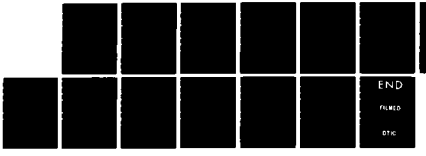
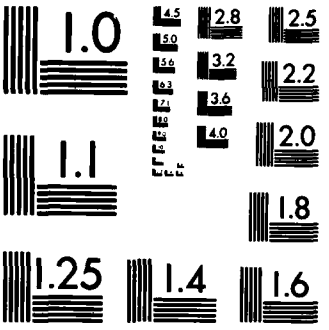
AD-A158 613 MONTE CARLO ANALYSIS OF NONLINEAR STATISTICAL MODELS I 1/1
THEORY REVISION.. (U) PURDUE UNIV LAFAYETTE IN SCHOOL OF
INDUSTRIAL ENGINEERING J J SWAIN ET AL. AUG 85
UNCLASSIFIED RM-85-10 N00014-79-C-0832 F/G 12/1 NL





(2)

MONTE CARLO ANALYSIS OF NONLINEAR STATISTICAL MODELS, I: THEORY

James J. Swain
Georgia Institute of Technology

Bruce W. Schmeiser
Purdue University

Research Memorandum 85-10
School of Industrial Engineering
Purdue University
West Lafayette, IN 47907

Technical Report J-84-11
School of Industrial and Systems Engineering
Georgia Institute of Technology
Atlanta, GA 30332

December 1984
Revised August 1985

This research was partially supported by Office of Naval Research Contract N00014-79-C-0832 at Purdue University and by National Science Foundation Grant ECS-8206498 at the Georgia Institute of Technology. The second author received time and resource support from the Operations Research Department at AT&T Bell Laboratories.

ABSTRACT

Parameter values of nonlinear statistical models are typically estimated from data using iterative numerical procedures. The resulting joint sampling distribution of the parameter estimators is often intractable, resulting in the use of approximators or Monte Carlo simulation to determine properties of the sampling distribution.

This paper develops methods, using linear and quadratic approximators as control variates, that reduce the variance of the Monte Carlo estimator by orders of magnitude. Estimation of means, higher order raw moments, variances, covariances, and percentiles is considered.

Keywords: Control variates; Monte Carlo; Nonlinear estimation; Simulation; Swindle; and Variance reduction.

1. INTRODUCTION

Statistical models are used in many fields to relate dependent variables to independent variables. For linear models with additive independent and normally distributed errors and for a few extensions, the statistical theory is well-established and fitting the model to data is straightforward. For nonlinear models or for models without normally distributed errors the theory is less complete and iterative numerical procedures are needed for fitting, which is the price paid for a better fitted model with associated better predictions and inference.

Both the precision of predictions and the statistical significance of inferences depend upon the sampling distribution of the estimators of the model parameters. This typically multivariate sampling distribution is often summarized by statistical properties such as means, higher order moments, variances, covariances, percentiles and quantiles.

For nonlinear models these properties are usually unobtainable analytically. Methods for approximating these properties have followed three approaches: asymptotic results, series expansions, and Monte Carlo sampling. The usual advantages and disadvantages of these approaches hold here: Asymptotic results are often easy to obtain but can have large error for small sample sizes; approximations based on series expansions have fixed accuracy and can be difficult to obtain, particularly in multivariate cases; and Monte Carlo sampling is easy to implement but requires large sample sizes to be accurate.

In this paper the efficiency of the Monte Carlo sampling approach is improved by using series expansion approximations as control variates. Roughly, the improvement results from sampling to estimate the error of the approximation rather than to estimate the property of interest directly. The resulting methods require few assumptions and are straightforward.

The theory developed here is oriented toward providing the methodology for efficient Monte Carlo study of statistical models. In particular, the results are oriented toward development of a computer package that, given only the description of the model of interest, automatically and efficiently evaluates the model. Such a methodology is important since the use of nonlinear models is growing, due to advances in both computer hardware and in software for fitting nonlinear models. But these computer advances, while adequate for fitting, often are not sufficient for the Monte Carlo studies necessary to study nonlinear models, particularly given the increasing use of microcomputers in interactive environments. In addition, automation is important since most analysts lack the time and/or training to develop variance reduction methods manually. Swain (1982) and Swain and Schmeiser (1983) reduce variance by orders of magnitude for a variety of models using the variance reduction methods developed here.

Section 2 introduces notation, terminology, and problem statements. Section 3 surveys approximation methods for nonlinear models. Monte Carlo sampling in the context of nonlinear statistical models, with emphasis on control variates for variance reduction, is developed in Section 4. Section 5 develops these ideas for the cases of estimating means, higher order moments, variances and covariances, and percentiles.

2. NOTATION, TERMINOLOGY, AND PROBLEM STATEMENTS

The following notation is used. Random variables are denoted by upper-case Roman letters, upper-case Greek letters, or lower-case Greek letters augmented with carets and bars. Realizations of random variables are denoted by lower-case letters, as are constants and functions. Variables are vectors or matrices with scalar components. Exceptions are specifically identified. Vectors are columns, with row vectors denoted by the transpose of a vector. Powers of a vector, V^k, denote componentwise exponentiation. The kth moment about the origin is denoted by $\mathrm{E}[V^k]$. $\mathrm{Cov}[V,W]$ denotes the $p \times q$ covariance matrix $\mathrm{E}[(V-\mathrm{E}[V])(W-\mathrm{E}[W])^T]$ of the p- component random vector V and the q-component random vector W. The variance of V, $\mathrm{Cov}[V,V]$, is also denoted by $\mathrm{Var}[V]$.

Consider the nonlinear q-dimensional response function defined over an r-dimensional domain,

$$\mathrm{E}[Y_i] = \eta^*(X_i^*; \theta^*)$$

where $Y_i = (Y_{i1}, Y_{i2}, \cdots, Y_{iq})$ is the measured response, $X_i^* = (X_{i1}^*, X_{i2}^*, \cdots, X_{ir}^*)$ is the actual design point, and $\theta^* = (\theta_1^*, \theta_2^*, \cdots, \theta_p^*)$ is the actual parameter value for the surface. Y_i has the associated error random variable

$$E_i = Y_i - \eta^*(X_i^*; \theta^*)$$

where $E_i = (E_{i1}, E_{i2}, \cdots, E_{iq})$. These errors arise from four sources: error in controlling X_i, error in measuring X_i, error in measuring Y_i, and random variation in the system. Reflecting these sources of error, the model can be specified by $\{\eta^*, \theta^*, G^*, H^*\}$, where G^* is the multivariate cdf associated with the design matrix and H^* is the multivariate cdf associated with the errors. There are three design matrices of interest: (1) X', the nominal design matrix, (2) X^*, the actual design matrix, and (3) X, the measured design matrix. Correspondingly, G^* has three components: (1) $G_{X'}^*$, (2) $G_{X^*}^*$, and (3) G_X^*. $G_{X^*}^*$ is conditional on the outcome of $G_{X'}^*$, and G_X^* is conditional on the outcomes of both $G_{X'}^*$ and $G_{X^*}^*$. Similarly, H^* is the multivariate cdf of the two types of error terms, E^* and E', whose individual cdfs are denoted by $H_{E^*}^*$ and $H_{E'}^*$. The error term E is the convolution of the (possibly dependent) error terms E^*, the random variation in the system, and E', the error in measuring Y. $H_{E^*}^*$ is conditional on the outcome of $G_{X^*}^*$. $H_{E'}^*$ is conditional on the outcome of G_X^* and $H_{E^*}^*$. These relationships among G^* and H^* are stated in algorithm form in the discussion of the Monte Carlo method at the beginning of Section 4.

A surface η is fitted by determining an estimate $\hat{\theta}$ of θ^* as a function of observed data (x, y), where $x = (x_1, x_2, \cdots, x_n)^T$ and $y = (y_1, y_2, \cdots, y_n)^T$, which arise from the actual model $\{\eta^*, \theta^*, G^*, H^*\}$. Since the actual model is unknown, the analyst must assume a family of models, denoted here by $\{\eta, T, G, H\}$, and referred to as the *unfitted assumed model*. G and H are analogous to G^* and H^* of the true model. The feasible region for $\hat{\theta}$ is denoted by $T \subseteq \mathbf{R}^p$, which arises from inherent properties of η or from the context of the application.

Denote the (typically iterative) procedure used to determine $\hat{\theta}$ by h_s, where s is a short notation for $s(\theta; \eta, G, H, x, y)$, a scalar function of the data and the unfitted

assumed model. Such procedures almost always depend upon the residuals $e(\theta) = (e_1(\theta), e_2(\theta), \cdots, e_n(\theta))^T$, where $e_i(\theta) = y_i - \eta(x_i; \theta)$ for $\theta \in T$. The two most common examples of s are the sum of squared residuals and the likelihood of the residuals. Procedures h_s minimize the former and maximize the latter.

Point estimation is then

> *POINT*: Given the unfitted assumed model $\{\eta, T, G, H\}$, data (x, y), and procedure h_s, determine $\hat{\theta}$.

In general, *POINT* is a nonlinear programming problem, with an objective function s and with constraints $\theta \in T$, that is solved by procedure h_s. Whether or not h_s results in the optimal nonlinear programming solution is irrelevant. However, excluded from discussion in this article are procedures that can return a different estimate on different applications to the same data with the same unfitted assumed model, such as sometimes occurs when the analyst provides initial values to the procedure.

Having solved *POINT*, the practitioner is faced with the problem underlying this paper: evaluating the quality of the estimator $\hat{\theta}$.

> *DIST*: Given the assumed unfitted model $\{\eta, T, G, H\}$ and procedure h_s used in *POINT*, and the true fitted model $\{\eta^*, \theta^*, G^*, H^*\}$, determine properties of the multivariate distribution of $\hat{\theta}$.

An important application of *DIST* is to study the effect of assuming the wrong model, which occurs when one or more of $\eta = \eta^*$, $G = G^*$, or $H = H^*$ do not hold. For example, $H \neq H^*$ when testing sensitivity of the sampling distribution to departures from the assumed error distribution H, which is typically independent, identically-distributed, zero-mean normal.

Unlike the linear case, in the general nonlinear case the distribution of $\hat{\theta}$ depends in a nontrivial way on θ^*. Although the researcher knows the value of θ^*, the practitioner, who cannot know θ^*, can only solve *DIST* for several values of θ^* in a region of interest.

DIST is a very general problem — much more general than typically considered in research on nonlinear models. Monte Carlo methods alone have the ability to analyze all special cases of this general problem. However, the extensive literature of approximations for subsets of this problem are useful as the basis for variance reduction techniques for Monte Carlo solutions, as developed in this paper.

3. APPROXIMATION METHODS

Approximations are crucial to both the estimation problem *POINT* and the distribution problem *DIST* in nonlinear models. The solution procedure h_s in *POINT* often uses low-order approximations as the basis of iterating to an improved solution. In *DIST*, approximations sometimes provide easy to compute, fairly accurate approximations to the distribution of $\hat{\theta}$. This section discusses a linear approximation

corresponding to the least squares estimator in the context of both *POINT* and *DIST*. This approximation and the analogous quadratic approximation are good, but not the only possible, choices for use in the control variate variance reduction techniques developed later in this article.

Consider a procedure h, based on minimizing the (weighted) sum of the squared residuals for a scalar ($q=1$) response,

$$s(\theta) = e^T(\theta)\, W\, e(\theta),$$

where W is a matrix of weights. When the residuals depend upon a nonlinear function of the data the normal equations require solution of simultaneous nonlinear equations. One approach is to linearize the residuals at the current iterate, $\theta^{(j)}$, using

$$e(\theta) \approx y - \eta(x, \theta^{(j)}) - F(\theta^{(j)})(\theta - \theta^{(j)})$$

where $F(\theta^{(j)})$ is the $n \times p$ Jacobian matrix of first derivatives $\partial \eta(x_i, \theta)/\partial \theta_k$ ($i = 1,2,\cdots,n$ and $k = 1,2,\cdots,p$) evaluated at $\theta = \theta^{(j)}$. This approximation leads to the Gauss update of the Newton algorithm

$$\theta^{(j+1)} = \theta^{(j)} + [F^T(\theta^{(j)})\, W\, F(\theta^{(j)})]^{-1} F^T(\theta^{(j)})\, W\, e(\theta^{(j)}) \ .$$

This approximation can also be viewed as using only the leading first-order terms in the second derivatives (the square brackets, above) in the Newton update.

There is little special about the choice of this linear approximator or about least squares. When higher derivatives can be computed, higher-order approximations are sometimes used for greater accuracy or for numerical stability. Series approximations can also be used to create algorithms to solve estimation problems with maximum likelihood and other loss functions.

Now consider linear approximations applied to *DIST*. For example, consider a weighted least squares estimation problem with $\eta = \eta^*$, a deterministic design matrix, and weight matrix $W^{-1} \equiv \mathrm{Var}[E]$. In the Gauss update formula, replacing θ^* with $\theta^{(i)}$ yields the residuals $E(\theta^*) \equiv E$ and the linear approximator Δ

$$\Delta = \theta^* + [F^T(\theta^*)\, W\, F(\theta^*)]^{-1} F^T(\theta^*)\, W\, E \ .$$

The approximator Δ has readily determined properties. When $E \sim N(0, W)$, then $\Delta \sim N(\theta^*, \mathrm{Var}[\Delta])$, where $\mathrm{Var}[\Delta] = [F^T(\theta^*)\, W\, F(\theta^*)]^{-1}$. The distribution of Δ is an approximation of the sampling distribution of $\hat{\theta}$, although it provides no information about the bias of $\hat{\theta}$ as an estimator of θ^*. Gallant (1975) uses such a linear approximator to construct approximators to the likelihood ratio test. The generalization for multiple responses ($q > 1$) is discussed in Bard (1974).

Again this linear approximator and least squares are just examples. In the multiparameter quadratic case considerable ingenuity (and perhaps further approximation) must be exercised to specify the approximator as a function of the appropriate derivatives and the error vector, E. Approximators are also often based on asymptotic properties. Both the least squares and maximum likelihood estimators have asymptotic normal distributions under general conditions (Chambers, 1977; Crowder, 1976; Wilks, 1962; and Wu, 1981). Shenton and Bowman (1977) discuss maximum likelihood in detail.

The linear approximation is useful in both *POINT* and *DIST*. In *POINT*, the realization δ approximates θ. In *DIST*, the distribution of the random variable Δ approximates the distribution of $\hat{\theta}$. Both are referred to as the ***delta approximation*** throughout the rest of this paper. The analogous quadratic approximation, with associated random variable denoted by Γ and realizations γ, is referred to as the ***gamma approximation***. The *gamma approximation* for *DIST* is based on results for the least squares estimator in Box (1971) and Clarke (1980), who provide expressions for the bias and variance matrix. The corresponding quadratic approximation for *POINT* is developed in Swain (1985), who extends Clarke's result to the specific context of Monte Carlo studies.

The accuracy of approximations is fixed by the true model $\{\eta^*, \theta^*, G^*, H^*\}$ and the unfitted assumed model $\{\eta, T, G, H\}$, although occasionally increased accuracy is possible by extending the order of the approximation. Measures of nonlinearity (Beale, 1960; Bates and Watts, 1980) give an indication of the accuracy of approximations, but the measures can be misleading, as when errors are nonnormal (Gillis and Ratkowsky, 1978).

4. MONTE CARLO SAMPLING AND CONTROL VARIATES

Unlike approximations, Monte Carlo methods can provide any level of accuracy with sufficient sampling effort. The most straightforward Monte Carlo simulation analysis, termed *direct* Monte Carlo, and the ideas underlying reducing Monte Carlo sampling error via control variates are discussed in this section.

Consider the direct Monte Carlo sampling analysis of *DIST*: m repeated observations of $\hat{\theta}$, denoted by $\hat{\theta}_{:u}$, $u = 1,2,\cdots,m$, are generated and the properties of the sampling distribution are estimated from the sample. In the direct experiment, quantities with subscript $i:u$ are scalar components of vector quantities denoted by subscripts $:u$. This direct experiment defines the class of problems considered in this article. Correlated errors are not excluded. Extensions such as multiplicative errors are straightforward but not considered.

The Direct Monte Carlo Experiment

1. For $u = 1,2,\cdots,m$
   a. Generate $x'_{:u}$ from G^*_X,
   b. Generate $x^*_{:u}$ from $G^*_{X^*}$ given $x_{:u}'$
   c. Generate $x_{:u}$ from G^*_X given $x_{:u}', x^*_{:u}$
   d. Generate $\epsilon^*_{:u}$ from H^*_E, given $x^*_{:u}$
   e. For $i = 1,2,\cdots,n$; calculate $y^*_{i:u} = \eta^*(x^*_{i:u}; \theta^*) + \epsilon^*_{i:u}$
   f. Generate $\epsilon^*_{:u}$ from $H^*_{E^*}$ given $x'_{:u}, x^*_{:u}, x_{:u}$, and $y^*_{:u}$
   g. For $i = 1,2,\cdots,n$; calculate $y_{i:u} = y^*_{i:u} + \epsilon^*_{i:u}$
   h. Solve *POINT* using h_u on $(x_{:u}, y_{:u})$ assuming $\{\eta, T, G, H\}$, which leads to $\hat{\theta}_{:u}$, possibly through the residuals $e(\theta) = y_{:u} - \eta(x_{:u}, \theta)$

2. Compute estimates of interest from $\hat{\theta}_{:s}$, $s = 1,2,\cdots,m$

The usual estimates of the marginal raw moments and the variance matrix of $\hat{\theta}$ are

$$\mu_k' = m^{-1} \sum_{s=1}^{m} (\hat{\theta}_{:s})^k \qquad k = 1,2,\cdots$$

and

$$\hat{\Sigma}_{\hat{\theta}} = (m-1)^{-1} \sum_{s=1}^{m} (\hat{\theta}_{:s} - \mu_1')(\hat{\theta}_{:s} - \mu_1')^T$$

respectively.

While the direct Monte Carlo method is conceptually simple to implement and provides consistent estimates of the desired distribution, standard error decreases only with $m^{1/2}$. Thus large Monte Carlo sample sizes are needed to obtain precise estimates with the direct experiment. Worse, each observation $\hat{\theta}_{:s}$ is expensive, not because of the sampling from G^* and H^*, but because h_s in step 2 is typically a nonlinear programming algorithm in p variables. Therefore, variance reduction techniques are important when using Monte Carlo methods to analyze *DIST*.

Variance reduction techniques are almost as old as Monte Carlo sampling; in fact the term *Monte Carlo* was once reserved for sampling methods in which some sort of swindle was used to reduce variance. Consequently, a large literature exists; Hartley (1977), Kahn (1956), McGrath and Irving (1973), and Wilson (1984) are interesting surveys. Examples of variance reduction in Monte Carlo studies in statistical settings include Andrews, Bickel, Hampel, Huber, Rogers, and Tukey (1972), Arnold, Bucher, Trotter, and Tukey (1956), Gallant (1980), Kleijnen (1977), Koehler (1981), Relles (1974), and Schruben and Margolin (1978). Nelson (1983) and Nelson and Schmeiser (1984a, 1984b, 1985) propose a framework for variance reduction.

The variance reduction methods developed in the next section use the delta or gamma approximations, discussed in the last section, as control variates. The remainder of this section reviews control variates, drawing from Lavenberg and Welch (1981), Cheng (1978), and Rubinstein and Marcus (1985). The method of control variates is treated in most Monte Carlo texts (Hammersley and Handscomb, 1964) and advanced simulation texts (Kleijnen, 1974; Fishman, 1978; Law and Kelton, 1982; Bratley, Fox, and Schrage, 1983).

Consider for now a primary estimator Z, a p random vector, and a control variate C, an l random vector, whose mean is known. The control variate estimator is

$$Z(B) = Z - B(C - \mathrm{E}[C])$$

where B is a $p \times l$ weight matrix. $Z(B)$ has the same expected value as $Z = Z(0)$ if B is a constant or is independent of C. The variance of $Z(B)$ is the matrix

$$\mathrm{Var}[Z(B)] = \mathrm{Var}[Z] + B\,\mathrm{Var}[C]\,B^T - B\,\mathrm{Cov}[C,Z] - \mathrm{Cov}[Z,C]\,B^T$$

whose determinant is minimized by the setting B equal to

$$B^* = \mathrm{Cov}[Z,C]\,\mathrm{Var}^{-1}[C] \ .$$

In addition to minimizing the generalized variance, $|\text{Var}[Z(B)]|$, this set of control weights also minimizes the trace of the variance matrix. Rubinstein and Marcus (1985) show that

$$|\text{Var}[Z(B^*)]| = |\text{Var}[Z(0)]| \prod_{j=1}^{p} (1-\rho_j^2)$$

where the ρ_j are the canonical correlations between Z and C (Anderson, 1958).

External control variates are random variables analogous to Z but from another system. This other system is simulated so as to induce a correlation between C and Z, thereby reducing the variance from that of Z to that of $Z(B)$. The correlation is induced by subjecting both systems to the same uniform (0,1) random number streams, generating random variates using the inverse distribution function, and coding the simulation to provide a monotonic transformation between the inputs and outputs — in this case between the inputs $x_{:s}', x_{:s}^*, x_{:s}'', e_{:s}^*, e_{:s}''$ and the output $\hat{\theta}_{:s}$.

Often when the control C is external, setting $B = I_p$, the p dimensional identity matrix, is close to optimal (Hammersley and Handscomb, 1964); but even in the external control variate case, estimating B^* is sometimes worthwhile when studying *DIST* (Swain, 1982; Swain and Schmeiser, 1984), especially when the errors are nonnormal (Swain, 1984). The most common estimator is the regression estimator, obtained by substituting sample values for the covariance and variance terms in the definition of B^*. Swain (1982) studies a variety of estimators of B^*.

Estimating B^* gives rise to issues of both bias and efficiency. Bias can be avoided by using a splitting estimator, as discussed in Kahn (1956), Tocher (1963), and Swain and Schmeiser (1983). Efficiency depends upon the choice and number of control variates employed, as discussed by Lavenberg and Welch (1981), Porta Nova (1985), Rubinstein and Marcus (1985), and Venkatraman and Wilson (1985). The loss in efficiency for each additional control variate used when B^* is estimated may or may not be offset by the additional information carried in the correlation between the new component of C and Z. Fortunately, in this sense, *DIST* gives rise to only a limited number of natural, high-quality control variates.

The control variate technique is not the only simulation swindle that is possible for *DIST*, but an advantage of control variates is that they can be applied simultaneously to several estimators. In contrast, many other variance reduction techniques lead to variance reduction for one estimator while leading to a variance increase for another estimator.

5. APPLICATION OF CONTROL VARIATES TO DIST

Control variate estimators for *DIST* are developed in this section based on the delta and gamma approximations. Estimation of arbitrary properties of the distribution of $\hat{\theta}$ is considered. Means, higher-order marginal moments, the variances and covariances, and percentiles of the sampling distribution of $\hat{\theta}$ are given special attention. The new algorithm is identical to the direct Monte Carlo method stated earlied, with the modification of step 2 to replace the direct estimators with the control variate estimators and the addition of step 11:

11. Compute $\delta_{:u}$ or $\gamma_{:u}$ from $x_{:u}$, $y_{:u}$

The definition of $\delta_{:u}$ and $\gamma_{:u}$ was given in Section 2. In the general case, defined by the direct Monte Carlo procedure of Section 4, some care and modification is necessary for implementing step 11.

When $\eta \neq \eta^*$, care should be taken to use η rather than η^* when computing F.

When $\eta \neq \eta^*$, θ^* is inappropriate in the definition of Δ, since for example the dimensionality of Δ may be different than that of θ^*. In this case, replace θ^* with $\hat{\theta}$, the estimator from the direct experiment. Then $\delta_{:u}$ can not be calculated until after all the direct sampling is completed. No extra sampling is required.

When the weight matrix W is unknown, replace W with $[\hat{\text{Var}}[E]]^{-1}$, where the estimate is again from the direct sampling experiment and calculation of $\delta_{:u}, u = 1,2, \cdots m$, is delayed until later.

When $\eta \neq \eta^*$, the design matrix X is not deterministic, or W is unknown the sampling distribution of Δ is unknown. In these and other situations in which the distribution of Δ is unknown, a supplementary Monte Carlo experiment is run to obtain the needed properties of Δ to negligible error. The observations of Δ are much cheaper to obtain than observations of $\hat{\Theta}$, so this supplementary run could obtain the negligible error with a large sample size. However, variance reduction can be applied to the supplementary experiment as well. The most obvious method is to use conditional expectations, which means that rather than calculating the properties of interest from the observed values of Δ, the sample average of the properties of interest is calculated.

There is an option to the supplementary experiment — perturb the direct Monte Carlo observations to obtain a control system for which the properties of Δ are known. For example, set the design matrix to the mean (or median) value, and transform nonnormal error terms to normality using

$$\epsilon_{:u}' = \text{Var}^{1/2}[E]\, \Phi^{-1}(H(\epsilon_{:u}))$$

where Φ is the p-dimensional $N(0, I)$ distribution function. Unfortunately, setting the design matrix to a constant decreases the correlation induced between the system. In addition, the first author has noted empirically that very nonnormal errors (e.g., uniform) also result in substantially less variance reduction. Therefore, the supplemental experiment seems to be the preferred general method for determining propoerties of Δ.

However, in many cases the supplemental run is unnecessary because the properties of Δ are known analytically. In particular, assume the design matrix is deterministic, $\eta = \eta^*$, and the errors are homogeneous normal with zero mean. The control variate estimators with the values of the known mean of Δ are given in the next four sections

for means, higher order marginal moments, the variance matrix, and percentiles.

5.1 Control Variate Estimators for the Mean

Since the estimators for nonlinear models are biased, estimating the mean of $\hat{\theta}$ is of interest. The direct estimator for the mean vector, $\mu' = \mathrm{E}[\hat{\theta}]$, is

$$\mu' = m^{-1} \sum_{s=1}^{m} \hat{\theta}_{:s} \quad ,$$

which is unbiased. The delta approximation sample mean is

$$\bar{\Delta} = m^{-1} \sum_{s=1}^{m} \Delta_{:s} \quad ,$$

which has mean θ^*. The gamma approximation sample mean is

$$\bar{\Gamma} = m^{-1} \sum_{s=1}^{m} \Gamma_{:s} \quad ,$$

whose mean is given by Box (1971), Clarke (1980), and Swain (1985). The general form of the control estimator for the mean is

$$\mu'(B) = \mu - B(C - \mathrm{E}[C]) \quad ,$$

where $C = \bar{\Delta}$ or $C = \bar{\Gamma}$.

5.2 Control Variate Estimators for Marginal Moments

Control variates for estimating $\mathrm{E}[\hat{\theta}^k]$ are developed in this section based on two generalizations of the approach just described for estimating the mean. The first, the *approximator power control*, is based on the sample average of appropriate powers of the delta or gamma approximators using the original parameterization. The second, the *power transformation control*, uses the estimators from Section 5.1 directly, except that the delta and gamma approximations arise from the new parameterization $\phi(k) \equiv \theta^k$ of the same surface η.

In the approximator power control,

$$\bar{\Delta}(k) = m^{-1} \sum_{s=1}^{m} \Delta_{:s}^{k} \quad ,$$

and

$$\bar{\Gamma}(k) = m^{-1} \sum_{s=1}^{m} \Gamma_{:s}^{k} \quad .$$

The approximator power control variate estimator for the moments is

$$\mu_k'(B) = \mu_k' - B(C - \mathrm{E}[C]) \quad .$$

where $C = \bar{\Delta}(k)$ or $C = \bar{\Gamma}(k)$. The mean $\mathrm{E}[\bar{\Delta}(k)]$ is provided in Appendix A. The mean $\mathrm{E}[\bar{\Gamma}(k)]$ is conceptually straightforward, but not algebraically appealing.

In the power transformation control,

$$\bar{\Delta}'(k) = m^{-1} \sum_{s=1}^{m} \Delta'_{:s}(k) \quad ,$$

and

$$\bar{\Gamma}'(k) = m^{-1} \sum_{s=1}^{m} \Gamma'_{:s}(k) \ ,$$

where

$$\Delta'_{:s}(k) = \phi^*(k) + [F^T(\phi^*(k)) W F(\phi^*(k))]^{-1} F^T(\phi^*(k)) \ W \ E_{:s}$$

where $F(\phi(k)) = F(\theta)\left[\frac{\partial \phi(k)}{\partial \theta}\right]^{-1}$.

Clarke (1980) derives the more complicated expressions for $\Gamma''_{:s}(k)$ as well as for any arbitrary transformation.

The equation for $\Delta'_{:s}(k)$ follows directly from the chain rule. The matrix of derivatives is the diagonal matrix whose ith diagonal element is $k^{-1}(\theta_i^*)^{(1-k)}$, so application of the chain rule is straightforward.

The power transformation control variate estimator is

$$\mu'_k(B) = \mu'_k - B \ (\ C - \mathrm{E} \ [\ C \ ]) \ .$$

where $C = \bar{\Delta}'(k)$ or $C = \bar{\Gamma}'(k)$. $\mathrm{E}\,[\bar{\Delta}'(k)] = (\theta^*)^k$. The mean for the gamma approximation estimator is given in Swain (1985).

5.3 Control Estimator for the Variance Matrix of $\hat{\theta}$

This section concentrates on control variates for estimating Var $[\hat{\theta}]$. Although variances and covariances could be estimated by differences of raw moments, such estimators are biased and have been found (Swain 1982) to have greater variance than the control variate estimators developed here. The two approaches of the last section reduce to the same method here, since the random variable of interest is $\phi(1) = \theta^1$.

Consider $\hat{\Sigma}_{\hat{\theta}}$ and $\hat{\Sigma}_C$, the usual unbiased estimators of the variance matrices Var $[\Theta]$ and Var $[C]$. Since these matrices are symmetric, let $S_{\hat{\theta}}$ and S_C be the $p(p+1)/2$ element vectors containing the lower triangular portions of $\hat{\Sigma}_{\hat{\theta}}$ and $\hat{\Sigma}_C$ stored rowwise. Let $\hat{S}_{\hat{\theta}}$ and $\hat{S}_C$ denote the direct Monte Carlo estimators. Then a control variate estimator for Var $[\hat{\theta}]$ is

$$\hat{S}_{\hat{\theta}}(B) = \hat{S}_{\hat{\theta}} - B(\hat{S}_C - \mathrm{E}\,[\hat{S}_C]) \ .$$

Either $\hat{S}_C = \hat{S}_\Delta$ or $\hat{S}_C = \hat{S}_\Gamma$, the sample estimators of Var $[\Delta]$ and Var $[\Gamma]$. When the delta approximation is used,

$$\mathrm{E}\,[\hat{\Sigma}_\Delta] = [F^T(\theta^*) \ W \ F(\theta^*)]^{-1} F^T(\theta^*) \ W \ \mathrm{Var}\,[E] \ W \ F(\theta^*) \ [F^T(\theta^*) \ W \ F(\theta^*)]^{-1} \ .$$

The mean of the gamma control variate is given in Swain (1985).

For the first time estimation of the control weights B is not straightforward via regression. Other than using the identity matrix, the other obvious option is to group the m Monte Carlo observations into k independent batches from which an estimator

can be formed.

5.4 Control Estimation of the Percentiles

The direct Monte Carlo estimator of $P(\hat{\theta} < \tau)$, where $P(\hat{\theta} < \tau)$ and τ are p-component vectors, is

$$P(\hat{\theta} < \tau) = m^{-1} \sum_{s=1}^{m} I_{\tau}(\hat{\theta}_{:s}) \ .$$

where $I_{\tau}(\hat{\theta}_{:s})$ is the p-component indicator function with elements

$$I_{\tau_i}(\theta_i) = \begin{cases} 1 & \theta_i < \tau_i \\ 0 & \text{otherwise} \end{cases} .$$

The delta and gamma approximation control variates are $\hat{P}(\Delta<\tau)$ and $\hat{P}(\Gamma<\tau)$, the observed fractions of the control samples less than τ. The Cornish-Fisher expansion can be used to approximate these probabilities for the gamma approximation control. The form of the control variate estimator is

$$\hat{P}(\hat{\theta}<\tau; B) = \hat{P}(\hat{\theta}<\tau) - B\ (\hat{P}(C<\tau) - P(C<\tau)$$

where $C = \Delta$ or $C = \Gamma$.

Indicator functions and control variate estimators are defined analogously for estimating probabilities of confidence intervals and regions.

Other control variate methods based on the delta or gamma approximations are possible. One method, suggested by Rothery (1982) in a different context, is to construct a 2x2 contingency table on $\hat{\theta}$ and Δ falling above and below their critical points τ. The maximum likelihood estimator is then the control variate. The efficiencies obtained by this method are comparable to the linear control formulation.

Estimation of the α^{th} quantile of $\hat{\theta}$, $\hat{\theta}(\alpha)$ satisfying $P(\hat{\theta}<\hat{\theta}(\alpha)) = \alpha$, is more difficult than estimation of percentiles, despite the close relationship between percentiles and quantiles. $\hat{\theta}(\alpha)$ can be estimated using interpolation based on percentile estimates, order statistics, or stochastic approximation. Control variate methods based on these estimators can be formed in the usual way, although again the mean of the control variable must be known or a supplemental experiment performed. Some details are given in Swain (1982).

APPENDIX: LINEAR POWER-CONTROL EXPECTATIONS

The k*th* power of the linear power control variate required in Section 5.2 is $\Delta^k = [\theta^* + J\,E]^k$, where componentwise exponentiation is indicated, and $J = [F^T(\theta^*)WF(\theta^*)]^{-1}\,F^T(\theta^*)\,W$ is the transformation for the linear approximation. The l*th* component of the vector Δ^k is

$$\Delta_i^k = [\theta_i^* + J_i E]^k = \sum_{j=0}^{k} \binom{k}{j} (\theta_i^*)^j\ (J_i E)^{k-j} \qquad i=1,2,\cdots,p\ , \tag{A.1}$$

where J_i is the l*th* row of J. If J is constant and $E \sim N(0, I\sigma^2)$, then

$$\mathbf{E}\,[(J_i E)^1] = 0,$$

$$\mathbf{E}\,[(J_i E)^2] = \sigma^2 \sum_{j=1}^{p} J_{ij}^2,$$

$$\mathbf{E}\,[(J_i E)^3] = 0,$$

and

$$\mathbf{E}\,[(J_i E)^4] = \sigma^4 \left[6 \sum_{j=1}^{p-1} \sum_{j'=j+1}^{p} J_{ij}^2 J_{ij'}^2 + 3 \sum_{j=1}^{p} J_{ij}^4 \right],$$

where J_{ij} is the jth component of J_i. Substitution into equation (A.1) yields for $i = 1,2,\cdots,p$

$$\mathbf{E}\,[\Delta_i] = \theta_i^*,$$

$$\mathbf{E}\,[\Delta_i^2] = (\theta_i^*)^2 + \sigma^2 \sum_{j=1}^{p} J_{ij}^2,$$

$$\mathbf{E}\,[\Delta_i^3] = (\theta_i^*)^3 + 3\theta_i^* \sigma^2 \sum_{j=1}^{p} J_{ij}^2,$$

and

$$\mathbf{E}\,[\Delta_i^4] = (\theta_i^*)^4 + 6(\theta_i^*)^2 \sigma^2 \sum_{j=1}^{p} J_{ij}^2 + \sigma^4 \left[6 \sum_{j=1}^{p-1} \sum_{j'=j+1}^{p} J_{ij}^2 J_{ij'}^2 + 3 \sum_{j=1}^{p} J_{ij}^4 \right].$$

REFERENCES

Anderson, T.W. (1958), *An Introduction to Multivariate Statistical Analysis*, John Wiley & Sons, New York.

Andrews, D.F., P.J. Bickel, F.R. Hampel, P.J. Huber, W.H. Rogers, J.W. Tukey (1972), *Robust Estimates of Location*, Princeton University Press, Princeton.

Arnold, H.J., B.D. Bucher, H.F. Trotter, and J.W. Tukey (1956), "Monte Carlo Techniques in a Complex Problem About Normal Samples," *Symposium on Monte Carlo Methods*, Meyer (Ed.), John Wiley & Sons, New York, pp. 80-88.

Bard, Y. (1974), *Nonlinear Parameter Estimation*, Academic Press, New York.

Bates, D.M., and D.G. Watts (1980), "Relative Curvature Measures of Nonlinearity," *Journal of the Royal Statistical Society B 42*(1), pp. 1-25.

Beale, E.M.L. (1960), "Confidence Regions in Non-linear Estimation," *Journal of the Royal Statistical Society B 22*(1), pp. 41-88.

Box, M.J. (1971), "Bias in Nonlinear Estimation," *Journal of the Royal Statistical Society B 33*, pp. 171-201.

Bratley, P., B.L. Fox, and L.E. Schrage (1983), *A Guide to Simulation*, Springer-Verlag, New York.

Chambers, J.M. (1977), *Computational Methods for Data Analysis,* John Wiley & Sons, New York.

Cheng, R.C.H. (1978), "Analysis of Simulation Experiments Under Normality Assumptions," *Journal of the Operational Research Society 29*(5), pp. 493-497.

Clarke, G.P.Y. (1980), "Moments of the Least Squares Estimators in a Non-linear Regression Model," *Journal of the Royal Statistical Society B 42*(2), pp. 227-237.

Crowder, M.J. (1976), "Maximum Likelihood Estimation for Dependent Observations," *Journal of the Royal Statistical Society B 38*(1), pp. 45-53.

Fishman, G.S. (1978), *Principles of Discrete Event Simulation,* John Wiley & Sons, New York.

Gallant, A.R. (1975), "Nonlinear Regression," *The American Statistician 29*(2), pp. 73-81.

Gallant, A.R. (1980), "Explicit Estimators of Parametric Functions in Nonlinear Regression," *Journal of the American Statistical Society 75*(369), pp. 182-193.

Gillis, P.R., and D.A. Ratkowsky (1978), "The Behaviour of Estimators of the Parameters of Various Yield-Density Relationships," *Biometrics 34,* p. 191.

Hammersley, J.M., and D.C. Handscomb (1964), *Monte Carlo Methods,* Chapman and Hall, London.

Hartley, H.O. (1977), "Solution of Statistical Distribution Problems by Monte Carlo Methods," *Statistical Methods for Digital Computers,* Vol. III, Enslein, Ralston, Wilf (Eds.), John Wiley & Sons, New York.

Kahn, H. (1956), "Use of Different Monte Carlo Sampling Techniques," from *Symposium on Monte Carlo Methods,* Meyer (Ed.), John Wiley & Sons, New York, pp. 146-190.

Kleijnen, J.P.C. (1974), *Statistical Techniques in Simulation, Part I,* Marcel Dekker, New York.

Kleijnen, J.P.C. (1977), "Robustness of a Multiple Ranking Procedure: A Monte Carlo Experiment Illustrating Design and Analysis Techniques," *Communications in Statistics B, Simulation and Computation 6*(3), pp. 235-262.

Koehler, K.J. (1981), "An Improvement of a Monte Carlo Technique Using Asymptotic Moments with an Application to the Likelihood Ratio Statistic," *Communications in Statistics B, Simulation and Computation 10*(4), pp. 343-357.

Lavenberg, S.S., and P.D. Welch (1981), "A Perspective on the Use of Control Variables to Increase the Efficiency of Monte Carlo Simulations," *Management Science 27*(3), pp. 322-335.

Law, A.M., and W.D. Kelton (1982), *Simulation Modeling and Analysis,* McGraw-Hill, New York.

McGrath, E., and D. Irving (1973), "Techniques for Efficient Monte Carlo Simulation, Vol. III, Variance Reduction," *ORLC Report* SAI-72-509-LJ.

Nelson, B.L. (1983), "Variance Reduction in Simulation Experiments: A Mathematical-Statistical Framework," unpublished Ph.D. dissertation, School of Industrial Engineering, Purdue University.

Nelson, B.L. and B.W. Schmeiser (1984a), "A Mathematical-Statistical Framework for Variance Reduction, Part I: Simulation Experiments," Research Memorandum 84-4, School of Industrial Engineering, Purdue University.

Nelson, B.L. and B.W. Schmeiser (1984b), "A Mathematical-Statistical Framework for Variance Reduction, Part II: Classes of Transformations," Research Memorandum 84-5, School of Industrial Engineering, Purdue University.

Nelson, B.L. and B.W. Schmeiser (1985), "Decomposition of Some Well-Known Variance Reduction Techniques," *Journal of Statistical Computation and Simulation*, forthcoming.

Porta Nova, A.M. (1985), "A Generalized Approach to Variance Reduction in Discrete-Event Simulation Using Control Variables," Ph.D. Dissertation, Department of Mechanical Engineering, The University of Texas, Austin, TX 78712.

Relles, D.A. (1970), "Variance Reduction Techniques for Monte Carlo Sampling from Student Distributions," *Technometrics 12*, pp. 499-515.

Rothery, P. (1982), "The Use of Control Variates in Monte Carlo Estimation of Power," *Applied Statistics 31*(2), pp. 125-129.

Rubinstein, R.Y., and R. Marcus (1985), "Efficiency of Multivariate Control Variates in Monte Carlo Simulation," *Operations Research 33*(3), pp. 661-677.

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*, pp. 504-520.

Shenton, L.R., and K.O. Bowman (1977), *Maximum Likelihood Estimation in Small Samples*, McMillan, New York.

Swain, J.J. (1982), "Monte Carlo Estimation of the Sampling Distribution of Nonlinear Parameter Estimators," Unpublished Ph.D. Thesis, School of Industrial Engineering, Purdue University.

Swain, J.J. (1985), "Quadratic Approximators for Monte Carlo Studies of Nonlinear Model Parameter Estimators," School of Industrial and Systems Engineering, Georgia Institute of Technology, Atlanta, Georgia.

Swain, J.J., and B.W. Schmeiser (1983), "Monte Carlo Estimation of the Sampling Distribution of Nonlinear Model Parameter Estimators," *Annals of Operations Research*, forthcoming.

Tocher, K.D. (1963), *The Art of Simulation*, D Van Nostrand Company, Princeton, New Jersey.

Venkatraman, S., and J.R. Wilson (1985), "The Efficiency of Control Variates in Multiresponse Simulation," School of Industrial Engineering, Purdue University.

Wilks, S.S. (1962), *Mathematical Statistics,* John Wiley & Sons, New York.

Wilson, J.R. (1984), "Variance Reduction Techniques for Digital Simulation," *American Journal of Mathematical and Management Sciences 4,* 3&4, 277-312.

Wu, C.F. (1981), "Asymptotic Theory of Nonlinear Least Squares Estimation," *Annals of Statistics 9,* pp. 501-513.

Unclassified

SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

REPORT DOCUMENTATION PAGE		READ INSTRUCTIONS BEFORE COMPLETING FORM
1. REPORT NUMBER 85-10	2. GOVT ACCESSION NO. AD-A158 613	3. RECIPIENT'S CATALOG NUMBER
4. TITLE (and Subtitle) Monte Carlo Analysis of Nonlinear Statistical Models, I: Theory		5. TYPE OF REPORT & PERIOD COVERED Technical Report 6. PERFORMING ORG. REPORT NUMBER 85-10
7. AUTHOR(s) James J. Swain Bruce W. Schmeiser		8. CONTRACT OR GRANT NUMBER(s) N00014-79-C-0832
9. PERFORMING ORGANIZATION NAME AND ADDRESS Purdue Research Foundation School of Industrial Engineering Purdue University, West Lafayette, IN 47907		10. PROGRAM ELEMENT, PROJECT, TASK AREA & WORK UNIT NUMBERS NR 042-477
11. CONTROLLING OFFICE NAME AND ADDRESS Office of Naval Research -- Code 411 800 North Quincy Street Arlington, VA 22217		12. REPORT DATE August 1985 13. NUMBER OF PAGES 17
14. MONITORING AGENCY NAME & ADDRESS(if different from Controlling Office)		15. SECURITY CLASS. (of this report) Unclassified 15a. DECLASSIFICATION/DOWNGRADING SCHEDULE

16. DISTRIBUTION STATEMENT (of this Report)

Approved for Public Release: Distribution Unlimited.

17. DISTRIBUTION STATEMENT (of the abstract entered in Block 20, if different from Report)

18. SUPPLEMENTARY NOTES

19. KEY WORDS (Continue on reverse side if necessary and identify by block number)

Control Variates
Simulation
Monte Carlo
Nonlinear Estimation
Variance Reduction

20. ABSTRACT (Continue on reverse side if necessary and identify by block number)

Parameters of nonlinear statistical models are typically estimated from data using iterative numerical procedures. The resulting joint sampling distribution of the parameter estimators is often intractable, resulting in the use of approximators or Monte Carlo simulation to determine properties of the sampling distribution. This paper develops methods, using linear and quadratic approximators as control variates, that reduce the variance of the Monte Carlo estimator by orders of magnitude. Estimation of means, higher order raw moments, variances, covariances, and percentiles is considered.

DD FORM 1 JAN 73 1473 EDITION OF 1 NOV 65 IS OBSOLETE
S/N 0102-LF-014-6601

Unclassified

SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

END

FILMED

10-85

DTIC