Proceedings of the 2017 Winter Simulation Conference
W. K. V. Chan, A. D'Ambrogio, G. Zacharewicz, N. Mustafee, G. Wainer, and E. Page, eds.

HISTORY OF IMPROVING STATISTICAL EFFICIENCY

Russell Barton

Smeal College of Business
Pennsylvania State University
University Park, PA 16802 USA

Marvin K. Nakayama

Computer Science Department
New Jersey Institute of Technology
Newark, NJ 07102, U.S.A.

Lee Schruben

IEOR Department
University of California
Berkeley, CA 94720 USA

ABSTRACT

Statistical efficiency has been a focus of research since the inception of discrete-event simulation modeling and analysis, with origins perhaps twenty years before the first Winter Simulation Conference. We review important work in the design of simulation experiments, variance reduction through dependence structures, and efficient rare-event simulation. The focus is on the early developments, although some recent innovations also receive mention.

1 INTRODUCTION AND ORIGINS

This paper is motivated by the 1969 Winter Simulation Conference paper "Statistical Methods for Improving Simulation Efficiency" by Donald P. Gaver, Jr. (Gaver 1969) and the paper "Variance Reduction Techniques for Digital Simulation" by James R. Wilson (Wilson 1984). These early papers provide a framework for methods of statistical efficiency applied in the discrete-event simulation context. Both categorize the set of methods into a few groups. We identify these groups as methods for design of simulation experiments (sample size, screening, efficient experiment designs); variance reduction /dependency induction (correlation induction via common and antithetic random variates, and control variables); and rare events (stratified and importance sampling, Russian Roulette, splitting). In many cases we cite Winter Simulation Conference papers: often the first work appeared there. Each of these areas is discussed in the sections below. First we present more detail on the historical context, which predates the Conference by as much as twenty years.

The origins of statistical efficiency in simulation depended on the convergence of three activities in the late 1940s and early 1950s: the drive to understand atomic fusion and fission via Monte Carlo models, the advent of digital computers, and recent developments in applied probability and statistics, including rare-event-probability estimation, queueing models, Monte Carlo methods, time series analysis, and the design of experiments. Monte Carlo studies to understand fission used analog computers in the 1940s, and continued with the first digital computer, ENIAC. The ENIAC architecture had significant problems (no memory storage, programming by patch cords), which led John von Neumann to propose a modern architecture for a digital computer (EDVAC) in the 1940's. After the conclusion of World War II, he returned to the Institute for Advanced Study in Princeton, where he oversaw a Princeton team, led by Col. Herman Goldstein, that produced the IAC Princeton Computer in collaboration with RCA Laboratories, also in Princeton. This computer was housed in the basement of the Institute for Advanced Study in

Princeton and was used for fusion simulations, freeway traffic modeling and evolutionary studies; see IAS Computer (1953). The von Neumann architecture was implemented in various forms at about the same time in other laboratories around the U.S. Nicholas Metropolis led a Los Alamos team to implement the architecture for fusion simulation studies as the Mathematical And Numerical Integrator And Calculator: MANIAC.

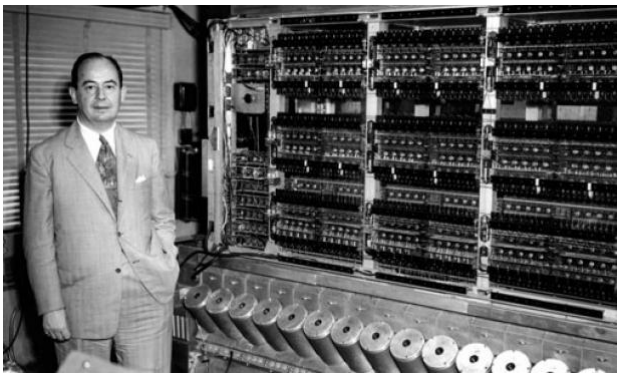

Figure 1: John von Neumann and the MANIAC computer, 1952 (Princeton2017).

Almost simultaneously to the digital computer development, there were advances in statistical efficiency for Monte Carlo methods by Herman Kahn 1948-52, published in Kahn and Harris (1949), Kahn (1950, 1954), and Kahn and Marshall (1953); and other work by Hammersley, Morton and Maulden (referenced in Harling 1958). Also, there were contemporaneous advances in time series analysis (Wiener 1949; Jenkins 1954), control variates (Fieller and Hartley 1954) and response surface methodology (Box and Wilson 1951) that are cited in early simulation analysis papers such as Burdick and Naylor (1966) and Fishman and Kiviat (1967).

RAND research contracts with the Air Force supported the early development of many methods of statistical efficiency. In an early RAND report, Kahn (1954) presents concepts of stratified and importance sampling, splitting, and control variates in a Monte Carlo setting. The RAND activity diffused to many other locations. Richard Conway spent time at RAND, working with Harry Markowitz, and later returning to Cornell to take courses in design of experiments, ranking and selection, decision theory and sequential methods from Jack Kiefer, Robert Bechhofer, Jacob Wolfowitz and Lionel Weiss. Given Conway's influence on many in our field, including Bill Maxwell, Phil Kiviat, and Lee Schruben, he might be considered a key progenitor of methods for statistical efficiency in simulation. He participated in perhaps the first discrete-event-simulation symposium which focused primarily on applications, including inventory systems, bus terminals, airport waiting lines, shop scheduling, financial models (Conway's presentation) and enterprise-wide systems modeling (Alberts 1957). He along with his students published an early discussion of issues in simulation, including common random numbers, response surface methodology, and batch means (Conway, Johnson, and Maxwell 1959). In a later publication (Conway 1963), he reviews statistical efficiency issues in initialization of runs for system comparisons (common initial conditions, common sequences of events) and for spaced batch means (recommending 10-20 batches) for single-run variance estimation of continuous-time statistics. Autocorrelation for discrete-time statistics is demonstrated.

George Fishman was another early developer of methods of statistical efficiency for simulation, and his work has been influential for many of the field's early researchers, including Bob Sargent, Lee Schruben, Jim Wilson, and others ("Computer Simulation Archive Home | Computer Simulation Archive" 2017). George Fishman joined Philip Kiviat at RAND, and worked there in 1964-65 on the development of efficient methods for simulation output analysis, at the request of Murray Geisler, chair of the Logistics Department. Kiviat was working with Markowitz on the development of the SIMSCRIPT simulation

language. Fishman was influenced by his economics training at MIT and Stanford, while Kiviat was influenced by Bechhofer at Cornell and Markowitz. Much of this early work on statistical analysis of simulation models is compiled in Fishman's book *Concepts and Methods in Discrete Event Digital Simulation* (Fishman 1973).

Hammersley and Handscomb (1964) summarize many variance reduction methods, including antithetic and control variates, stratified, sampling, and more. By the first WSC, the "Conference on Applications of Simulation using the General Purpose Simulation System (GPSS), in 1967, the area of statistical efficiency in simulation had more than ten years of development. Nelson and Schmeiser's decomposition approach to variance reduction presents a unifying perspective at the early period of variance reduction research. Nelson's 1985 Winter Simulation Conference tutorial references Purdue technical reports containing the research findings (Nelson and Schmeiser 1985; Nelson 1985). In another important unifying work, Peter Glynn and Ward Whitt provide a comprehensive theoretical characterization of efficiency as of twenty-five years ago, at the end of this early period of methodological development (Glynn and Whitt 1992). They define estimation efficiency of a mean output measure as 1/((output mean squared error)×(total simulation run length)).

We now explore each of the focus areas in more detail: design of simulation experiments (Section 2), variance reduction/dependency induction (Section 3), and rare-event techniques (Section 4). We concentrate on the historical origins, but in some cases important contributions are relatively recent. We are only able to present a small window on the vast accomplishments of many researchers over many decades, and we apologize for omissions of key work that we either missed or lacked room to include.

2 THE DESIGN OF SIMULATION EXPERIMENTS

The importance of careful experiment design was identified in the earliest years of our field. An early paper by Donald Burdick and Thomas Naylor provides a framework for the design and analysis of simulation experiments (Burdick and Naylor 1966).

1. Problem Formulation
2. Data Collection and Processing
3. Formulation of a Mathematical/Probability Model
4. Estimating Model Parameters
5. Evaluation of Model and Parameter Estimates
6. Computer Program Specification, Coding and Verification
7. Computer Model Validation
8. Experimental Design
9. Analysis of Simulation Data.

Predating the first Winter Simulation Conference, this framework retains relevance today, and places the design of experiments in the context of the overall process for a simulation study. This section highlights historical contributions in three areas: sample-size selection for estimating effects, screening designs, and efficient multivariate designs. Designs for optimization are topics that are not included in this review.

Experimental effort in discrete-event simulation differs from effort in physical experiments. The development of design-of-experiments techniques originated in an agricultural setting, where effort was usually proportional to the number of samples in the experiment, a product of the number of experimental conditions, and the number of replications. There was a fixed timeframe, a season, for experimentation. For simulation experiments, the experimental effort depends on the number of samples times the computational effort per sample in the single-processor case, and the number of processors times the average or maximum computational effort per processor in the multiple-processor case. Further, since the data are dynamic, it is possible to compute the interesting quantities using a single sample, if autocorrelation is properly taken into account.

There has been much progress since Frank's comment in the 1968 conference (Frank 1968): "More often than not, the analyst is more concerned with 'debugging' than analysis. When he does finally pass

through the mystical state of model validation, he performs further black magic by experimenting in a manner that would disgust any student of statistics." Good early references for the design of simulation experiments are Fishman (1973) and Kleijnen (1974).

2.1 Sample Size for Estimation

We consider first the simple case of estimating a mean for an output measure of a simulation. Fishman and Kiviat (1965) use spectral methods to characterize the variance of a sample mean resulting from a single long run. The variance of the sample mean is a function of the spectral density at zero frequency. They compare sample sizes required for equivalent-precision estimates of mean queue length using a single long run vs. multiple replications with initialization-bias deletion. For the example studied, they found that the replication approach required up to 50% more computing effort. This work was extended to determine sample size requirements for confidence intervals of specified width in (Fishman 1967; Fishman and Kiviat 1965, 1967; Fishman 1968a). Heidelberger and Welch (1981) use regression to estimate the spectral density at zero frequency more efficiently than Fishman and Kiviat's method.

Hauser, Barish and Ehrenfeld (1966) examine sample size for single runs using autocorrelation estimates via a two-stage approach (requiring stopping and restarting the run). They also determine the required number of batches in a batch-means procedure using either a two-stage or a sequential approach. Fishman (1971) proposes a sample-size method based on an autoreressive moving average (ARMA) model for the output series. Some systems exhibit statistically independent behavior across *regenerative periods*, such as the periods between times when the system is empty and idle. Sample sizes for regenerative estimation are discussed in early papers by Kabak (1968) and Crane and Lemoine (1976). Fishman (1978) explores sample-size issues for batch means. A sequential batch-means procedure with provable properties was presented at the 1976 Winter Simulation Conference and is detailed in Law and Carson (1979). For statistically efficient estimation, Schmeiser (1982) shows that the number of batches should lie typically between 10-30. Nelson (1989) recommends 30-60 batches when used with one to five control variates (Nelson 1989). Sample sizes for batch means and overlapping batch means continues to be studied, with relatively recent results in (Song and Chih 2008; Tafazzoli et al. 2008; Tafazzoli, Steiger, and Wilson 2011).

While computer simulation models admit advantageous strategies not possible with physical experiments, practical implementation of strategies have been incompatible with software capability, particularly in the early years. Frank, in the first *Proceedings*, notes the difficulty in stopping and restarting a GPSS model in mid-run (Frank 1968). Similar practical difficulties arise with estimation via the regenerative method.

The work on statistical efficiency of estimates from one run or from multiple replications, and associated sample-size issues has been a focus of simulation research since Conway, Johnson and Maxwell (1959). While the variance of the sample mean for a single run can be greater or less than that for the same effort split into multiple replications, the mean squared error is generally less (Cheng 1976). The difference in efficiency between a single run and a small set (say 10) of replications with the same total effort is generally small, however (Whitt 1991).

Before consideration of estimation, experiments are necessary for verification and validation of the simulation model. A comprehensive summary of research in design of verification and validation design appears in (Kleijnen and Sargent 2000).

2.2 Sample Size for Comparing Two or More Systems

Early sample size procedures for comparing systems were based on existing statistical methods. Conway, Johnson, and Maxwell (1959) discuss a sequential procedure for the following hypotheses formulation to compare the means μ_1 and μ_2 of two systems (Sobel and Wald 1949; Bechhofer 1955):

$$H_1: \mu_2 - \mu_1 \geq a,$$
$$H_2: -a < \mu_2 - \mu_1 < a,$$

$$H_3: \mu_2 - \mu_1 \leq -a.$$

Fishman (1968b) presents one of the earliest studies particularly relevant for simulation experimentation. Since the sample size corresponds approximately to computing effort, and runs can be made sequentially, he identifies efficient two-stage methods for system comparisons that employ unequal run lengths for the systems, with or without common random numbers. The result is a reduction in computing effort of up to 45%.

Statistically efficient methods for ranking and selection were suggested by Conway (1963), and reaffirmed by Burdick and Naylor (1966). Edward Dudewicz (1976) presents a sample-size rule for ranking-and-selection procedures to guarantee a probability of correct selection given a standard deviation of the output, the number of systems, and the threshold difference. Goldsman, Nelson, and Schmeiser (1991) address sample size determination both for selecting the best system and for multiple comparisons against the best. The paper was reprised in the 2007 Winter Simulation Conference as a Landmark Paper.

2.3 Multivariable Experiment Design

From the earliest years of investigation into efficient statistical methods, multivariable problems were of interest. Conway, Johnson, and Maxwell (1959) note: "Simulation is just a type of experimental investigation. … It will be indeed unfortunate if these new areas do not profit by the lessons and methods of older experimental sciences. … While in principle simulation can be used to investigate the effect of any factors, conditions, procedures and interactions of which the investigator can conceive, in practice this results in factorial experiments whose dimensions dwarf the most powerful computer and the most lavish budget, so that the experimental designs actually used are rather modest."

The earliest publications for multivariable simulation experiment designs drew directly from standard statistical methods. Jacoby and Harrison (1962) describe a large number of multivariable designs appropriate for simulation studies. These include fractional-factorial, nested and compacted hypercubes, random balance designs, Plackett-Burman and other orthogonal designs, split-plot designs, and sequential bifurcation designs. A. L. Frank suggests factorial and response surface designs in a paper published in the first Winter Simulation Conference proceedings (Frank 1968). Early work by Corynen (1975), Mihram (1972), Fishman (1973), and Kleijnen (1974) present mathematical formalisms for the design and conduct of multivariable simulation experiments, taking into account the special nature of experiments using computer models. Biles (1979) is an early comprehensive review of simulation experiment design that was presented at the 1979 Winter Simulation Conference. Important references for choosing sample size for multivariable experiments are (Kleijnen 1987; Nelson 1992); the first is Kleijnen's text and the second is a Winter Simulation Conference paper.

2.4 Screening

The need for efficient screening designs in simulation was recognized well before the first Winter Simulation Conference. Jacoby and Harrison review fractional-factorial, random balance and sequential bifurcation experiment designs for settings where the number of simulation runs must be small relative to the number of input variables (Jacoby and Harrison 1962). Kleijnen (1975) suggests resolution III and IV designs as well as group-screening methods. Montgomery and Weatherby (1979) is a detailed description of fractional-factorial and supersaturated designs for screening that was presented at the 1979 Winter Simulation Conference. This work contains a bibliography of important early research in screening and in simulation experiment design more generally.

While sequential bifurcation was identified early on as an important screening method and demonstrated by Jacoby and Harrison (1962), interest by the discrete-event community was relatively late. Bettonvil and Kleijnen (1997) present a slight variation of this method for deterministic simulation based on Bettonvil's 1990 thesis. Russell Cheng (1997) extends their method to the stochastic setting. Hong Wan,

Bruce Ankenman and Barry Nelson modified the method to incorporate tests for significance and power, calling the modification *controlled sequential bifurcation* (Wan, Ankenman, and Nelson 2006). For the case when the signs of effects are not known, a two-phase approach beginning with a fractional factorial design (FF-CSB) can be effective. This was presented at the 2006 Winter Simulation Conference, with extended results in Sanchez, Wan, and Lucas (2009). Work on sequential bifurcation continues to the present. Bruce Ankenman, Russell Cheng and Susan Lewis recently developed an enhancement, the Anscombe Fully Sequential Bifurcation Method, that employs sequential stopping rules for the number of simulation replications needed at each step, for screening based on influence on the mean or on the variance of the output (Ankenman, Cheng, and Lewis 2015). The statistical efficiency of the method dominates earlier methods.

Simultaneous sinusoidal variation of multiple input parameters, each at different frequencies, provides a way to do screening within one or two simulation runs. First presented at the 1981 Winter Simulation conference, Schruben and Cogliano cite that work in their 1987 *CACM* paper (Schruben and Cogliano 1981; 1987). There has been continued refinement and extension of the method; see for example (Morrice and Schruben 1993; Morrice 1995; Sanchez, Moeeni, and Sanchez 2006). Schruben's doctoral advisee Sheldon Jacobson and co-authors have extended the frequency domain strategy to designs for constructing response surface models (Jacobson 1989; Jacobson, Buss, and Schruben 1991).

2.5 Efficient Experiment Designs for Response Models

Early papers on multivariable simulation experiment design promoted designs from the statistical literature with known efficiency properties. These were typically full-factorial, fractional-factorial, or response surface designs. The potential strategy of blocking a factorial or response surface design based on common random numbers was recognized early. Conway, Johnson, and Maxwell (1959) note: "Since a sequence of random numbers can be reproduced, it is possible to compare alternatives using identical sequences. Such an experimental procedure could perhaps be considered the limiting case of the blocking concept commonly employed in experimental designs – blocks as homogenous as possible are selected to reduce the variability of the results." They note the difficulty of analysis using common random numbers for comparison of more than two alternatives at a time.

Schruben and Margolin (1978) were the first to develop an efficient experiment design strategy and analysis approach that specifically took advantages of the simulation setting. Much research ensued, including the discussion papers that immediately followed Schruben and Margolin's paper. Subsequent works on efficient response surface designs with correlation induction include Nozari, Arnold, and Pegden (1987), Donohue, Houck, and Myers (1993), and Tew and Wilson (1992).

Sequential designs such as those proposed by Donohue, Houck, and Myers (1993) provide efficiency improvement in multivariable designs. Much focus has been on simulation optimization, but there have been important developments for response surface designs used for characterization. Cheng and several coauthors propose an efficient design strategy when variance is heterogeneous across design points (Cheng and Kleijnen 1999; Cheng, Kleijnen, and Melas 2000; Lamb and Cheng 2002). Jack Kleijnen, Susan Sanchez, Thomas Lucas and Thomas Cioppa provide an overview of experiment design strategies in Kleijnen et al. (2005) that includes brief discussion of designs for non-traditional response models. Figure 1 in that paper gives recommended designs for a comprehensive set of model scenarios.

2.6 DOE for Non-traditional Models

Experiment designs for general response models preceded the development of many of the models in use today. In the late 1980's, Sacks, Welch, Mitchell, and Wynn presented a very flexible response surface model based on spatial correlation, with a Bayesian interpretation (Sacks et al. 1989; Currin et al. 1991; Mitchell and Morris 1992). The model is based on work by Krige (1951), and such response surface models are often called kriging models. Latin hypercube designs were developed as a stratified sampling method for estimating the mean of a function of several random variables (see equation (1) below) in McKay,

Beckman, and Conover (1979), and Sacks et al. (1989) recommend their exploration for fitting spatial correlation models. In two Winter Simulation Conference papers, Barton provided an early overview of nontraditional models, including spatial correlation, generalized linear models, frequency domain and wavelet models, splines, radial basis functions, and kernel models (Barton 1992, 1994). Experiment design issues were discussed in the second paper. Variations on the Latin hypercube design remain popular for fitting nontraditional models, but such models present an opportunity for new kinds of sequential designs. Kleijnen and van Beers (2004) estimated prediction variance using jackknifing and sequentially added design points at locations with largest prediction variance. Although the "nugget effect" provided a method for non-interpolating kriging models, it was not until the 2008 Winter Simulation Conference that the simulation community saw a formal statistical method for stochastic kriging appropriate for the output of discrete event simulation models. The subsequent journal paper includes details on the model, experiment design process, and fitting (Ankenman, Nelson, and Staum 2010). Experiment designs using common random numbers were found to be ineffective with this model (Chen, Ankenman, and Nelson 2012). Efficient experiment designs when using the model for input uncertainty characterization appear in Barton, Nelson, and Xie (2013).

2.7 Models for Statistical Efficiency

We mention only briefly that there has been important research in employing special statistical models for gaining efficiency. These include ARMA models for the output stream (Fishman 1971; Schriber and Andrews 1984), indirect estimation (Carson and Law 1977; Glynn and Whitt 1989), and standardized time series (Schruben 1983).

3 VARIANCE REDUCTION TECHNIQUES

Variance reduction techniques (VRTs) include a variety of sampling and/or modeling "tricks" with the intent of increasing the precision of some statistical estimators without decreasing their accuracy (by too much) or requiring (too much) extra effort. There is a vast literature on VRTs: econpapers.repec.org identifies over 25000 research papers with "Variance Reduction Techniques" in their titles or keywords (Econpapers.repec.org 2017). A search of the Winter Simulation Conference Archives (http://informs-sim.org) for "Variance Reduction" returns several hundred links. A sample of VRT papers spanning the five decades of the Winter Simulation Conference includes Lavenberg and Welch (1978), Nelson (1985), Cheng (1986), L'Ecuyer (1994), Glasserman, Heidelberger, and Shahabuddin (2000), and Dong and Nakayama (2014).

Section 1 of this paper presented different ways VRTs have historically been categorized. Section 2 of this paper set the stage for these next two sections: VRTs typically are embedded within an overall simulation experiment. In this section, we focus on the historical development of VRTs where dependences are induced among samples from the simulation output. Section 4 presents the history of VRTs for rare-event probability estimation.

3.1 Variance Reduction and Simulation Effort

Suppose that we want to estimate a parameter θ, where for example $\theta = E(Y)$ for some simulation response Y. Consider two different simulation studies producing two unbiased estimators $\hat{\theta}^{(1)}$ and $\hat{\theta}^{(2)}$, where the amount of simulation effort to obtain $\hat{\theta}^{(1)}$ with a variance of σ_1^2 is C_1, and the amount of simulation effort to obtain $\hat{\theta}^{(2)}$ with a variance of σ_2^2 is C_2. Then if $C_1/C_2 < \sigma_2^2/\sigma_1^2$, we say that the estimator $\hat{\theta}^{(1)}$ is more efficient than $\hat{\theta}^{(2)}$. Simulation effort, following Owen (2017), includes intangibles such as modeling, execution, analysis, and time to decision. Kahn (1954) describes C_i as simulation *cost*, including designing, programming verification and validation of the model, and the cost of computer run time. This is more meaningful for simulation analysis than simply comparing sample sizes, usually as the number of simulation runs or output observations. It is possible to change the simulation model and the

computational effort at any time while running a simulation model, and so the effort in a simulation experiment can be determined continuously, e.g. as simulated time (Schruben 2010). The modeling world-view *Weltanschauung* (Nance and Sargent 2002) used in a simulation model can also make orders-of-magnitude differences in total effort (Roeder et al. 2002) and can facilitate or impede the use of VRTs.

In this Section, we focus on VRTs that sample dependent data, and measure that dependency by the correlations among the observations. A large number of techniques for generating dependent samples have been developed: common random numbers for trying to induce positively dependent samples, antithetic variates for trying to generate negatively dependent samples, control variates, conditional sampling, and some more exotic methods such as using copulae, and others. The specific techniques used to create dependencies is not our focus here, and the simulation effort required in using different methods can differ.

3.2 The Monte Carlo Swindle

Variance reduction techniques are described by McLaughlin and Tukey (as quoted in Gross 1973) as a restatement of a problem which reduces the amount of computation needed to achieve a desired precision in the results. This is sometimes called the colorful, but misleading, name of being a Monte Carlo "swindle" (Kafadar 2006). To illustrate how one might try to profit from dependent sampling, consider two different simulation experiments, producing responses Y_1 and Y_2, with common variance σ^2. These responses are generically denoted by a function

$$Y = \eta(y),$$

where $y = (y_1, y_2, \ldots, y_r)'$ is generated in a simulation run. For example, y may represent the vector of generated interarrival and service times in a queue, and $\eta(y)$ computes, for the realization y, the average waiting time of a particular customer. Now suppose that $Y_1 = \eta(y^{(1)})$ and $Y_2 = \eta(y^{(2)})$, where $y^{(1)}$ and $y^{(2)}$ are sampled in a dependent manner. The hope is that this sample dependency will carry over to produce dependent simulation responses Y_1 and Y_2 (which can be ensured when η has some good behavior). In one of our two experiments, the two responses will have a positive correlation, and in the other they will be negatively correlated. Two statistics that might be of interest are the sample average response and the (normalized) response difference,

$$\hat{\theta}_{\text{avg}} = (Y_1 + Y_2)/2 \quad \text{and} \quad \hat{\theta}_{\text{diff}} = (Y_1 - Y_2)/2.$$

In either of the two experiments, if the induced correlation of the responses is ρ, the variances of these unbiased estimators are

$$\text{Var}\left[\hat{\theta}_{\text{avg}}\right] = (1 + \rho)\sigma^2/2 \quad \text{and} \quad \text{Var}\left[\hat{\theta}_{\text{diff}}\right] = (1 - \rho)\sigma^2/2.$$

We see that for negatively correlated responses ($\rho < 0$), the variance of the average is reduced, but the variance of the difference is increased *by exactly the same amount*. The opposite is true if the responses are positively correlated ($\rho > 0$). There is no net reduction in the sum of the variances of the two statistics from dependent sampling, so the swindle is in fact a variance trade. This example illustrates the result in Chen, Ankenman, and Nelson (2012): inducing positively correlated responses tends to increase the variance of location estimators.

This variance invariance property is more general: in a general simulation experiment producing k simulation responses, Y_j, $j = 1, 2, \ldots, k$, with n linear statistics of interest, $\hat{\theta}_p$, $p = 1, 2, \ldots, n$, having weighting functions x_j^p, we have that

$$\hat{\theta}_p = \sum_{j=1}^{k} x_j^p Y_j, \text{ or using matrix notation } \widehat{\boldsymbol{\theta}} = \boldsymbol{XY},$$

where the p^{th} row of the $n \times k$ matrix $\boldsymbol{X}$ is $\boldsymbol{x}^p = (x_1^p, x_2^p, \ldots, x_k^p)$ and $\boldsymbol{Y} = (Y_1, Y_2, \ldots, Y_k)'$. For any set of orthonormal $\boldsymbol{x}^p$, $p = 1, 2, \ldots, n$, the sum of the variances of the entries in $\widehat{\boldsymbol{\theta}}$ is a constant, irrespective of the correlations induced in the simulation output. This means that if we try to use dependent sampling to reduce the variance of some statistics, we will increase the variance of some other independently estimable statistics by the same amount. It is up to the experimenter to drive a good variance bargain by inducing correlations to inflate the variances of statistics that are not of particular interest. This suggest the strategy of adding *uninteresting* factors (like the artificial effect of the induced correlations) to saturate an experiment and then induce dependencies that maximize the sum of the variances of this set of uninteresting estimators, as described by Schruben (1979).

This variance invariance property shows up in some apparently unrelated results. For example: when discussing the impact of a monotonic response on the effectiveness of antithetic variates, Owen (2013) (in the paragraph after equation 8.4) shows that antithetic sampling eliminates the variance when a response is odd, but doubles the variance if a response is even, with no net reduction (Owen's odd and even quantities are orthogonal). Cheng (1984) shows an effective method for dealing with difficult response functions $\eta(\cdot)$, illustrated with two case studies.

3.3 Dependency Induction and Control Variates

This section gives a brief overview of several methods for intentionally inducing dependencies in a sample. Some have changed little since they were described in the seminal book by Hammersley and Handscomb (1964): antithetic sampling, common random numbers, and control variates. There have been some relatively recent innovations such as copulae sampling (Ehrlichman and Henderson 2008).

The simplest method of introducing positive and negatively correlated samples is through the reuse and modification of the pseudo-random number streams and their antithetic streams as inputs to simulations of the same or different experimental design points.

To induce positive dependencies, the most popular method is to use the same sets of pseudo-random numbers in different simulation runs. This is known as using common random numbers. As noted earlier, this often reduces the errors for comparisons. Common random numbers have been used successfully in a wide range of applications and are incorporated in many simulation languages. See Schruben (2010) for some extensions, and methods for keeping random number steams synchronized between runs, such as the innovative transient entity "tattooing" suggestion by Kelton (1999).

There are a number of methods proposed for inducing negative correlation: the simplest is to run a pair of simulations, using a set of pseudo-random numbers used to generate random variates in one run and subtracting the first-run pseudo-random numbers from 1 in the second run to create *antithetic variates* (Hammersley and Morton 1956). As mentioned earlier in the reference by Owen, monotonicity is sufficient to generate negative correlations, but not necessary. Further, the degree of induced correlation can be small even with a monotonic $\eta(\cdot)$. Latin hypercube sampling (McKay, Beckman, and Conover 1979) has been shown to induce stronger negative dependencies in simulated activity networks (Avramidis and Wilson 1993; Tew and Wilson 1992). Other innovative methods of inducing correlations have been proposed by Page (1965), Fishman and Huang (1980), and Burt, Gaver, and Perlas (1970).

The concept of control variates unifies geometrically many dependency induction methods (Szechtman and Glynn 2002). The basic form of control variates is to combine an estimator with a control estimator that has (asymptotic) zero-mean. Inducing dependencies between the raw estimator and the control estimator, in certain circumstances, can lead to a variance reduction. The control variates can be sampled from another simulation whose expected value is (asymptotically) known (external control variates) or an estimator of a quantity with known mean in the same simulation model (internal control variates), such as the sample mean for input interarrival and/or service times. The weighting of the raw estimator with the

control estimator can be adaptive, changed as more is learned about the variances and covariances of these estimators during a run. See for example, Kim and Henderson (2004).

3.4 Combining Dependency Induction Methods

An abstract framework for variance reduction in a simulation experiment was developed by Nelson and Schmeiser (Nelson 1985, 1987; and Nelson and Schmeiser 1983). Correlation induction using common and antithetic random number streams was integrated into the overall design of a simulation experiment as discussed earlier in Section 2.5.

This can be generalized using the variance invariance property given in Section 3.2. This includes the Schruben-Margolin strategy as a special case that was extended to control variates in Schruben (1979). The property can be generalized to include any methods for inducing positive or negative dependencies, for example in Avramidis and Wilson (1993), using Latin hypercube sampling to induce correlations; in Tew and Wilson (1994), where the focus is on simulating activity networks and validation; and in the paper mentioned in Section 2.5 by Nozari, Arnold, and Pegden (1987), where they present some statistical analysis techniques.

We close this section by noting that the dependent sampling VRTs discussed in this section have had mixed results. Many of the negative outcomes can be viewed as trying to pull off a variance swindle but instead making a poor variance trade.

4 RARE-EVENT SIMULATION

While the VRTs discussed in Section 3 can be effective in reducing variance in many settings, they may not be sufficient to analyze rare events. For example, we may be interested in estimating the probability of an extremely rare but highly important event. For a civil commuter aircraft, regulations may specify that the probability of a catastrophic failure during an average-length flight (about 8 hours) is less than 10^{-9}. In particle shielding, a physicist or nuclear engineer may want to ensure that the probability that a neutron penetrates a shield is very small, in the range of 10^{-6} to 10^{-10}.

To understand the difficulties arising in rare-event problems, suppose that we want to estimate the probability $\theta = P(B)$ of a (rare) event B with $\theta = 10^{-8}$. With simple random sampling (SRS), we perform n independent and identically distributed simulation runs, and let I_j be the indicator of event B occurring on the jth run, i.e., $I_j = 1$ (resp., 0) if B occurs (resp., does not happen) on the jth run. The SRS estimator is then $\hat{\theta}_n = (1/n)\sum_{j=1}^{n} I_j$. Simply observing a single occurrence of B requires on average a sample size $n = 1/\theta = 10^8$. To obtain a reasonably accurate estimate of θ, we need n much larger. For example, the SRS estimator $\hat{\theta}_n$ of θ has standard deviation $\sqrt{\theta(1-\theta)/n} \approx 10^{-4}/\sqrt{n}$. Even though in absolute terms the estimator's standard deviation is quite small, it is orders of magnitude larger (for moderate n) than the value θ we are estimating. In particular, define the *relative error* of estimator $\hat{\theta}_n$ as the relative expected half width of a 95% confidence interval for θ; i.e., the relative error of $\hat{\theta}_n$ is $1.96\sqrt{\theta(1-\theta)}/(\theta\sqrt{n}) \approx 1.96/\sqrt{\theta n} = 1.96 \times 10^4/\sqrt{n}$. Thus, obtaining a relative error of $\delta = 10\%$ requires a sample size $n = \left(\frac{1.96}{\delta}\right)^2 \frac{(1-\theta)}{\theta} \approx 4 \times 10^{10}$.

The VRTs in Section 3 may be able to reduce somewhat the necessary sample size, but they often are not designed to analyze such vanishingly infrequent events. Instead, these types of problems require special techniques specifically built for this purpose. The main simulation methods for analyzing rare events are *importance sampling* (IS) and *splitting*, and we will discuss their historical developments in Sections 4.1 and 4.2, respectively. Many of the techniques for rare-event simulation arose during the early days of Monte Carlo related to particle transport. For more details on simulation methods for rare-event problems, see the survey articles by Heidelberger (1995), Nicola, Shahabuddin, and Nakayama (2001), Juneja and

Shahabuddin (2006), and Blanchet and Lam (2012), and the books by Lux and Koblinger (1991), Bucklew (2004), and Rubino and Tuffin (2009).

4.1 Importance Sampling

Hammersley and Handscomb (1964) state (p. 42), "Never sample from a distribution merely because it arises in the physical context of a problem, for we may be able to use a better distribution in the computations and still get the right answer." This idea forms the basis of importance sampling: change the sampling distributions driving the original system so that the rare events of interest occur more frequently. Because the system now fundamentally differs from the original one, we have to multiply the results by a correction factor, known as the likelihood ratio, to recover unbiased estimates. Specifically, suppose our goal is to estimate $\theta = E[Y]$ for some $\Re$-valued random variable $Y = h(X)$, where X is an $\Re^d$-valued random variable with density function f, and $h: \Re^d \to \Re$ is a given function. Thus, we have that

$$\theta = E_f[h(X)] = \int h(x)f(x)dx, \tag{1}$$

where E_f denotes the expectation when X has density f. For example, in the special case when θ is the probability that X lies in a (rare) set A, we have that $h(X) = I(X \in A)$, where $I(\cdot)$ is the indicator function. In general, the SRS approach to estimate θ first generates n independent observations $X_1, X_2, \ldots, X_n$ from density f, and then computes $\hat{\theta}_n = (1/n)\sum_{j=1}^{n} h(X_j)$ as an unbiased estimator of θ.

To implement importance sampling, let g be another density function on $\Re^d$ such that $g(x) > 0$ whenever $h(x)f(x) > 0$. We then re-express θ as

$$\theta = \int h(x)f(x)dx = \int h(x)\frac{f(x)}{g(x)}g(x)dx = E_g[h(X)L(X)], \tag{2}$$

where $L(x) \equiv f(x)/g(x)$ is the *likelihood ratio*, and E_g denotes expectation when X has density g. The second and third steps in (2) apply a so-called "change of measure" because the expectation E_g is now taken with respect to the measure corresponding to g rather than the original one f. By (2), we can obtain an unbiased estimator of θ by generating independent observations $X_1, X_2, \ldots, X_n$ from density g, and then averaging the $h(X_j)L(X_j)$; i.e., the IS estimator of θ is $\tilde{\theta}_n = (1/n)\sum_{j=1}^{n} h(X_j)L(X_j)$. The key to applying IS is choosing g so that $\tilde{\theta}_n$ has smaller variance than the SRS estimator $\hat{\theta}_n$, and the approach has been developed for a wide spectrum of stochastic systems (e.g., Glynn and Iglehart 1989). A poor choice of g can lead to a variance increase, or even infinite variance.

Some early papers on IS call the approach "quota sampling" (Goertzel 1949a, Goertzel 1949b). A footnote in Goertzel and Kahn (1949) states, "Importance sampling is the term used herein for what has in several reports been called 'quota sampling.' … We believe 'importance sampling' to be more descriptive." A footnote in Kahn (1950) observes, "We had previously called this process 'quota sampling.' This last terminology is a little misleading as it is not identical with the quota sampling of the statisticians." Kahn (1950) further writes, "We are trying to sample every part of the phase space that we are studying, with a frequency proportional to the importance of that part of the phase space times the probability of getting into it. For this reason the name 'importance sampling' has been suggested by G. Goertzel for this procedure."

Pinpointing an exact origin for IS has been elusive. In a technical report describing IS, Goertzel (1949b) writes, "This document is intended to be a working report for the ORNL Shielding Work Session (Summer 1949) and does not have the complete set of references a final report would have. None-the-less, I feel it desirable to remark that the work herein is a logical continuation of some work of H. Kahn and T. E. Harris (various RAND Reports). Further, many of the thoughts herein arose during conferences between Mr. Kahn and me during the days June 15–17" (1949). A footnote of Kahn (1950) states, "It was during some

conversations at the Oak Ridge National Laboratory during the summer of 1949 that it was decided that finding the optimum importance function was probably always equivalent to solving the adjoint problem. Present at these conversations were H. Feshbach, F. Friedman, G. Goertzel, and H. Kahn."

When the response function h in (1) is nonnegative, the optimal choice of the IS density g in (2) is g_* defined as $g_*(x) = h(x)f(x)/\theta$. In this case, note that

$$h(x)L(x) = h(x)\frac{f(x)}{h(x)f(x)/\theta} = \theta,$$

so just a single sampled observation will result in the exact value θ we want to estimate. Although this approach will produce an IS estimator with zero variance, it cannot be implemented in practice because it requires knowing the quantity θ that we are trying to estimate. Kahn and Harris (1949) and Goertzel (1949b) are early references noting the optimal g_*, and they also observe that designing g to mimic g_* can lead to a "good" change of measure.

While the preliminary work on importance sampling focuses primarily on empirical evaluation of particular changes of measure, some later studies examine theoretical properties by considering various asymptotic regimes. Here, we introduce a *rarity parameter* $\epsilon > 0$, and we consider a sequence of models parameterized by ϵ, where the rare event of interest becomes rarer as ϵ shrinks. For example, in a system with a single queue having a large but finite capacity, we may want to estimate the probability that the queue overflows before emptying, and we let the queue capacity be $1/\epsilon$. As another example, consider a highly dependable Markovian system comprising a collection of highly reliable components, each with specified failure and repair rates, where failure rates are much smaller than repair rates. The system, which fails when particular combinations of components are simultaneously down, may further include *failure propagation*, where the failure of one component can trigger others to also fail instantaneously with certain probabilities. Shahabuddin (1994) parameterizes the model by assuming that in the resulting continuous-time Markov chain, failure transitions (i.e., transitions in which one or more components fail) have rates that are positive powers of ϵ, while repair transitions (with one or more components getting repaired) have constant rates, independent of ϵ. In both the queueing example and the reliability problem, the rare-event performance measure θ being estimated is therefore a function of ϵ, so we write it as θ_ϵ. We then analyze the theoretical behavior (e.g., the relative error) of the simulation estimator of θ_ϵ as $\epsilon \to 0$.

4.1.1 Queueing Systems

One of the earliest works establishing asymptotic properties of IS estimators is Siegmund (1976). For a given interval $[a, b]$ with $a < 0 < b$, this paper considers estimating the probability that a one-dimensional random walk first exits the interval by exceeding b, and it establishes asymptotic efficiency of an IS estimator as $b \to \infty$ (or equivalently, $b = 1/\epsilon$ with $\epsilon \to 0$). The approach is based on large-deviations theory, and later papers apply this idea to study insurance and queueing problems. For example, Asmussen (1985) devises an IS method to efficiently estimate the ruin probability of an insurance company, which receives premiums and pays out claims, and ruin occurs when the firm's reserves first become negative. In queueing, the rare-event goal is often to estimate the probability that a large but finite queue overflows before the system empties; e.g., see Cottrell, Fort, and Malgouyres (1983), Parekh and Walrand (1989), Sadowsky and Bucklew (1990), and Sadowsky (1991).

Initial IS schemes employ state-independent changes of measure, but Glasserman and Kou (1995) and Glasserman and Wang (1997) show that the approach can produce asymptotically inefficient estimators. To overcome this issue, Dupuis and Wang (2004, 2007) devise state-dependent importance samplers for light-tailed systems (e.g., interarrival- and service-time distributions have moment generating functions that exist in a neighborhood of 0), exploiting ideas from control theory.

When input distributions have heavy tails (e.g., tail distribution functions decrease polynomially), the way in which the rare event of interest typically occurs fundamentally differs from the light-tailed case. For this setting, early papers noting the difficulties in IS caused by heavy-tailed increments include Asmussen, Binswanger, and Hojgaard (2000) and Juneja and Shahabuddin (2002).

4.1.2 Highly Reliable Markovian Systems

In the 1980s, IBM produced large-scale, fault-tolerant computer systems with high-dependability requirements. Such systems are used, e.g., by overnight-delivery companies to track packages, and by banks to handle financial transactions. The systems must achieve extremely small unavailability, say less than 10^{-6}. To study the dependability of designs of the systems, Ambuj Goyal of the IBM Thomas J. Watson Research Center led a team of researchers to create the System Availability Estimator (SAVE) tool (Goyal et al. 1986), a software package used by system designers to evaluate the dependability of a proposed system. SAVE builds a Markov chain model from a user-input high-level description of a system as a collection of components, each of which can fail and be repaired. SAVE then computes various dependability measures, including the steady-state unavailability (SSU), the mean time to failure (MTTF), and the unreliability, which is the probability that the system fails before a specified time horizon. The tool allows for various component interactions, such as operational dependencies (e.g., an operational component becomes dormant and thus has a different failure rate when other specified components are failed), repair dependencies (i.e., a failed component cannot be repaired while other specified components are down), and failure propagation.

The initial version of SAVE employed analytical (i.e., non-simulation) techniques to numerically solve for the dependability measures of the Markov chain. But the size of the model grows exponentially in the number of components in the system; e.g., a simple system with m distinct components, each of which can be up or down, has at least 2^m states in its Markov chain model. This severely limits the size of the systems that analytical techniques can handle.

To avoid the state-space explosion, Conway and Goyal (1987) enhanced SAVE to also employ simulation, incorporating the regenerative method (Crane and Iglehart 1974) and discrete-time conversion (Hordijk, Iglehart, and Schassberger 1976). But these methods by themselves are inadequate to accurately estimate the dependability measures for the rare-event problems, so Conway and Goyal (1987) further apply importance sampling. In particular, the technique of *failure biasing*, originally devised by Lewis and Bohm (1984) to study nuclear reactors, proved to be extremely effective at reducing variance in the Markov chain simulations of SAVE.

In 1987 and 1988, Perwez Shahabuddin, then a Ph.D. student in operations research at Stanford University, was a summer intern at IBM Watson, mentored by Goyal, Phil Heidelberger, and Steve Lavenberg. Shahabuddin worked on extending SAVE's simulation capabilities, which led to several important advances in IS methodology. Exploiting the regenerative structure, one can express dependability measures such as the SSU using a ratio representation, and Goyal, Heidelberger, and Shahabuddin (1987) find that independently estimating the numerator and denominator using different sampling measures (a technique they call *measure-specific IS*) can be more effective than estimating both using the same approach. A ratio representation of the MTTF proves to be extremely useful as it isolates the rare event of interest in the ratio's denominator, allowing efficient estimation via measure-specific IS (Shahabuddin et al. 1988).

While the original version of failure biasing (later called *simple failure biasing*) works well when all failure transitions out of a state have rates of the same magnitude, the approach may perform poorly for unbalanced systems (failure transitions out of a state have rates of different orders of magnitude). This led Shahabuddin (1994) to devise *balanced failure biasing* (BFB), which is implemented in the SAVE package (Goyal et al. 1992). Shahabuddin (1994) establishes that BFB produces estimators of the SSU and MTTF

having bounded relative error; i.e., the BFB estimators have relative errors that remain bounded as $\epsilon \to 0$, where we recall that ϵ is the rarity parameter used in specifying the rates of failure transitions.

4.2 Splitting

Splitting is another simulation technique useful for analyzing rare events. The method embeds the rare set of states of interest in a sequence of larger, less-rare sets, and reinforces sampling paths that get "closer" to the desired rare set, and terminates those that move further away. The idea of splitting was introduced in Kahn and Harris (1949), but they credit John von Neumann with devising the method. Specifically, in describing three approaches (splitting, importance sampling, and conditional Monte Carlo) for rare-event simulations, they write, "The first one we will discuss is the splitting technique, mentioned by Dr. von Neumann. In one method of applying this, one [defines] regions of importance in the space being studied, and, when the sampled particle goes from a less important to a more important region, it is split into two independent particles, each one-half the weight of the original. If you go from a more important to a less important region, you double the weight of the sampled neutron but play a game of chance, with probability one-half of winning to decide if the history is to be continued. The purpose of this is to spend most of the time studying the important rather than the typical particles, but to do it in unbiased fashion." The idea of randomly terminating a generated history is called Russian roulette, which Kahn (1954) notes, "John von Neumann and Stanley Ulam are responsible for both the sampling technique and its name."

Splitting (under various names) has been applied to study rare events in many settings, including particle transport (Booth 2009); telecommunications, where it has been called RESTART (Repetitive Simulation Trials After Reaching Thresholds; Villén-Altamirano and Villén-Altamirano 1991, 1994), and structural reliability, where it is known as subset simulation (Au and Beck 2001). Glasserman et al. (1998, 1999) and Dean and Dupuis (2009) analyze theoretical properties of splitting.

5 CONCLUSION

The foundational contributions to statistical efficiency in simulation experiments described here span nearly sixty years. Yet this continuing series of developments provide a guidepost to the future. We expect that simulation researchers are not finished with improvements to statistical efficiency - in rare event simulation, variance reduction /dependency induction and design of simulation experiments.

ACKNOWLEDGMENTS

We appreciate the efforts of the reviewers in helping us to improve this paper. This work has been supported in part by the National Science Foundation under Grant No. CMMI-1537322. Any opinions, findings, and conclusions or recommendations expressed in this material are those of the authors and do not necessarily reflect the views of the National Science Foundation.

REFERENCES

"A History of Early Computing at Princeton - Alan Turing Centennial." 2017. Accessed April 25. https://www.princeton.edu/turing/alan/history-of-computing-at-p/.

Alberts, W. E. 1957. "Report to the Eighth A.I.I.E. National Convention on the System Simulation Symposium, New York, May 16-17, 1957." *Operations Research* 5: 851–58.

Ankenman, B. E., R. C. H. Cheng, and S. M. Lewis. 2015. "Screening for Dispersion Effects by Sequential Bifurcation." *ACM Transactions on Modeling and Computer Simulation* 25:2:1-2:27.

Ankenman, B. E., B. L. Nelson, and J. Staum. 2010. "Stochastic Kriging for Simulation Metamodeling." *Operations Research* 58:371–82.

Asmussen, S. 1985. "Conjugate Processes and the Simulation of Ruin Problems." *Stochastic Processes and their Applications*20:213–229.

Asmussen, S., K. Binswanger, and B. Hojgaard. 2000. "Rare Events Simulation for Heavy-tailed Distributions." *Bernoulli* 6:303–322.

Au, S.-K., and J. L. Beck. 2001. "Estimation of Small Failure Probabilities by Subset Simulation". *Probabilistic Engineering Mechanics* 16:263–277.

Avramidis, A. N., and J. R. Wilson. 1993. "Integrated Variance Reduction Strategies." In *Proceedings of the 1993 Winter Simulation Conference*, edited by G. W. Evans, M. Mollaghasemi, E. C. Russell, and W. E. Biles, 445-454. Piscataway, New Jersey: Institute of Electrical and Electronics Engineers, Inc.

Avramidis, A. N., and J. R. Wilson. 1993. "A Splitting Scheme for Control Variates." *Operations Research Letters* 14:187–198.

Avramidis, A. N., and J. R. Wilson. 1998. "Correlation-Induction Techniques for Estimating Quantiles in Simulation Experiments." *Operations Research* 46:574-59.

Barton, R. R. 1992. "Metamodels for Simulation Input-Output Relations." In *Proceedings of the 1992 Winter Simulation Conference*, edited by J. J. Swain, D. Goldsman, R. C. Crain, and J. R. Wilson, 289-299. Piscataway, New Jersey: Institute of Electrical and Electronics Engineers, Inc.

Barton, R. R. 1994. "Metamodeling: A State of the Art Review." In *Proceedings of the 1994 Winter Simulation Conference*, edited by J. D. Tew, S. Manivannan, D. A. Sadowski, and A. F. Seila, 237-244. Piscataway, New Jersey: Institute of Electrical and Electronics Engineers, Inc.

Barton, R. R., B. L. Nelson, and W. Xie. 2013. "Quantifying Input Uncertainty via Simulation Confidence Intervals." *INFORMS Journal on Computing* 26:74–87.

Bechhofer, R. 1955. "A Three Decision Problem Concerning the Mean of a Normal Population." Presentation at the December 30, 1955 American Statistical Society Meeting, New York, NY.

Bettonvil, B., and J. P. C. Kleijnen. 1997. "Searching for Important Factors in Simulation Models with Many Factors: Sequential Bifurcation." *European Journal of Operational Research* 96:180–94.

Biles, W. E. 1979. "Experimental Design in Computer Simulation." In *Proceedings of the1979 Winter Simulation Conference*, edited by H. J. Highland, M. G. Spiegel, and R. Shannon, 3-9. Piscataway, New Jersey: Institute of Electrical and Electronics Engineers, Inc.

Blanchet, J., and H. Lam. 2012. "State-Dependent Importance Sampling for Rare-Event Simulation: An Overview and Recent Advances." *Surveys in Operations Research and Management Science* 17:38–59.

Booth, T. 2009. "Particle Transport Applications." Chapter 10, in *Rare Event Simulation using Monte Carlo Methods*, edited by G. Rubino and B. Tuffin, 215–242. Chichester, UK: Wiley.

Box, G. E. P., and K. B. Wilson. 1951. "On the Experimental Attainment of Optimum Conditions." *Journal of the Royal Statistical Society. Series B (Methodological)* 13:1–45.

Box, G. E. P. 1992. "George's Column." *Quality Engineering*, 5:321–330.

Bucklew, J. A. 2004. *Introduction to Rare Event Simulation*. Berlin: Springer.

Burdick, D. S., and T. H. Naylor. 1966. "Design of Computer Simulation Experiments for Industrial Systems." *Communications of the ACM* 9:329–339.

Burt J. M. Jr., D. P. Gaver, M. Perlas. 1970. "Simple Stochastic Networks: Some Problems and Procedures." *Naval Research Logistics* 17:439–445.

Carson, J. S., and A. M. Law. 1977. "Conservation Equations and Variance Reduction in Queueing Simulations." In *Proceedings of the 1977 Winter Simulation Conference*, edited by H. J. Highland, R. G. Sargent, and J. W. Schmidt, 186–189. Piscataway, New Jersey: Institute of Electrical and Electronics Engineers, Inc.

Chen, X., B. E. Ankenman, and B. L. Nelson. 2012. "The Effects of Common Random Numbers on Stochastic Kriging Metamodels." *ACM Transactions on Modeling and Computer Simulation* 22:7:1-7:20.

Cheng, R. C. H. 1976. "A Note on the Effect of Initial Conditions on a Simulation Run." *Operational Research Quarterly* 27:467–470.

Cheng, R. C. H. 1997. "Searching For Important Factors: Sequential Bifurcation Under Uncertainty." In *Proceedings of the 1997 Winter Simulation Conference*, edited by S. Andradottir, K. J. Healy, D. H. Withers, and B. L. Nelson, 275-280. Piscataway, New Jersey: Institute of Electrical and Electronics Engineers, Inc.

Cheng, R. C. H. 1985. "Generation of Multivariate Normal Samples with Given Sample Mean and Covariance Matrix." *Journal of Statistical Computation and Simulation* 21:39–49.

Cheng, R. C. H. 1986. "Variance Reduction Methods." In *Proceedings of the 1986 Winter Simulation Conference*, edited by J. R. Wilson, J. O. Henriksen, and S. D. Roberts, 60-68. Piscataway, New Jersey: Institute of Electrical and Electronics Engineers, Inc.

Cheng, R. C. H., J. P. C. Kleijnen, and V. B. Melas. 2000. "Optimal Design of Experiments with Simulation Models of Nearly Saturated Queues." *Journal of Statistical Planning and Inference* 85:19–26.

Cheng, R. C. H., and J. P. C. Kleijnen. 1999. "Improved Design of Queueing Simulation Experiments with Highly Heteroscedastic Responses." *Operations Research* 47:762–777.

Cochran, W. G. 1977. *Sampling Techniques* (3rd Ed). New York: John Wiley & Sons.

"Computer Simulation Archive Home | Computer Simulation Archive." 2017. https://d.lib.ncsu.edu/computer-simulation/.

Conway, A. E., and A. Goyal. 1987. "Monte Carlo Simulation of Computer System Availability/Reliability Models." In *Digest of Papers: 17th Annual International Symposium on Fault-Tolerant Computing (FTCS-17)*, edited by F. Cristian and J. Goldberg, 230–235. Washington, D.C.: IEEE Computer Society Press.

Conway, R. W. 1963. "Some Tactical Problems in Digital Simulation." *Management Science* 10: 47-61.

Conway, R. W., B. M. Johnson, and W. L. Maxwell. 1959. "Some Problems of Digital Systems Simulation." *Management Science* 6:92–110.

Corynen, G. C. 1975. "Experimental Inquiries: A Mathematical Approach." In *Proceedings of the 1975 Winter Simulation Conference*, 121-143. Piscataway, New Jersey: Institute of Electrical and Electronics Engineers, Inc.

Cottrell, M., J.-C. Fort, and G. Malgouyres. 1983. "Large Deviations and Rare Events in the Study of Stochastic Algorithms." *IEEE Transactions on Automatic Control* 28:907–920.

Crane, M. A., and D. L. Iglehart. 1974. "Simulating Stable Stochastic Systems, II: Markov Chains." *Journal of the ACM* 21:114–123.

Crane, M. A., and A. J. Lemoine. 1976. "A Tutorial on the Regenerative Approach and Discrete-Event Simulations." In *Proceedings of the 1976 Winter Simulation Conference*, edited by R. G. Sargent, H. J. Highland, and T. J. Schriber, 41-47. Piscataway, New Jersey: Institute of Electrical and Electronics Engineers, Inc.

Currin, C., T. Mitchell, M. Morris, and D. Ylvisaker. 1991. "Bayesian Prediction of Deterministic Functions, with Applications to the Design and Analysis of Computer Experiments." *Journal of the American Statistical Association* 86:953–963.

Dean, T., and P. Dupuis. 2009. "Splitting for Rare Event Simulation: A Large Deviation Approach to Design and Analysis." *Stochastic Processes and their Applications* 119: 562–587.

Dion, M., and P. L'Ecuyer. 2010. "American Option Pricing with Randomized Quasi-Monte Carlo Simulations." in the *Proceedings of the 2010 Winter Simulation Conference*, edited by B. Johansson, S. Jain, J. Montoya-Torres, J. Hugan, and E. Yücesan, 2705-2720. Piscataway, New Jersey: Institute of Electrical and Electronics Engineers, Inc.

Dong, H., and M. K. Nakayama. 2014. "Constructing Confidence Intervals for a Quantile Using Batching and Sectioning when Applying Latin Hypercube Sampling." In *Proceedings of the 2014 Winter Simulation Conference*, edited by A. Tolk, S. Y. Diallo, I. O. Ryzhov, L. Yilmaz, S. Buckley, and J. A. Miller, 640-651. Piscataway, New Jersey: Institute of Electrical and Electronics Engineers, Inc.

Dupuis, P., and H. Wang. 2004. "Importance Sampling, Large Deviations, And Differential Games." *Stochastics and Stochastic Reports* 76:481-508.

Dupuis, P., and H. Wang. 2007. "Subsolutions of an Isaacs Equation and Efficient Schemes of Importance Sampling." *Mathematics of Operations Research* 32:723–757.

Donohue, J. M., E. C. Houck, and R. H. Myers. 1993. "A Sequential Experimental Design Procedure for the Estimation of First- and Second-Order Simulation Metamodels." *ACM Transactions on Modeling and Computer Simulation* 3:190–224.

Dudewicz, E. J. 1976. "Statistics in Simulation: How to Design for Selecting the Best Alternative." In *Proceedings of the 1976 Winter Simulation Conference*, edited by R. G. Sargent, H. J. Highland, and T. J. Schriber, 67-71. Piscataway, New Jersey: Institute of Electrical and Electronics Engineers, Inc.

Econopapers.repec.org 2017. http://econpapers.repec.org/scripts/search.pf?kw= 764%20 variance%20 reduction%20techniques ;iframes=no

Ehrlichman, S. M. T., and S. G. Henderson. 2008. "Comparing Two Systems: Beyond Common Random Numbers." *Proceedings of the 2008 Winter Simulation Conference*, edited by S. J. Mason, R. R. Hill, L. Mönch, O. Rose, T. Jefferson, and J. W. Fowler, 245-251. Piscataway, New Jersey: Institute of Electrical and Electronics Engineers, Inc.

Fedorov, V. V. 1969. *Theory of Optimal Experiments* (in Russian), English translation 1970, New York, NY: Academic Press.

Fieller, E. C., and H. O. Hartley. 1954. "Sampling with Control Variables." *Biometrika* 41:494–501.

Fishman, G. S. 1967. "Problems in the Statistical Analysis of Simulation Experiments: The Comparison of Means and the Length of Sample Records." *Communications of the ACM* 10:94–99.

Fishman, G. S. 1968a. "Estimating Reliability in Simulation Experiments." In *Proceedings of the Second Conference on Applications of Simulation*, edited by A. Ockene and J. Reitman, 6–10. Piscataway, New Jersey: Institute of Electrical and Electronics Engineers, Inc.

Fishman, G. S. 1968b. "The Allocation of Computer Time in Comparing Simulation Experiments." *Operations Research* 16:280–295.

Fishman, G. S. 1971. "Estimating Sample Size in Computing Simulation Experiments." *Management Science* 18:21-38.

Fishman, G. S. 1973. *Concepts and Methods in Discrete Event Digital Simulation*. New York: John Wiley & Sons Inc.

Fishman, G. S. 1978. "Grouping Observations in Digital Simulation." *Management Science* 24:510–521.

Fishman, G. S., and B. D. Huang. 1980. "Antithetic Variates Revisited." Tech. Rep. 80-4, Operations Research and Systems Analysis, University of North Carolina, Chapel Hill, NC.

Fishman, G. S., and P. J. Kiviat. 1965. "Spectral Analysis of Time Series Generated by Simulation Models,." RM-4393-PR. RAND Corporation, Santa Monica, CA.

Fishman, G. S., and P. J. Kiviat. 1967. "The Analysis of Simulation-Generated Time Series." *Management Science* 13:525-557.

Frank, A. L. 1968. "The Use of Experimental Design Techniques in Simulation." In *Proceedings of the Second Conference on Applications of Simulation*, edited by A. Ockene and J. Reitman, 11-12. Piscataway, New Jersey: Institute of Electrical and Electronics Engineers, Inc.

Gaver, D. P. 1969. "Statistical Methods for Improving Simulation Efficiency." In *Proceedings of the Third Conference on the Applications of Simulation*, edited by P. J. Kiviat and A. Ockene, 38–46. Piscataway, New Jersey: Institute of Electrical and Electronics Engineers, Inc.

Glasserman, P., P. Heidelberger, and P. Shahabuddin. 2000. "Variance Reduction Techniques for Value-at-Risk with Heavy-Tailed Risk Factors." In *Proceedings of the 2000 Winter Simulation Conference*, edited by J. A. Joines, R. R. Barton, K. Kang, and P. A. Fishwick, 604-609. Piscataway, New Jersey: Institute of Electrical and Electronics Engineers, Inc.

Glasserman, P., P. Heidelberger, P. Shahabuddin, and T. Zajic. 1998. "A Large Deviations Perspective on the Efficiency of Multilevel Splitting." *IEEE Transactions on Automatic Control* AC-43:1666–1679.

Glasserman, P., P. Heidelberger, P. Shahabuddin, and T. Zajic. 1999. "Multilevel Splitting for Estimating Rare Event Probabilities." *Operations Research* 47:585–600.

Glasserman, P., and S. Kou. 1995. "Analysis of an Importance Sampling Estimator for Tandem Queues." *ACM Transactions on Modeling and Computer Simulation* 5:22–42.

Glasserman, P., and Y. Wang. 1997. "Counterexamples in Importance Sampling for Large Deviations Probabilities." *Annals of Applied Probability* 7:731–746.

Glynn, P. W., and D. L. Iglehart. 1989. "Importance Sampling For Stochastic Systems." *Management Science* 35:1367–1393.

Glynn, P. W., and W. Whitt. 1989. "Indirect Estimation Via L = λW." *Operations Research* 37:82–103.

Glynn, P. W., and W. Whitt. 1992. "The Asymptotic Efficiency of Simulation Estimators." *Operations Research* 40: 505-520.

Goertzel, G. 1949a. "A Proposed Particle Attenuation Problem." Technical Report ORNL-423, December 1949. Oak Ridge National Laboratory, Oak Ridge, TN.

Goertzel, G. 1949b. "Quota Sampling and Importance Functions in Stochastic Solution of Particle Problems." Technical Report ORNL-434, June, 1949. Oak Ridge National Laboratory, Oak Ridge, TN.

Goertzel, G., and H. Kahn. 1949. "Monte Carlo Methods for Shield Computation." Technical Report ORNL-429, December 1949. Oak Ridge National Laboratory, Oak Ridge, TN.

Goldsman, D., B. L. Nelson, and B. Schmeiser. 1991. "Methods for Selecting the Best System." In *Proceedings of the1991 Winter Simulation Conference*, edited by B. L. Nelson, W. D. Kelton, and G. M. Clark, 177–186. Piscataway, New Jersey: Institute of Electrical and Electronics Engineers, Inc.

Goyal, A., P. Heidelberger, and P. Shahabuddin. 1987. "Measure Specific Dynamic Importance Sampling for Availability Simulations." In *Proceedings of the 1987 Winter Simulation Conference*, edited by A. Thesen, H. Grant, and W. D. Kelton, 351–357. Piscataway, New Jersey: Institute of Electrical and Electronics Engineers, Inc.

Goyal, A., W. C. Carter, E. De Souza e Silva, S. S. Lavenberg, and K. S. Trivedi. 1986. "The System Availability Estimator." In *Digest of Papers: 16th Annual International Symposium on Fault-Tolerant Computing (FTCS-16)*, 84–89. Washington, D.C.: IEEE Computer Society Press.

Goyal, A., P. Shahabuddin, P. Heidelberger, V. Nicola, and P. W. Glynn. 1992. "A Unified Framework for Simulating Markovian Models of Highly Dependable Systems." *IEEE Transactions on Computers* C-41:36–51.

Gross, A. M. 1973. "A Monte Carlo Swindle for Estimators of Location." *Journal of the Royal Statistical Society Series C* 22:347–353.

Hammersley, J. M., and D. C. Handscomb. 1964. *Monte Carlo Methods*. London: Methuen.

Hammersley, J. M., and K. W. Morton. 1956. "A New Monte Carlo Technique: Antithetic Variates." *Mathematical Proceedings of the Cambridge Philosophical Society* 52(3):449–475.

Harling, J. 1958. "Simulation Techniques in Operational Research." *OR* 9:9–21.

Hauser, N., N. N. Barish, and S. Ehrenfeld. 1966. "Design Problems in a Process Control Simulation." *Journal of Industrial Engineering* 17:79–86.

Heidelberger, P. 1995. "Fast Simulation of Rare Events in Queueing and Reliability Models." *ACM Transactions on Modeling and Computer Simulation* 5:43–85.

Heidelberger, P., and P. D. Welch. 1981. "A Spectral Method for Confidence Interval Generation and Run Length Control in Simulations." *Communications of the ACM* 24:233–245.

Hordijk, A., D. L. Iglehart, and R. Schassberger. 1976. "Discrete-Time Methods for Simulating Continuous-Time Markov Chains." *Advances in Applied Probability* 8:772–788.

IAS Computer. 1953. "P-377_The_History_And_Development_Of_The_IAS_Computer_Mar53.pdf." http://bitsavers.trailing-edge.com/pdf/rand/

Jacobson, S. H. 1989. "Oscillation Amplitude Considerations in Frequency Domain Experiments." In *Proceedings of the 1989 Winter Simulation Conference*, edited by E. A. MacNair, K. J. Musselman, and P. Heidelberger, 406–410. Piscataway, New Jersey: Institute of Electrical and Electronics Engineers, Inc.

Jacobson, S. H., A. H. Buss, and L. W. Schruben. 1991. "Driving Frequency Selection for Frequency Domain Simulation Experiments." *Operations Research* 39:917–924.

Jacoby, J. E., and S. Harrison. 1962. "Multi-Variable Experimentation and Simulation Models." *Naval Research Logistics Quarterly* 9:121–136.

Jenkins, G. M. 1954. "Tests of Hypotheses in the Linear Autoregressive Model: 1. Null Hypothesis Distributions in the Yule Scheme." *Biometrika* 41:405–419.

Juneja, S., and P. Shahabuddin. 2002. "Simulating Heavy Tailed Processes Using Delayed Hazard Rate Twisting. *ACM Transactions on Modeling and Computer Simulation* 12(2):94–118.

Juneja, S., and P. Shahabuddin. 2006. "Rare Event Simulation Techniques: An Introduction and Recent Advances". In *Elsevier Handbooks in Operations Research and Management Science: Simulation*, edited by S. G. Henderson and B. L. Nelson, 291–350. Amsterdam: Elsevier.

Kabak, I. W. 1968. "Stopping Rules for Queuing Simulations." *Operations Research* 16:431–437.

Kafadar, K. 2006. "Monte Carlo Swindle." In *Encyclopedia of Statistical Sciences*. New York: John Wiley and Sons.

Kahn, H. 1950. "Modifications of the Monte Carlo Method". In *Proceedings, Seminar on Scientific Computation, November 1949*, edited by I. A. S. Department, 20–27. New York, NY: IBM.

Kahn, H. 1954. *Applications of Monte Carlo*. RM-1237-AEC.RAND Corporation, Santa Monica, CA. 19 April 1954. Revised 27 April 1956. Available at https://www.osti.gov/scitech/biblio/4353680.

Kahn, H., and T. E. Harris. 1949. "Estimation of Particle Transmission by Random Sampling." In *Monte Carlo Method*, Volume 12 of *Applied Mathematics Series*, 27–30: National Bureau of Standards.

Kahn, H. and A. Marshall. 1953. "Methods of Reducing Sample Size in Monte Carlo Computations." *Journal of the Operations Research Society of America* 1:263–278.

Kelton, W. D. 1999. "Designing Simulation Experiments." In *Proceedings of the 1999 Winter Simulation Conference*, edited by P. A. Farrington, H. B. Nembhard, D. T. Sturrock, and G. W. Evans, 33-38. Piscataway, New Jersey: Institute of Electrical and Electronics Engineers, Inc.

Kim, S., and S. G. Henderson. 2004. "Adaptive Control Variates." In *Proceedings of the 2004 Winter Simulation Conference*, edited by R. G. Ingalls, M. D. Rossetti, J. S. Smith, and B. A. Peters, 621-629. Piscataway, New Jersey: Institute of Electrical and Electronics Engineers, Inc.

Kleijnen, J. P. C. 1974. *Statistical Techniques in Simulation*. Vol. 9, Statistics, Textbooks and Monographs. New York: Marcel Dekker Inc.

Kleijnen, J. P. C. 1975. "Screening Designs for Poly-Factor Experimentation." *Technometrics* 17 (4): 487–493.

Kleijnen, J. P. C. 1987. *Statistical Tools for Simulation Practitioners*. New York: Marcel Dekker Inc.

Kleijnen, J. P. C and W. C. M. van Beers. 2004. "Application-Driven Sequential Designs for Simulation Experiments: Kriging Metamodelling." *Journal of the Operational Research Society* 55:876–883.

Kleijnen, J. P. C., S. M. Sanchez, T. W. Lucas, and T. M. Cioppa. 2005. "State-of-the-Art Review: A User's Guide to the Brave New World of Designing Simulation Experiments." *INFORMS Journal on Computing* 17:263-289.

Kleijnen, J. P. C., and R. G. Sargent. 2000. "A Methodology for Fitting and Validating Metamodels in Simulation." *European Journal of Operational Research* 120:14–29.

Krige, D. G. 1951. "A Statistical Approach to Some Basic Mine Valuation Problems on the Witwatersrand." *Journal of the Chemical, Metallurgical and Mining Society of South Africa* 52:119–139.

L'Ecuyer, P. 1994. "Efficiency Improvement and Variance Reduction." In *Proceedings of the 1994 Winter Simulation Conference*, edited J. D. Tew, S. Manivannan, D. A. Sadowski, and A. F. Seila, 122–132. Piscataway, New Jersey: Institute of Electrical and Electronics Engineers, Inc.

Lamb, J. D., and R. C. H. Cheng. 2002. "Optimal Allocation of Runs in a Simulation Metamodel with Several Independent Variables." *Operations Research Letters* 30:189–194.

Lavenberg, S. S., and P. D. Welch. 1978. "Variance Reduction Techniques." In *Proceedings of the 1978 Winter Simulation Conference*, edited by H. J. Highland, L. G. Hull, and N. R. Nielsen, 167-170.Piscataway, New Jersey: Institute of Electrical and Electronics Engineers, Inc.

Law, A. M., and J. S. Carson. 1979. "A Sequential Procedure for Determining the Length of a Steady-State Simulation." *Operations Research* 27:1011–1025.

Lemieux, C., and P. L'Ecuyer 2000. "Using Lattice Rules for Variance Reduction in Simulation." In *Proceedings of the 2000 Winter Simulation Conference*, edited by J. A. Joines, R. R. Barton, P. A. Fishwick, and K. Kang, 509-516. Piscataway, New Jersey: Institute of Electrical and Electronics Engineers, Inc.

Lewis, E. E., and F. Bohm. 1984. "Monte Carlo Simulation of Markov Unreliability Models." *Nuclear Engineering and Design* 77:49–62.

Lux, I., and L. Koblinger. 1991. *Monte Carlo Particle Transport Methods: Neutron and Photon Calculations*. Boca Raton, FL: CRC Press.

McKay, M. D., R. J. Beckman, and W. J. Conover. 1979. "A Comparison of Three Methods for Selecting Values of Input Variables in the Analysis of Output from a Computer Code." *Technometrics* 21:239–245.

Mihram, G. A. 1972. *Simulation: Statistical Foundations and Methodology*. New York: Academic Press.

Mitchell, T. J., and M. D. Morris. 1992. "The Spatial Correlation Function Approach to Response Surface Estimation." In *Proceedings of the 1992 Winter Simulation Conference*, edited J. J. Swain, D. Goldsman, R. C. Crain, and J. R. Wilson, 565–571. Piscataway, New Jersey: Institute of Electrical and Electronics Engineers, Inc.

Montgomery, D. C., and G. Weatherby. 1979. "Factor Screening Methods in Computer Simulation Experiments." In *Proceedings of the 1979 Winter Simulation Conference*, edited by H. J. Highland, M. G. Spiegel, and R. Shannon, 347–358. Piscataway, New Jersey: Institute of Electrical and Electronics Engineers, Inc.

Morrice, D. J. 1995. "A Comparison of Frequency Domain Methodology and Conventional Factor Screening Methods." *Operations Research Letters* 17:165–174.

Morrice, D. J., and L. W. Schruben. 1993. "Simulation Factor Screening Using Harmonic Analysis." *Management Science* 39:1459–1476.

Nance, R. E., and R. G. Sargent 2002. "Perspectives on the Evolution of Simulation." *Operations Research* (50th Anniversary Issue) 50:161-172.

Nelson, B. L. 1985. "A Decomposition Approach to Variance Reduction." In *Proceedings of the 1985 Winter Simulation Conference*, edited by D. T. Gantz, G. C. Blais, and S. L. Solomon, 23-32. Piscataway, New Jersey: Institute of Electrical and Electronics Engineers, Inc.

Nelson, B. L. 1987. "A Perspective on Variance Reduction in Dynamic Simulation Experiments." *Communications in Statistics: Simulation and Computation* 16:385-426.

Nelson, B. L. 1989. "Batch Size Effects on the Efficiency of Control Variates in Simulation." *European Journal of Operational Research* 43:184–196.

Nelson, B. L. 1992. "Designing Efficient Simulation Experiments." In *Proceedings of the 1992 Winter Simulation Conference*, edited J. J. Swain, D. Goldsman, R. C. Crain, and J. R. Wilson, 126-132. Piscataway, New Jersey: Institute of Electrical and Electronics Engineers, Inc.

Nelson, B. L. and B. Schmeiser. 1983. "Variance Reduction: Basic Transformations." In *Proceedings of the 1983 Winter Simulation Conference*, edited by S. D. Roberts, J. Banks, and B. W. Schmeiser, 225-258. Piscataway, New Jersey: Institute of Electrical and Electronics Engineers, Inc.

Nelson, B. L., and B. W. Schmeiser. 1985. "Decomposition of Some Well-Known Variance Reduction Techniques. Revision." RM-84-6. School of Industrial Engineering, Purdue University.

Nicola, V. F., P. Shahabuddin, and M. K. Nakayama. 2001. "Techniques for Fast Simulation of Models of Highly Dependable Systems." *IEEE Transactions on Reliability* 50:246–264.

Parekh, S., and J. Walrand. 1989. "Quick Simulation of Rare Events in Networks." *IEEE Transactions on Automatic Control* 34:54–66.

Nozari, A., S. F. Arnold, and C. D. Pegden. 1987. "Statistical Analysis for Use with the Schruben and Margolin Correlation Induction Strategy." *Operations Research* 35:127–139.

Owen, A. B. 2017. "Monte Carlo Theory, Methods and Examples." http://statweb.stanford.edu/~owen/mc.

Page, E. S. 1965. "On Monte Carlo Methods in Congestion Problems, Simulation of Queueing Systems." *Operations Research* 13:300-305.

Princeton. 2017. https://paw.princeton.edu/sites/default/files/styles/feature_image/public/images/content/Turing_NeumannMANIAC.jpg?itok=Z7sThQ_i

Roeder, T., S. Fischbein, M. Janakiram, and L. W. Schruben. 2002. "Resource-Driven and Job-Driven Simulations." *Proceedings of the 2002 International Conference on Modeling and Analysis of Semiconductor Manufacturing*, edited by G. T. Mackulak, A. Schomig, J. Fowler, and D.-B. Perng, 78-83. Arizona State University.

Rubino, G., and B. Tuffin. 2009. *Rare Event Simulation using Monte Carlo Methods*. Chichester, UK: John Wiley.

Sadowsky, J. S. 1991. "Large Deviations Theory and Efficient Simulation of Excessive Backlogs in a GI/GI/m Queue." *IEEE Transactions on Automatic Control* 36:1383–1394.

Sadowsky, J., and J. Bucklew. 1990. "On Large Deviations Theory and Asymptotically Efficient Monte Carlo Estimation." *IEEE Transactions on Information Theory* 36:579–588.

Sacks, J., W. J. Welch, T. J. Mitchell, and H. P. Wynn. 1989. "Design and Analysis of Computer Experiments." *Statistical Science* 4:409–423.

Sanchez, S. M., F. Moeeni, and P. J. Sanchez. 2006. "So Many Factors, So Little Time…Simulation Experiments in the Frequency Domain." *International Journal of Production Economics* 103:149–165.

Sanchez, S. M., H. Wan, and T. W. Lucas. 2009. "Two-Phase Screening Procedure for Simulation Experiments." *ACM Transactions on Modeling and Computer Simulation* 19:1–24.

Schmeiser, B. 1982. "Batch Size Effects in the Analysis of Simulation Output." *Operations Research* 30:556–568.

Schriber, T. J., and R. W. Andrews. 1984. "ARMA-Based Confidence Intervals for Simulation Output Analysis." *American Journal of Mathematical and Management Sciences* 4:345–373.

Schruben, L. W. 1979. "Designing Correlation Induction Strategies for Simulation Experiments." Chapter 16 in *Current Issues in Computer Simulation*, edited by N. R. Adam and A. Dogramaci. Amsterdam: Elsevier.

Schruben, L. W. 1983. "Confidence Interval Estimation Using Standardized Time Series." *Operations Research* 31:1090–1108.

Schruben, L. W. 2010. "Simulation Modeling for Analysis." ACM *Transactions on Modeling and Computer Simulation* 20:1-22.

Schruben, L. W., and V. J. Cogliano. 1981. "Simulation Sensitivity Analysis: A Frequency Domain Approach." In *Proceedings of the 1981 Winter Simulation Conference,* edited by T. Oren, C. M. Delfosse, and C. M. Shub, 455–459. Piscataway, New Jersey: Institute of Electrical and Electronics Engineers, Inc.

Schruben, L. W., and V. J. Cogliano. 1987. "An Experimental Procedure for Simulation Response Surface Model Identification." *Communications of the ACM* 30:716–730.

Schruben, L. W., and B. H. Margolin. 1978. "Pseudorandom Number Assignment in Statistically Designed Simulation and Distribution Sampling Experiments." *Journal of the American Statistical Association* 73:504–520.

Shahabuddin, P. 1994. "Importance Sampling for Highly Reliable Markovian Systems." *Management Science* 40:333–352.

Shahabuddin, P., V. F. Nicola, P. Heidelberger, A. Goyal, and P. W. Glynn. 1988. "Variance Reduction in Mean Time to Failure Simulations." In *Proceedings of the 1988 Winter Simulation Conference*, edited

by M. A. Abrams, P. L. Haigh, and J. C. Comfort, 491–499. Piscataway, New Jersey: Institute of Electrical and Electronics Engineers, Inc.

Siegmund, D. 1976. "Importance Sampling in the Monte Carlo Study of Sequential Tests." *Annals of Statistics* 4:673–684.

Sobel, M., and A. Wald. 1949. "A Sequential Decision Procedure for Choosing One of Three Hypotheses Concerning the Unknown Mean of a Normal Distribution." *The Annals of Mathematical Statistics* 20: 502–522.

Song, W. T., and M. Chih. 2008. "Implementable MSE-Optimal Dynamic Partial-Overlapping Batch Means Estimators for Steady-State Simulations." In *Proceedings of the 2008 Winter Simulation Conference*, edited by S. Mason, R. Hill, L. Moench, O. Rose, T. Jefferson, and J. Fowler, 426–435. Piscataway, New Jersey: Institute of Electrical and Electronics Engineers, Inc.

Szechtman, R., and P. W. Glynn. 2002. "Some New Perspectives on the Method of Control Variates." In *Monte Carlo and Quasi-Monte Carlo Methods,* edited by K. T. Fang, F. J. Hickernell, and H. Niederreiter, 27-49. Berlin: Springer-Verlag.

Tafazzoli, A., N. M. Steiger, and J. R. Wilson. 2011. "N-Skart: A Nonsequential Skewness- and Autoregression-Adjusted Batch-Means Procedure for Simulation Analysis." *IEEE Transactions on Automatic Control* 56:254–264.

Tafazzoli, A., J. R. Wilson, E. K. Lada, and N. M. Steiger. 2008. "Skart: A Skewness- and Autoregression-Adjusted Batch-Means Procedure for Simulation Analysis." In *Proceedings of the 2008 Winter Simulation Conference*, edited by S. Mason, R. Hill, L. Moench, O. Rose, T. Jefferson, and J. Fowler, 387-395. Piscataway, New Jersey: Institute of Electrical and Electronics Engineers, Inc.

Tew, J. D., and J. R. Wilson. 1992. "Validation of Simulation Analysis Methods for the Schruben-Margolin Correlation-Induction Strategy." *Operations Research* 40:87–103.

Tew, J. D. and J. R. Wilson. 1994. "Estimating Simulation Meta-Models Using Combined Correlation-Based Variance Reduction Techniques." *IIE Transactions*26:2–16.

Villén-Altamirano, M., and J. Villén-Altamirano. 1991. "RESTART: A Method For Accelerating Rare Event Simulations." In *Queueing, Performance and Control in ATM*, edited by J. W. Cohen and C. D. Pack. New York: Elsevier Science.

Villén-Altamirano,M., and J. Villén-Altamirano. 1994. "RESTART:A Straightforward Method for Fast Simulation of Rare Events." In *Proceedings of the 1994 Winter Simulation Conference*, edited J. D. Tew, S. Manivannan, D. A. Sadowski, and A. F. Seila, 282-289. Piscataway, New Jersey: Institute of Electrical and Electronics Engineers, Inc.

Wan, H., B. E. Ankenman, and B. L. Nelson. 2006. "Controlled Sequential Bifurcation: A New Factor-Screening Method for Discrete-Event Simulation." *Operations Research* 54:743–755.

Whitt, W. 1991. "The Efficiency of One Long Run versus Independent Replications in Steady-State Simulation." *Management Science* 37:645–666.

Wiener, N. 1949. *Extrapolation, Interpolation, and Smoothing of Stationary Time Series: With Engineering Applications*. Cambridge, MA: Technology Press of the Massachusetts Institute of Technology.

Wilson, J. R. 1984. "Variance Reduction Techniques for Digital Simulation." *American Journal of Mathematical and Management Sciences* 4: 277–312.

AUTHOR BIOGRAPHIES

RUSSELL R BARTON is a distinguished professor of supply chain and information systems and of industrial engineering at the Pennsylvania State University. He currently serves as senior associate dean for research and faculty in the Smeal College of Business. He received a B.S. degree in electrical engineering from Princeton University and M.S. and Ph.D. degrees in operations research from Cornell University. He serves as vice president of sections and societies on the INFORMS Board of Directors. He is a fellow of IISE, a senior member of IEEE and a Certified Analytics Professional®. His research interests include

applications of statistical and simulation methods to system design and to product design, manufacturing and delivery. His email address is rbarton@psu.edu.

MARVIN K. NAKAYAMA is a professor in the Department of Computer Science at the New Jersey Institute of Technology. He received his Ph.D. in operations research from Stanford University and a B.A. in mathematics-computer science from U.C. San Diego. He won second prize in the 1992 George E. Nicholson Student Paper Competition sponsored by INFORMS and is a recipient of a CAREER Award from the National Science Foundation. A paper he co-authored received the Best Theoretical Paper Award for the 2014 Winter Simulation Conference. He served as the simulation area editor for the *INFORMS Journal on Computing* from 2007–2016, and is an associate editor for *ACM Transactions on Modeling and Computer Simulation*. His research interests include simulation, modeling, statistics, risk analysis, and energy. His email address is marvin@njit.edu.

LEE SCHRUBEN is a Chancellor's Professor in the Industrial Engineering and Operations Research faculty at the University of California, Berkeley. Before moving to Berkeley he was on the Operations Research and Information (née Industrial) Engineering faculty at Cornell for 23 years, where he held the Andrew Schultz Professorship of Industrial Engineering. Prior to that, he was on the faculty in the Colleges of Medicine and Engineering at the University of Florida, Gainesville. He has a B.S. degree from Cornell Engineering, an M.S. in statistics from the University of North Carolina, and a Ph.D. from Yale. His email address is LeeS@berkeley.edu.