

# Variance reduction

Timo Tiihonen

2018

## Variance reduction techniques

The most efficient way to improve the accuracy and confidence of simulation is to try to reduce the variance of simulation results.

We will discuss four different approaches

- ▶ use of common random variates
- ▶ use of antithetic random variates
- ▶ replacing random variables with their expectations
- ▶ use of control variables

## Common random variates

Let us study the difference between two system variants ( $A$  and  $B$ ). The simulation outputs are  $X_A$  and  $X_B$ . We are interested in  $X_A - X_B$  and its expectation  $\alpha_A - \alpha_B$ .

The point estimate is  $\bar{X}_A - \bar{X}_B$ . If the simulations are independent, the variance of the point estimate (of difference) is

$$\text{Var}(\bar{X}_A - \bar{X}_B) = \text{Var}(\bar{X}_A) + \text{Var}(\bar{X}_B) \approx (S_A^2 + S_B^2)/N$$

## Common random variates

If simulations are mutually dependent,

$$\text{Var}(\bar{X}_A - \bar{X}_B) = \text{Var}(\bar{X}_A) + \text{Var}(\bar{X}_B) - 2\text{Cov}(\bar{X}_A, \bar{X}_B)$$

So, if the simulation experiments are positively correlated, the variance of the result (difference) is smaller than in uncorrelated case.

Positive correlation can result from using same random numbers for both variants (arrival times, service times etc generated using same seed values for both variants).

## Common random variates

- ▶ Requires that we can fully control the starting of random streams in every simulation run.
- ▶ Works best when both variants use the random variates in the same way.
- ▶ For example, all client related entities are generated at one place independently of clients future path in the system (easy and difficult clients come in same order but are routed differently).

## Antithetic variates

Consider a single simulation experiment (not a difference of two variants). We use random streams  $U_i$ , that give  $N$  instances of output  $X_i$ . We then make  $N$  additional experiments with random streams  $U_i^*$  with outputs  $X_i^*$ . We choose as point estimate

$$Y = (\bar{X} + \bar{X}^*)/2$$

for which  $E(Y) = E(X) = E(X^*)$  and

$$\text{Var}(Y) = \frac{1}{4}(\text{Var}(\bar{X}) + \text{Var}(\bar{X}^*) + 2\text{Cov}(\bar{X}, \bar{X}^*))$$

or  $\frac{1}{2}(\text{Var}(\bar{X}) + \text{Cov}(\bar{X}, \bar{X}^*))$ .

## Antithetic variates

If  $Cov(\bar{X}, \bar{X}^*) < 0$ , the variance is reduced more than when doing  $N$  independent additional trials.

Negative correlation may result from

- ▶ replacing  $U$  by  $1 - U$  in Unif(0,1) generator systematically.
- ▶ by exchanging the seed values between the arrival time and service time generators etc.

This results to making  $N$  independent simulations with  $2N$  simulation runs, so that the extreme phenomena occur in more symmetric way and hence the variance is likely to be smaller.

## Replacing with expectations

Often the observation of interest is a sum of a random amount of random variates. If the expectation of one of the random variates contributing to the result is known, this can be used in the analysis. Example: Utilization rate can be defined as

$$\sum_{i=1}^{N_{serv}} s_i / T$$

where  $s_i$  is the service time of individual client,  $N_{serv}$  the number of clients served. However, we know  $E(s_i)$  (simulation parameter). Let us denote it by  $t_s$ . Then also

$$N_{serv} t_s / T$$

is an estimate for utilization rate.



## Replacing with expectations

If the system is in equilibrium, then on average clients served = clients arrived - clients lost. Once again we know expectation for clients arrived (simulation parameter  $t_a$  interarrival time/client) .

Thus on average  $N_{serv} = T/t_a - N_{lost}$ .

Hence we get for utilization rate the estimate

$$t_s/t_a - N_{lost}t_s/T$$

This has the same expectation but often smaller variance than the original observation.

## Replacing with expectations

Lost clients can be analyzed further. We lose clients only when they arrive to a full system. Hence

$$E(N_{lost}) = E(T_{full})/t_a$$

So we can monitor the times when the system is full to get estimate

$$t_s/t_a(1 - T_{full}/T)$$

Different estimates do not require new simulation runs (only deeper analysis of the system).

It is possible to create several observations and to select the ones that give smallest variances.

## Control variables

Often the most powerful way to reduce variance is to use control variables.

Control variable is a random variate with known expectation that has strong correlation with the variate of interest. The difference between the control and original variate is to be observed.

Let us monitor  $X_i$  with unknown expectation  $E(X_i) = \alpha$ . In addition we observe  $Y_i$  with known expectation  $E(Y_i) = \beta$ .

Then the variable  $V = \bar{X} + \beta - \bar{Y}$  gives an unbiased estimate for  $\alpha$ ,  $E(V) = \alpha + \beta - \beta = \alpha$ .

$$\text{Var}(V) = \text{Var}(\bar{X}) + \text{Var}(\bar{Y}) - 2\text{Cov}(\bar{X}, \bar{Y})$$

If  $\text{Cov}(\bar{X}, \bar{Y}) > \frac{1}{2} \text{Var}(\bar{Y})$  then  $V$  leads to more accurate estimate to  $\alpha$  than  $X$  (smaller confidence interval).

## Control variables

As control variable  $Y$  we can use simulation output of a simpler model variant. This requires simulation and theoretical analysis of another model which may be costly and difficult.

Another option is to use as  $Y$  some model data that is known (expectation) and explains well the quantity of interest). These variables are called concomittant control variables. These do not require additional simulation models or deep analysis.

In the utilization rate example, for example, the arrived clients (times average service time) is a control variable for observed total service time.