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

# VARIANCE REDUCTION OF QUANTILE ESTIMATES VIA NONLINEAR CONTROLS

Peter A. W. Lewis
Richard L. Ressler
Naval Postgraduate School
Monterey, CA 93943

## ABSTRACT

Linear controls are a well known technique for achieving variance reduction in computer simulation. Unfortunately the effectiveness of a linear control depends upon the correlation between the statistic of interest and the control which is often low. Since statistics are often nonlinear functions of the control this implies that nonlinear controls offer a means for improvement over linear controls. Nonlinear controls have had success in increasing the variance reduction over a linear control. This current work focuses on the use of nonlinear controls for reducing the variance of quantile estimates. The paper begins with a short discussion of linear controls. It describes nonlinear controls and the possibility for improved performance. The final sections discuss quantiles as controls and the potential of nonlinear controls for variance reduction in quantile estimation.

## 1. LINEAR CONTROLS

In the usual scenario one conducts a simulation to estimate an unknown quantity $\mu$ using a random variable $Y$ as the estimator. The simulation consists of replicating $n$ samples of $Y$, i.e. $Y_i, i = 1, \ldots, n$ and using these to estimate $\mu$. Linear control schemes rely on the availability of a random variable $C$, with known expected value, which is correlated with $Y$. $C$ is the control variable.

The standard linear control scheme for a single control uses the linear additive combination

$$Y' = Y - \beta(C - \mathrm{E}[C])$$

to minimize the variance of the controlled estimate $Y'$. The value of $\beta$ which maximizes the variance reduction is $\beta = \mathrm{cov}(Y, C)/\mathrm{var}(C)$. For the multiple control case when $C$ and $\beta$ are vectors, $\underline{C}$ and $\underline{\beta}$, control equation becomes

$$Y' = Y - \underline{\beta}^T(\underline{C} - \mathrm{E}[\underline{C}]) \tag{1}$$

and the optimal values for $\underline{\beta}$ are the canonical correlation coefficients. A common measure of effectiveness of a linear control scheme using the optimal value for $\beta$ is

$$\frac{\mathrm{var}(Y')}{\mathrm{var}(Y)} = \left(1 - \rho^2\right) \tag{2}$$

where $\rho$ is the correlation between $Y$ and $C$. For multiple controls $\rho$ is replaced by the multiple correlation coefficient. (2) implies that one should choose controls that are "highly" correlated with $Y$.

## 2. NONLINEAR CONTROLS

One can generalize (1) to $Y'' = Y - C''$ where $C''$ is some function of $\underline{C}$, and a set of parameters $\beta^*$, i.e. $C'' = h(\underline{C}, \beta^*)$. $C''$ could include linear as well as nonlinear terms each with their own parameters. $C''$ will be referred to as the control function. A control function with terms that are nonlinear in the unknown parameters is a nonlinear control. In some simulations possible control variables may have low correlation with $Y$. Two of the possible sources for the low correlation are:

1. there is in fact very little structural relationship between $Y$ and the control i.e. a bivariate scatter plot of $Y$ versus $C$ would look patternless, or
2. the structural relationship between $Y$ and $C$ is of a nonlinear form which is poorly approximated by a straight line.

In the first case, a nonlinear control may or may not offer improvement over the linear control. In the second case, a nonlinear control can offer substantial improvement in variance reduction as in Lewis, Ressler and Wood (1989).

A simple example will show the potential benefits of nonlinear transformations. Let $X$ be a Normal (0,1) random variable and let $Y = X^2$.

$$\mathrm{cov}(Y, X) = \mathrm{E}\left[X^3\right] - \mathrm{E}\left[X^2\right]\mathrm{E}[X] = 0$$

and $Y$ and $X$ are uncorrelated. Now allow the nonlinear transformation $C = X^p$ with $p = 2$. $C$ is a $\chi_1^2$ random variable with mean 1 and variance 2. It follows that

$$\operatorname{cov}(Y, C) = \operatorname{var}\left(X^2\right) = 2 \Longrightarrow \rho(Y, C) = \frac{2}{2} = 1.$$

Therefore when evaluating a potential control, one should ask: *Can this control be transformed so it will have a "high" correlation with Y*.

Let $\theta(Y)$ and $\phi(C)$ be mean-zero functions of random variables $Y$ and $C$ such that $\operatorname{var}(\theta(Y)) = 1$ and $\operatorname{var}(\varphi(C)) < \infty$. Breiman and Friedman (1985) showed that transformations which maximize the correlation between $\theta(Y)$ and $\phi(C)$ exist and are, in the bivariate case, the conditional expected values :

$$\theta(Y) = \frac{\mathrm{E}[\phi(C) \mid Y]}{\|\mathrm{E}[\phi(C) \mid Y]\|}$$

where $\|\cdot\| = \left\{\mathrm{E}\left[\cdot^2\right]\right\}^{1/2}$, and

$$\phi(C) = \mathrm{E}[\theta(Y) \mid C].$$

For multivariate $\underline{C}$, transformations of the components of $\underline{C}$ which maximize the correlation of $Y$ with a linear combination of the transformed components also exist. These transformations are also optimal for variance reduction and may be linear or nonlinear. Transforming $Y$ is beyond the scope of this paper so we will keep $\theta(Y) = Y$.

Lancaster (1966) has shown that if $Y$ and $\underline{C}$ have a multivariate normal distribution, the $h(\underline{C}, \beta^*)$ which maximizes the correlation between $Y$ and $h(\underline{C}, \beta^*)$ over all square summable $h(\underline{C}, \beta^*)$ is the additive linear control scheme (1) using the cannonical correlation coefficients for $\beta^* = \beta$. Whenever the joint distribution of $Y$ and $\underline{C}$ is not multivariate normal, a nonlinear control offers the possibility for improvement over a linear control.

Analytically determining the optimal transformation requires the joint distribution of $Y$ and $\underline{C}$ which is unknown. For linear controls a standard technique is to use the sample estimates of the canonical correlation coefficients. This is equivalent to solving a multiple least squares regression of $Y - \overline{Y}$ on $\underline{C}$. For a nonlinear control a workable alternative to using the analytical conditional expected value is to approximate the optimal nonlinear relationship with a nonlinear transformation $\hat{h}(\underline{C}, \beta^*)$. This can be done using any of several parametric transformations such as the scaled power transformation $\hat{h}(C, \beta^*) = (C^p - 1)/p$ where $p$ is an unknown parameter which must be estimated. The complete control equation could look like:

$$Y' = Y - \beta\left\{\frac{C^p - 1}{p} - \mathrm{E}\left[\frac{C^p - 1}{p}\right]\right\} \qquad (3)$$

where $\beta$ and $p$ are the parameters to be estimated. These parameters can be estimated using nonlinear least-squares regression of $Y - \overline{Y}$ on $C'$.

A key component of using controls for variance reduction is that the expected value of the transformed control must be known exactly or approximately. The expected value of the transformed control is subtracted off of each term as in (1) and (3) so that if $C$ is an unbiased estimator, the control function will be unbiased with mean zero. A difficulty with nonlinear transformations is analytically computing the expected value of the transformed control. In the estimation of quantiles, this difficulty can be greatly reduced.

## 3. QUANTILE ESTIMATION

### 3.1. Quantiles and Monotone Transformations

Let $\mathcal{Z}$ be a random variable with a continuous cumulative distribution function $F_{\mathcal{Z}}(z)$ which is strictly increasing whenever $0 < F_{\mathcal{Z}}(z) < 1$ and let $\mathcal{Z}$ have density function $f_{\mathcal{Z}}(z)$. Define the $\alpha$th quantile of $\mathcal{Z}$, $\mathcal{Z}_\alpha$, as the unique solution for $z$ in the equation

$$F_{\mathcal{Z}}(z) = \alpha \text{ for } 0 < \alpha < 1.$$

Given a simulation sample of $\mathcal{Z}$ of size $n$ with order statistics $Z_{(1)}, \ldots, Z_{(n)}$, define a nonparametric estimator of the $\alpha$th quantile, $\widehat{Z}_\alpha(n)$, as in Lewis and Orav (1988) as follows:

$$\widehat{Z}_\alpha(n) = \begin{cases} Z_{(n\alpha)} & \text{if } n\alpha \text{ is an integer} \\ Z_{(\lfloor n\alpha+1 \rfloor)} & \text{if } n\alpha \text{ is not an integer} \end{cases}$$

where $\lfloor w \rfloor$ denotes the integral part of $w$. Let $\hat{z}_\alpha(n)$ be the sample realization of $\widehat{Z}_\alpha(n)$. The $(n)$ will be dropped when the sample size is clear from the context. For a given $n$ and $\alpha$, $\widehat{Z}_\alpha(n)$ is an order statistic, so the form of its distribution is known. Unfortunately its distribution depends on the unknown distribution of the underlying $\mathcal{Z}$.

Let $\mathcal{Y}$ and $\mathcal{C}$ be random variables whose distribution and density functions meet the criteria listed for $\mathcal{Z}$ above. To emphasize the fact that we are controlling quantile estimates and not means, we will use $\widehat{Y}_\alpha$ and $\widehat{C}_\alpha$ for $Y$ and $C$. Even though the corresponding quantile of $\mathcal{C}$ is not necessarily the best control for the $\alpha$th quantile of $\mathcal{Y}$, it is usually the first choice.

When using a quantile estimator $\widehat{C}_\alpha$ as a control, the underlying distribution of $\mathcal{C}$ is known and thus the distribution of $\widehat{C}_\alpha$. Using the probability integral transform $U_{(r)} = F_{\mathcal{C}}^{-1}(C_{(r)})$ David (1970) arrives at the following results for when the $r$th order statistic is the estimator from a sample of size $n$ for the $\alpha$th quantile:

$$\begin{aligned} \mathrm{E}\left[C_{(r)}\right] &= F_C^{-1}\left(\tfrac{r}{n+1}\right) - \\ &\frac{1}{2}\left(\frac{\alpha(1-\alpha)}{n+2}\right)\frac{f_C'\left(\frac{r}{n+1}\right)}{f_C^3\left(\frac{r}{n+1}\right)} + o\left(\frac{1}{n^2}\right) \end{aligned} \quad (4)$$

and

$$\operatorname{var}\left(C_{(r)}\right) = \frac{\alpha(1-\alpha)}{(n+2) f_C^2\left(\frac{r}{n+1}\right)}. \quad (5)$$

(4) shows that the estimator $\widehat{C}_\alpha$ is only asymptotically unbiased. When using a quantile estimate as a control, although the biased expected value could be subtracted off in the control function, usually the asymptotic expected value of the estimator, i.e. the actual quantile value is used. The $o(1/n)$ bias this produces isn't the only bias in the control function since in practice, when the parameters of the control function have to be estimated using the same data that generates $\hat{y}_\alpha$ and $\hat{c}_\alpha$, bias is introduced into $C''$. Bias in the control function becomes important as the sample size or the estimated controlled variance decreases.

Since the distribution of $\widehat{C}_\alpha$ is known, it may be possible to compute the expected value of $\hat{h}(\widehat{C}_\alpha, \beta^*)$. It is important to note that the random variable being transformed is the quantile estimator and not the underlying $C$. For example, if $C$ has a uniform (0,1) distribution the random variable being transformed, $\widehat{C}_\alpha$, has a beta distribution which is less tractable. The asymptotic expected value of this beta distributed random variable could be computed but it can be easily approximated using the monotone transformation property of quantiles.

Quantiles have the property that under strictly monotone transformations of the underlying random variable, the quantiles transform monotonely as well. For example, let $\hat{h}() = g()$ be a strictly monotone function and let $\mathcal{D} = g(C)$. If $\Pr(C \leq c) = \alpha$ and $\Pr(\mathcal{D} \leq d) = \alpha$, then

$$\Pr(\mathcal{D} \leq d) = \Pr(C \leq g^{-1}(d)) = \alpha \Longrightarrow d = g(c).$$

The quantile estimator transforms monotonely as well. If $C_{(r)}$ is the estimator for the $\alpha$th quantile of $C$, the estimator for the $\alpha$th quantile of $g(C)$, $g(C)_{(r)}$ will equal $g(C_{(r)})$. One can transform the quantile estimates directly and use $g(C_\alpha)$ as the asymptotic expected value in $C''$. This eliminates much of the analytical difficulty when using nonlinear transformations of quantile estimates.

### 3.2. Methods for Computing the Controlled Estimate

There are several methods for generating the controlled estimate. For any method to be considered complete in addition to the point estimate, it must provide an estimate of the variance of the point estimate. For each of the methods, the elements of $\beta^*$ have to be estimated. To estimate the parameters via nonlinear least square regression, one needs more than one estimate of both $\widehat{Y}_\alpha$ and $\widehat{C}_\alpha$. Multiple estimates can be generated by sectioning the data. The general procedure for a size $n$ data set follows.

1. Separate the $n$ samples of $Y$ and $C$ into $v$ sections of length $l$ where $v \times l = n$.
2. Compute $\hat{y}_{\alpha,i}(l), i = 1, \ldots, v$ and $\hat{c}_{\alpha,i}(l), i = 1, \ldots, v$.
3. Use these $v$ pairs of estimates as data for the regression.

Let $\widehat{\beta}^*(v, l)$ be the estimated parameters based on $v$ estimates from $l$-sized samples.

The first method is a straightforward sectioning method. A subscript $S$ denotes an estimator based on sectioning. Each of the $\hat{y}_\alpha(l)$ is controlled using the estimated parameters. The final controlled point and variance estimators are

$$\widehat{Y}'_{S,\alpha}(n) = \overline{\widehat{Y}'}_{S,\alpha}(l)$$

$$= \frac{1}{v}\sum_{i=1}^{v}\left(\widehat{Y}_{\alpha,i}(l) - \hat{h}\left(\widehat{C}_{\alpha,i}(l), \widehat{\beta}^*(v,l)\right)\right)$$

and

$$S_S^2 = \frac{1}{v(v-1)}\sum_{i=1}^{v}\left(\widehat{Y}'_{S,\alpha,i}(l) - \widehat{Y}'_{S,\alpha}(n)\right)^2$$

The initial idea behind the second method was to use a point estimator of

$$\widehat{Y}'_{R,\alpha}(n) = \widehat{Y}_\alpha(n) - \hat{h}\left(\widehat{C}_\alpha(n), \widehat{\beta}^*(v,l)\right).$$

The goal was to improve the bias characteristics of $\widehat{Y}'_{S,\alpha}(n)$. If $Y$ were a mean and $\hat{h}()$ was linear, method 1 and 2 would be identical. Unfortunately, as a quantile estimator, $\widehat{Y}'_{R,\alpha}(n)$ has no readily available variance estimate due to the dependence on the underlying distribution as in (5). Independent replications could be used to generate multiple realizations of $\widehat{Y}'_{R,\alpha}(n)$ and take the sample average as the estimate. For comparison's sake, the sample size must remain at $n$. Therefore the second method separates the $n$ samples into $k$ "independent replications" each of sample size $m$ where $k \times m = n$. A subscript $R$ will denote the replications method estimators which are:

$$\widehat{Y}'_{R,\alpha}(n) = \overline{\widehat{Y}'}_{R,\alpha}(m)$$

$$= \frac{1}{k}\sum_{i=1}^{k}\left(\widehat{Y}_{\alpha,i}(m) - \hat{h}\left(\widehat{C}_{\alpha,i}(m), \widehat{\beta}^*(v, m/v)\right)\right),$$

and

$$S_R^2 = \frac{1}{k(k-1)}\sum_{i=1}^{k}\left(\widehat{Y}'_{R,\alpha,i}(m) - \overline{\widehat{Y}'}_{R,\alpha}(m)\right)^2.$$

Two possible alternatives for attempting to reduce bias as well as estimate the variance are the jackknife (Efron and Gong, 1983) and splitting (Beale, 1985). Lavenberg, Moeller and Welch (1982) examined the use of the jackknife for producing confidence intervals for a linearly controlled estimate of the mean under the assumption that $Y$ and $C$ had multivariate normal distribution. They found that the jackknifed confidence interval was usually larger and more computationally expensive than the standard linear control based confidence interval. Nelson (1988) analyzed at the performance of several estimation methods when the normality assumption was violated and compared the methods to the standard linear control of the mean. He found that the jackknife was usually dominated by the splitting estimator.

### 3.3. Research Issues

Selecting a particular method and the parameters of the method such as $n$, $v$ or $k$, requires consideration of bias, computational aspects versus effectiveness and the effects of inducing normality. Several research issues are involved. A complete description of the sources, magnitudes and effects of bias is necessary. For a fixed sample size, computing a nonlinear controlled estimate is more expensive. In some situations, it can improve the achieved variance reduction enough to justify its use.

Weiss (1964) proved under mild conditions that sample quantiles from a multivariate population have an asymptotic multivariate normal distribution where the covariance is a function of the multivariate distribution of the underlying population. A key research issue is the interplay between $n$ and $v$ and the rate at which $v$ should increase as $n$ increases. If $v$ stays fixed while $n$ goes to infinity, then the asymptotic multivariate normal distribution of the quantile estimates will eventually negate the usefulness of a nonlinear control. It is not clear at what $n$ and $v$ combination this will begin to be significant.

For a fixed sample size $n$ there is a trade-off between wanting $v$ large and wanting $l$ large. The bias and variance of the sectioned estimates are both decreasing functions of $l$. The larger $l$ the smaller the range over which $\hat{h}$ needs to approximate the true conditional expected value. Under generally applicable conditions outlined in Gallant (1975), the parameters being estimated in the nonlinear regression are asymptotically normally distributed with decreasing variance as $v$ increases. Guidelines for selecting $v$ as a function of $n$ need to be determined.

## 4. SUMMARY

Nonlinear controls have been effective in improving the variance reduction over linearly controlled estimates of the mean. Controlling quantiles with nonlinear controls is analytically tractable if the nonlinear transformations are limited to strictly monotonic functions. The performance of the available methods with respect to bias, variance estimation and the selection of parameters will be discussed during the talk.


## ACKNOWLEDGMENTS

The research of P.A.W.Lewis was supported at the Naval Postgraduate School under an Office of Naval Research Grant.


## REFERENCES


Beale, E.M.L. (1985). Regression: A bridge between analysis and simulation. *The Statistician* **34** , 141-154.

Breiman, Leo and Friedman, Jerome H. (1985) Estimating optimal transformations for multiple regression and correlation. *Journal of the American Statistical Association* **80**, 391, 580-619.

Efron, Bradley and Gong, Gail (1983). A leisurely look at the Bootstrap, the Jackknife and Cross-Validation. *The American Statistician* **37**, 1, 36 - 48.

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

Lancaster, H.O. (1966). Kolmogorov's remark on the Hotelling canonical correlations. *Biometrika* **53**, 585-588.

Lavenberg, Stephen, S. and Welch, Peter D. (1981). A perspective on the use of control variables to increase the efficiency of monte carlo simulations. *Management Science* **27**, 3, 322-335.

Lewis, P. A. W. and Orav, E. J. (1989). *Simulation Methodology for Statisticians, Operations Analysts and Engineers.* Wadsworth & Brooks/Cole, Pacific Grove, California.

Lewis, P.A.W., Ressler, Richard L. and Wood, R. Kevin (1989). Variance reduction using nonlinear controls and transformations. *Communications in Statistics* **18B**, 2.

Nelson, Barry (1988). Control variate remedies. Working Paper Series No. 1988-004, Department of Industrial and Systems Engineering, The Ohio State University, Columbus, Ohio.

Weiss, Lionel (1963). On the asymptotic joint normality of quantiles from a multivariate distribution. *Journal of Research of the National Bureau of Standards* **68B**, 2, 65-66.

## AUTHORS' BIOGRAPHIES

PETER A. W. LEWIS is Distinguished Professor of Statistics and Operations Research in the Department of Operations Research at the Naval Postgraduate School in Monterey, California. He holds an A.B. degree (1954) from Columbia University, B.S. and M.S. degrees (1955 and 1957) from Columbia School of Engineering, and a Ph. D. in Statistics from the University of London in 1964. He is co-author, with D.R. Cox, of the book 'The Statistical Analysis of Series of Events' (1966), editor of the Volume entitled 'Stochastic Point Processes: Statistical Analysis, Theory and Applications' (1972) and co-author, with E.J.Orav, of the book 'Simulation Methodology for Statisticians, Operations Analysts and Engineers' (1989). He has also published a software package 'Enhanced Simulation and Statistics Package' (1989) with E.J.Orav and L.Uribe. He is the author of over eighty refereed papers in technical journals, and is a Fellow of the American Statistical Association and the Institute of Mathematical Statistics.

Peter A.W. Lewis
Code 55LW
Naval Postgraduate School
Monterey, CA 93943
(408) 646-2283

RICHARD L. RESSLER is a Ph.D. student at the Naval Postgraduate School. He received his B.A. in biochemistry from the University of Pennsylvania in 1978 and his M.S. in Operations Research from the Naval Postgraduate School in 1987.

Richard L. Ressler
Code 55
Naval Postgraduate School
Monterey, CA 93943
(408) 646-2676