ON THE ANALYSIS OF UNDERSAMPLING PROBLEMS IN MONTE CARLO CALCULATIONS

Wilson J. Vieira
Instituto de Pesquisas Energéticas e Nucleares - IPEN
Paul N. Stevens
University of Tennessee, Knoxville

SUMMARY

Undersampling conditions can often lead to an estimated solution which can be significantly smaller than the true solution — but the estimate of the standard deviation may seem acceptably small. Various techniques devised to flag undersampling in Monte Carlo calculations were investigated. The calculation of F values, of the variance of the variance, and of the figure of merit, can be helpful but were not conclusive. A particle contribution distribution histogram which is output at the end of each batch provided a very effective way of detecting undersampling.

INTRODUCTION

The first question adressed in the assessment of a Monte Carlo calculation is: "How can the quality of an estimate be guaranteed?" Assuming that the mathematical modeling is a suitable representation of the real problem, this question is commonlly answered with the estimation of the standard deviation of the mean. However, if the the population sample does not adequatly represent the true population, the estimate should be considered meaningless. This condition is associated with insufficient sampling (undersampling) which occurs in problems where the effect of interest is determined primarily by rare events and a very large population sample is required to achieve a good estimate. This problem was pointed out by Gelbard [1] and Cramer et al. [2], both of whom considered that many of the problems associated with statistical uncertainty were still unresolved.

The problem of undersampling is particularly important when point-detector estimators are used because the contributions to the outer detectors from collisions near the source region are by their nature very small and can be essentially of the same magnitude. If a sufficient number of particles is not sampled to include enough particles which experience rare events, the final results will be unrealistically small while their standard deviations may indicate seemingly acceptable results. In this work, this condition will be called underestimation.

THEORETICAL CONSIDERATIONS

The statistical behavior of estimators commonly used in Monte Carlo calculations can be better understood with the consideration of Analysis-of-Variance theories. Consider a population sample of I groups each with J elements ($N = I \times J$) as shown in Table 1. The total mean μ is estimated by

$$\overline{x} = \frac{1}{I}\sum_{i=1}^{I}\overline{x}_i = \frac{1}{N}\sum_{n=1}^{N} x_n. \qquad (1)$$

It is possible to derive various estimators for the population variance σ^2. The first one S_w^2 is called the "variance of mean square within", which can be estimated using

$$S_i^2 = \frac{1}{J-1}\sum_{j=1}^{J}(x_{ij} - \overline{x}_i)^2 \qquad (2)$$

so that

$$S_w^2 = \frac{1}{I}\sum_{i=1}^{I} S_i^2. \qquad (3)$$

Another estimator for the population variance S_b^2 is called "variance of the mean square between", which is based on the fact that for a sufficient number of elements, the group means are normally distributed with a variance equal to the population variance divided by J

$$\frac{\sigma^2}{J} = \frac{1}{I-1}\sum_{i=1}^{I}(\overline{x}_i - \overline{x})^2, \qquad (4)$$

Table 1: A Standard Analysis-of-Variance Table.

Group						
1	2	...	i	...	I	
x_{11}	x_{21}	...	x_{i1}	...	x_{I1}	
x_{12}	x_{22}	...	x_{i2}	...	x_{I2}	
⋮	⋮	⋮	⋮	⋮	⋮	
x_{1j}	x_{2j}	...	x_{ij}	...	x_{Ij}	
⋮	⋮	⋮	⋮	⋮	⋮	
x_{1J}	x_{2J}	...	x_{iJ}	...	x_{IJ}	
$\overline{x}_1$	$\overline{x}_2$	...	$\overline{x}_i$	...	$\overline{x}_I$	$\overline{x}$

and

$$S_b^2 = \frac{J}{I-1}\sum_{i=1}^{I}(\bar{x}_i - \bar{x})^2. \quad (5)$$

This is the batch estimate of the population variance and is the procedure used in standard versions of the MORSE Monte Carlo code. Finally, a third estimator for the population variance is designated as S_t^2 and is called the "variance of the mean square total",

$$S_t^2 = \frac{1}{N-1}\sum_{i=1}^{I}\sum_{j=1}^{J}(x_{ij} - \bar{x})^2. \quad (6)$$

Any of the equations 3, 5, or 6 can be used to calculate an estimate for the population variance S^2 and estimates for the variance of mean $S_{\bar{x}}^2$ are given by:

$$S_{\bar{x}}^2 = \frac{1}{N}S_w^2 = \frac{1}{N}S_b^2 = \frac{1}{N}S_t^2. \quad (7)$$

The fractional standard deviation (FSD) is the estimate of precision calculated in the MORSE code and is given by:

$$FSD = \frac{\sqrt{S_b^2}}{\bar{x}}. \quad (8)$$

The Utilization of F Tests. It is expected that the group means are normally distributed, but suppose that one of the groups yielded a much larger value for its mean. This would suggest that rare events were sampled in that particular group — and that the other groups may be subject to underestimation. Therefore, this condition could give an estimate of the sample size necessary to sample enough rare events to prevent undersampling and underestimation.

A statement of the F test for the equality of the group means would be:

$$\begin{aligned} &H_0 : \mu_1 = \mu_2 = \ldots = \mu_I \\ &H_1 : \text{at least one is different,} \end{aligned} \quad (9)$$

and it is performed by calculating

$$F_{(I-1)(N-I)} = \frac{S_b^2}{S_w^2}, \quad (10)$$

which, as can be demonstrated, has the expected value

$$\frac{\sigma^2 + \frac{J}{I-1}\sum_{i=1}^{I}\alpha_i^2}{\sigma^2}, \quad (11)$$

where $\alpha_i = \mu_i - \mu$. When the null hypothesis is true, the test yields a value of one.

Coefficient of Variation of the Standard Deviation. Another procedure that provides measures of reliability of the various estimators is the concept of relative error. The coefficient of variation V is defined as the square root of the relative variance of the population V^2, which is given by:

$$V^2 = \frac{S^2}{\bar{x}^2} = \frac{1}{N-1}\sum_{i=1}^{N}\frac{(x_i - \bar{x})^2}{\bar{x}^2}. \quad (12)$$

Analogously, the relative variance of the variance is given by:

$$V_{S^2}^2 = \frac{\sigma_{\sigma^2}^2}{(\sigma^2)^2}. \quad (13)$$

It is possible to demonstrate[3] that Equation 13 can be expressed as

$$V_{S^2}^2 = \frac{1}{N}\left(\beta - \frac{N-3}{N-1}\right), \quad (14)$$

where β is the kurtosis of the distribution.

It can also be demonstrated that the coefficient of variation of the standard deviation V_S is related with the coefficient of variation of the variance V_{S^2} simply by $V_S = V_{S^2}/2$. A reasonable value for V_S is a subject for concern. The confidence limits for the standard deviation do not need to be the same as the confidence limits for the mean. Because V_S depends upon the kurtosis of the distribution, the significance of the confidence limits of the standard deviation varies for different distributions. A distribution with a large kurtosis is characterized by a large value for V_S which can be more reliable than an undersampled distribution with a small kurtosis and a much smaller V_S .

The Figure of Merit. The figure of merit $\sigma_{\bar{x}}^2 T$ is widely accepted as a measure of the calculational efficiency and also as a indicator that the solution has achieved the asymptotic $1/\sqrt{N}$ behavior as predicted by the central limit theorem. The figure of merit can also be expressed as

$$FOM = \frac{1}{\sigma_{\bar{x}}^2 T}, \quad (15)$$

where $\sigma_{\bar{x}}^2$ is the variance of the mean and T is the total computation time. A larger FOM indicates a more efficient calculation.

Considering t as the average computation time per particle, Equation 15 can be rewritten as

$$FOM = \frac{N}{\sigma^2 N t} = \frac{1}{\sigma^2 t}. \quad (16)$$

Therefore, when σ^2 becomes constant, i.e. a sufficient number of particles have been sampled, the figure of merit also becomes constant. However, for highly skewed distributions with a small proportion of very large contributions, the behavior of $\sigma_{\bar{x}}^2$ is typically not $1/\sqrt{N}$ and may assume either faster or slower rates of convergence. The figure of merit will become stable only after a sufficient number of particles have been sampled. Also, if undersampling is severe, the figure of merit will appear essentially constant over a wide range of sample sizes thus giving false indications about the solution.

The Sampling Distribution and Rare Events.

The reasoning for the examination of the sampling distribution in the analysis of the accuracy of Monte Carlo calculations is that: if the population distribution contains important rare events, these rare events have to be sufficiently sampled so that the estimates of the mean and variance continue to approach their asymptoptically true values.

To obtain information about the sampling distribution, a counter was introduced into the MORSE code that records the number of particles that have the value of their contributions within a given range or channel. Therefore, it is possible to know how many particles account for specific fractions of the total response. This counter is collapsed from 188 intervals to 10 percentage intervals which would indicate if a small number of particles are responsible for a large percentage of the total value of the response. This provision turn out to be very useful in the analysis and identification of undersampling because of its ability to isolate the contributions due to rare events from background contributions.

It is possible to use any previously known information about the population distribution to enhance the probability of scoring rare (important) events. This procedure is commonly accomplished by variance reduction techniques such as importance function biasing, survival biasing, stratified sampling etc. The variance can also be reduced by using statistical estimators such as the last-flight and the next-event estimators. However, the utilization of these techniques does not guarantee that important rare events will be sampled, which again enphasizes the importance of observing the behavior of the sampling distribution.

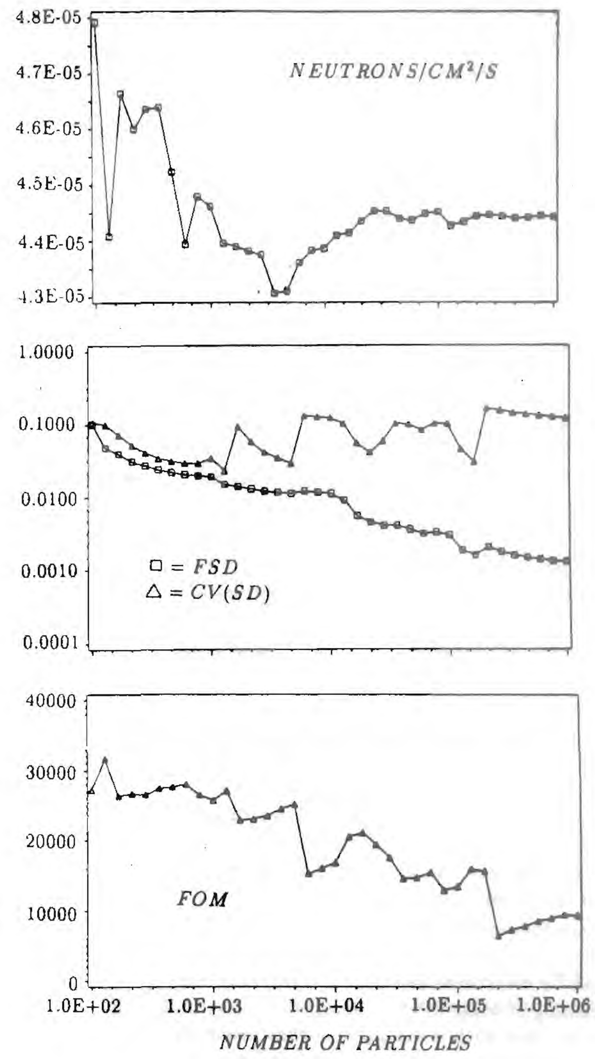

Figure 1: Neutron Flux, FSD, $CV(SD)$, and FOM Behavior for Detector 2 - NESXE.

SELECTED RESULTS

To test the various techniques, a problem was chosen which consists of a 1-meter radius concrete sphere with 10 detectors positioned at 10-cm intervals from the center of the sphere. An isotropic monoenergetic (14-Mev) neutron point source is located at the center of the sphere. All responses are for the first group only (from 8.18 to 15.0 MeV).

The utilization of a next-event surface crossing estimator (NESXE) posed no problems in the solution. Figure 1 show the evolution of the flux, FSD, and FOM with increasing number of particles for Detector 2. It is interesting to observe the asymtoptic behavior of these statistical parameters over a wide range of sample sizes.

An important fact suggested from Figure 1 is that the coefficient of variation of the standard deviation V_S which is denoted by $CV(SD)$ generally follows but is almost always higher than the standard deviation. Also, the changes experienced by the $CV(SD)$ is often much more peaked than that of the standard deviation because of the changes in the kurtosis of the distribution. An increase in the $CV(SD)$ may reflect either a small increase or even a decrease in the standard deviation. Analyzing the FOM behavior it is possible to see that this parameter can reject perfectly acceptable solutions.

Table 2 shows the F tests results for a calculation comprising of 10 batches each having 10^4 particles. It is possible to see that even for extremely low standard deviations the tests present meaningless results which could be explained by the fact that the particle contribution distribution is not a normal distribution as required for the application of F-tests.

The utilization of point-detector estimators (PDE) significantly increases the degree of difficulty in the solution of this problem. In fact, the effects of undersampling can be easily seen when the point detector solutions are compared with solutions using the next-event estimator.

Table 2: F Values for Sample Problem 1 with Next-Event Estimator.

Detector	Neutron Flux	FSD	F Value
1	3.8398D-04	0.00327	1.45084
2	4.4619D-05	0.00526	0.64334
3	8.9053D-06	0.00679	0.84945
4	2.2075D-06	0.01134	0.98125
5	6.2544D-07	0.01497	0.37831
6	1.8826D-07	0.02405	1.99945
7	6.0974D-08	0.03847	0.31681
8	2.0435D-08	0.04535	0.88189
9	6.3437D-09	0.06413	0.81423
10	2.1429D-09	0.10191	1.16697

Figure 2 shows the behavior of the neutron flux, FSD and $CV(SD)$, and the figure of merit with increasing values for the sample size for Detector 2. Underestimation is observed for sample sizes up to 200,000 particles and the FOM exhibits stable plateaus, therefore, failing to detect the underestimated solutions.

Table 3 shows the particle distribution estimates for 10^5 and 10^6 particles (PDE calculation) and Table 4 shows the particle distribution relative to the value of their contributions. In the case of Detector 7 (10^5 particles) only one particle accounts for more than 50% of the total response. The number of particles in the percentage intervals is not exact because some channels can have enough particles to cover more than one 10 per cent bin. Therefore, in the case of Detector 2 for example, 69700 particles can account for up to 30% of the solution, 14634 for up to 20%. However, this description is sufficient to illustrate the undersampling condition. Based on these results and the discussion above, the solutions for detectors 1, 2, 3, and 4 can be considered reliable. Detectors 5 and 6 represent the beginning of what will be called the "breakdown region", which is associated with a sample size that provides either only a few or no important rare event making the standard deviations – which can be very small – experience large increases. Obviously, all detectors beyond this point will be undersampled unless additional particles are processed.

An important fact, as seen in Table 4, is that the particle contribution table always included a small number of particles which accounted for large fractions of the solutions for all the undersampled detectors. Therefore, the histogram is able to detect the undersampling condition for both the breakdown region and the region of severe undersampling. In deep-penetration problems, the explanation for this behavior is that even in the severe undersampling condition, there are some particles that make contributions only large enough to make them stand out from the background contributions, and as the sample size increases important rare events are sampled and these contributions will eventually move to the intermediate channels. One possible exception can be seen in the case of Detector 2 which presented a very heavy contribution in the third batch that raised the value of the flux estimate by 14.4 per cent while the standard deviation jumped from 2.2 to 5.5 per cent. In this case the histogram presented a high number of particles for all percentage intervals in Batch 2, therefore, failing to detect the anomaly in the solution.

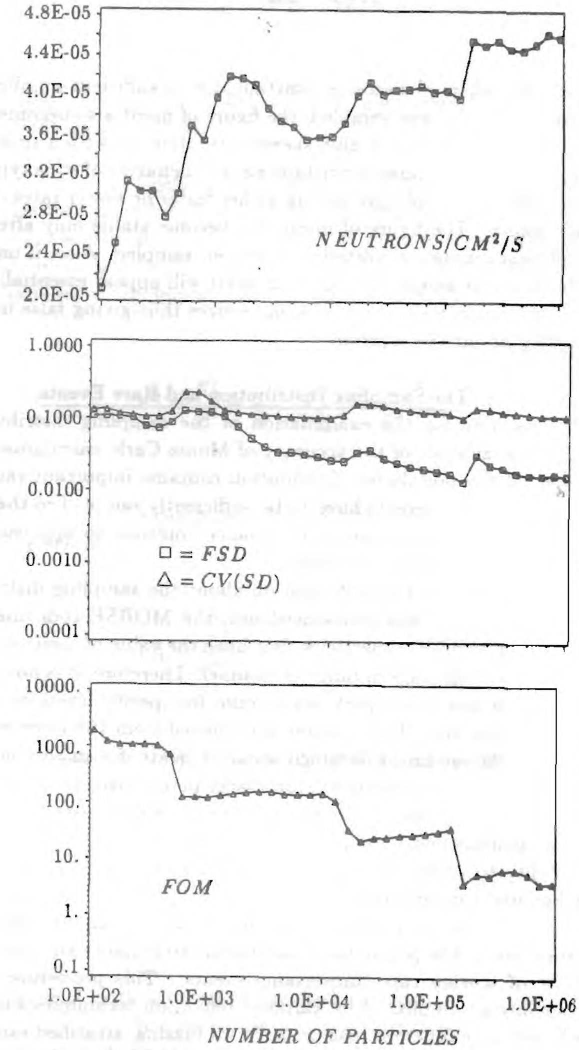

Figure 2: Neutron Flux, FSD, $CV(SD)$, and FOM Behavior for Detector 2 – PDE.

Table 3: Particle Distribution Estimates - PDE.

Det	Neutron Flux (10^5 particles)	FSD	Neutron Flux (10^6 particles)	FSD
1	3.8195D-04	0.024	3.9006D-04	0.009
2	4.0657D-05	0.032	4.5916D-05	0.030
3	1.0318D-05	0.190	9.0757D-06	0.051
4	2.2010D-06	0.105	2.1204D-06	0.058
5	5.8018D-07	0.164	7.0483D-07	0.097
6	1.3670D-07	0.140	1.8347D-07	0.163
7	1.0456D-07	0.651	5.0838D-08	0.151
8	3.6075D-08	0.400	1.7408D-08	0.216
9	4.4715D-09	0.288	5.5855D-09	0.298
10	8.1909D-10	0.221	1.0233D-09	0.216

Table 4: Particle Distribution - PDE.

	Particle Distribution in Percentage Bins Relative to Increasing Value of Contributions									
Det	10.0	20.0	30.0	40.0	50.0	60.0	70.0	80.0	90.0	100.0
					10^5 Particles					
1	82140	0	0	0	6767	7358	717	735	251	49
2	69700	0	14634	7215	3000	1316	971	68	103	5
3	78956	13447	1962	1753	289	144	23	13	1	0
4	81208	12320	2213	405	198	43	21	6	5	0
5	90177	4818	1100	165	76	18	7	3	1	1
6	88529	6765	885	246	37	34	2	4	2	0
7	96335	288	15	3	1	0	0	0	0	0
8	96700	30	0	1	1	1	0	0	0	0
9	96124	619	63	10	0	1	2	0	1	0
10	96140	703	86	21	0	3	2	0	1	0
					10^6 Particles					
1	822591	0	0	0	66835	73435	7097	8228	1869	333
2	696863	0	146584	102522	8582	11123	3570	870	201	18
3	790633	80272	66264	19624	6252	1790	840	281	73	9
4	782995	147859	23780	6419	2527	667	362	30	38	4
5	922034	35882	4794	1585	161	140	10	19	1	2
6	927492	33989	3311	456	270	43	48	1	4	0
7	955274	9730	1164	469	101	39	19	8	3	1
8	964274	3122	424	95	34	4	4	1	1	0
9	966562	2324	125	40	15	1	1	1	0	0
10	967635	2078	295	65	27	8	4	2	0	0

CONCLUSIONS

After these studies the following conclusions were made:

1. The standard deviation is a necessary but insufficient parameter to guarantee the reliability of an estimated effect-of-interest.
2. Estimated parameters such as the figure of merit, the coefficient of variation of the standard deviation, and F-tests, can be helpful but also proved to be inconclusive.
3. The sampling distribution histogram is a powerful tool in the study of problems whose solutions are determined by important rare events and therefore subject to undersampling.
4. Graphics of the asymtoptic behaviors of the estimates of the various parameters provides another powerful and fast means of analysis.

The major conclusion of this work is that Monte Carlo calculations should be accompanied by other formal means to guarantee the quality of an estimate rather than relying on the standard deviation alone. And, the observation of the sampling distribution is a very promising means for this purpose. Another and potentially very important application of sampling distribution behavior techniques is in the study of importance function biasing and other variance reduction procedures.

Acknowledgments. This research is part of the dissertation work of the first author[4] who was supported by the CNPq of the Brazilian government.

REFERENCES

[1] E. L. Gelbard. Unfinished Monte Carlo business. In *Proceedings ANS/ENS Topical Meeting of Advances in Mathematical Methods for the Solution of Nuclear Engineering Problems*, 1981.

[2] S. N. Cramer, J. Gonnord, and J. S. Hendricks. Monte Carlo techniques for analyzing deep-penetration problems. *Nuclear Science and Engineering*, 92, 1986.

[3] M. H. Hansen, W. N. Hurwitz, and W. G. Madow. *Sample Survey Methods and Theory*. John Wiley and Sons, Inc, New York, 1953.

[4] W. J. Vieira. *A General Study of Undersampling Problems in Monte Carlo Calculations*. PhD thesis, University of Tennessee, Knoxville, December 1989.