SANDIA REPORT

SAND2016-4679
Unlimited Release
Printed May 2016

A Control Variate Method for Probabilistic Performance Assessment: Improved Estimates for Mean Performance Quantities of Interest

Robert J. MacKinnon and Kristopher L. Kuhlman

Prepared by
Sandia National Laboratories
Albuquerque, New Mexico 87185-0747

Sandia National Laboratories is a multi-program laboratory managed and operated by Sandia Corporation, a wholly owned subsidiary of Lockheed Martin Corporation, for the U.S. Department of Energy's National Nuclear Security Administration under contract DE-AC04-94AL85000.

Sandia National Laboratories

Available to DOE and DOE contractors from
U.S. Department of Energy
Office of Scientific and Technical Information
P.O. Box 62
Oak Ridge, TN 37831

Telephone: (865) 576-8401
Facsimile: (865) 576-5728
E-Mail: reports@osti.gov
Online ordering: http://www.osti.gov/scitech

Available to the public from
U.S. Department of Commerce
National Technical Information Service
5301 Shawnee Rd
Alexandria, VA 22312

Telephone: (800) 553-6847
Facsimile: (703) 605-6900
E-Mail: orders@ntis.gov
Online order: http://www.ntis.gov/search

SAND2016-4679
Unlimited Release
Printed May 2016

A Control Variate Method for Probabilistic Performance Assessment: Improved Estimates for Mean Performance Quantities of Interest

Robert J. MacKinnon and Kristopher L. Kuhlman
Applied Systems Analysis & Research Department
Sandia National Laboratories
P.O. Box 5800, MS-0747
Albuquerque, New Mexico 87185-0747

ABSTRACT

We present a method of control variates for calculating improved estimates for mean performance quantities of interest, $E(PQI)$, computed from Monte Carlo probabilistic simulations. An example of a PQI is the concentration of a contaminant at a particular location in a problem domain computed from simulations of transport in porous media. To simplify the presentation, the method is described in the setting of a one-dimensional elliptical model problem involving a single uncertain parameter represented by a probability distribution. The approach can be easily implemented for more complex problems involving multiple uncertain parameters and in particular for application to probabilistic performance assessment of deep geologic nuclear waste repository systems. Numerical results indicate the method can produce estimates of $E(PQI)$ having superior accuracy on coarser meshes and reduce the required number of simulations needed to achieve an acceptable estimate.

ACKNOWLEDGMENTS

We would like to thank Daniel J. Clayton for his technical review of this report and his constructive recommendations for its improvement.

CONTENTS

FIGURES

TABLES

NOMENCLATURE

Cov	covariance
DGR	deep geologic (nuclear waste) repository
PA	performance assessment
PQI	performance quantity of interest
SNL	Sandia National Laboratories
Var	variance

1. INTRODUCTION

Obtaining reliable performance predictions of natural and engineered systems is an important endeavor in many disciplines. Natural and engineered systems include deep geologic nuclear waste repositories, hydrocarbon reservoirs, nuclear power reactors, carbon sequestration operations, and the earth's climate. In many natural systems, particularly deep geologic repository (DGR) systems for nuclear waste disposal, significant epistemic and aleatory uncertainties make reliable predictions challenging because of the difficulty in gathering required data, understanding subsurface processes and predicting future events. Data uncertainties are significant and arise because DGR systems can be km-scale and heterogeneous. Data are practicably accessible only at a relatively limited set of characterization borehole locations, at the repository or underground research facility if one is present, or derived from indirect methods such as geophysical methods.

There are different computational approaches in practice today for making complex system performance predictions in the presence of uncertainty. These approaches include perturbation-, stochastic differential equation-, and Monte Carlo-based methods. These methods have advantages and disadvantages depending on the structure of the problem being analyzed and the nature of its uncertainties. DGR systems are more amenable to Monte Carlo-based methods because these systems involve multiple coupled thermal, hydrological, mechanical, and chemical processes and numerous uncertainties. As examples, the 2008 Performance Assessment (PA) of the proposed Yucca Mountain repository included 9 primary coupled computational models and approximately 400 uncertain parameters [1]. The 1996 performance assessment of the Waste Isolation Pilot Plant included 9 primary computational models and approximately 60 uncertain parameters [2].

PA is the required regulatory approach for assessing DGR compliance with quantitative radiological safety criteria. Because PA simulations are used as a basis for making regulatory decisions regarding repository safety, it is important that the overall error and uncertainty in performance predictions be understood. Overall error and uncertainty arise from the following major modeling activities [1, 3]:

1. Selection of the mathematical models providing an abstraction of the physical processes and events of interest;
2. Identification of appropriate parameters and data defining the models;
3. Use of physical observations and measurements, including data from the literature, laboratory, and field to validate and calibrate the models;
4. Development of a computational model through discretization of the mathematical model and its implementation on a computer;
5. Identification of specific goals of PA simulations and the performance quantities of interest; and
6. Quantification of uncertainties in the predictions, including sensitivity analysis.

The first three activities are closely linked with model validation: "The process of determining if a mathematical model of a physical event represents the actual physical event with sufficient accuracy" [4]. The fourth activity is evaluated as an element of model verification: "The process of determining if a computational model obtained by discretizing a mathematical model of a physical event represents the mathematical model with sufficient accuracy" [4]. Activities 5 and 6 are linked with both verification and validation. We emphasize that all six activities are in general iterative and evolve in complexity as knowledge about a specific DGR site increases and the computational modeling capabilities mature.

We undertake the work presented herein within the context of these six activities. In particular, this work includes elements of activities 4, 5, and 6 and is an initial step towards formulating the assessment of numerical discretization error on system performance quantities of interest. This assessment will include both the error in a performance quantity of interest (PQI) due to errors in approximations to the governing equations (e.g., spatial and temporal discretization errors), and the effect of approximating uncertainty in input parameters (statistical error). In DGR systems, typical PQIs include, as examples, the time-dependent concentration of a radionuclide at a location in the host geosphere away from the DGR, and the peak temperature at the wall of a waste emplacement tunnel.

Prior theoretical research exists for simple systems demonstrating how to compute the individual contributions of spatial and temporal truncation error and Monte Carlo truncation error to the total error in predictive probabilistic simulations [5, 6]. A goal of our research is to extend and tailor this work for large-scale DGR simulations. As a first step we address the approximation of the expectation of a PQI, $E(PQI)$, from a Monte Carlo simulation and develop a control variate approach [6, 7] to reduce its variance, improve its accuracy, and reduce the required number of simulations needed to achieve an acceptable estimate. Our choice of the control variate approach is based on two factors.

First, based on our past experience with DGR simulations, PQIs are strongly correlated to the random variable inputs to the simulations. Past sensitivity analyses bear this out but also indicate that a PQI is typically sensitive to a subset of the uncertain inputs. In the Yucca Mountain simulations noted previously it was found that PQIs were typically only strongly correlated to a few (< 5) of the ~ 400 uncertain parameters [8]. The control variate technique provides a simple variance reduction method once a sensitivity analysis has been completed and we have identified the uncertain variables highly correlated with the PQI. In this study we examine a simple case involving a one-dimensional elliptic differential equation with one random input variable and provide example results below. Importantly, the method straightforwardly extends to multiple random variables.

Second, we desired a variance reduction method that is a post-processing method and therefore non-intrusive to the probabilistic simulations. Multi-level Monte Carlo methods are post-processing methods that have been developed in recent years for probabilistic simulations and uncertainty quantification [9, 10, and references therein]. These methods have been effectively applied to elliptic transport problems and require at least two levels of spatial resolution. Unlike the control variate method presented herein, these methods do not directly exploit the known characteristics of the uncertain random parameters.

This report is organized as follows. We first present a simple two-dimensional elliptic model problem that we use to describe the different numerical approximation errors and implementation of the control variate technique. We then present analytical and numerical results for a one-dimensional problem. Two probability density functions are considered for a single random variable, a uniform distribution and a triangular distribution. Two appendices are included that present an example problem with an exact variance reduction and numerical results.

1.1. Model Problem

Consider the following model problem in domain Ω, a d-dimensional domain in R^d, with boundary $\partial\Omega$

$$Lu(x) = f(x) \quad x \in \Omega \tag{1a}$$

$$u(x) = u_0 \quad x \in \Omega \tag{1b}$$

$$u(x) = \Gamma \quad \partial\Omega, \tag{1c}$$

where the solution u is a scalar quantity that is a function of spatial position x and f is a forcing function for u. The operator L has the form

$$L = \nabla \cdot [\beta(x)\nabla u] - \nabla \cdot [v(x)u]. \tag{1d}$$

L is a linear differential spatial operator that contains imprecisely known (or uncertain) model parameters β and v, where β and v are assumed to be independent and a scalar and diagonal tensor, respectively. It is assumed $\beta(x)$ is strictly positive with $0 < \beta(x) < \beta_{max} \; \forall \; x \in \Omega$, where β_{max} is an upper bound.

The numerical solution to problem (1) can be written as

$$U = g(\omega), \tag{2}$$

where we have dropped the variable x to simplify the notation, g represents the numerical model implemented to solve (1) and ω is used to denote model parameter fields $\omega = [\beta(x), v(x)]$. In this study, we use a Monte-Carlo sampling procedure to evaluate the uncertainty in U resulting from the uncertainty in model parameters ω and to obtain the expected value and variance of U. The uncertainty in each of the elements $\omega = [\beta, v]$ is characterized by distributions D_β and D_v. Model predictions $U = [U_1(\omega_1),\ldots,U_{NR}(\omega_{N_R})]$ are calculated for each Monte Carlo input sample $\omega = [\omega_1, \ldots \omega_{N_R}]$ consistent with D_β and D_v, where N_R is the number of sample input vectors and hence the number of realizations of model predictions. The numerical solution U in general contains error contributions from both spatial discretization and temporal numerical approximations, and from the statistical Monte Carlo approximation.

The expected value of $u(\omega)$ is given by

$$E[u(\omega)] = \int_S u(\omega)m(\omega)\, d\omega, \tag{3}$$

where $m(\omega)$ represents the joint density function for $u(\omega)$ resulting from parameter input distributions D_β and D_v. The exact solution u in Equation (3) can be expressed in terms of the numerical solution U and associated spatial numerical error by introducing

$$u(\omega) = U(\omega) + e_h(\omega), \tag{4}$$

hence

$$E[u(\omega)] = \int_S [U(\omega) + e_h(\omega)] m(\omega)\, d\omega \tag{5}$$

Approximating the integral in (3) using a Monte Carlo method with N_R independent realizations, the expected value can then be written as

$$\begin{aligned} E[u(\omega)] &= \frac{1}{N_R}\sum_{j=1}^{N_R} U(\omega) + e_{MC}[U(\omega)] + E[e_h(\omega)] \\ &= \bar{E}[U(\omega)] + e_{MC}[U(\omega)] + E[e_h(\omega)], \end{aligned} \tag{6}$$

where e_{MC} is the error in the Monte Carlo approximation to the expectation of $U(\omega)$ in (5), referred to herein as statistical error, and the last term is the spatial discretization error.

The error in the expectation $E[u(\omega)]$ is

$$\varepsilon = E[u(\omega)] - \bar{E}[U(\omega)] = e_{MC}(U) + E[e_h(\omega)]. \tag{7}$$

In this study we consider a simpler version of the model problem (1) to examine the characteristics of the control variate method and the statistical error term $e_{MC}(U)$ in (7). This simpler version will include a limited set of cases where Ω is one dimensional, β is a deterministic constant independent of x, and v is an epistemic uncertain parameter independent of x. In a follow-on study we will examine both error terms in (7) in one- and two-dimensions.

The statistical error for the mean of the approximate solution U in (7) is bounded by

$$|e_{MC}(S)| \le e^*_{MC} = z_{1-\frac{\alpha}{2}} \frac{|S|}{\sqrt{N_R}}, \tag{8a}$$

where *S* is the sample standard deviation

$$S = \sqrt{\frac{\sum_{j=1}^{N_R} (U_j - \bar{U})^2}{N_R - 1}}. \tag{8b}$$

In (8a) $z_{1-\alpha/2}$ is the critical value at $N_R - 1$ degrees of freedom corresponding to a Student's *t* distribution upper-tail area of $\alpha/2$. For example, for $N_R = 10$ and confidence (or probability) levels of 50%, 95%, and 99% critical values are $z_{0.25} = 0.703$, $z_{0.025} = 2.262$, and $z_{0.005} = 3.25$, respectively. For comparison, the corresponding values for $N_R \to \infty$ are $z_{0.25} = 0.674$, $z_{0.025} = 1.96$, and $z_{0.005} = 2.576$.

The error in (8a) can be reduced either by reducing the variance of U, i.e., S^2, by increasing the number of realization N_R, or by a combination of the two. The control variate method presented herein reduces the variance of U for a given N_R.

1.1.1. Performance Quantities of Interest

As noted in the introduction we are typically interested in evaluating a quantity of interest (PQI) at a specific location in the problem domain.

The standard estimate for the mean of the output function $PQI = PQI(U)$ is given by

$$\bar{PQI} = \frac{1}{N_R}\sum_{i=1}^{N_R} PQI_i = \frac{1}{N_R}\sum_{i=1}^{N_R} PQI(U_i). \tag{9a}$$

We use the control variate reduction technique to reduce the error associated with (9a) by instead using the control variate estimator

$$\hat{PQI} = \frac{1}{N_R}\sum_{i=1}^{N_R} PQI_i^* = \frac{1}{N_R}\sum_{i=1}^{N_R} PQI^*(U_i), \tag{9a}$$

where

$$PQI_i^* = PQI_i^* + c_1(g_{1i} - G_1) + \ .... \ + c_k(g_{ki} + G_k) \tag{10}$$

for control parameter $c_k \in R$, and k uncertain parameters or properties $g_j, j = 1,2, ...k$, and where $\boldsymbol{g}_j$ are vectors of length N_R having elements g_{ji}, $i = 1, \dots N_R$ and

$$G_k = E[g_k]. \tag{11}$$

The variance of the control estimator is

$$Var(PQI^*) = Var(PQI) + 2\sum_{i=1}^{k} c_i\, Cov(PQI, g_i) + \sum_{i=1}^{k}\sum_{j=1}^{k} c_i c_j\, Cov(g_i, g_j). \tag{12}$$

The coefficients c_i can be determined by taking the partial derivative of (12) with respect to each c_i, set each resulting equation to 0, and solve the system of k equations for k unknowns. Each c_i is estimated by using sampled and computed quantities g_i and PQI, respectively, from the probabilistic assessment simulations. Equation (10) is then used to determine the control variate estimate or improved estimate of the mean.

Considering a single control variate, equation (10) reduces to

$$PQI^* = PQI + c_1(g_1 - G_1) \tag{13}$$

and (12) becomes

$$Var(PQI^*) = Var(PQI) + 2c_1\, Cov(PQI, g_1) + c_1^2\, Var(g_1). \tag{14}$$

Choosing c_1 so that the $Var(PQI^*)$ is minimized, c_1 is found to be

$$c_1 = -\frac{Cov(PQI, g_1)}{Var(g_1)}. \tag{15}$$

An alternate method for estimating coefficients c_i is linear regression analysis [7]. Equation 15 can be written as

$$c_1 = -\frac{\sum_{i=1}^{N_R}\left(PQI_i - \bar{PQI}\right)\left(g_{1i} - G_1\right)}{\sum_{i=1}^{N_R}\left(g_{1i} - G_1\right)}. \tag{16}$$

Now consider the linear regression model

$$PQI = A + Bg_1 + \epsilon, \tag{17}$$

where ϵ is a random error term with mean 0 and variance σ^2. The usual least square estimates of A and B are

$$\hat{A} = \bar{PQI} - \hat{B}\bar{g}_1 \tag{18a}$$

and

$$\hat{B} = \frac{\sum_{i=1}^{N_R}\left(PQI_i - \bar{PQI}\right)\left(g_{1i} - G_1\right)}{\sum_{i=1}^{N_R}\left(g_{1i} - G_1\right)}. \tag{18b}$$

Comparing (18b) with (16) we find that $\hat{B} = -c_1$ and from (13) and (17)

$$\hat{PQI} = \bar{PQI} + c_1\left(\bar{g}_1 - G_1\right) = \bar{PQI} - \hat{B}\left(\bar{g}_1 - G_1\right) = \hat{A} + \hat{B}G_1. \tag{19}$$

Therefore from (19) the control variate estimate $\hat{PQI}$ is equal to PQI given by the linear regression equation evaluated at $g_1 = G_1$. Furthermore the regression estimate of the variance σ^2 is

$$\hat{\sigma}^2 = Var(PQI - \hat{B}g_1) = Var(PQI + c_1 g_1) \tag{20}$$

Hence, the variance of the control estimator is

$$Var\,(\hat{PQI}) = \frac{\hat{\sigma}^2}{N_R} = Var\big(\bar{PQI} + c_1(\bar{g}_1 - G_1)\big). \tag{21}$$

The linear regression approach also applies when multiple control variates are used as in (10) by using a multiple regression model. See Appendix B for a comparison of results between $\hat{B}$ and c_1.

One advantage of using the linear regression approach to estimating control variates and the variance in $\hat{PQI}$ is that linear regression is often used to perform a sensitivity analysis as part of system performance assessment and therefore the use of this technique in control variates is a natural extension of this analysis. In addition, linear regression packages are readily available.

1.1.2. One-Dimensional Problem

A one-dimensional problem is presented here to illustrate and test the implementation of the control variate method. Consider the following elliptic homogeneous problem with Dirichlet boundary conditions

$$\frac{\partial^2 u}{\partial x^2} - v\frac{\partial u}{\partial x} = 0 \qquad 0 \le x \le 1 \tag{22}$$

$$u(0) = 1 \tag{23a}$$

$$u(1) = 0 \tag{23b}$$

where we assume $v(\omega) \ge 0$ and constant but with uncertainty described by probability density function D_v. The test cases considered in this study include uniform and triangular density functions. The analytical solution (24) to model problem (22) and (23) is illustrated in Figure 1.

$$u(x) = \frac{e^{vx} - e^{v}}{1 - e^{v}} \tag{24}$$

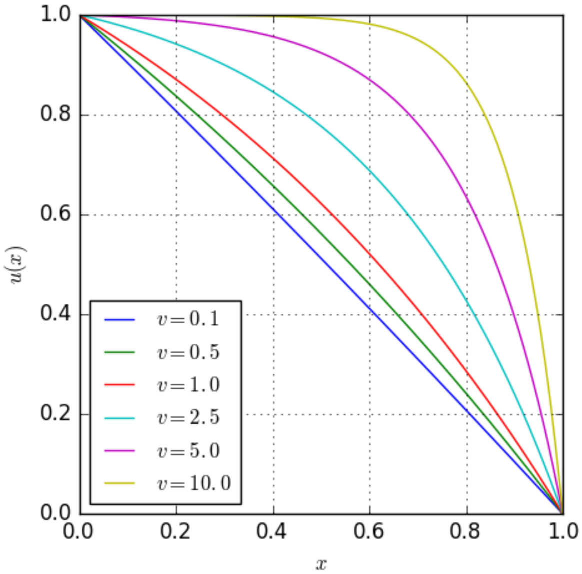

Figure 1. Behavior of analytical solution for a range of v.

In this study we examine the performance quantity of interest $PQI(x) = u(x)$ at $x = 0.9$.

For model problem (22) and (23) we consider a single control variate that corresponds to the uncertain coefficient $v(\omega_j)$, where $j = 1,2, \ldots N_R$

$$PQI_j^* = PQI_j + c_v(v_j - G_v), \tag{25}$$

where G_v is the mean of the probability density function D_v, and PQI_j^* is the value of the controlled performance measure for realization j. The improved mean value of the PQI at $x = 0.9$ is

$$\hat{PQI}(0.9) = \frac{1}{N_R}\sum_{j=1}^{N_R} PQI_j^*(0.9) = \frac{1}{N_R}\sum_{j=1}^{N_R}(PQI_j + c_v v_j) - c_v G_v \tag{26}$$

The coefficient c_v is given by

$$c_v = -\frac{Cov(PQI, v)}{Var(v)}, \tag{27}$$

where equation (27) would be evaluated in a post-processing manner after generating N_R realizations of PQI_j

$$c_v = \frac{\sum_{j=1}^{N_R}\left(PQI_j - \bar{PQI}\right)(v_j - G_v)}{\sum_{j=1}^{N_R}(v_j - G_v)^2} \tag{28}$$

with

$$\bar{PQI} = \frac{1}{N_R}\sum_{j=1}^{N_R} PQI_j. \tag{29}$$

For the model problem considered here it is possible to derive an analytical expression for the exact mean of $PQI(0.9)$ for a simple D_v such as the uniform density function $D_v = 1/(c-a)$. The triangular density function is

$$D_v(x) = \begin{cases} \dfrac{2(x-a)}{(c-a)(pk-a)} & x < pk \\ \dfrac{2(c-x)}{(c-a)(c-pk)} & x \geq pk \end{cases} \tag{30}$$

where pk is the location of the peak (i.e., mode) of the distribution. This mean is given by

$$E[u(x)] = \int_a^c \frac{e^{xv} - e^v}{1 - e^v} D_v dv, \tag{31}$$

the variance of v is given by

$$E[Var(v)] = \int_a^c (v - \bar{v})^2 D_v dv, \tag{32}$$

and the covariance of v and $u(x)$ is

$$E\{Cov[u(x),v]\} = \int_a^c \left[\frac{e^{xv} - e^v}{1 - e^v} - E[u(x)]\right][v - \bar{v}]D_v dv. \tag{33}$$

2. NUMERICAL RESULTS

We numerically integrate Equations 31 through 33 for three uniform distributions: $U\{a = 0.1, c = 1\}$, $U\{a = 0.1, c = 5\}$, $U\{a = 0.1, c = 10\}$, and one triangular distribution $Tr\{a = 0.1, pk = 7, c = 10\}$ to obtain the results in Table 1.

Table 1. Numerically Integrated Expected Values for *x* = 0.9

Distribution	*E*[*u*(0.9)]	E[Var(v)]	*E*{*Cov*[*u*(0.9),*v*]}	c_v
U{0.1, 1}	0.12685	0.06750	0.0034521	-0.051142
U{0.1, 5}	0.24602	2.00083	0.12207	-0.061010
U{0.1, 10}	0.38606	8.16750	0.45309	-0.055475
Tr{0.1, 7, 10}	0.42764	4.29500	0.23840	-0.055505

The results presented in Figure 2 show the simple mean ($\bar{PQI}$ given by a black line, each line representing a $\bar{PQI}$ for a different random seed), obtained from averaging PQI_i for a set of random inputs i = 1, .. N_R, and the variate controlled estimate given by (26) ($\hat{PQI}$ given by corresponding solid color lines). The four subplots give results consistent with the four distributions listed in Table 1. While details of the resulting statistics differ for the different seeds as expected, the improvement from applying the variance reduction method consistently improves the estimate of PQI across distributions and sample sizes. For the narrower distributions the improvement in approximately two orders of magnitude reduction in error for all N_R, and one order of magnitude improvement for the wider uniform and triangular distributions across all N_R. Note that the narrow distribution (red in leftmost subplot) had the smallest variance, while the widest uniform distribution (blue) has the highest variance. Also, all results in Figure 2 converge to their corresponding exact solution at a rate of ~ $1/\sqrt{N_R}$ as expected from Equation (8a) and as shown on the log-log plots with a slope of ~ -1/2. Appendix provides the statistics for the plots in Figure 2.

Figure 2 also shows that the control variate technique gives the same level of error in the variate controlled mean that would be obtained by computing a larger number of realizations and using the simple mean (e.g., moving horizontally from left to right at a particular error value and small N_R (< 100) to a larger N_R ($\sim 10{,}000$ realizations)).

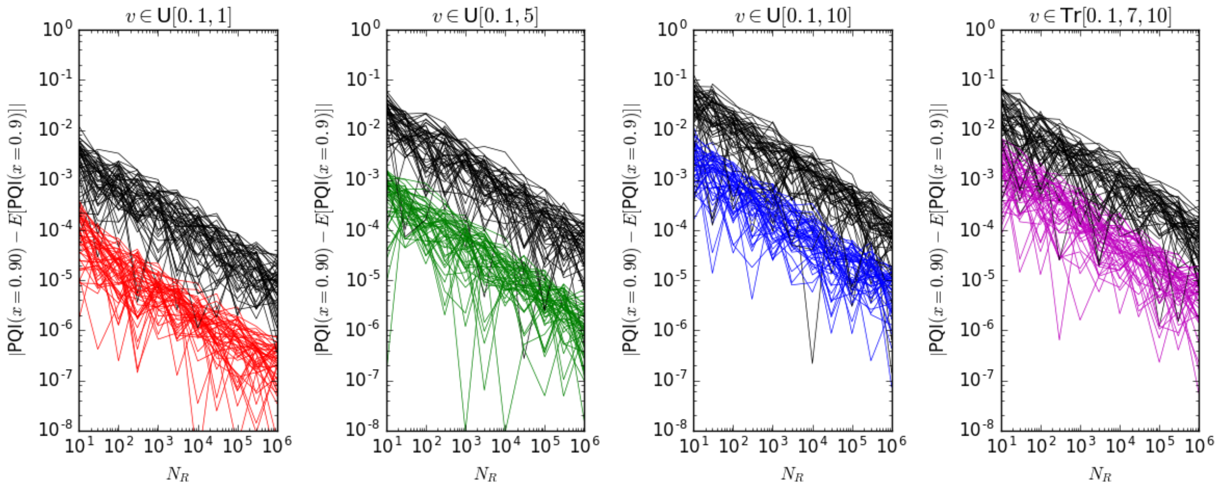

Figure 2. PQI(x=0.9) for Four Distributions in Table 1 across Sample Sizes N_R for 50 different random seeds.

Figure 3 shows representative horsetail plots of all solutions for each distribution in the $N_R = 100$ case (using the first random seed out of 50 random seeds). The integration of (26) via quadrature (i.e., using the quadpack library available through scipy - https://www.scipy.org) is given as the heavier red line. Each subplot shows the analytical solution (Equations 22 and 23) across the domain $0 \le x \le 1$ for the vector of 100 v values given from the distribution.

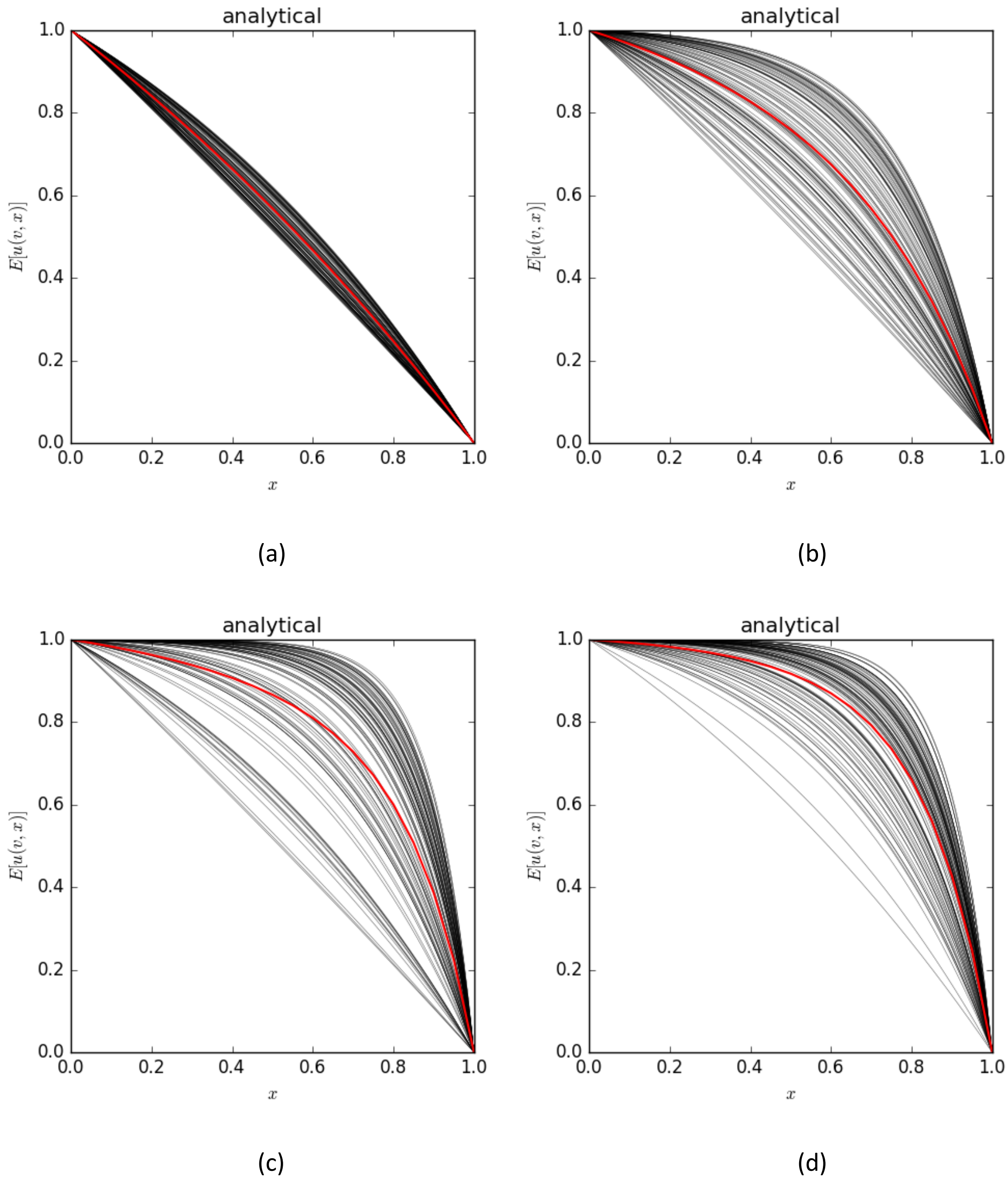

Figure 3. Horsetail Plots Showing Analytical Solution Results for N_R=100 and Uniform Distributions (a) U[0.1,1], (b) U[0.1,5], (c) U[0.1,10], and Triangular Distribution (d) U[0.1,7, 10].

3. CONCLUSIONS

This report presents a preliminary study of the method of control variates for calculating an improved estimate for the expectation of a performance quantity of interest $E(PQI)$ computed from Monte Carlo probabilistic simulations. Numerical results indicate that the method can produce estimates of $E(PQI)$ having superior accuracy on coarser meshes and reduce the required number of simulations needed to achieve an acceptable estimate. The method is well suited for large-scale performance assessment analyses because it can easily be implemented for problems involving a large number of uncertain parameters and it is a post-processing method and therefore non-intrusive to the probabilistic simulations. Future work will include investigations of multidimensional problems involving several uncertain parameters and actual performance assessment simulations. These investigations will examine prediction intervals for $E(PQI)$ and the relative errors of $E(PQI)$ and individual contributions of spatial and temporal truncation error from numerical approximations to the governing partial differential equations.

4. REFERENCES

1. SNL (Sandia National Laboratories), *Total System Performance Assessment for Yucca Mountain License Application*, MDL-WIS-PA-000005, Rev 00, AD 01, Sandia National Laboratories, Las Vegas, NV, March 2008.
2. Meacham, P.G., Anderson, D.R., Bonano, E.J. and Marietta, M.G., *Sandia National Laboratories Performance Assessment Methodology for Long-Term Environmental Programs: The History of Nuclear Waste Management*, SAND2011-8270, Sandia National Laboratories, Albuquerque, NM, November 2011.
3. Oden, J.T., Prudhomme, S., Control of modeling error in calibration and validation processes for predictive stochastic models, *International Journal for Numerical Methods in Engineering*, *87(1-5)*:262-272, September 2010.
4. Babuska, I.I., Nobile, F., Oden, J.T., Tempone, R., *Reliability, Uncertainty, Estimates Validation and Verification, Workshop on the Elements of Predictability*, J. Hopkins University, Baltimore, MD, November 2003.
5. Babuska, I., Tempone, R., Zouraris, G., Solving elliptic boundary value problems with uncertain coefficients by the finite element method: the stochastic formulation, *Computer Methods in Applied Mechanics and Engineering*, *194(12)*:1251-1294, 2005.
6. Nelson, B.L., On control variate estimators, *Computers & Operations Research 14(3)*:219-225, 1987.
7. Szechtman, R., *Control Variates Techniques for Monte Carlo Simulation*, Proceedings of the 2003 Winter Simulation Conference, 2003.
8. Helton, J.C., Hansen, C.W., Sallaberry, C.J., Expected dose and associated uncertainty and sensitivity analysis results for all scenario classes in the 2008 performance assessment for the proposed high-level radioactive waste repository at Yucca Mountain Nevada, *Reliability Engineering and System Safety 122*:421-435, February 2014.
9. Giles, M.B., Multilevel Monte Carlo path simulation, *Operations Research, 56(3)*:607-617, 2008.
10. Charrier, J., Scheichl, R., Teckentrup, A.L., Finite element error analysis of elliptic PDEs with random coefficients and its application to multilevel Monte Carlo methods, *SIAM Journal on Numerical Analysis, 51(1)*:322-352, 2013.

APPENDIX A. EXAMPLE PROBLEM

The example presented here shows that in the case where the exact solution u is known or calculated by the numerical method that the control variate techniques will yield an exact expectation for the $PQI = u(x_i)$ at point x_i, for $0 \le x_i \le 1$.

$$\frac{\partial^2 u}{\partial x^2} - v\frac{\partial u}{\partial x} = f \qquad 0 \le x \le 1 \tag{A1}$$

$$f = 2v(vx - 1) \tag{A2}$$

$$u(0) = 1 \tag{A3a}$$

$$u(1) = 1 - v \qquad a \le v \le b \tag{A3b}$$

with a and b constants. The solution to (A1) through (A3) is

$$u = 1 - vx^2. \tag{A4}$$

We assume that uncertain v is given by a uniform density function

$$D_v = \frac{1}{b - a} \tag{A5}$$

and probability distribution

$$F_v = \frac{v - a}{b - a}. \tag{A6}$$

The expectation and variance of v is given by

$$E(v) = \frac{a + b}{2} \tag{A7}$$

and

$$Var(v) = \frac{(b - a)^2}{12}. \tag{A8}$$

The enhanced or controlled PQI^* is

$$\begin{aligned} PQI_j^* &= PQI_j + c_v(v_j - E[v]) \\ &= 1 - x^2 v_j + c_v(v_j - E[v]). \end{aligned} \tag{A9}$$

If $c_v = x^2$, then the exact result is given

$$PQI_j^* = 1 - E[v]. \tag{A10}$$

This result can also be determined from (A9) by inspection. Since (A10) yields a constant then it follows that $E(PQI_j^*)$ is exact.

APPENDIX B.: NUMERICAL EXAMPLE RESULTS

Appendix B lists the statistics for plots in Figure 2, showing how the estimates of $Var(v)$, $Cov(Q,v)$, PQI, c_v, and $\hat{PQI}$ converge toward their exact values as the sample size increases. The numerically computed results also converge as illustrated in the rightmost columns.

uniform distribution: 0.1 <= v <= 1.0 seed:1411193087

N	var(v)	cov[Q(x),v]	PQI	cv	PQIhat	|	slope	intercept	r^2_val	std_err
10	0.0920328	0.0047003	0.1224125	-0.0510720	0.1270130	|	0.0510720	0.0989234	0.9995529	0.0003819
30	0.0722313	0.0036876	0.1272948	-0.0510533	0.1268764	|	0.0510533	0.0987971	0.9994949	0.0002169
100	0.0689005	0.0035340	0.1292520	-0.0512910	0.1268615	|	0.0512910	0.0986515	0.9993803	0.0001290
300	0.0684014	0.0035085	0.1284726	-0.0512928	0.1268546	|	0.0512928	0.0986436	0.9994379	0.0000705
1000	0.0667399	0.0034093	0.1267527	-0.0510838	0.1268460	|	0.0510838	0.0987499	0.9994280	0.0000387
3000	0.0666961	0.0034071	0.1265992	-0.0510841	0.1268457	|	0.0510841	0.0987494	0.9994377	0.0000221
10000	0.0677884	0.0034648	0.1267221	-0.0511126	0.1268513	|	0.0511126	0.0987394	0.9994393	0.0000121
30000	0.0680877	0.0034812	0.1268202	-0.0511289	0.1268529	|	0.0511289	0.0987320	0.9994371	0.0000070
100000	0.0675568	0.0034546	0.1268661	-0.0511368	0.1268501	|	0.0511368	0.0987248	0.9994392	0.0000038
300000	0.0675827	0.0034561	0.1268428	-0.0511383	0.1268502	|	0.0511383	0.0987241	0.9994387	0.0000022
1000000	0.0675408	0.0034541	0.1268534	-0.0511416	0.1268499	|	0.0511416	0.0987221	0.9994391	0.0000012
True:	0.0675000	0.0034521	0.1268497	-0.0511420						

==

uniform distribution: 0.1 <= v <= 5.0 seed:1411193087

N	var(v)	cov[Q(x),v]	PQI	cv	PQIhat	slope	intercept	r^2_val	std_err
10	2.0073546	0.1238244	0.2416460	-0.0616853	0.2455266	0.0616853	0.0882290	0.9998472	0.0002696
30	1.8700348	0.1155911	0.2534395	-0.0618123	0.2453677	0.0618123	0.0877465	0.9998328	0.0001511
100	2.0087038	0.1227814	0.2501162	-0.0611247	0.2459611	0.0611247	0.0900931	0.9993512	0.0001573
300	1.9983496	0.1221054	0.2465775	-0.0611031	0.2459657	0.0611031	0.0901528	0.9993157	0.0000926
1000	2.0070287	0.1226063	0.2486283	-0.0610885	0.2460014	0.0610885	0.0902259	0.9993471	0.0000494
3000	2.0188573	0.1232042	0.2463171	-0.0610267	0.2460215	0.0610267	0.0904034	0.9992894	0.0000297
10000	1.9956744	0.1217065	0.2456199	-0.0609851	0.2460335	0.0609851	0.0905215	0.9992762	0.0000164
30000	2.0025217	0.1221386	0.2451714	-0.0609924	0.2460228	0.0609924	0.0904922	0.9992824	0.0000094
100000	2.0073458	0.1224704	0.2460245	-0.0610111	0.2460250	0.0610111	0.0904466	0.9992953	0.0000051
300000	2.0066791	0.1224242	0.2459474	-0.0610084	0.2460256	0.0610084	0.0904543	0.9992934	0.0000030
1000000	2.0031026	0.1222079	0.2459731	-0.0610093	0.2460232	0.0610093	0.0904494	0.9992929	0.0000016
True:	2.0008333	0.1220706	0.2460220	-0.0610099					

==

uniform distribution: 0.1 <= v <= 10.0 seed:1411193087

N	var(v)	cov[Q(x),v]	PQI	cv	PQIhat	slope	intercept	r^2_val	std_err
10	8.3458865	0.4596259	0.3801683	-0.0550721	0.3853206	0.0550721	0.1072063	0.9955070	0.0013081
30	7.8556679	0.4386254	0.3858923	-0.0558355	0.3862329	0.0558355	0.1042635	0.9941197	0.0008115
100	6.9742665	0.3874481	0.4035476	-0.0555540	0.3876733	0.0555540	0.1071258	0.9946715	0.0004107
300	7.9742281	0.4409414	0.3968190	-0.0552958	0.3861108	0.0552958	0.1068670	0.9950028	0.0002270
1000	8.0457852	0.4464711	0.3842223	-0.0554913	0.3861715	0.0554913	0.1059404	0.9952416	0.0001215
3000	8.2449704	0.4576389	0.3809896	-0.0555052	0.3858485	0.0555052	0.1055471	0.9954443	0.0000686
10000	8.0810368	0.4487960	0.3848168	-0.0555369	0.3861575	0.0555369	0.1056960	0.9953949	0.0000378
30000	8.1991872	0.4549739	0.3856996	-0.0554901	0.3860275	0.0554901	0.1058024	0.9953944	0.0000218
100000	8.1618570	0.4526394	0.3866702	-0.0554579	0.3860692	0.0554579	0.1060068	0.9953383	0.0000120
300000	8.1579699	0.4525623	0.3860866	-0.0554749	0.3860758	0.0554749	0.1059277	0.9953609	0.0000069
1000000	8.1610378	0.4527619	0.3859124	-0.0554785	0.3860696	0.0554785	0.1059034	0.9953744	0.0000038
True:	8.1675000	0.4530894	0.3860604	-0.0554747					

==

triangle distribution: 0.1 <= v <= 10.0 (peak=7.0) seed:1411193087

N	var(v)	cov[Q(x),v]	PQI	cv	PQIhat	slope	intercept	r^2_val	std_err
10	1.1548942	0.0647796	0.4109397	-0.0560914	0.4333764	0.0560914	0.1136555	0.9985833	0.0007470
30	4.1009968	0.2213913	0.4406383	-0.0539848	0.4277857	0.0539848	0.1200725	0.9931886	0.0008449
100	4.2706408	0.2358519	0.4300553	-0.0552264	0.4277693	0.0552264	0.1129791	0.9939301	0.0004360
300	4.6311031	0.2544533	0.4370673	-0.0549444	0.4268079	0.0549444	0.1136247	0.9943566	0.0002398
1000	4.5654297	0.2535479	0.4247725	-0.0555365	0.4271528	0.0555365	0.1105949	0.9949997	0.0001246
3000	4.3813075	0.2429508	0.4273735	-0.0554517	0.4274901	0.0554517	0.1114156	0.9947613	0.0000735
10000	4.3129460	0.2390303	0.4288595	-0.0554216	0.4275705	0.0554216	0.1116675	0.9946865	0.0000405
30000	4.3015577	0.2386473	0.4286005	-0.0554793	0.4276068	0.0554793	0.1113749	0.9947114	0.0000234
100000	4.2740504	0.2371906	0.4280964	-0.0554955	0.4276691	0.0554955	0.1113448	0.9946852	0.0000128
300000	4.2860414	0.2377893	0.4281159	-0.0554799	0.4276396	0.0554799	0.1114040	0.9947026	0.0000074
1000000	4.2973168	0.2384860	0.4277764	-0.0554965	0.4276279	0.0554965	0.1112980	0.9947297	0.0000040

True: 4.2950000 0.2383953 0.4276379 -0.0555053

DISTRIBUTION

1	MS0747	Robert J. MacKinnon	6224
1	MS0747	Kristopher L. Kuhlman	6224
1	MS0747	S. David Sevougian	6224
1	MS0747	Glenn E. Hammond	6224
1	MS0747	Paul Mariner	6224
1	MS0747	Daniel J. Clayton	6223
1	MS0736	Evaristo J. Bonano	6220
1	MS0899	Technical Library	9536 (electronic copy)

Sandia National Laboratories