

# Comparing Variance Reduction Techniques for Monte Carlo Simulation of Trading and Security in a Three-Area Power System

Magnus Perninge  
School of Electrical Engineering  
Royal Institute of Technology  
Stockholm, Sweden  
Email: magnus.perninge@ee.kth.se

Mikael Amelin  
School of Electrical Engineering  
Royal Institute of Technology  
Stockholm, Sweden  
Email: mikael.amelin@ee.kth.se

Valerijs Knazkins  
School of Electrical Engineering  
Royal Institute of Technology  
Stockholm, Sweden  
Email: valerijs.knazkins@ee.kth.se

**Abstract**—A variance reduction technique is a method used to reduce the variance of a Monte Carlo Simulation. In this paper four of the most commonly used variance reduction techniques are tested to estimate the trade of between trading and security in a three-area electric power system. The comparison is made by regarding both the variance reduction and the bias induced by the method.

## I. INTRODUCTION

When performing a Monte Carlo Simulation the outcome will be a stochastic variable, say  $X$ , on some given probability space  $(\Omega, \mathfrak{F}, P)$ . The variance of  $X$  on  $\Omega$  is given by

$$\text{Var}(X) \equiv \int_{\Omega} (X(\omega) - m_X)^2 dP(\omega), \quad (1)$$

where

$$m_X \equiv \int_{\Omega} X dP, \quad (2)$$

is the expectation of  $X$  over  $\Omega$ . If the method used is unbiased then  $m_X$  as defined above will be the true sought value. The probability of obtaining  $m_X$  as a result from a simulation is  $P(X^{-1}(m_X))$  which will often be zero (a.s.). The variance tells us how broad the distribution of  $X$  will be. As we increase the number of samples in our simulation, the variance will decrease as  $1/n$ , where  $n$  is the number of samples, hence giving us a series, converging almost surely [1]. Variance reduction techniques are used to decrease the variance of each sample yielding a larger probability of obtaining a good estimate from our simulation.

In this paper four methods for reducing the variance of Monte Carlo Simulations will be tested and compared to one another and to the standard Simple Sampling. The comparison is made on the problem of finding the Trade-Off between Trading and Security of a benchmark three-area power system.

## II. VARIANCE REDUCTION TECHNIQUES

The variance reduction techniques tested on our problem will be: Dagger Sampling, Importance Sampling, Stratified Sampling and the Control Variates method [2]. These methods

are arguably the most commonly used techniques for increasing the probability of obtaining good estimates from Monte Carlo Simulations.

### *Dagger Sampling*

Dagger sampling is commonly used when the sought expected value is a function of a number of r.v. living in  $\{1, 2\}$  (e.g. the reliability of a large system of components). In dagger sampling a single random number is used to generate a large number of outcomes. Assume that we have a system of two components in series each with a probability  $q_i$  of failure. If these probabilities are small, say 0.001 each, the risk of not getting a single system failure from 1000 scenarios of a Monte Carlo Simulation of the system using Simple Sampling is more than 10%. If on the other hand we were to use Dagger Sampling to generate the outcomes we start by partitioning the interval  $[0,1]$  into  $\lfloor 1/q_i \rfloor$  segments  $\{\alpha_k\}$  of length  $q_i$  and possibly a reminder  $\beta$  of length  $1 - q_i \lfloor 1/q_i \rfloor$  for  $i = 1, 2, \dots$ . We then generate two random numbers  $x_1$  and  $x_2$  from a  $U(0,1)$  distributed s.v. To generate outcomes for the Monte Carlo Simulation we find the  $k$  such that  $x_i \in \alpha_k$  and let outcome number  $k$  be a failure of component  $i$ . If  $x_i \in \beta$  no outcome is a failure. Hence there will be no risk of not getting a single system failure when performing a Monte Carlo Simulation on the system described above.

### *Importance Sampling*

When using Importance Sampling for variance reduction the probability measure of the random variables is changed to reduce the variance of the Monte Carlo Simulation. To get an unbiased estimate the resulting randomized cases must then be weighted before taking the mean. Suppose that we change the probability of failure from 0.001 to 0.1 for each component in the two-component system described above. This would yield a probability of  $1 - (1 - 0.1)^2 = 0.19$  of system failure, instead of the probability  $1.999 \cdot 10^{-3}$  of failure in the original system. If we carry out our simulations on the new system and then weight the results with

$$w(s) = \frac{\text{Probability of state } s \text{ in the original system}}{\text{Probability of state } s \text{ in the alternated system}}, \quad (3)$$

the expected value of the estimate will still be the same as in the original case. However, the risk of not getting enough cases from some of what might very often be the most important outcomes will be much smaller.

### Stratified Sampling

If there is a natural way to partition the sample space  $\Omega$  of our probability space  $(\Omega, \mathfrak{S}, P)$ , into  $m$  subsets  $\alpha_1, \alpha_2, \dots, \alpha_m$ , (such that  $\alpha_i \cap \alpha_j = \emptyset \forall (i, j), i \neq j, \bigcup_{i=1}^m \alpha_i = \Omega, \alpha_i \in \mathfrak{S} \forall i$ ) called strata, and a partition  $n_1, n_2, \dots, n_m$  of  $n$  such that

$$\sum_{i=1}^m (P(\alpha_i))^2 \frac{1}{n_i} \text{Var}(g(X)|X \in \alpha_i) < \frac{1}{n} \text{Var}(g(X)), \quad (4)$$

then the variance will be reduced if we divide the sampling procedure into  $m$  different Monte Carlo Simulations in which we constrain the r.v  $X$  to live on  $\alpha_1, \alpha_2, \dots, \alpha_m$  respectively, with p.m. induced by the conditional probability  $P(*|\alpha_i)$ , and then weight the resulting estimates together using the weights  $P(\alpha_i)$ . When a partition is made we can use Neyman allocation to find the ultimate choice of  $n_1, n_2, \dots, n_m$ . Let  $\sigma_h = \sqrt{\text{Var}(X_h)}$  be the variance of  $X$  on  $\alpha_h$  and  $\omega_h = P(\alpha_h)$ , for  $h = 1, \dots, m$ . Then

$$n_h = n \frac{\omega_h \sigma_h}{\sum_{k=1}^m \omega_k \sigma_k} \quad (5)$$

is the partition of  $n$  that minimizes the variance of our estimate. This principle is called Neyman allocation.

### Control Variates

When using the control variates method our knowledge of the system is used to get a better estimate. Assume that the objective of our Monte Carlo Simulation is to estimate the mean of some r.v.  $Y = g(X) : X(\Omega) \rightarrow S$ , where  $X$  is a r.v. on some given probability space  $(\Omega, \mathfrak{S}, P)$ . Here, of course,  $E(g(X))$  has to be well defined. By simplifying the model  $g(*)$  to get a new r.v.  $Z = \tilde{g}(X) : X(\Omega) \rightarrow S'$ , with mean  $m_Z$ , that can be calculated analytically, we can, instead of sampling  $Y$ , sample the r.v.  $Y - Z$  and then add  $m_Z$  to the estimate of  $E(Y - Z)$ . Now, since  $\text{Var}(Y - Z) = \text{Var}(Y) + \text{Var}(Z) - 2\text{Cov}(Y, Z)$ , a good choice of  $\tilde{g}$  will result in a dramatically reduced variance when estimating  $m_X$ . A warning to keep in mind is that an error in computing  $m_Z$  will lead to a bias when estimating  $Y$ .

## III. PROBLEM DESCRIPTION

The problem on which the variance reduction techniques will be tested and compared to one another is finding the trade-off between trading and security in a benchmark three area electric power system. The system, shown in Figure 1, consists of three areas, each area connected to the other two with  $l_{ij}$

lines each  $(i, j) \in S$ , where  $S = \{(i, j) \in \{1, 2, 3\}^2 | i \neq j\}$ . All the lines have a transfer capacity of  $\kappa$  MW each, and an availability probability  $p_l$ . If we try to transfer more than the capacity of a line the connection will be turned off. Area  $i$  contains  $kv_i$  power plants. Let  $J_i = \{1, 2, \dots, kv_i\}$ , giving us a total of  $\sum kv_i$  power plants to support the loads placed in each area. Power plant  $j$  in area  $i$  has production cost  $C_{ij}$  €/MWh, production capacity  $\bar{G}_{ij}$  and availability probability  $p_{ij}$ , for  $(i, j) \in V$  where  $V = J_1 \times J_2 \times J_3$ . The loads  $D_i, i \in I = \{1, 2, 3\}$ , have a known joint probability distribution. We assume that all the actors on the electricity market have perfect information, and thus that the cheapest power plants are always used. The problem to be analyzed is what happens to the expected benefit of trading and the security of the system when a set of transfer limits  $K_{ij}, (i, j) \in S$ , is used. All the frequency regulation is placed in Area 1 and the transferring capacity that is not used by the trading can be used to regulate the frequency. I.e. if there is a failure in plant 1 in Area 3, the production in Area 1 will increase to cover the new demand.

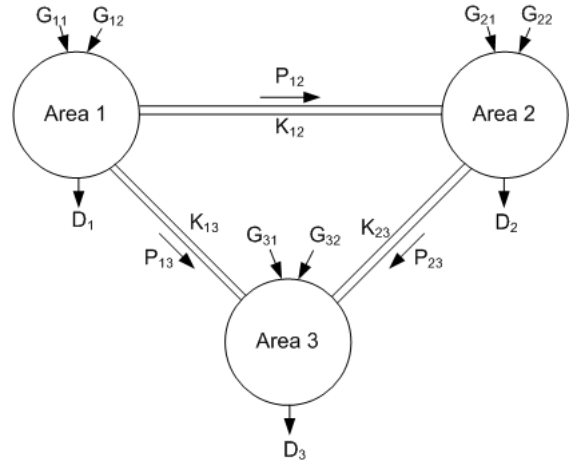


Fig. 1. The three-area system with 2 power plants in each area

### The Value of Trading

From a publicly economic point of view, maximizing the value of trading will be the same as minimizing the expected production cost plus the expected value of the cost originating from having to disconnect load [3]. The cost of disconnecting load is  $C_{DC}$  €/MWh and is the same for all loads.

### The Security of the System

As a measure of the security of the system we use the probability of a major disturbance causing loads to be disconnected. Real time regulation is assumed. This means that if we lose a transmission line, for example, the power plants will (if possible) immediately change their production to fit the new need. If the transfer limit is set too high the failure of one line will, however, lead to cascading failures.

The optimization problem of covering the demand of the market will look as follows:

$$\underset{\substack{G_{ij} \\ (i,j) \in V}}{\text{minimize}} \sum_{(i,j) \in V} C_{ij} G_{ij} + \sum_{i \in I} C_{DC} D_{DCi} \quad (6)$$

subject to:

$$\sum_{j \in J_i} G_{ij} - D_i + D_{DCi} = \sum_{j \in I - \{i\}} P_{ij}, \quad \forall i \in I, \quad (7)$$

$$P_{ij} \leq K_{ij}, \quad \forall (i, j) \in S, \quad (8)$$

$$0 \leq G_{ij} \leq \bar{G}_{ij}, \quad \forall (i, j) \in V, \quad (9)$$

$$P_{ij} = \frac{U_i^2 R}{|Z|^2} + \frac{U_i U_j}{|Z|^2} \{X \sin(\theta_i - \theta_j + x_{ij}) - R \cos(\theta_i - \theta_j + x_{ij})\}, \quad \forall (i, j) \in S. \quad (10)$$

$P_{ij}$ , is the active power flow from area  $i$  towards area  $j$  and is constrained by the load flow equations (3.5). In (3.5)  $U_i$  and  $U_j$  is the per unit values of the voltages in area  $i$  and area  $j$  respectively,  $\theta_i$  and  $\theta_j$  is the phase angle in area  $i$  and area  $j$  respectively, and  $x_{ij}$  is the angular changes induced by the phase shifters in area  $i$  used to shift the phase on the line towards area  $j$ . A physical bound might have to be added here as one of the constraints to the problem.  $Z = R + iX$  is the impedances of the lines used in the system.

#### IV. SIMPLIFIED MODEL

For our simplified model, which will be used to test the control variates method and importance sampling to get the value of trading, we assume that there are no losses in the system, i.e.  $P_{ij} = P_{ji} \quad \forall (i, j) \in S$ , and that all transfers for which  $P_{ij} \leq K_{ij}$  are allowed. We also assume that no failures of either the power plants or the lines will ever occur. With these assumptions made it is actually possible to find an analytic solution to the problem of finding the value of trading for a certain load vector  $D = [D_1, D_2, D_3]^T$ , and constraints  $K_{ij}, (i, j) \in S$ .

The solution is based on finding the production in every power plant as a function  $G_{ij} : R_+^3 \rightarrow R_+$ ,  $G_{ij}(D)$  = "Production in power plant  $j$  of Area  $i$  when the load is  $D = [D_1, D_2, D_3]^T$ ", for  $(i, j) \in V$ . Let us start with the cheapest power plant which is situated in say Area 1. We want to use this power plant as much as possible since it is the cheapest. We cannot, however, export more than  $K_{12}$  MWh/h to Area 2 and  $K_{13}$  MWh/h to Area 3. Therefore

$$G_{11} = \bar{G}_{11} \wedge (D_1 + K_{12} \wedge (D_2 + P_{23}^+) + K_{13} \wedge (D_3 + P_{32}^+)) \quad (11)$$

where

$$P_{23}^+ = K_{23} \wedge ((D_3 - K_{13}) \vee 0) \quad (12)$$

is what could be sent to Area 3 via Area 2 and

$$P_{32}^+ = ((K_{13} - D_3) \vee 0) \wedge (K_{32} \wedge (D_2 - K_{23}) \vee 0) \quad (13)$$

is what could be sent to Area 2 via Area 3.

We now move to the next power plant in order of production cost and so on until the functions  $G_{ij}$  are found for all  $(i, j) \in V$ . This method can be generalized to cover problems with power plants having nonlinear but finitely piecewise linear production costs by splitting one power plant into several plants, one plant for each linear part of the production cost.

#### V. CASE STUDY

As the Case Study on which our variance reduction techniques will be tested we use the following:

- In Area 1 there are 2 hydro power plants with production costs 10 €/MWh and 15 €/MWh respectively, and production capacities 100 MW and 100 MW respectively.
- In Area 2 there are 2 nuclear power plants with production costs 20 €/MWh and 25 €/MWh respectively, and production capacities 40 MW and 100 MW respectively.
- In Area 3 there are 2 heat power plants with production costs 30 €/MWh and 40 €/MWh respectively, and production capacities 30 MW and 100 MW respectively.
- All power plants have an availability probability of 99%.
- All lines have an availability probability of 99%.
- The load  $D = [D_1, D_2, D_3]^T$  is  $N(m, \Sigma)$ -distributed with mean  $m = [50, 50, 50]^T$  MW and covariance matrix  $\Sigma = \begin{pmatrix} 10 & 9 & 9 \\ 9 & 10 & 9 \\ 9 & 9 & 10 \end{pmatrix}$
- The cost of disconnecting load is 100 €/MWh.
- As for the power system quite ordinary values are used and the phase shifters can shift the angles to fit all purposes.
- $K = \begin{pmatrix} - & 50 & 50 \\ 50 & - & 25 \\ 50 & 25 & - \end{pmatrix}$

*How does the comparison work?*

To compare different variance reduction techniques to one another we use the measure "time taken by the simulation"  $\times$  "variance of the resulting estimate". This measure is reasonable since variance decays as  $1/n$  and  $t \propto n$ , where  $n$  is the amount of samples and  $t$  is the time taken by the simulation. In this paper all the methods tested take approximately the same time when using the same  $n$ . While Importance Sampling, Stratified Sampling and Dagger Sampling are methods that affect our way of generating random variables, the Control Variates method uses the same way of randomizing the state variables as Simple Sampling. Therefore the Control Variates method can be combined with other methods to reduce variance even further. We came to the conclusion that Stratified Sampling and Dagger Sampling should be used only when generating the system state (Dagger Sampling cannot be used in any other way on this problem). Importance Sampling on the other hand will be used when randomizing both the loads and the system state. All in all six set of methods were tested, as depicted in the table below:

	Loads		System State				
	IS	SS	IS	DS	Str	SS	CVM
1		*				*	
2	*					*	
3		*				*	*
4		*	*				*
5		*		*			*
6		*			*		*

where IS refers to Importance Sampling, SS is Simple Sampling, DS is Dagger Sampling, Str is Stratified Sampling and CVM is the Control Variates method.

### Simplified model

With the values given above there is practically no risk that any area will run out of production capacity and the power plants are aligned so that Area 1 contains all the cheapest plants, Area 2 the next two plants in order of production cost and Area 3 the most expensive plants. Therefore we can model the two power plants in Area  $i$  as one power plant  $G_i$  and then let  $G_{i1} = H(\bar{G}_{i1} - G_i)G_i$  and  $G_{i2} = H(G_i - \bar{G}_{i1})(G_i - \bar{G}_{i1})$ , for all  $i \in I$ , where  $H$  is the heaviside function. The  $G_i$ 's are then:

$$G_1 = D_1 + K_{12} \wedge (D_2 + P_{23}^+) + K_{13} \wedge (D_3 + P_{32}^+), \quad (14)$$

$$G_2 = 0 \vee (D_2 + P_{23}^+ - K_{12} - P_{23}^-), \quad (15)$$

$$G_3 = 0 \vee (D_3 - P_{23}^+ - K_{13}). \quad (16)$$

### Setting 1

This method is straightforward and no explanations are needed. It gave an expected production cost of 1795 €/period and a variance of 103, when taking  $n = 500$  samples. However, we never get any disconnection of loads, which implies that this method gives us a so called cardinal error (when all power plants are down, load will undeniably have to be disconnected), i.e. we miss some important cases.

### Setting 2

When using Importance Sampling there is a certain strategy that actually yields a variance of 0. If we want to sample the function  $X = g(Y) : Y(\Omega) \rightarrow S$  of the r.v.  $Y$  with well known distribution function  $f_Y$  and instead take samples from a r.v.  $Z$  with distribution function

$$f_Z(\psi) = \frac{\hat{g}(\psi)}{m_X} f_Y(\psi) \quad (17)$$

and, then weight the resulting samples with  $w = f_Y/f_Z$ , we get an estimate of  $m_X$  with variance 0. However, this requires knowing

$$m_X = \int_{\mathbb{R}_+^3} \hat{g}(Y) dF_Y, \quad (18)$$

which is what we wanted in the first place. Using our approximation  $\hat{g}$  of  $g$  and its analytically computable mean value

gives us a method for reducing the variance of our estimate. The time needed to compute the integral (5.5) is neglected, since this process only has to be done once. Unfortunately this method did not work well on this system.

### Setting 3

We now turn our attention to the Control Variates method to see if it can do what Importance Sampling could not. Here as with Importance Sampling we have to calculate the expected value (5.4), and for the same reasons as before we disregard the time taken by this procedure. However this method seems to be amazingly good and the variance of the estimate is only 0.2375. To get that small variance with Simple Sampling we would have to take approximately  $n = 200000$  samples.

### Setting 4

Having seen what the Control Variates method can do to the variance, we keep using this method and let the loads be randomized as with Simple Sampling. In Setting 3 we use Importance Sampling to randomize the system state. We change the probabilities of component failure from 1% to 20% and then weight the resulting scenarios. This was quite effective and the result was that the probability of system failure was estimated to  $3.8731 * 10^{-7}$ , but this still seems somewhat small considering that the probability of three or more components not working is  $2.0562 * 10^{-4}$ .

### Setting 5

This method did not work well at all, even when using it together with Importance Sampling (raising the probabilities of failure and then using Dagger Sampling) did not give a much better result than in Setting 3.

### Setting 6

When partitioning the sample space into 13 different strata  $\alpha_0, \dots, \alpha_{12}$  such that all states with  $i$  failures belong to strata  $\alpha_i$  and then using the Control Variates method to get the value of trading, we obtained the best result with a variance of 2.3469 when estimating the estimated cost of supporting the loads in the system and an estimated probability of failure equal to  $2.2271 * 10^{-4}$  and with variance  $1.7450 * 10^{-8}$ . Using the same arguments as with Setting 4, this seems to be a reasonable result. The partition of  $n$  was not made using Neyman allocation as described in Section 2. This because we want to take two estimates from every simulation, both the cost of supporting the loads and the risk of having to disconnect loads. When having to weigh these variances against one another it seemed easier and more efficient to try a few partitions to see which gave the best results.

The most efficient method was thus to use Simple Sampling techniques for randomizing the loads and then Stratified Sampling to randomize the system state. Using this technique, we made the plot (Fig 2) showing the cost of supporting the

loads as a function of  $x$  when

$$K = x \times \begin{pmatrix} - & 1 & 1 \\ 1 & - & 0.5 \\ 1 & 0.5 & - \end{pmatrix}$$

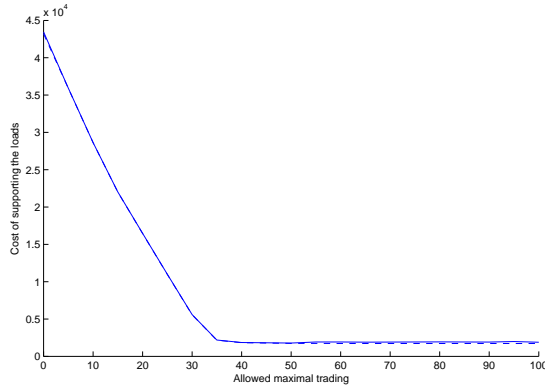


Fig. 2. Cost of supporting the loads

where the dashed line is the mean of cost of supporting the loads in the simplified system.

If we look closer at this figure, we see that in the interval  $x > 50$  the two lines become separated. This follows from the larger probability of the failure of a line causing the system to trip and loads to be disconnected.

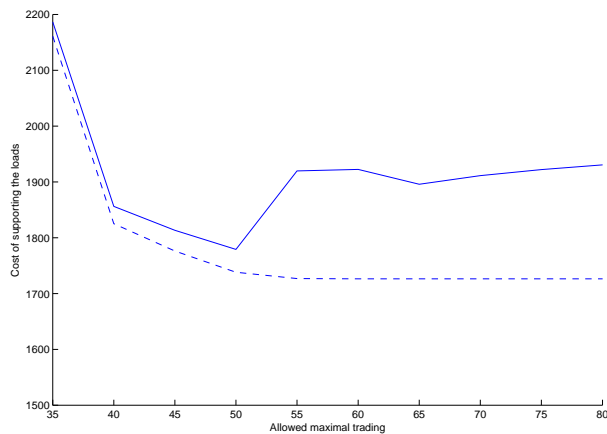


Fig. 3. Cost of supporting the loads

As for the probability of system failure, Figure 4 shows us the relation between  $x$  and the probability of having to disconnect loads.

## VI. CONCLUSIONS

When using Monte Carlo methods to simulate the Trade-Off between Trading and Security in a Three-Area Power

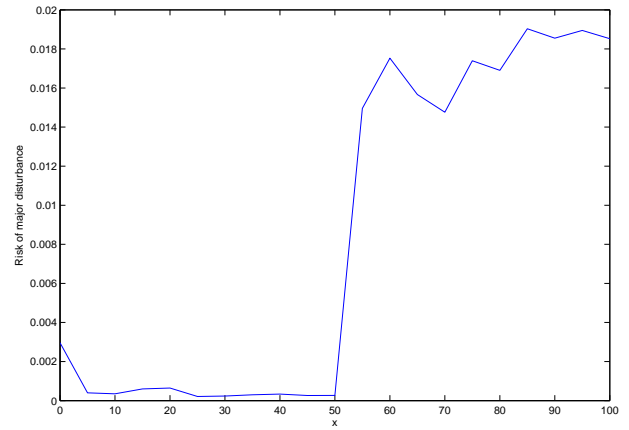


Fig. 4. Risk of system failure

System, a good choice of variance reduction technique can dramatically increase the efficiency of the simulation. A poor choice might, however, not give that much of an improvement. Some techniques, like Simple Sampling, give a result that seems very good but actually miss some very important scenarios, yielding a cardinal error. For our case study, using Stratified Sampling together with the Control Variates method gave the best results. These techniques can, of course, be used on many different problems in power system analysis.

## REFERENCES

- [1] Kai Lai Chung, *A Course in Probability Theory*, 3rd ed. San Diego, California: Academic Press, 2001.
- [2] Reuven Y. Rubinstein, *Simulation and The Monte Carlo Method*, New York: John Wiley & Sons, Inc., 1981.
- [3] Mikael Amelin, PhD Thesis *On Monte Carlo Simulation and Analysis of Electricity Markets*, (Avaliable on [www.eps.ee.kth.se](http://www.eps.ee.kth.se)), Stockholm: Universitetsservis US-AB, 2004.