Proceedings of the 1989 Winter Simulation Conference
E.A. MacNair, K.J. Musselman, P. Heidelberger (eds.)

AUGMENTING LINEAR CONTROL VARIATES USING TRANSFORMATIONS

James J. Swain
Georgia Institute of Technology
Atlanta, GA 30332

ABSTRACT

Linear control variates provide a convenient means for variance reduction in a wide variety of simulation and Monte Carlo sampling problems. The effectiveness of control variates can often be improved using appropriate transformations which increase the correlation between the primary variate and the transformed control variate. Possible transformations include polynomials, the Box-Cox family of power transformations, and the inverse-normal probability transformation. Here we consider generalized transformations using cubic splines which provide a virtually unrestricted source of transformations. To illustrate the transformation methodology, a nonlinear regression problem is used. In the illustration, the transformed control is compared to the control when estimating the mean of the sampling distribution of the estimated parameters.

1 INTRODUCTION

Control variates are used in many statistical sampling experiments to improve the accuracy of the solution. For instance, an estimate of the mean μ of a p-dimensional variate Y may be desired. A direct estimator is $\bar{Y}$. This estimator can often be improved provided that a q-dimensional *control variate* C is available with the property that C has a known expectation (without loss of generality we may take the expectation to be zero) and is correlated with Y. The well-known controlled estimate of μ is

$$\bar{Y}(B) = \bar{Y} - B\bar{C} \qquad (1)$$

where B is a p by q weight matrix to be specified. The variance of this estimator is readily shown to be

$$\begin{aligned} Var[\bar{Y}(B)] &= Var[\bar{Y}] + BVar[\bar{C}]B^T \\ &\quad -BCov[\bar{Y},\bar{C}] - Cov[\bar{C},\bar{Y}]B^T \qquad (2) \end{aligned}$$

which can be minimized by the choice, $B^* = Cov[\bar{Y},\bar{C}]Var^{-1}[\bar{C}]$. For further details, see Venkatraman and Wilson (1986).

There is a close relation between control variates and linear regression; both rely on the same general theory, though there are subtle differences as well, as Cheng (1978) points out. When the regression of Y on C is linear, B^* provides the least variance estimate of μ. Since $Cov[Y,C]$ is generally unknown, B must be estimated. The standard estimator $\hat{B}$ of B^* is obtained by linear regression of the components of Y upon the controls C. This may introduce a bias in $\hat{\mu}$, since $\hat{B}$ and $\bar{C}$ are uncorrelated only if Y and C are multivariate normal or the conditional distribution of Y given C is independent of C, Cheng (1978). Nelson (1988) discusses these threats to control variate theory and recommends remedies for practice.

As with regression modelling, the residual variation can be reduced by the inclusion of further terms (for instance, polynomials of the controls, or the inclusion of additional controls), or through appropriate transformations of the controls. In a typical regression analysis it is also possible to transform the dependent variate Y, but this is rarely done in simulation studies, since the properties of Y are of direct interest, not those of a transformed variate. We may freely transform C, however, provided that the correlation between it and Y is not decreased and the expectation of the transformed control is known or at least computable.

Transformations are discussed in the next section. In the third section an application involving a sampling distribution problem in nonlinear regression is introduced to illustrate the construction and effectiveness of the proposed transformations. A brief discussion about future work concludes the paper.

2 TRANSFORMATIONS

There are several general approaches that are used in transformations. Classical regression analysis suggests the use of power transformations, such as the Box-Cox family. These transformations have the property of shrinking or expanding the control by varying amounts, either to improve its correlation with Y or in order to improve its normality. When the distribution of the control is known, an inverse-normal probability transformation can also be used for this purpose. Finally, non-parametric transformations are possible in which the form the transformation is determined directly from the data, without restriction other than smoothness.

2.1 Parametric Transformations

Power transformations are often used in standard regression practice to improve the fit of an empirical model. Box and Draper (1987) provide a good introduction. The Box-Cox family provide a parametric version of power models, including the log transformation as a special case, and suitable for estimation to obtain an "optimal" transformation. Since the response variable Y is not transformed, the problem is roughly one of linear regression variable selection (if variables are to be added), or nonlinear regression (if an existing control is to be transformed), since the parameters of the transform would enter the model nonlinearly.

Transformations of this type often arise in statistical sampling, where polynomial or power approximations of a statistic are the basis of control variates. One practical difficulty arises in the evaluation of the expectations, particularly when more than the first moment is required. Further moments can be obtained numerically, provided the dimensionality of the inputs is not too large. For instance, in a nonlinear regression problem, Swain and Goldsman (1989) are only able to derive expectations up to the third moment for a quadratic control variate, while Swain (1988b) resorts to numerical quadrature to obtain the necessary expectations of a similar control variate.

In certain cases, distributional transformations such as the logit and inverse normal arise naturally. For instance, many statistics are approximately normal for a wide range of inputs. Swain (1988a) illustrates this approach for a problem with small sample sizes. One difference between the distributional transformations and the power transformations is what happens in the middle of the domain of the transformations: power transformations monotonically stretch or shrink the variate, while a prcbability transform may stretch the ends and shrink the middle. More general transformations than the normal could be used; the Johnson family of distributions is one possibility.

2.2 Nonparametric Approaches

The classical approaches considered so far consist of a transformational form, usually indexed by unknown parameters, which may then be set by judgement or estimation to specify the precise transformation to use. While this may replace a linear estimation problem with a nonlinear one, the additional computational effort is generally not substantial. An alternative approach is *nonparametric regression* where the response is considered to be additive in smooth functions not specified in advance. These *computer intensive methods*, Thisted (1988), generally impose few restrictions on the regression function in advance, aside from imposing a roughness penalty. This approach provides considerable flexibility at the expense of additional computational effort.

For purposes of illustration, natural cubic splines are used in this paper to demonstrate the possibilities for determining a transformation. The cubic spline has an implicit smoothness, since both first and second derivatives are continuous. A somewhat more general approach is discussed in Lewis, Ressler, and Wood (1989), where piecewise power-transformed variables are used as the controls. The transformations are nonparametric in the sense that the break points in the approximation are selected by the algorithm. They also use the alternating conditional expectations (ACE) method of Breiman and Friedman (1985) as a basis for comparison, since in principle ACE permits any nonparametric transformation. In general, ACE permits transformations of both response and regressor variables, though in their example transformations of Y were not used.

Nonparametric regression can be used to form a control variate, provided that the expectation of the new control can be evaluated. In many simulations both Y and C are functions of n-dimensional inputs U, where $n \gg \max(p,q)$, so that a potential n-dimensional integration problem arises. However, provided the controls C can be constructed or transformed to a known distribution as a function of the inputs U, the dimensionality can be reduced to a feasible level, and the expectations provided analytically or numerically.

3 APPLICATION

The sampling distribution of parameters obtained via nonlinear regression is difficult to determine for small samples. Characteristics of the sampling distribution, such as the mean, variance, and other moments, and quantiles, provide ways to evaluate the accuracy and precision of estimation. Control variates based upon linear approximations are introduced in the next section, and a method for obtaining a transformed control, without further sampling, is introduced in the following subsection.

3.1 Nonlinear Parameter Estimation

In the nonlinear regression problem, a response $W = (W_1, \ldots, W_n)$ is a nonlinear function of unknown parameters θ. The individual responses are given by

$$W_i = \eta(x_i; \theta_0) + E_i \tag{3}$$

for errors $E = (E_1, \ldots, E_n)$, generally assumed to be independently and identically normally distributed with mean 0 and common variance σ^2. Under these assumptions, the nominal value θ_0 of the parameters θ can be estimated via least-squares, with $\hat{\theta}$ minimizing the sum of the squared residuals,

$$S(\theta) = \sum_{i=1}^{n}(w_i - \eta(x_i; \theta))^2 \tag{4}$$

In a sampling experiment, R replications of n-observations each are obtained, from which R estimates, $\hat{\theta}_i, i = 1, \ldots, R$ can be computed. The moments and quantiles of the $\hat{\Theta}$ distribution can be estimated using appropriate sample statistics; for instance, the mean is estimated using $\bar{Y} = R^{-1}\sum_{i=1}^{R}\hat{\theta}_i$. That is, Y is some function of the estimated $\hat{\theta}_i$ (which depend in turn upon the inputs, E).

Control variates are readily constructed in this case, and are documented in Swain and Schmeiser (1989, 1988), for instance. A linear approximator Δ is computed using

$$\Delta = \theta_0 + [F^T(\theta_0)F(\theta_0)]^{-1}F^T(\theta_0)E \tag{5}$$

where the errors E are the ones used to generate W, θ_0 is the nominal value of θ to be estimated, and $F(\theta_0)$ is the n by p Jacobian matrix of first derivatives, $\| \partial\eta(x_i; \theta)/\partial\theta_j \|$, with the derivatives evaluated at θ_0. The approximator is constructed by making a linear approximation of $\eta(x; \theta)$ at $\theta = \theta_0$ as a function of Δ, and solving the resultant linear least-squares problem for Δ. Note that in a Monte Carlo experiment θ_0 is known. When the errors E are $\mathrm{Nor}[0, \sigma^2 I]$, Δ has a normal distribution with mean θ_0 and variance matrix $\sigma^2[F^T(\theta_0)F(\theta_0)]^{-1}$.

Linear approximation is widely used in the analysis of nonlinear models, both as the basis of optimization and in the approximation of the statistical properties of the estimated parameters. Nonlinear models can be classified according to the divergence between the properties of their estimators and the properties of the approximating linear model. The divergence arises from two sources – *intrinsic nonlinearity* due to the curvature of the nonlinear response surface (the linear response surface is planar) and *parameter effects nonlinearity* due to the uneven spacing between coordinate lines on the response surface (coordinate lines are equally spaced in the linear case). These effects have been quantified by Beale (1960) and most recently, using differential geometry, by Bates and Watts (1980). In a survey of models, Bates and Watts report that parameter effects nonlinearity is typically the more significant problem when models diverge from linearity.

3.2 Construction of Transformed Controls

The linear approximation control variate Δ is effective in Monte Carlo studies involving nonlinear parameter estimators, Swain and Schmeiser (1988). In many of the cases studied the optimal control weight matrix is very close to the identity matrix, I. Use of a constant weight matrix eliminates the difficulties of computing the weights and the bias problem caused when $\bar{C}$ and $\hat{B}$ are estimated using the same sample. The loss of efficiency is generally less than 5%.

Transformations can be used to further improve the effectiveness of linear control variates. In certain cases with non-normal errors, transformations of the errors E to normality has been effective, Swain (1988a), but in general the relation between $\hat{\theta}$ and the errors or the linear approximator is complicated and not easily characterized. Thus recourse to nonparametric functions is a useful alternative.

Three approaches may be used to obtain the δ, θ pairs for the regression. One could obtain a preliminary, independent sample of $\hat{\Theta}$ and Δ, from which p univariate transformations would be obtained. This reduces the overall effectiveness of the their use, since it involves additional sampling, but in a large study could be effective. Additionally, one can examine the correspondence between the two response surfaces in terms of their "coordinates". For instance, any point on the nonlinear response surface can be linearly projected onto the linear response plane and its δ coordinates determined by a second linear transformation. By following the changes in one θ coordinate (with the others fixed), one could obtain values of δ that match each one. A slightly more difficult approach is to start with values of δ and to determine the values of θ that most closely correspond using nonlinear least-squares. The contrast between this approach and the sampling approach discussed initially (they both involve nonlinear least-squares) can be likened to the difference between "unplanned" and "designed" regression experiments: both do about the same thing, but the latter can do it more efficiently. In this case, since the solution to the last point will be close to the next solution (in terms of θ parameters), the least squares estimation should progress quite rapidly.

Whatever approach is used, the new approximator, π, can be constructed using a nonparametric regression of the θ upon the δ. It appears that in cases of low and moderate nonlinearity (where the linear approximation works well), any of the approaches will suffice. For cases of higher nonlinearity the third approach is preferable. As a practical matter, the experimenter will be observing the control variate, so it makes sense to design the experiment in the units to be observed. In addition, in cases of high nonlinearity, the coordinate lines on the nonlinear response surface arch and bend, and in extreme cases (far from the nominal point $\eta(X;\theta_0)$) can re-cross δ coordinate lines, rendering the inverse mapping impossible. Thus a matching based upon δ coordinates works best in practice.

3.3 Monte Carlo Results

To demonstrate the potential of the transformed controls, two nonlinear models are used as examples. These are the B and C reactant models of Guttman and Meeter (1965). These two models represent the normalized concentrations of two reactants B and C in the decomposition $A \rightarrow B \rightarrow C$, with initial A concentration of 1, and no B or C. Under linear decomposition the two responses are:

$$\eta_B = \frac{\theta_1}{\theta_1 - \theta_2}(e^{-\theta_2 t} - e^{-\theta_1 t}) \tag{6}$$

$$\eta_C = 1 - \frac{1}{\theta_1 - \theta_2}(\theta_1 e^{-\theta_2 t} - \theta_2 e^{-\theta_1 t}) \tag{7}$$

While the two models appear similar, the B response model possesses moderate nonlinearity, while the C response model has high parameter effects nonlinearity. The high nonlinearity is due to a negative correlation between the two parameters and insensitivity in the response parameters to the value of the θ_1.

Two experiments are performed using these models. For purposes of illustration, the mean of the parameter estimator $\hat{\Theta}$ is estimated, though in practice other statistics about the estimator would be desired. In both cases the nominal value of the parameters is $\theta_0^T = (1.4, 0.4)$. The simulations were performed using the Fortran code provided by the Ratkowsky (1983) text, with random numbers from the IMSL generator GGUBS.

The B response model has fairly moderate nonlinearity, and the linear approximator performs very effectively for this model. The experimental design (in the time variable, t) is $X^T = (0.5, 1., 2., 4., 6.)$ for the nonlinear estimation problem, and $\sigma^2 = .0004$ for the error distribution. In this experiment, a replication consisted of 100 sets of 5 y_i observations, from which 100 estimates $\hat{\theta}$ were obtained, together with 100 observations of the two controls, Δ and Π. From these observations a replicate of the estimated means and variances are then obtained for the direct and control estimates. These observations were replicated 100 times to obtain the average variances and from them the average efficiencies. For instance, based on these 10,000 regressions, the direct estimator $\bar{\theta}$ and the identity controls $\bar{\theta}(\delta) = \bar{\theta} - (\bar{\delta} - E[\bar{\delta}])$ and $\bar{\theta}(\pi) = \bar{\theta} - \bar{\pi}$ have the following variances:

$$Var[\bar{\theta}] = \begin{bmatrix} .0039 & .00017 \\ .00017 & .00021 \end{bmatrix}$$

$$Var[\bar{\theta}(\delta)] = \begin{bmatrix} .000014 & .00000053 \\ .00000053 & .00000037 \end{bmatrix}$$

$$Var[\bar{\theta}(\pi)] = \begin{bmatrix} .0000071 & .00000053 \\ .00000053 & .00000013 \end{bmatrix}$$

The determinants of these three quantities are respectively 1.1E-6, 5.1E-12, and 6.8E-13. The use of Δ controls alone results in an efficiency of about 200,000, and the transformed controls increases this approximately an order of magnitude to 1.6 million. Both components of the mean are estimated more precisely with the transformed controls than with the Δ controls so that the relevant confidence intervals would be 30 % to 40 % shorter than with the Δ controls alone.

Because of the higher nonlinearity, the C response model has lower efficiencies. The design for the regression experiment is $X^T = (1, 2, 4, 6, 8)$ and the error variance is $\sigma^2 = .0001$. The same procedure is repeated, with 100 replications, each replication consisting of 100 regressions. A number of difficulties were encountered in this illustration. The range of the Δ approximator includes values that are infeasible, since the responses in equations (6) and (7) do not apply when $\theta_1 = \theta_2$, and attention is restricted to the region $\theta_1 > \theta_2$. However, the sampled δ include values outside the region in which the transformation is fit, requiring some method of extrapolation. The method applied here is to extrapolate linearly from the first (or last) two data pairs for that coordinate. This choice performed more effectively than using linear extrapolation based on the spline coefficients. The averaged variance matrices for the three estimators were:

$$Var[\bar{\theta}] = \begin{bmatrix} .026 & -.0026 \\ -.0026 & .00033 \end{bmatrix}$$

$$Var[\bar{\theta}(\delta)] = \begin{bmatrix} .00023 & .000018 \\ .000018 & .0000064 \end{bmatrix}$$

$$Var[\bar{\theta}(\pi)] = \begin{bmatrix} .00013 & -.0000010 \\ -.0000010 & .0000020 \end{bmatrix}$$

The determinants of these three quantities are respectively 1.7E-5, 1.2E-8, and 2.7E-9. Thus the use of Δ controls alone results in an efficiency of about 1,500, and the transformed controls increases this approximately half an order of magnitude to 6,500. The large parameter effects nonlinearity could be the reason that the transformed controls are not more successful, and it is likely that a transformation that uses both components of Δ would be more successful in this case. Even so, the components of the mean are again estimated about 30 % more precisely with the transformed controls than with the Δ controls alone.

4 CONCLUSION

The use of nonparametric transformations is promising as a method of improving the effectiveness of linear control variates. Further improvements appear possible when multivariate transformations are used using nonparametric regression or the method of Lewis, Ressler, and Wood (1989). In addition, these methods should be of even greater use when used to estimate statistics such as quantiles, where the relation between the control and the primary variate is less likely to straight.

REFERENCES

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

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

Box, G.E.P., and N.R. Draper (1987). *Empirical Model-Building and Response Surfaces.* John Wiley, New York.

Breiman, L., and J.H. Friedman (1985). Estimating Optimal Transformations for Multiple Regression and Correlation. *Journal of the American Statistical Association*, **80**, pp. 580-597.

Cheng, R.C.H. (1978). Analysis of Simulation Experiments Under Normality Assumptions. *Journal of the Operational Research Society*, **20**, pp. 493-497.

Guttman, I. and D.A. Meeter (1965). On Beale's Measure of Nonlinearity. *Technometrics*, **7**, pp. 623-637.

Lewis, P.A.W., R.L. Ressler, and R.K. Wood (1989). Variance Reduction Using Nonlinear Controls and Transformations. *Communications in Statistics*, **B**, to appear.

Nelson, B.L. (1988). Control-Variate Remedies. Working Paper Series 1988-004, Department of Industrial and Systems Engineering, The Ohio State University, Columbus, OH.

Ratkowsky, D.A. (1983). *Nonlinear Regression Modeling.* Dekker, New York.

Swain, J.J. (1988a). Control Variates for Monte Carlo Analysis of Nonlinear Statistical Models, IV: Nonnormal Error Distributions. *Communications in Statistics*, **B 17**, pp. 251-274.

Swain, J.J. (1988b). Higher Order Approximators for Monte Carlo Study of Nonlinear Regression Estimators. Technical Report J-88-2, School of Industrial and Systems Engineering, Georgia Institute of Technology, Atlanta, GA.

Swain, J.J., and D.M. Goldsman (1989). Moments of Second Order Polynomials with Simulation Applications. *Institute of Industrial Engineering Transactions*, to appear.

Swain, J.J., and B.W. Schmeiser (1989). Control Variates for Monte Carlo Analysis of Nonlinear Statistical Models, I: Overview. *Communications in Statistics*, **B 18**, to appear.

Swain, J.J., and B.W. Schmeiser (1988). Control Variates for Monte Carlo Analysis of Nonlinear Statistical Models, II: Raw Moments and Variances. *Journal of Statistical Computation and Simulation*, **29**, pp. 251-274.

Thisted, R.A. (1988). *Elements of Statistical Computing.* Chapman and Hall, New York.

Venkatraman, S., and J.R. Wilson (1986). The Efficiency of Control Variates in Multiresponse Simulation. *Operations Research Letters*, **5**, pp. 37-42.

AUTHOR'S BIOGRAPHY

JAMES J. SWAIN is an Assistant Professor in the School of Industrial and Systems Engineering at the Georgia Institute of Technology. From 1977 to 1979 he was a Systems Analyst in the Management Information Department of Air Products and Chemicals, Allentown, PA. He received a BA in Liberal Studies in 1974, a BS in Engineering Science in 1975, and an MS in Mechanical Engineering in 1977 from the University of Notre Dame. He received his Ph.D. in Industrial Engineering from Purdue University in 1982. His current research interests include the anlysis of nonlinear regression models, Monte Carlo variance reduction methods in statistical problems, and numerical methods. He is a member of ASA, ORSA, SCS, and TIMS.

James J. Swain

School of ISYE

Georgia Tech

Atlanta, GA 30332

(404) 894-3025

jswain@gtri01.bitnet