(n^{-1})$. Theorem 5.2 also provides a result about unconditional variance, which may be easier to interpret.

THEOREM 5.2. *Suppose that Assumptions (A1)–(A6) and (C1) hold, and π is the stationary probability measure of $\{X_n : n = 1, 2, \ldots\}$. For any target state $x \in \mathcal{X}$, if the function σ_x^2 defined in Equation (5) is in $L_1(\mathcal{X}, \mathcal{B}(\mathcal{X}), \pi)$, then*

$$\lim_{n\to\infty} nr\mathbb{V}ar\left[\widehat{\mu}_{n,r}^{ILR}(x)\middle| X_1, \ldots, X_n\right] = \int_{\mathcal{X}} \sigma_x^2(x')d\pi(x'). \tag{21}$$

If, furthermore, the target-x-sample-X_n variance process $\{\sigma_x^2(X_n), n = 1, 2 \ldots\}$ is uniformly integrable, then

$$\lim_{n\to\infty} nr\mathbb{V}ar\left[\widehat{\mu}_{n,r}^{ILR}(x)\right] = \int_{\mathcal{X}} \sigma_x^2(x')d\pi(x'). \tag{22}$$

PROOF. By the conditional independence of the random vectors $\{Y_k^{(j)} : j = 1, \ldots, r, k = 1, \ldots, n\}$ given $X_1, \ldots, X_n$, we have

$$\mathbb{V}\text{ar}\left[\widehat{\mu}_{n,r}^{ILR}(x)\middle| X_1, \ldots, X_n\right] = \frac{1}{n^2r^2}\sum_{k=1}^{n}\sum_{j=1}^{r}\mathbb{V}\text{ar}\left[F\left(Y_k^{(j)}\right)\frac{h\left(Y_k^{(j)}; x\right)}{h\left(Y_k^{(j)}; X_k\right)}\right] = \frac{1}{n^2r}\sum_{k=1}^{n}\sigma_x^2\left(X_k\right). \tag{23}$$

Therefore, by ergodicity of $\{X_n : n = 1, 2, \ldots\}$, we obtain Equation (21):

$$\lim_{n\to\infty} nr\mathbb{V}\text{ar}\left[\widehat{\mu}_{n,r}^{ILR}(x)\middle| X_1, \ldots, X_n\right] = \lim_{n\to\infty}\frac{1}{n}\sum_{k=1}^{n}\sigma_x^2\left(X_k\right) = \int_{\mathcal{X}}\sigma_x^2(x')d\pi(x'), \quad \text{a.s. and in } L_1.$$

To establish Equation (22), consider that

$$\begin{aligned}\mathbb{V}\text{ar}\left[\widehat{\mu}_{n,r}^{ILR}(x)\right] &= \mathbb{E}\left[\mathbb{V}\text{ar}\left[\widehat{\mu}_{n,r}^{ILR}(x)\middle| X_1, \ldots, X_n\right]\right] + \mathbb{V}\text{ar}\left[\mathbb{E}\left[\widehat{\mu}_{n,r}^{ILR}(x)\middle| X_1, \ldots, X_n\right]\right] \\ &= \mathbb{E}\left[\mathbb{V}\text{ar}\left[\widehat{\mu}_{n,r}^{ILR}(x)\middle| X_1, \ldots, X_n\right]\right] + \mathbb{V}\text{ar}\left[\mu(x)\right] \\ &= \mathbb{E}\left[\frac{1}{n^2r}\sum_{k=1}^{n}\sigma_x^2\left(X_k\right)\right],\end{aligned}$$

using Equation (23). Therefore,

$$\lim_{nr\to\infty} nr\mathbb{V}\text{ar}\left[\widehat{\mu}_{n,r}^{ILR}(x)\right] = \lim_{n\to\infty}\mathbb{E}\left[\frac{1}{n}\sum_{k=1}^{n}\sigma_x^2\left(X_k\right)\right] = \mathbb{E}\left[\lim_{n\to\infty}\frac{1}{n}\sum_{k=1}^{n}\sigma_x^2\left(X_k\right)\right] = \int_{\mathcal{X}}\sigma_x^2(x')d\pi(x'),$$

where the exchange of limit and expectation holds by uniform integrability, and ergodicity of $\{X_n : n = 1, 2, \ldots\}$ justifies the last step. □

COROLLARY 5.3. *If the corresponding conditions stated in Theorem 5.2 hold, then*

$$\lim_{n\to\infty} nr\mathbb{V}ar\left[\widehat{\mu}_{n,r}^{*}(x)\middle| X_1, \ldots, X_n\right] \le \int_{\mathcal{X}}\sigma_x^2(x')d\pi(x'), \text{ and} \tag{24}$$

$$\lim_{n\to\infty} nr\mathbb{V}ar\left[\widehat{\mu}_{n,r}^{*}(x)\right] \le \int_{\mathcal{X}}\sigma_x^2(x')d\pi(x'), \tag{25}$$

where the asterisk may represent WLR or MLR.

PROOF. Due to Propositions 2.3 and 2.5, under the appropriate conditions stated in Theorem 5.2, the right sides of Equations (21) and (22) serve as upper bounds for $nr\mathbb{V}\text{ar}[\widehat{\mu}_{n,r}^{*}(x)|X_1, \ldots, X_n]$ and $nr\mathbb{V}\text{ar}[\widehat{\mu}_{n,r}^{*}(x)]$, respectively. □

By Theorem 5.2 and Corollary 5.3, the conditional variance of any of the three LR estimators, evaluated at a fixed target state x, given the state history $X_1, \ldots, X_n$, goes to zero at the rate $O(n^{-1})$. Because all three LR estimators are unbiased, it follows that they are consistent as $n \to \infty$.

Theorem 5.2 and Corollary 5.3 show the asymptotic superiority of the LR estimators to standard Monte Carlo (SMC), as the number n of repeated experiments increases. Recall, from the discussion of Equation (3), that the conditional variance of the SMC estimator for the current expected performance $\mu(X_n)$, given the state history, is $\sigma^2(X_n)/r$. This does not converge to zero as $n \to \infty$, assuming that sampling variances are positive. Yet the LR estimators converge to the true value $\mu(x)$ as $n \to \infty$. This means that we can obtain arbitarily high accuracy without increasing the budget r per experiment, merely by reusing output from experiments that are repeated at each time step with budget r. Under the assumptions of Theorem 5.2, reusing old simulation output is highly effective in the sense that the variance of the LR estimators converges as $O((nr)^{-1})$, which is the standard rate of convergence for Monte Carlo in terms of the computational budget nr expended on all experiments that were ever run.

6 NUMERICAL EXAMPLES

In this section, we use two numerical examples to illustrate green simulation, demonstrate its value, and compare our four green simulation estimators with each other and with standard Monte Carlo. First is a reinsurance example of pricing catastrophe bonds. For this example, we verify the conditions of Theorem 5.2 in Appendix A, which shows that these conditions are applicable to a realistic example. The experiment results conform to the theoretical predictions in that the MLR estimator was superior to the ILR estimator, which was superior to the SMC estimator. The EWLR estimator performed similarly to MLR when the number of experiments was small, but it was not even as good as the ILR estimator when many experiments were run. The GSK estimator succeeded in reducing mean squared error (MSE) as more experiments were run, but its MSE was more than those of the LR estimators. The second example involves measuring the credit risk of a loan portfolio. In this example, the conditions of Theorem 5.2 do not hold, but the experiment results show that green simulation can still deliver valuable results in such a situation. Although the ILR, EWLR, and GSK estimators were not successful in this example, the MLR estimator was superior to the SMC estimator.

6.1 Catastrophe Bond Pricing with Compound Losses

A catastrophe bond ("CAT bond") is an important reinsurance contract that helps insurance companies to hedge against losses from catastrophic events (Munich Re Geo Risks Research 2015). This example is relevant beyond insurance; for example, senior tranches of structured financial instruments are essentially economic catastrophe bonds, because they suffer credit losses only in the event of an economic catastrophe (Coval et al. 2009). Simulation of CAT bonds can be computationally intensive, because it involves fairly rare events in a complex geophysical models. Specifically, the example illustrates a simple simulation for pricing hurricane CAT bonds.

In practice, reinsurance contracts are subject to periodic renewals. This is the source of the repeated experiments: the same CAT bond is priced every period, using the same hurricane simulation model, with parameters X_n updated to reflect the current climatological forecast. In this example, the period is semi-annual. The state process $\{X_n : n = 1, 2, \ldots\}$ is modeled by an ergodic Markov chain. Given the state X_n, we can simulate hurricanes that take place during the lifetime of the CAT bond and the resulting total insured loss Y_n underlying the CAT bond. Finally, we compute the payoff of the CAT bond per dollar invested,

$$F(Y_n) = \mathbf{1}_{\{Y_n \le K\}} + p\mathbf{1}_{\{Y_n > K\}},$$

where $\mathbf{1}_{\{\cdot\}}$ is the indicator function, K is the trigger level, and $p \in [0, 1)$ is the fraction of face value that is received if insured losses exceed the trigger level. The fair price of this CAT bond can be obtained in terms of the expected payoff $\mu(X_n) = \mathbb{E}\left[F\left(Y_n\right)|X_n\right]$. We consider a hurricane CAT bond with lifetime 10 years, trigger level $K = 25$ million dollars, and recovery fraction $p = 0.5$.

In this example, we use a simplified version of the model of Dassios and Jang (2003). The insured loss is modeled as a compound random variable: $Y_n = \sum_{i=1}^{M_n} Z_n^i$, where M_n denotes the number of claims and Z_n^i denotes the ith claim size. In this model, $Z_n^{(i)}$, $i = 1, 2 \ldots$ are i.i.d. and independent of M_n. This is a popular loss model due to its flexibility and suitability for many practical applications (Klugman et al. 2012), yet it can provide mathematical tractability. Let the probability mass function of M_n be $p(m; \lambda_n)$ and the probability density of Z_n^i be $f(z; \theta_n)$, where λ_n and θ_n are parameters determined by the state X_n. Specifically, we take M_n to be Poisson with mean λ_n and Z_n^i to be exponential with mean θ_n. In this model, the expected payoff is

$$\mu(X_n) = \mathbb{E}\left[F\left(Y_n\right)|X_n\right] = \mathbb{E}\left[p + (1-p)\mathbf{1}_{\{Y_n \le K\}}\right] = p + \sum_{m=1}^{\infty} p(m; \lambda_n) F(K; \theta_n, m), \tag{26}$$

where $F(K; \theta, m)$ is the Gamma cumulative distribution function with scale parameter θ and shape parameter m. From 1981 to 2010, the average number of major hurricanes was 2.7 per decade and the average cost per hurricane was about \$5,000 million (Blake and Gibney 2011). Therefore, we set up a stochastic model of the states $\{X_n : n = 1, 2, \ldots\}$ and a transformation $(\lambda, \theta) = \psi(x)$, so λ_n is usually around 2.7 and θ_n is usually around 5 (measured in thousands of millions of dollars).

In this example, the ergodic Markov chain driving the parameters of the loss model is a stationary AR(1) process with state space $\mathcal{X} = \mathbb{R}^2$, given by

$$X_n = \mu_\infty + \varphi X_{n-1} + \varepsilon_n,$$

where $\{\varepsilon_n : n = 1, 2, \ldots\}$ is an i.i.d. sequence of bivariate normal random vectors with mean zero and variance $\mathrm{diag}(\sigma_\varepsilon^2)$, and the parameters are

$$\mu_\infty = \begin{bmatrix} 0 \\ 0 \end{bmatrix}, \quad \varphi = \begin{bmatrix} 0.6 \\ 0.5 \end{bmatrix}, \quad \text{and} \quad \sigma_\varepsilon^2 = \begin{bmatrix} 0.8^2 \\ 0.5^2 \end{bmatrix}.$$

The state space $\mathcal{X} = \mathbb{R}^2$ is inappropriate for parameters that must be non-negative, because they represent an expected number of hurricanes and an expected loss per hurricane. We introduce a transformation $\psi : \mathbb{R}^2 \mapsto (\underline{\lambda}, \bar{\lambda}) \times (\underline{\theta}, \bar{\theta})$ so the parameters $(\lambda, \theta) = \psi(x)$ lie between plausible lower and upper bounds. The transformation is sigmoidal and maps $x = [x^1, x^2]$ to

$$(\lambda, \theta) = \psi(x) = \left[\underline{\lambda} + \frac{\bar{\lambda} - \underline{\lambda}}{1 + e^{-x^1}}, \quad \underline{\theta} + \frac{\bar{\theta} - \underline{\theta}}{1 + e^{-x^2}}\right].$$

In particular, we took $(\underline{\lambda}, \bar{\lambda}) = (2, 4)$ as the range for expected number of hurricanes per decade and $(\underline{\theta}, \bar{\theta}) = (4, 6)$ as the range for expected loss per hurricane. To give a picture of the variability of the parameters for repeated experiments in this example, Figure 2 shows histograms of the parameters λ and θ resulting from sampling from the stationary distribution of the AR(1) process.

To clarify the model, Algorithm 3 shows how the standard Monte Carlo simulation works. The density function is an essential element for implementing the three LR methods in Algorithms 1 and 2. The density function for this example, $h(y, x)$, where y is the compound loss and $x = (\lambda, \theta)$, is shown in Equation (27) in the Appendix.

To investigate the effectiveness of green simulation, we performed a sequence of 100 repeated simulation experiments (i.e., $n = 1, 2, \ldots, 100$) with $r = 100$ replications each. Using the same simulation output $\{Y_n^{(j)} : n = 1, 2, \ldots, 100, j = 1, 2, \ldots, 100\}$ from the same random sample paths $\{X_n : n = 1, 2, \ldots, 100\}$, we evaluated the SMC, ILR, EWLR, and MLR estimators at each period

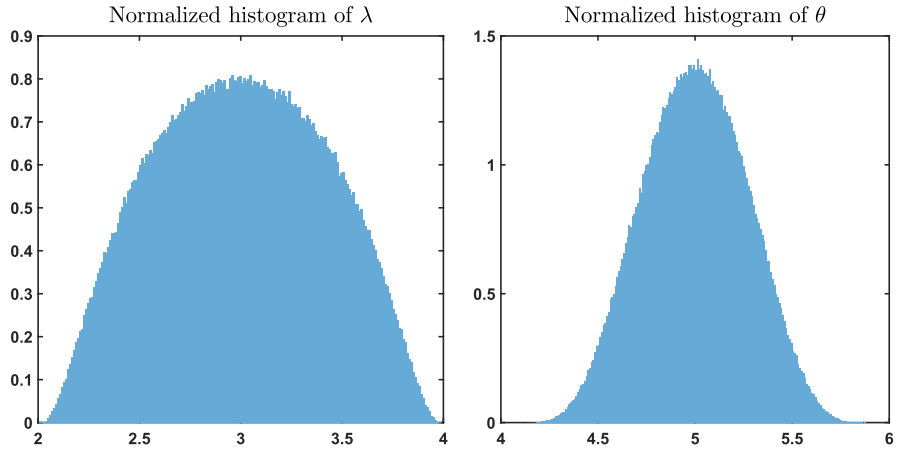

Fig. 2. Histograms of parameters sampled based on the stationary distribution of the AR(1) process.

ALGORITHM 3: Standard Monte Carlo Simulation for Catastrophe Bond Pricing

Initialization: Generate initial state X_0 from the stationary distribution of the AR(1) process
for $n = 1, 2, \ldots,$ **do**
  Generate state X_n from AR(1) process, conditional on X_{n-1};
  Set parameters $(\lambda_n, \theta_n) \leftarrow \psi(x_n)$;
  Set $\widehat{\mu}_{n,r}^{SMC}(X_n) \leftarrow 0$;
  for $j = 1, \ldots, r$ **do**
    Simulate number of hurricanes $M_n^{(j)} \sim p(\cdot; \lambda_n)$;
    for $i = 1, \ldots, M_n^{(j)}$ **do**
      Simulate loss of the ith hurricane $Z_n^{(i,j)} \sim f(\cdot; \theta_n)$;
    end
    Set $Y_n^{(j)} \leftarrow \sum_{i=1}^{M_n} Z_n^{i,j}$;
    if $Y_n^{(j)} \leq K$ **then**
      Set $F(Y_n^{(j)}) \leftarrow 1$;
    else
      Set $F(Y_n^{(j)}) \leftarrow p$;
    end
    Set $\widehat{\mu}_{x_n}^{SMC}(X_n) \leftarrow \widehat{\mu}_{x_n}^{SMC}(X_n) + F(Y_n^{(j)})$;
  end
  Set $\widehat{\mu}_{n,r}^{SMC}(X_n) \leftarrow \widehat{\mu}_{n,r}^{SMC}(X_n)/r$ and output;
end

$n = 1, 2, \ldots, 100$, in each of three states: the current state X_n, the central state $x_{mi} = (0, 0)$ corresponding to the moderate parameters $\lambda = 3$ and $\theta = 5$, and an extreme state $x_{hi} = (\infty, \infty)$ corresponding to extreme parameters $\lambda = 4$ and $\theta = 6$[1]. Starting from $n = 5$, we also evaluated the GSK estimator in these three states.

[1]Strictly speaking, the extreme state x_{hi} does not belong to the state space $\mathcal{X} = \mathbb{R}^2$, but the parameter vector $(\lambda, \theta) = (4, 6)$ is a limit point of the range of the transformation ψ that maps states to parameters.

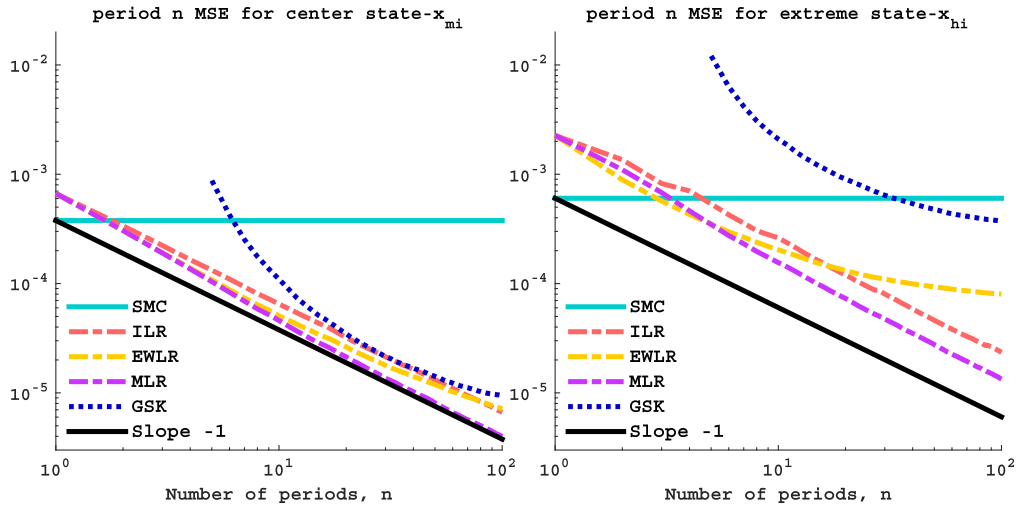

Fig. 3. Log-log plots of estimated MSEs of fixed-state estimators for CAT bond pricing example.

For the purpose of accurately estimating the unconditional MSEs of the all these estimators, we performed such a sequence of experiments 10,000 times. These 10,000 macro-replications of the sequence of experiments have independent sample paths and simulation output. The estimated MSE of a fixed-state estimator $\widehat{\mu}(x)$ of $\mu(x)$ was $\sum_{k=1}^{10,000}(\widehat{\mu}^{(k)}(x) - \mu(x))^2/10{,}000$, where $\widehat{\mu}^{(k)}(x)$ is the value of the estimator on the kth macro-replication. Likewise, the estimated MSE of a current-state estimator $\widehat{\mu}(X_n)$ of $\mu(X_n)$ was $\sum_{k=1}^{10,000}(\widehat{\mu}^{(k)}(X_n^{(k)}) - \mu(X_n^{(k)}))^2/10{,}000$. Due to using 10,000 macro-replications, the standard errors of these estimated MSEs are less than 1% of the corresponding estimated MSE. Recall that MSE equals variance for the SMC and LR estimators, which are unbiased.

Figure 3 is a log-log plot of the estimators' MSEs for two fixed states, x_{mi} and x_{hi}, for each experiment $n = 1, 2, \ldots, 100$. The fixed-state SMC estimators $\widehat{\mu}_{n,r}^{SMC}(x_{mi})$ and $\widehat{\mu}_{n,r}^{SMC}(x_{hi})$ were generated by sampling according to $h(\cdot; x_{mi})$ and $h(\cdot; x_{mi})$, respectively, in experiments distinct from the experiments described in the preceding paragraph. The SMC variance forms a horizontal line, because, for each experiment n, $\widehat{\mu}_{n,r}^{SMC}(x_{mi})$ and $\widehat{\mu}_{n,r}^{SMC}(x_{hi})$ use a fixed number r of replications drawn from the sampling distribution associated with the fixed state x_{mi} or x_{hi}. In addition, a black solid line with slope -1 and intercept equal to the SMC variance is plotted for reference. This line shows the variance of an SMC estimator with nr replications, which is the cumulative number of replications simulated up through the nth experiment. We compare the lines for green simulation estimators against the solid lines. When they go below the horizontal line, they have lower variance than a SMC simulation with r replications, the budget for a single experiment. When they go near the line with slope -1, they have variance nearly as low as a SMC simulation with nr replications, the cumulative budget for all experiments so far. That is a major success for green simulation, because it shows that reusing old simulation output is nearly as effective as generating new simulation output in the current experiment.

We see from Figure 3 that the MLR estimator has lower variance than the ILR estimator, which is consistent with Equation (17). In this example, the gap between them grows to be substantial as the number n of repeated experiments increases: for $n = 100$, the ratio between ILR and MLR variances is about 1.7 for both x_{mi} and x_{hi}. Initially, for very small n, the LR estimators have higher variances than an SMC estimator based on a simulation in the fixed state x_{mi} or x_{hi}. The cause is the probability that none of the states visited so far, $X_1, X_2, \ldots, X_n$, were near x_{mi} or x_{hi}. This

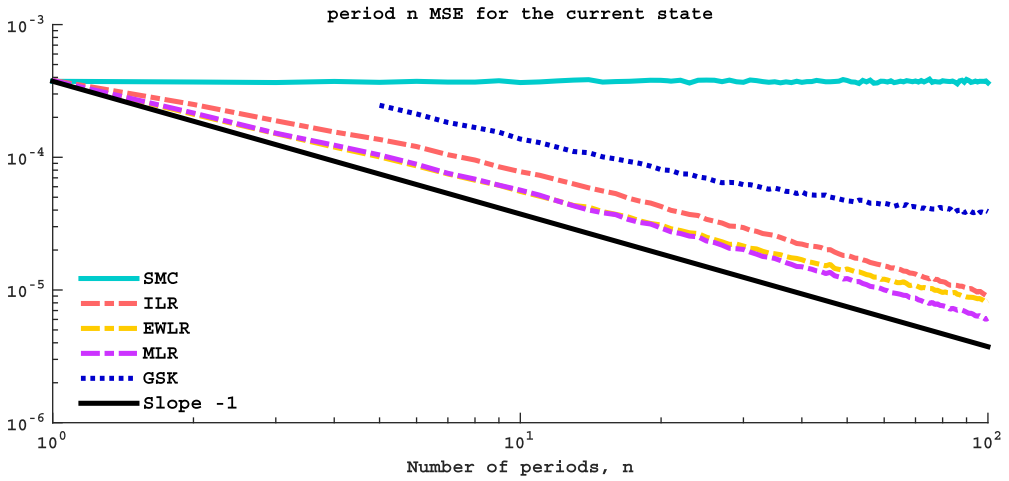

Fig. 4. Log-log plot of estimated MSEs for current-state estimators for CAT bond pricing example.

event is more likely for the extreme state x_{hi} than for the center state x_{mi}, which is why the higher variances persist longer for the extreme state (for the first five experiments) than for the center state (only for the first experiment). The variances of the green simulation soon become lower than those of SMC and continue to decrease as the outputs of more experiments are reused. After 100 experiments, the MLR estimator's variance is over 45 times smaller than the SMC estimator's variance for x_{hi} and over 95 times smaller for x_{mi}. In other words, by using 10,000 total replications simulated in 100 simulation experiments, with simulation based on parameters corresponding to $X_1, X_2, \ldots, X_{100}$ and not x_{mi}, the MLR estimator achieves higher accuracy in estimating $\mu(x_{mi})$ than standard Monte Carlo with 9,500 replications simulated based on parameters corresponding to x_{mi}. Comparing to the black solid reference line, we see that the green simulation estimators' variances eventually decrease approximately at a rate of n^{-1}, as discussed in connection with Theorem 5.2. For small n, the EWLR estimator has variance comparable to that of the MLR estimator, but it is not competitive for large n. In the extreme state, the EWLR variance grows much larger than the ILR variance for large n, and it does not appear to converge as $O(n^{-1})$. The MSE of the GSK estimator is larger than that of the LR estimators. However, its MSE does decrease as n increases. For the center state, after 10 experiments the GSK MSE is comparable to the ILR variance. Because x_{hi} is always outside the range of $X_1, \ldots, X_n$, SK must extrapolate when it estimates $\mu(x_{hi})$. Extrapolation is challenging for SK, so it is a success for GSK that, for the extreme state, its MSE becomes lower than the SMC variance after about 40 experiments, and continues to decrease.

Next, we consider the variance of the SMC and the green simulation estimators for the current-state expected performance $\mu(X_n)$. Figure 4 shows their variances. For the first experiment ($n = 1$), there is no stored simulation output from a previous experiment, so the LR estimators are the same as the SMC estimator. In this example, the green simulation estimators' MSEs decrease from the beginning and are always less than the SMC variance. Otherwise, the performance of the green simulation estimators is similar to what was seen when the state was fixed. After 100 experiments, the MLR estimator's variance is over 61 times smaller than the SMC estimator's variance.

6.2 Credit Risk Management

In simulation for financial risk management, experiments using the same simulation model are performed periodically, as often as daily. Information observed in the markets is used to update

parameters that affect risk, and the simulation model is run again with new parameters. In this example, the risk management simulation measures the credit risk exposure of a portfolio containing loans to companies with listed equity. The asset values of these debtor companies are observable and serve as parameters in the risk model: the lower the asset value of a debtor company, the more likely it is to default in the future. Thus, in our setting of repeated experiments, the current state X_n contains the current asset values of all debtor companies.

In this example, we work with a structural model of default based on the influential work of Merton (1974); for an exposition, see McNeil et al. (2005), for example. At any period n, the asset value of a company equals the sum of its equity and debt values. The equity value can be observed in the stock market and the debt value can be observed from public records, so the asset value can be computed. For simplicity of exposition, we assume that debt remains constant. In this model, the asset value follows geometric Brownian motion, and a company defaults when its asset value falls below its debt. Because geometric Brownian motion is not an ergodic process, Theorem 5.2 and Corollary 5.3 do not apply. As n increases, the current state X_n tends to drift further away from an earlier state, such as X_1. Therefore, intuition suggests that the benefit of reusing old simulation output would diminish over time. We consider this example to show that green simulation nonetheless delivers valuable results.

We consider a loan portfolio whose composition remains constant over time. Many loan portfolios are dynamic: as outstanding loans are being repaid, new loans are being initiated. Despite the dynamic nature of such portfolios, there are lending businesses in which portfolios retain a similar composition over time. For example, in the business of accounts receivable, customers may place regular, periodic orders of the same size, each resulting in payment due in 90 days. An investment fund may target a loan portfolio with fixed characteristics such as maturity and portfolio weights on different types of loans.

In our simplified example, we consider risk management of the value at time horizon $t = 0.5$ years of a portfolio in which there are two loans, both having maturity $T = 5$ years. Simulation experiments are repeated weekly, that is, with a period of $\Delta t = 1/52$ years. In the nth experiment, the quantity $\mu(X_n)$ being estimated is the conditional probability, given the current asset values X_n, that the cost of defaults and anticipated defaults after t years will exceed a threshold $\kappa = 6$. The random vector Y_n that we simulate in the nth experiment is the asset values $S_t = [S_{t,1}, S_{t,2}]$ at time t, given the current asset values $S_0 = [S_{0,1}, S_{0,2}] = X_n$. For each company $i = 1, 2$, the marginal distribution of the asset return $S_{t,i}/S_{0,i}$ is lognormal, determined by the drift η_i and volatility ς_i of the geometric Brownian motion for asset value. Specifically, $\eta_1 = 15\%$, $\eta_2 = 10\%$, $\varsigma_1 = 30\%$, and $\varsigma_2 = 20\%$. The joint distribution of the asset returns $S_{t,1}/S_{0,1}$ and $S_{t,2}/S_{0,2}$ is specified by a Student t copula with 3 degrees of freedom and correlation 0.5. The initial asset values of the two companies are $X_{0,1} = 100$ and $X_{0,2} = 90$, and their debt is $D_1 = D_2 = 85$. The loss if company i defaults is denoted a_i, and $a_1 = 5$ and $a_2 = 4$. The discount rate is $r = 5\%$. The cost of a default or anticipated default by company i, as of time t, is

$$L_i = \ell_i(S_{t,i}) = \begin{cases} a_i & \text{if } S_{t,i} < D_i \\ a_i e^{-r(T-t)} \Phi\left(\frac{\ln(S_{t,i}/D_i) + (r - \varsigma_i^2/2)(T-t)}{\varsigma_i \sqrt{T-t}} \right) & \text{if } S_{t,i} \geq D_i \end{cases}.$$

The first line of the formula represents the loss if company i defaults at time t. The second line of the formula represents a risk-neutral conditional expectation, as of time t, of the discounted loss at time T if company i defaults then. The portfolio's loss is $L_1 + L_2$, the sum of losses over the debtor companies. The simulation output $F(Y_n)$ is 1 if $L_1 + L_2 > \kappa$, and 0 otherwise. To clarify the model, Algorithm 4 shows how the standard Monte Carlo simulation works.

ALGORITHM 4: Standard Monte Carlo Simulation for Credit Risk Management

for $n = 1, 2, \ldots,$ **do**
  Sample state X_n from bivariate lognormal distribution with Student-t copula, based on time increment Δt, conditional on X_{n-1};
  Set $\widehat{\mu}_{n,r}^{SMC}(X_n) \leftarrow 0$;
  for $j = 1, \ldots, r$ **do**
    Sample state $Y_n^{(j)}$ from bivariate lognormal distribution with Student t copula, based on time increment t, conditional on X_n;
    Set $L_n^{(j)} \leftarrow \ell_1(Y_{n,1}^{(j)}) + \ell_2(Y_{n,2}^{(j)})$;
    if $L_n^{(j)} \leq \kappa$ **then**
      Set $F(Y_n^{(j)}) \leftarrow 1$;
    else
      Set $F(Y_n^{(j)}) \leftarrow 0$
    end
    Set $\widehat{\mu}_{x_n}^{SMC}(X_n) \leftarrow \widehat{\mu}_{x_n}^{SMC}(X_n) + F(Y_n^{(j)})$;
  end
  Set $\widehat{\mu}_{n,r}^{SMC}(X_n) \leftarrow \widehat{\mu}_{n,r}^{SMC}(X_n)/r$ and output;
end

To investigate the effectiveness of green simulation, we performed a sequence of 52 simulation experiments repeated weekly (i.e., $n = 1, 2, \ldots, 52$) with $r = 1{,}000$ replications each. Using the same simulation output $\{Y_n^{(j)} : n = 1, 2, \ldots, 52, j = 1, 2, \ldots, 100\}$ from the same random sample path $\{X_n : n = 1, 2, \ldots, 52\}$, we evaluated the SMC and green simulation estimators at each period $n = 1, 2, \ldots, 100$, in the current state X_n. For the purpose of accurately estimating the unconditional variances of the all these estimators, we performed such a sequence of experiments 10,000 times. These 10,000 macro-replications of the sequence of experiments have independent sample paths and simulation output. The estimated variance of a current-state estimator $\widehat{\mu}(X_n)$ of $\mu(X_n)$ was $\sum_{k=1}^{10,000}(\widehat{\mu}^{(k)}(X_n^{(k)}) - \mu(X_n^{(k)}))^2/10{,}000$. These estimated variances appear in a log-log plot in Figure 5, along with vertical error bars representing their 95% approximate-normal confidence intervals.

Figure 5 shows behavior for the MLR estimator similar to what was seen for the previous example in Figure 4. The MLR estimator's variance is less than the SMC estimator's variance, and it decreases as the number n of repeated experiments increases. By $n = 52$, it is over 17 times smaller than the SMC estimator's variance. However, in Figure 5, we see effects of the non-ergodic nature of the state process $\{X_n : n = 1, 2, \ldots\}$. Due to the positive drifts of the asset prices, debtor companies tend to become less likely to default, so the SMC variance decreases slightly as the number of periods n increases, instead of forming a straight line as in Figure 4. The dramatic effect is on the behavior of the ILR and EWLR estimators. Their variances decrease over the first 6 experiments, but after about 20 experiments, their variances increase again. Eventually, their variances exceed the SMC estimator's variance. Apparently, the difference between sampling distributions for state X_1 and X_{52} is likely to become so large that the use of likelihood ratios in the ILR estimator Equation (6) inflates variance. In such a situation, Proposition 2.3 states that the WLR estimator is no worse than the SMC estimator. However, Proposition 2.3 does not apply to the EWLR estimator, whose weights involve variance estimates. When a large variance is underestimated, the associated weight is too large, leading to inflated variance. The failure of the ILR and EWLR estimators

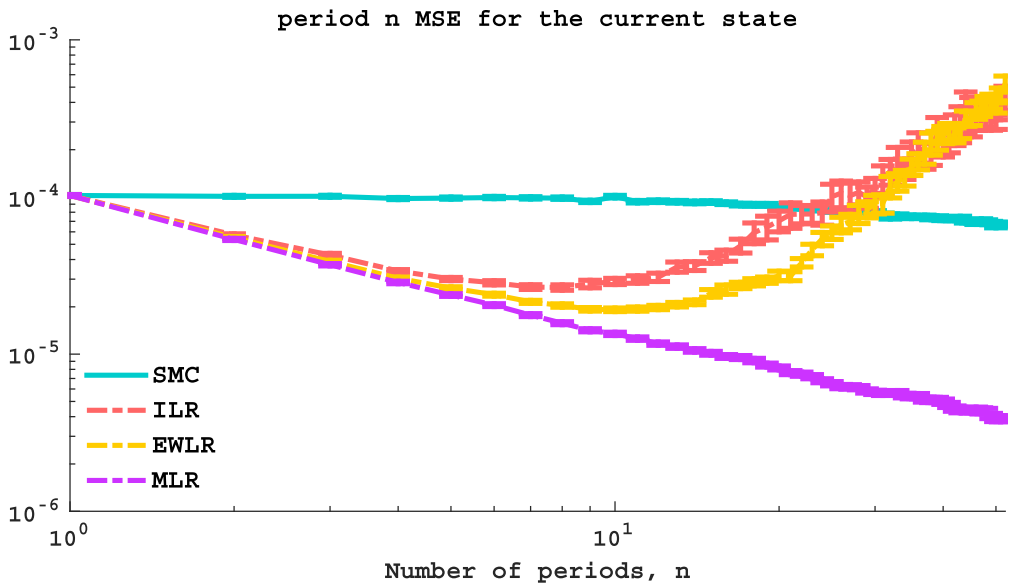

Fig. 5. Log-log plot with error bars for estimated MSEs for current state estimators for credit risk example.

and the success of the MLR estimator in this example demonstrate the practical importance of the MLR estimator. The evident disutility of some of the old simulation output in this example also raises a future research question in green simulation: how to determine which old simulation output is worthwhile to reuse in estimating the expected performance in the current state.

The GSK estimator performed poorly in this example, so we omitted it from Figure 5. The reason for the poor performance of GSK in this example is as follows. In this example, there is a strong correlation between the conditional expectation $\mu(X_n)$ and conditional variance $\sigma^2(X_n)$ of the simulation output. Furthermore, $\sigma^2(X_n)$ can be very small and difficult to estimate, so the estimated conditional variance $\widehat{\sigma}^2(X_n)$ can even be zero. In SK, simulation output with lower estimated variance receives higher weight; if lower estimated variance is correlated with lower estimated mean, bias can result (Staum 2009). This SK bias is large in this example, causing the poor performance of GSK.

7 CONCLUSIONS AND FUTURE RESEARCH

In Section 5, we established theorems about the convergence of green simulation estimators as the number of repeated experiments increases. We tested their practical performance for small and moderate numbers of repeated experiments in two examples in Section 6. In the example of Section 6.1, the conditions of Theorem 5.2 held. Our ILR, EWLR, MLR, and GSK estimators were successful in significantly reducing variance compared to standard Monte Carlo, but MLR was best. In the example of Section 6.2, the conditions of Theorem 5.2 did not hold. The MLR estimator was successful in significantly reducing variance, whereas the ILR, EWLR, and GSK estimators had problems. The GSK estimator is more widely applicable than the LR estimators. However, in the examples we investigated, it was not competitive with the LR estimators. Among the LR estimators, we recommend the MLR estimator for doing green simulation in the setting where simulation experiments are repeated with changes to parameters of distributions of random variables generated in the simulation. The variance reduction achieved by the LR estimators depends on several aspects of the particular example: the stochastic process that describes changing parameters, the particular distributions whose parameters change, and the number of repeated experiments. Under

the conditions of Theorem 5.2, as the number of repeated experiments increases, the LR estimators eventually become greatly superior to standard Monte Carlo. Our experiment results suggest that green simulation is extremely promising: in the only two examples that we investigated, the MLR estimator achieved variance lower than standard Monte Carlo by factors of 17 and 61 after a moderate number of repeated experiments.

Because green simulation is a new paradigm, there are several good directions for future research. Here, we call attention to a few that are most relevant to this article.

Some future research topics are relevant to the specific methods proposed in this article. We found the MLR estimator to be satisfactory for our purposes. However, further enhancements have been considered in the literature on importance sampling. For example, Hesterberg (1995) investigated different ways to normalize weights, and Owen and Zhou (2000) proposed to use likelihoods that appear in the MLR estimator as control variates. At the end of Section 6.2, the experiment results raised the question of which old simulation output to reuse in green simulation. This question is worthy of investigation in connection with the methods proposed in this article and also with other methods. In general, there are two possible drawbacks to using all of the old simulation output. One is that if the amount of old simulation output is extremely large, reusing more of it generates diminishing returns in terms of improved estimator quality compared to the computational cost of reusing it. To limit the computational cost of reuse, one could employ hard thresholding: reusing only that subset of the old simulation output that is deemed to be sufficiently relevant for estimating the expected performance for the target state. The selection criterion could be, for example, the similarity of the sampling density $h(\cdot; X_k)$ to the target density $h(\cdot; x)$ for LR estimators, or the distance from the previous state X_k to the target state x. The other drawback to using all of the old simulation output is that some of the old simulation output makes an estimator worse if it is reused than if it is not reused, as seen in Section 6.2. To cope with this problem, one may employ hard thresholding or soft thresholding, meaning assigning weight to simulation output in proportion to its perceived relevance. For example, the WLR estimator can be viewed as a soft thresholding method, which gives smaller weights to outputs that are associated with high variances. The theoretical and practical benefit of various green simulation methods could be enhanced by good rules for selecting the subset of old simulation output to reuse.

We focused on showing that when old simulation output is reused well, it provides greater accuracy when combined with a new simulation experiment than would be achieved by the same new simulation alone. We analyzed how the accuracy of an answer to the current question improves as the number of repeated experiments increases. However, it might be possible to answer the current question sufficiently accurately with no further experiment. If a new simulation experiment was indeed required, then one could design it in light of the current question and the currently available information. This leads to future research in experiment design not from a blank slate. Also, one might consider possible future questions when designing the current experiment, in light of knowledge of the state process.

APPENDIX

A VERIFYING THE CONDITIONS OF THEOREM 5.2 FOR THE CATASTROPHE BOND PRICING EXAMPLE

In this appendix, we verify the conditions of Theorem 5.2 and for the catastrophe bond (CAT bond) pricing example. In this example, the underlying state process $\{X_n : n = 1, 2, \ldots\}$ is AR(1), so it is ergodic. Recall that the loss in the CAT bond pricing example is given by $Y_n = \sum_{i=1}^{M_n} Z_n^i$. Conditional on $X_n = x$, and denoting $\phi(x) = (\lambda, \theta)$, the number M_n of claims is Poisson distributed with mean λ and independent of $\{Z_n^i : i = 1, \ldots, M_n\}$, which are independent random variables,

exponentially distributed with mean θ. The conditional distribution of Y_n given X_n places probability mass $h(0;x) = e^{-\lambda}$ on $y = 0$ and has probability density

$$h(y;x) = \sum_{m=1}^{\infty} \frac{\lambda^m}{m!} e^{-\lambda} \frac{y^{m-1} e^{-y/\theta}}{\Gamma(m)\theta^m} = \sqrt{\frac{\lambda}{\theta y}} e^{-\lambda - y/\theta} I_1\left(2\sqrt{\frac{\lambda y}{\theta}}\right) \tag{27}$$

for $y > 0$, where Γ is the Gamma function and I_1 is the modified Bessel function of the first kind of order 1.

We will first establish two lemmas that are useful for verifying the conditions of Theorem 5.2 for the CAT bond pricing example. Define the domain $\mathcal{K} = (0,\infty) \times (0,\infty) \times \mathbb{R}$ and the functions $A : \mathcal{K} \mapsto \mathbb{R}$ and $a : \mathcal{K} \times (0,\infty) \mapsto \mathbb{R}$ by

$$A(k) = \int_0^\infty a(k,y)dy, \quad \text{where} \quad a(k,y) = \frac{e^{-k_1 y + k_3\sqrt{y}}}{\sqrt{y}} I_1(k_2\sqrt{y}) \geq 0. \tag{28}$$

Lemma A.1. *The function A is continuous on $\mathcal{K}$.*

Proof. For any $k, k' \in \mathcal{K}$, we have

$$|A(k) - A(k')| \leq \int_N^\infty a(k,y)dy + \int_N^\infty a(k',y)dy + \int_0^N |a(k,y) - a(k',y)|\, dy.$$

For any $k \in \mathcal{K}$ and $\epsilon > 0$, we will show that there exist $N > 0$ and $\delta > 0$ such that the right side is bounded above by ϵ if $\|k - k'\|_2 < \delta$. First, we will show that, for any $k \in \mathcal{K}$ and $\epsilon > 0$, there exists $N_1 > 0$ such that $\int_{N_1}^\infty a(k,y)dy \leq \epsilon/3$. Applying the same argument to $k' \in \mathcal{K}$, there exists N_2 such that $\int_{N_2}^\infty a(k',y)dy \leq \epsilon/3$. We then let $N = \max\{N_1, N_2\}$. Finally, we show that for any $k \in \mathcal{K}$, $\epsilon > 0$, and $N > 0$, there exists some $\delta > 0$ such that $\int_0^N |a(k,y) - a(k',y)|\, dy \leq \epsilon/3$ if $\|k - k'\|_2 \leq \delta$.

First, it is proved by Luke (1972) that $\Gamma(\nu+1)(2/y)^\nu I_\nu(y) < \cosh(y)$ for $y > 0$ and $\nu > -1/2$. Taking $\nu = 1$ in this inequality, and observing that $\cosh(y) < e^y$, we have $I_1(y) < (y/2)e^y$. Let $\widetilde{k}_1 = k_1/2 > 0$. Then,

$$\begin{aligned} a(k,y) &= \frac{e^{-k_1 y + k_3\sqrt{y}}}{\sqrt{y}} I_1(k_2\sqrt{y}) \\ &< \frac{e^{-k_1 y + k_3\sqrt{y}}}{\sqrt{y}} \left(\frac{k_2\sqrt{y}}{2} e^{k_2\sqrt{y}}\right) \\ &\leq \frac{k_2 e^{(k_2+k_3)^2/(4\widetilde{k}_1)}}{2} e^{-\widetilde{k}_1 y} =: Ce^{-\widetilde{k}_1 y}, \end{aligned}$$

where the second inequality holds, because $(k_2 + k_3)\sqrt{y} - \widetilde{k}_1 y \leq (k_2+k_3)^2/(4\widetilde{k}_1)$, and the constant $C > 0$ is defined for ease of notation. Therefore, $\int_N^\infty a(K,y)dy \leq C\int_N^\infty e^{-\widetilde{k}_1 y}dy = C(e^{-N\widetilde{k}_1}/\widetilde{k}_1)$. Take

$$N_1 = -\frac{\ln[\epsilon\widetilde{k}_1/3C]}{\widetilde{k}_1} = -\frac{2\ln[(\epsilon k_1)/(6C)]}{k_1}.$$

Then, $\int_{N_1}^\infty a(K,y)dy \leq \epsilon/3$.

The function I_1 is a solution of Bessel's differential equation, so it is continuous on $(0,\infty)$. Consequently, the function $\widetilde{a} : \mathcal{K} \times (0,\infty) \mapsto \mathbb{R}$ defined as $\widetilde{a}(k,y) := e^{-k_1 y + k_3\sqrt{y}} I_1(k_2\sqrt{y}) = a(k,y)\sqrt{y}$ is continuous on $\mathcal{K} \times [0,\infty)$. Choose any $\delta_0 > 0$ and define the compact neighborhood $\mathcal{N}_k(\delta_0) := \{k' \in \mathcal{K} : \|k - k'\|_2 \leq \delta_0\}$. The function $\widetilde{a}$ is continuous on the compact set $\mathcal{N}_k(\delta_0) \times [0,N]$. Therefore, it is uniformly continuous on $\mathcal{N}_k(\delta_0) \times [0,N]$. Consequently, there exists $\delta \in (0,\delta_0]$ such that, for all $(k,y), (k',y') \in \mathcal{N}_k(\delta_0) \times [0,N]$ that satisfy $\|(k,y) - (k',y')\|_2 \leq \delta$, we have $|\widetilde{a}(k,y) - \widetilde{a}(k',y')| \leq \epsilon/(6\sqrt{N})$. Therefore, for any $k' \in \mathcal{K}$ such that $\|k - k'\|_2 \leq \delta$, we have

$$\int_0^N |a(k,y) - a(k',y)|\, dy = \int_0^N \frac{1}{\sqrt{y}} |\widetilde{a}(k,y) - \widetilde{a}(k',y)|\, dy \leq \frac{\epsilon}{6\sqrt{N}} \int_0^N \frac{1}{\sqrt{y}} dy = \frac{\epsilon}{3}. \qquad \square$$

Define the domain $\mathcal{K}^B = (0,\infty)\times(0,\infty)\times(0,\infty)$ and the functions $B:\mathcal{K}^B\mapsto\mathbb{R}$ and $b:\mathcal{K}^B\times(0,\infty)\mapsto\mathbb{R}$ by

$$B(k)=\int_0^\infty b(k,y)dy,\quad \text{where}\quad b(k,y)=\frac{e^{-k_1y}}{\sqrt{y}}\frac{[I_1(k_2\sqrt{y})]^2}{I_1(k_3\sqrt{y})}\geq 0.$$

Lemma A.2. *If $\bar{\mathcal{K}}\subset\mathcal{K}$ is compact, then $\sup\{A(K)|K\in\bar{\mathcal{K}}\}<\infty$. If $\bar{\mathcal{K}}^B\subset\mathcal{K}^B$ is compact, then $\sup\{B(K)|K\in\bar{\mathcal{K}}^B\}<\infty$.*

Proof. Because A is continuous in $\mathcal{K}$, by Lemma A.1, and $\bar{\mathcal{K}}\subset\mathcal{K}$ is compact, it follows that $\sup\{A(K)|K\in\bar{\mathcal{K}}\}=\max\{A(K)|K\in\bar{\mathcal{K}}\}<\infty$.

For any $k\in\mathcal{K}^B$, let $k^*=\max\{k_2,k_3\}$ and $k_*=\min\{k_2,k_3\}$. Then it follows from Theorem 2.1 of Laforgia (1991) that

$$b(k_1,k_2,k_3,y)<\frac{e^{-k_1y}}{\sqrt{y}}\left[e^{2(k^*-k_*)\sqrt{y}}\frac{k^*}{k_*}\right]I_1(k_2\sqrt{y})=a(k_1,k_2,2(k^*-k_*),y).$$

Therefore, $B(k_1,k_2,k_3)\leq A(k_1,k_2,2(k^*-k_*))$ for any $k\in\mathcal{K}^B$. Moreover, the compactness of $\bar{\mathcal{K}}^B$ implies the compactness of the set

$$\mathcal{K}^*:=\left\{(k_1,k_2,2(k^*-k_*))|(k_1,k_2,k_3)\in\bar{\mathcal{K}}^B,k^*=\max\{k_2,k_3\},k_*=\min\{k_2,k_3\}\right\},$$

which is a subset of $\mathcal{K}$. Therefore, $\sup\{B(K)|K\in\bar{\mathcal{K}}^B\}\leq\sup\{A(K)|K\in\mathcal{K}^*\}<\infty$. □

Proposition A.3. *In the catastrophe bond example, if $\bar{\lambda}\geq\underline{\lambda}>0$ and $2\underline{\theta}>\bar{\theta}\geq\underline{\theta}>0$, then for any target state $x\in\mathbb{R}^2$, $\int_{\mathbb{R}^2}\sigma_x^2(x')d\pi(x')<\infty$ and the sequence $\{\sigma_x^2(X_n),n=1,2\ldots\}$ is uniformly integrable.*

Proof. Consider any target state $x\in\mathbb{R}^2$ and any sampling state $x'\in\mathbb{R}^2$. The likelihood ratio is $\ell_x(y;x')=h(y;x)/h(y;x')$. Because the simulation output $F(Y_n)$ is between 0 and 1, for all n, the target-x-sample-x' variance $\sigma_x^2(x')$ defined in Equation (5) satisfies

$$0\leq\sigma_x^2(x')\leq\mathbb{E}\left[(F(Y_n)\ell_x(Y_n;x'))^2|X_n=x'\right]\leq\mathbb{E}\left[(\ell_x(Y_n;x'))^2|X_n=x'\right].$$

To establish the desired conclusions, it suffices to show that this conditional second moment has a finite upper bound over $x'\in\mathbb{R}^2$, for any fixed $x\in\mathbb{R}^2$. Denote $(\lambda,\theta)=\varphi(x)>0$ and $(\lambda',\theta')=\varphi(x')$. We have

$$\mathbb{E}\left[(\ell_x(Y_n;x'))^2|X_n=x'\right]=\left(\frac{e^{-\lambda}}{e^{-\lambda'}}\right)^2e^{-\lambda'}+\int_0^\infty\left(\frac{h(y;x)}{h(y;x')}\right)^2h(y;x')dy.$$

The first term is bounded above by $e^{\bar{\lambda}-2\lambda}$. For the second term, we have

$$\begin{aligned}\int_0^\infty\left(\tfrac{h(y;x)}{h(y;x')}\right)^2h(y;x')dy&=\int_0^\infty\frac{\left[\sqrt{\frac{\lambda}{\theta y}}e^{-\lambda-\frac{y}{\theta}}I_1\left(2\sqrt{\frac{\lambda y}{\theta}}\right)\right]^2}{\sqrt{\frac{\lambda'}{\theta' y}}e^{-\lambda'-\frac{y}{\theta'}}I_1\left(2\sqrt{\frac{\lambda' y}{\theta'}}\right)}dy\\&=\sqrt{\tfrac{\lambda^2\theta'}{\lambda'\theta^2}}e^{\lambda'-2\lambda}\int_0^\infty\tfrac{1}{\sqrt{y}}e^{-\left(\frac{2\theta'-\theta}{\theta\theta'}\right)y}\frac{\left[I_1\left(2\sqrt{\frac{\lambda y}{\theta}}\right)\right]^2}{I_1\left(2\sqrt{\frac{\lambda' y}{\theta'}}\right)}dy\\&=\sqrt{\tfrac{\lambda^2\theta'}{\lambda'\theta^2}}e^{\lambda'-2\lambda}B\left(\tfrac{2\theta'-\theta}{\theta\theta'},2\sqrt{\tfrac{\lambda}{\theta}},2\sqrt{\tfrac{\lambda'}{\theta'}}\right).\end{aligned}\tag{29}$$

For all $x'\in\mathbb{R}^2$, we have that $(\lambda',\theta')=\varphi(x')$ is in a compact set $[\underline{\lambda},\bar{\lambda}]\times[\underline{\theta},\bar{\theta}]$ that does not contain zero. On this set, the arguments of B are all bounded, therefore

$$\left\{\left(\frac{2\theta'-\theta}{\theta\theta'},2\sqrt{\frac{\lambda}{\theta}},2\sqrt{\frac{\lambda'}{\theta'}}\right)|(\lambda',\theta')\in[\underline{\lambda},\bar{\lambda}]\times[\underline{\theta},\bar{\theta}]\right\}=\bar{\mathcal{K}}^B$$

is compact. Thus, it follows from the second claim of Lemma A.2 that Equation (29) has a finite upper bound over $x' \in \mathbb{R}^2$. □

REFERENCES

Bruce Ankenman, Barry L. Nelson, and Jeremy Staum. 2010. Stochastic kriging for simulation metamodeling. *Operat. Res.* 58, 2 (2010), 371–382.

Russell R. Barton, Barry L. Nelson, and Wei Xie. 2014. Quantifying input uncertainty via simulation confidence intervals. *INFORMS J. Comput.* 26, 1 (2014), 74–87.

Richard J. Beckman and Michael D. McKay. 1987. Monte Carlo estimation under different distributions using the same simulation. *Technometrics* 29, 2 (1987), 153–160.

George D. Birkhoff. 1931. Proof of the ergodic theorem. In *Proceedings of the National Academy of Sciences*, Vol. 17. National Academy of Sciences, 656–660.

Eric S. Blake and Ethan J. Gibney. 2011. The Deadliest, Costliest, and Most Intense United States Tropical Cyclones from 1851 to 2010 (and Other Frequently Requested Hurricane Facts). Retrieved July 25, 2017 from http://www.nhc.noaa.gov/dcmi.shtml.

Joshua D. Coval, Jakub W. Jurek, and Erik Stafford. 2009. Economic catastrophe bonds. *Amer. Econ. Rev.* 99, 3 (June 2009), 628–666.

Angelos Dassios and Ji-Wook Jang. 2003. Pricing of catastrophe reinsurance and derivatives using the Cox process with shot noise intensity. *Fin. Stochast.* 7, 1 (January 2003), 73–95.

Mingbin Feng and Jeremy Staum. 2015. Green simulation designs for repeated experiments. In *Proceedings of the 2015 Winter Simulation Conference*, L. Yilmaz, W. K. V. Chan, I. Moon, T. M. K. Roeder, C. Macal, and M. D. Rossetti (Eds.). IEEE Press, Piscataway, NJ, 403–413.

Michael Fu and others. 2015. *Handbook of Simulation Optimization.* Vol. 216. Springer, New York.

Paul Glasserman and Xingbo Xu. 2014. Robust risk measurement and model risk. *Quant. Fin.* 14, 1 (2014), 29–58.

Tim Hesterberg. 1988. *Advances in Importance Sampling*. Ph.D. Dissertation. Stanford University.

Tim Hesterberg. 1995. Weighted average importance sampling and defensive mixture distributions. *Technometrics* 37, 2 (1995), 185–194.

Jack P. C. Kleijnen and Reuven Y. Rubinstein. 1996. Optimization and sensitivity analysis of computer simulation models by the score function method. *Eur. J. Operat. Res.* 88 (1996), 413–427.

Stuart A. Klugman, Harry H. Panjer, and Gordon E. Willmot. 2012. *Loss Models: From Data to Decisions* (4 ed.). John Wiley & Sons.

Andrea Laforgia. 1991. Bounds for modified Bessel functions. *J. Comput. Appl. Math.* 34, 3 (1991), 263–267.

Pierre L'Ecuyer. 1990. A unified view of the IPA, SF, and LR gradient estimation techniques. *Manage. Sci.* 36, 11 (1990), 1364–1383.

Pierre L'Ecuyer. 1993. Two approaches for estimating the gradient in functional form. In *Proceedings of the 25th Conference on Winter Simulation.* ACM, New York, 338–346.

Yudell L. Luke. 1972. Inequalities for generalized hypergeometric functions. *J. Approx. Theory* 5, 1 (1972), 41–65.

Alvaro Maggiar, Andreas Wächter, Irina S. Dolinskaya, and Jeremy Staum. 2015. A Derivative-Free Algorithm for the Optimization of Functions Smoothed via Gaussian Convolution Using Multiple Importance Sampling. Retrieved May 27, 2017 from http://www.optimization-online.org/DB_HTML/2015/07/5017.html.

Luca Martino, Victor Elvira, David Luengo, and Jukka Corander. 2015. An adaptive population importance sampler: Learning from uncertainty. *IEEE Trans. Signal Process.* 63, 16 (2015), 4422–4437.

Alexander J. McNeil, Rüdiger Frey, and Paul Embrechts. 2005. *Quantitative Risk Management: Concepts, Techniques, and Tools.* Princeton University Press, Princeton, NJ.

Robert C. Merton. 1974. On the pricing of corporate debt: The risk structure of interest rates. *J. Fin.* 29, 2 (1974), 449–470.

Sean Meyn and Richard L. Tweedie. 2009. *Markov Chains and Stochastic Stability* (2 ed.). Cambridge University Press, New York, NY.

Munich Re Geo Risks Research. 2015. *Loss Events Worldwide 1980–2014, 10 Costliest Events Ordered by Overall Losses.* Technical Report. Munich Re.

Esa Nummelin. 2004. *General Irreducible Markov Chains and Non-Negative Operators.* Cambridge Tracts in Mathematics, Vol. 83. Cambridge University Press.

Art B. Owen and Yi Zhou. 2000. Safe and effective importance sampling. *J. Amer. Statist. Assoc.* 95, 449 (2000), 135–143.

Reuven Y. Rubinstein and Alexander Shapiro. 1993. *Discrete Event Systems: Sensitivity Analysis and Stochastic Optimization by the Score Function Method.* John Wiley & Sons, New Jersey.

Peter Salemi, Jeremy Staum, and Barry L. Nelson. 2013. Generalized integrated Brownian fields for simulation metamodeling. In *Proceedings of the 2013 Winter Simulation Conference: Simulation: Making Decisions in a Complex World*, R. Pasupathy, S.-H. Kim, R. Hill, A. Tolk, and M. E. Kuhl (Eds.). IEEE Press, 543–554.

Jeremy Staum. 2009. Better simulation metamodeling: The why, what, and how of stochastic kriging. In *Proceedings of the 2015 Winter Simulation Conference*, M. D. Rossetti, B. Johansson R. R. Hill, A. Dunkin, and R. G. Ingalls (Eds.). IEEE Press, Piscataway, NJ, 119–133.

Eric Veach. 1997. *Robust Monte Carlo Methods for Light Transport Simulation.* Ph.D. Dissertation. Stanford University.

Eric Veach and Leonidas J. Guibas. 1995. Optimally combining sampling techniques for Monte Carlo rendering. In *Proceedings of the 22nd Annual Conference on Computer Graphics and Interactive Techniques (SIGGRAPH'95).* ACM, New York, NY, 419–428.

Wei Xie, Barry L. Nelson, and Russell R. Barton. 2014. A Bayesian framework for quantifying uncertainty in stochastic simulation. *Operat. Res. 62*, 6 (2014), 1439–1452.

Received October 2016; revised May 2017; accepted July 2017