Regression Estimation in Simulation

Anton C. M. Hopmans; Jack P. C. Kleijnen

The Journal of the Operational Research Society, Vol. 31, No. 11. (Nov., 1980), pp. 1033-1038.

Stable URL:
http://links.jstor.org/sici?sici=0160-5682%28198011%2931%3A11%3C1033%3AREIS%3E2.0.CO%3B2-R

The Journal of the Operational Research Society is currently published by Operational Research Society.

J. Opl Res. Soc. Vol. 31, pp. 1033 to 1038
Pergamon Press Ltd 1980. Printed in Great Britain

0160-5682/80/1101-1033$02.00/0

Regression Estimation in Simulation

ANTON C. M. HOPMANS

Dr Neher Laboratories, PTT, Leidschendam, Netherlands

and

JACK P. C. KLEIJNEN

Department of Business & Economics, Tilburg University, Tilburg, Netherlands

In simulation an input variable like interarrival time is sampled, and hence its average deviates from its known expectation. This information can be used to improve the estimated simulation response: regression sampling or control variate technique. The usual crude estimator is shown to be biased. If local linearity holds, then the regression estimator becomes unbiased. Moreover its variance becomes smaller under mild conditions. The assumption of local linearity is an alternative to the normality assumption of other authors. This paper further emphasizes the difference between results *ex ante* (unconditional) and *ex post* (given the experimental input values). A telephone-exchange simulation provides a case study.

INTRODUCTION

LINEAR regression estimation or Control Variates is a well-known variance reduction technique in simulation. Its statistical properties, however, are not well understood. Recently Cheng[1] derived some properties assuming multivariate normality. Further properties under the normality assumption were derived by Lavenberg *et al.*[2] The present paper derives a similar variance expression using the mathematical assumption of local linear approximation instead of the statistical assumption of normality. Moreover this paper emphasizes a new aspect, namely, the "crude" estimator (without regression) gives a biased answer. Both the "unconditional" and the "conditional" variances and biases are derived. Finally the regression technique is applied to a real-life queuing situation (a PTT telephone exchange) instead of an academic queuing example. In this case study, the total simulation run is cut into subruns of predetermined length instead of replicating each run a number of times. Readers who find the statistical derivations in this article not so straightforward, can write to the authors for a more detailed version.[3]

REGRESSION ESTIMATION

In a queuing simulation one may compute the average arrival rate of customers per (sub)run, say $\underline{x}$ (stochastic variables are underlined). Then $\underline{y}_i$, the simulation response of run i ($i = 1, \ldots, n$), may be represented as

$$\underline{y}_i = f(\underline{x}_i) + \underline{v}_i \tag{1}$$

where the noise $\underline{v}_i$ has zero expected value and constant variance σ_v^2. Note that for a fixed x-value $\underline{y}$ can still vary, depending on the order in which customers arrive and depending on other stochastic inputs. The value x_i will not be exactly equal to its known expected value μ. In the case study, if $x_i > \mu$ $\underline{y}$ is expected to exceed its expectation η: $y_i > \eta$. Therefore $\underline{y}$ is regressed on $\underline{x}$:

$$\underline{y}_i = \beta_0 + \beta_1 \underline{x}_i + \underline{u}_i \, (i = 1, \ldots, n) \tag{2}$$

where $\underline{u}$, the noise around the assumed *linear* model, may have non-zero expected value. To obtain the estimators $\hat{\underline{\beta}}_0$ and $\hat{\underline{\beta}}_1$ Least Squares is applied. Least Squares is a mathematical procedure and unlike Maximum Likelihood it is not a statistical procedure. In practice one is interested in the response η for $x = \mu$ (desired input), not for $x = \bar{x}$ (accidentally sampled average). Therefore (3) is used:

$$\hat{\underline{y}}_\mu = \hat{\underline{\beta}}_0 + \hat{\underline{\beta}}_1 \mu. \tag{3}$$

This regression model is very simple except for the stochastic character of the independent variable $\underline{x}$. Why is the *linear* approximation very realistic when applying control variates to simulation?

(1) Mathematics shows that a function can be approximated by a linear function *locally*. Hence $f(x)$ is replaced by $(\beta_0 + \beta_1 x)$ in the neighbourhood of $x = \mu$. Linear approximation might be inadequate when x is varied systematically away from μ as in experimental design. In simulation, however, the expected value of $\underline{x}_i$ remains μ and the length of the runs can be taken so long as to make the linear approximation hold. Moreover in practice "obvious" outliers should be rejected. Hence an additional advantage of control variates is its check on "wild" random number streams.

(2) Statistical theory shows that if $(\underline{y}, \underline{x})$ is bivariate *normal* then the linear model holds.[1,2] This assumption is realistic if $\underline{y}$ and $\underline{x}$ are based on many observations so that a central limit theorem applies. For instance, even if individual waiting times are autocorrelated, their average or quantile[4] is asymptotically normal under mild technical conditions.

Note that the above regression model uses a single (control) variate $\underline{x}$. Extensions to multiple variables are straightforward: either add other variables like service time (besides arrival time) or add higher order terms like $\underline{x}^2$. Elsewhere,[5] however, it has been suggested that increasing the number of control variates may very well be inefficient because the variance of the estimated parameters $\hat{\underline{\beta}}$ increases for fixed sample size n; see also Lavenberg *et al.*[2]

VARIANCE COMPARISON

The variances of the crude estimator $\bar{\underline{y}}$ and the regression estimator $\hat{\underline{y}}_\mu$ are examined for two cases, namely the general case $f(x)$ and the special case of a linear approximation. Runs (or subruns) are supposed to be independent so that the $\underline{y}_i$ are independent. Further $\underline{x}$ and $\underline{v}$ are assumed to be independent; see (1). For the crude estimator one obtains:

$$\operatorname{var}(\bar{\underline{y}}) = \operatorname{var}\left\{\frac{1}{n}\sum f(\underline{x}_i) + \frac{1}{n}\sum \underline{v}_i\right\} = \frac{1}{n}\operatorname{var}\{f(\underline{x})\} + \frac{\sigma_v^2}{n}. \tag{4}$$

The first component is the variation *along* the curve $f(x)$, and the second component is the variance around $f(x)$ *given* x. If $f(x)$ is assumed to be linear then (4) reduces to

$$\operatorname{var}(\bar{\underline{y}}) = \frac{1}{n}\operatorname{var}(\beta_0 + \beta_1 \underline{x}) + \frac{1}{n}\sigma_v^2 = \beta_1^2 \operatorname{var}(\underline{x})/n + \sigma_v^2/n. \tag{5}$$

So $\operatorname{var}(\bar{\underline{y}})$ increases when the control variate can vary much from run to run, and when the response reacts strongly to changes in x (high β_1).

For *given* $\mathbf{x}' = (x_1, \ldots, x_n)$ it is well-known that

$$\operatorname{var}(\hat{\underline{y}}_\mu | \mathbf{x}) = \sigma_v^2\left\{\frac{1}{n} + \frac{(\mu - \bar{x})}{\sum (x_i - \bar{x})^2}\right\} \tag{6}$$

where $\sigma_v^2 = \operatorname{var}(\underline{y}|x)$ is assumed to be *constant* over the (narrow) range of x; (6) does not assume normality of $\underline{y}$ or "correctness" of the fitted model (absence of specification error so that $\underline{u} = \underline{v}$). If and only if (2) is a correct model—$f(x)$ in (1) is linear—then an unbiased estimator of σ_v^2 (given $\mathbf{x}$) follows from the residuals $\underline{e} = \underline{y} - \hat{\underline{y}}$:

$$\hat{\underline{\sigma}}_v^2 = \sum_{i=1}^{n} \{\underline{y}_i - (\hat{\underline{\beta}}_0 + \hat{\underline{\beta}}_1 x_i)\}^2/(n-2). \tag{7}$$

From the conditional expectation $\mathscr{E}(\hat{\underline{\sigma}}_v^2 | \mathbf{x}) = \sigma_v^2$ immediately follows $\mathscr{E}(\hat{\underline{\sigma}}_v^2) = \sigma_v^2$.

Combining the basic formula

$$\operatorname{var}(\underline{y}) = \underset{\underline{x}}{\mathscr{E}}\{\operatorname{var}(\underline{y}|x)\} + \underset{\underline{x}}{\operatorname{var}}\{\mathscr{E}(\underline{y}|x)\} \tag{8}$$

with (3) and (6) yields

$$\operatorname{var}(\underline{\hat{y}}_\mu) = \underset{\underline{x}}{\mathscr{E}}\left[\sigma_v^2\left\{\frac{1}{n} + \frac{(\mu - \bar{x})^2}{\sum(x_i - \bar{x})^2}\right\}\right] + \underset{\underline{x}}{\operatorname{var}}[\mathscr{E}(\hat{\beta}_0 + \hat{\beta}_1\mu|\mathbf{x})]$$

$$= \frac{\sigma_v^2}{n} + \frac{\sigma_v^2}{n}\mathscr{E}\left[\frac{(\mu - \underline{\bar{x}})^2}{\underline{s}_x^2}\right] + \underset{\underline{x}}{\operatorname{var}}\{\underline{\hat{\beta}}_0(\mathbf{x}) + \underline{\hat{\beta}}_1(\mathbf{x})\mu\} \tag{9}$$

where $\underline{s}_x^2$ denotes the (biased) estimator of $\operatorname{var}(\underline{x}) = \sigma_x^2$:

$$\underline{s}_x^2 = \sum_1^n (\underline{x}_i - \underline{\bar{x}})^2/n \tag{10}$$

and $\hat{\beta}_0(\mathbf{x}) = \mathscr{E}(\underline{\hat{\beta}}_0|\mathbf{x})$, etc.

Equation (9) may be compared to (6). If $\underline{x}$ varies much then in general $\operatorname{var}(\underline{\bar{y}})$ increases: y moves along $f(x)$. For the second term in (9) the approximation for large n might be used:

$$\mathscr{E}\left[\frac{(\mu - \underline{\bar{x}})^2}{\underline{s}_x^2}\right] \approx \frac{\mathscr{E}(\mu - \underline{\bar{x}})^2}{\mathscr{E}(\underline{s}_x^2)} \approx \frac{\operatorname{var}(\underline{\bar{x}})}{\operatorname{var}(\underline{x})} = \frac{1}{n}. \tag{11}$$

The last term in (9) refers to changes in the regression parameters $\hat{\beta}_0$ and $\hat{\beta}_1$ as $\mathbf{x}$ changes. More insight results if $f(x)$ is approximated linearly. If the regression model is correctly specified, then

$$\mathscr{E}(\underline{\hat{\beta}}_0 + \underline{\hat{\beta}}_1\mu|\mathbf{x}) = \beta_0 + \beta_1\mu \tag{12}$$

independently of $\mathbf{x}$. Consequently in (9) the variance component vanishes and (9) reduces to

$$\operatorname{var}(\underline{\hat{y}}_\mu) = \frac{\sigma_v^2}{n} + \frac{\sigma_v^2}{n}\mathscr{E}\left[\frac{(\mu - \underline{\bar{x}})^2}{\underline{s}_x^2}\right]. \tag{13}$$

Equation (13) can be compared to (5):

(a) If *approximation* (11) is used then $\underline{\hat{y}}_\mu$ has a smaller variance than $\underline{\bar{y}}$ if

$$\rho^2 > 1/(n + 1). \tag{14}$$

To obtain (14) substitute the well-known relations $\beta_1^2 = \rho^2 \operatorname{var}(\underline{y})/\operatorname{var}(\underline{x})$ and $\operatorname{var}(\underline{y}) = \sigma_v^2/(1 - \rho^2)$ into (5). Condition (14) is met if ρ^2 approaches unity, still assuming large n.

(b) If $\underline{x}$ is assumed *normal* ($\underline{y}$ possibly not normal) then the following *exact* result is easily derived.

$$\operatorname{var}(\underline{\hat{y}}_\mu) = \frac{\sigma_v^2}{n}\left[1 + \frac{1}{(n - 1)}\mathscr{E}(\underline{F})\right] = \frac{\sigma_v^2}{n}\left(1 + \frac{1}{n - 3}\right) \tag{15}$$

which leads to a similar condition as (14): n becomes $(n - 3)$.

(c) *Given* the experimental inputs (x not stochastic), (13) or (6)—and also (9)—shows that regression gives a smaller variance if one of the following conditions holds:

(i) s_x^2 is large: a large spread in $\mathbf{x}$ permits more accurate prediction, as is known from experimental design theory,
(ii) $\bar{x}$ is close to μ: the confidence bands in regression are less tight farther away from the center of gravity (in simulation μ is known),
(iii) the response y reacts strongly to changes in x, as measured by β_1.

These are three intuitively acceptable conditions!

Note that (13) is identical to the result Lavenberg *et al.* obtained assuming normality; Chen's printed result (his equation 9) is wrong. When the variance is used to construct a *confidence interval* around $\hat{y}_\mu$, then the normality assumption yields an exact interval based on the t-statistic with $n - 2$ degrees of freedom.[1,2] If no normality is assumed then

this t-statistic is conjectured to remain valid because of the robustness of the t-statistic. If the practitioner does not trust this robustness he may use jackknifing.[2,5]

BIAS COMPARISON

From (1) it follows that

$$\mathcal{E}(\underline{y}) = \mathop{\mathcal{E}}_{\underline{x}}\{\mathcal{E}(\underline{y}|x)\} = \mathcal{E}\{f(\underline{x})\} \neq f\{\mathcal{E}(\underline{x})\} = f(\mu) \tag{16}$$

where the inequality holds unless $f(x)$ is linear. Because the n pairs $(\underline{y}, \underline{x})$ are independent (independent runs or subruns) one obtains

$$\mathcal{E}(\bar{\underline{y}}|\mathbf{x}) = \frac{1}{n}\sum \mathcal{E}(\underline{y}_i|x_i) = \frac{1}{n}\sum f(x_i) \neq f(\bar{x}) \tag{17}$$

and

$$\mathcal{E}(\bar{\underline{y}}) = \mathcal{E}\{f(\underline{x})\} \neq f(\mu) \tag{18}$$

where $f(\mu)$ is to be estimated! So the crude estimator shows both conditional bias given $\mathbf{x}$ (*ex post*) and overall bias (*ex ante*). Note that Least Squares implies $\hat{y}_{\bar{x}} = \bar{y}$. If $f(x)$ is *linear* (locally) then

$$\mathcal{E}(\bar{\underline{y}}|\mathbf{x}) = \beta_0 + \beta_1\bar{x} \tag{19}$$

and

$$\mathcal{E}(\bar{\underline{y}}) = \beta_0 + \beta_1\mu. \tag{20}$$

Though the overall bias is zero, the actual simulation experiment realizes $\mathbf{x}$ so that the simulation measures the response at the input value $\bar{\mathbf{x}}$, not at the value of interest μ.

If $f(x)$ is non-linear, then the regression estimator remains biased because this estimator approaches $f(x)$ linearly. In contrast to the crude estimator the regression estimator measures the approximated $f(x)$ at $x = \mu$, not at $x = \bar{x}$. If $f(x)$ is linear (locally), then the regression model is correct and the estimators $\hat{\underline{\beta}}_0$ and $\hat{\underline{\beta}}_1$ are unbiased for any realization $\mathbf{x}$; see (12). Hence regression yields an unbiased estimator of the response at $x = \mu$.

CASE STUDY

In a telephone exchange a telephone call (customer) gets a busy signal (blocking) if all lines (servers) are occupied. The simulation estimates the steady-state probability of blocking. The system is technically called a "grading" and forms an intricated network of servers.[6] To estimate the standard error of the simulated blocking probability the total run is divided into 15 subruns of 10,665 calls each (after discarding the initial part of the run). Table 1 shows x, the average arrival rate of calls per subrun with $\mathcal{E}(\underline{x}) = \mu = 28.7000$ (unit of time is mean service time) and y, the estimated blocking probability (in percents). Regression yields $\hat{\underline{\beta}}_0 = -20.1978$ and $\hat{\underline{\beta}}_1 = 0.7544$ so that $\hat{y} = 1.45$ whereas $\bar{y} = 1.54$. Next the sample analogue of (6) and (13) is used together with (7):

$$\hat{\text{var}}(\hat{\underline{y}}_\mu) = \frac{\hat{\sigma}_v^2}{n}\left\{1 + \frac{(\mu - \bar{x})^2}{s_x^2}\right\} = \frac{0.032}{15}\left\{1 + \frac{(0.119)^2}{0.180}\right\} = 0.0023 \tag{21}$$

so that the standard error is 0.048. Note that, assuming normality, (15) yields a result differing only in the fifth decimal. Instead of (5) one simply computes

$$\hat{\text{var}}(\bar{\underline{y}}) = \hat{\text{var}}(\underline{y})/n = \frac{\sum(y_i - \bar{y})^2}{(n-1)n} = 0.1325/15 = 0.0090 \tag{22}$$

so that the standard error is 0.095. Because the desired blocking probability is 1%, both estimates show significantly worse blocking. Note that $\hat{\sigma}_y^2 = 0.1325$ whereas

TABLE 1. CASE STUDY

Arrival rate x_i	Blocking y_i
28.7555	1.472105
28.7038	1.359587
28.3434	1.087670
28.9439	1.819034
29.0016	1.584623
29.4470	2.334740
28.7991	1.612752
28.4577	1.293952
28.4132	1.097046
28.7132	1.584623
28.5388	1.181435
28.2693	1.143929
28.9262	1.715893
29.1376	1.959681
29.8376	1.922175

$\hat{\sigma}_v^2 = 0.03219$ or $\hat{\rho} = 0.870$. The efficiency gain through regression, estimated by the ratio of (22) and (21), is 3.84; the ratio of the standard errors is 1.96 (the variance determines the number of subruns required for fixed accuracy; for a fixed sample size n the length of the confidence interval depends on the standard deviation).

CONCLUSION

Before the actual simulation experimentation one might decide whether to use regression, considering either condition (14), namely $\rho^2 > 1/(n+1)$ (approximation for large n), or the condition $\rho^2 > 1/(n-2)$ (exact result for normal input $\underline{x}$). If one conjectures that ρ^2 is small then one can save the extra work of regression analysis. In the case study $n = 15$ so that the conditions are met if $\rho > 0.28$.

Both bias and variance are unconditional *expected* values. The actual bias of the crude estimator and the standard error of the regression estimator depend on the realized x values (besides the parameter ρ or β_1 which depends on the simulated system). Hence after the *ex ante* decision whether to do regression at all (depending on $\rho^2 > 1/(n+1)$ or $\rho^2 > 1/(n-2)$) one can look at the outcome of the regression analysis. The (conditional) bias of the crude estimator can then be estimated as $\hat{\beta}_1(\bar{x} - \mu)$, assuming either local linearity or bivariate normality of $(\underline{y}, \underline{x})$. The standard errors of the two estimates $\bar{y}$ and $\hat{y}_\mu$ are computed from (21) and (22). One can *explain* why $\hat{y}_\mu$ has, say, smaller standard error: (i) the x values turned out to have a sizable spread measured by s_x^2, and (ii) the realized average $\bar{x}$ happens to be not too far away from the point of interest μ, and (iii) both (i) and (ii) would not have helped if $\underline{y}$ did not react to $\underline{x}$: $\hat{\sigma}_v^2$ would be as large as σ_y^2 if ρ would be zero (remember $\sigma_v^2 = (1 - \rho^2)\sigma_y^2$). Sizable variance reductions are possible, in the case study a factor 3.84. Note that the confidence intervals depend on the degrees of freedom $(n-1)$ or $(n-2)$, and these intervals require that the point estimators be unbiased.

ACKNOWLEDGEMENTS

The authors benefited from the comments by an anonymous referee on the technical details and the style of the previous version. They thank Mr B. Sanders (PTT) who performed the experimentation with the simulation model.

REFERENCES

[1] R. C. H. CHENG (1978) Analysis of simulation experiments under normality assumptions. *Opl Res. Q.* **29**, 493–497.

[2] S. S. LAVENBERG, T. L. MOELLER and P. D. WELCH (1978) *Statistical results on multiple control variables with application to variance reduction in queuing network simulation* RC 7432 (No. 31855), IBM Thomas J. Watson Research Center, Yorktown Heights, New York.

[3] A. C. M. HOPMANS and J. P. C. KLEIJNEN (1977) Regression estimators in simulation. Katholieke Hogeschool, Tilburg, Netherlands.
[4] H. CRAMER (1946) *Mathematical Methods of Statistics.* Princeton Univ. Press, Princeton.
[5] J. P. C. KLEIJNEN (1975) *Statistical Techniques in Simulation.* Marcel Dekker, New York.
[6] A. C. M. HOPMANS and J. P. C. KLEIJNEN (1979) Importance sampling in systems simulation: a practical failure? *Math. & Comp. Sim.* **21,** 209–220.