International Conference on Mathematics, Computational Methods & Reactor Physics (M&C 2009)
Saratoga Springs, New York, May 3-7, 2009, on CD-ROM, American Nuclear Society, LaGrange Park, IL (2009)

RATE OF CONVERGENCE OF ANTITHETIC TRANSFORMS IN MONTE CARLO SIMULATION

Michael S. Milgram
Consulting Physicist, Geometrics Unlimited, Ltd.,
Box 1484, Deep River, Ont. Canada. K0J 1P0
mike@geometrics-unlimited.com

Dagan R. and Becker B.
Institute for Reactor Safety (IRS)
Forschungszentrum Karlsruhe GmbH
Postfach 3640, D-76021 Karlsruhe, Germany
bjoern.becker@irs.fzk.de; dagan@irs.fzk.de

ABSTRACT

Various antithetic transforms are used to explore how variance can be reduced in random walk simulation, using numerical experimentation on typical models and antithetic sampling. It is shown that the use of non-scoring tallies is a helpful tactic to reduce variance in certain problems by introducing anti-correlation, but that in other problems it is equivalent to well-known means of achieving the same effect. The use of transforms that introduce negative weights is explored and found to be beneficial, but only for a certain class of problems. The convergence rate of alternative strategies is compared.
.

Key Words: Monte Carlo, simulation, antithetic, anti-correlation, Markov

1. INTRODUCTION

The method of antithetic variates in Monte Carlo simulation is based upon the observation that the variance associated with the average of two samples t_1 and t_2 is given by

$$\mathrm{Var}([t_1+t_2]/2) = \{\mathrm{var}(t_1) + \mathrm{var}(t_2) + 2*\mathrm{covar}(t1,t2)\}/4$$

so if t_1 and t_2 are anti-correlated, the covariance term becomes negative, and the total variance is reduced. In a previous[1] paper (I), the possible use of antithetic variates in Monte Carlo simulation was investigated, and encouraging results were found, but several questions remained unanswered. In particular, it was noted that one antithetic sampling strategy naturally suggests the tactic of pairing non-scoring and scoring tallies to obtain variance reduction via the resulting anti-correlation, and the natural question that arose was whether this tactic is better than the (orthodox) alternative of attempting to eliminate non-scoring tallies altogether. One purpose of this note is to investigate that question by formulating an abstract model of a typical Monte Carlo simulation model. The main motivation is to explore a range of typical Monte Carlo simulation problems with a view to determining the means by which antithetic variates can be utilized as a useful, and practical tool in reducing variance.

2. SIMULATION OF A SIMULATION

2.1. Generalization of a Monte Carlo Simulation – Representing the Physics

Consider Figure 1. Here, we simulate a typical Monte Carlo simulation where a random walk is generated in typical fashion and a tally of interest is eventually scored. Concentrate on an item of interest belonging to the underlying simulation (physics) by considering situations in which, somewhere along the (Markov) decision tree, there exists a random number (κ) that *"triggers"*, in a general sense, whether a random walk is likely to eventually score in a tally bin of particular significance, or otherwise **sharply and predictably** defines the subsequent random walk as a function of itself. Section 3 deals with the former possibility; Sections 4 and 5 illustrate the latter.

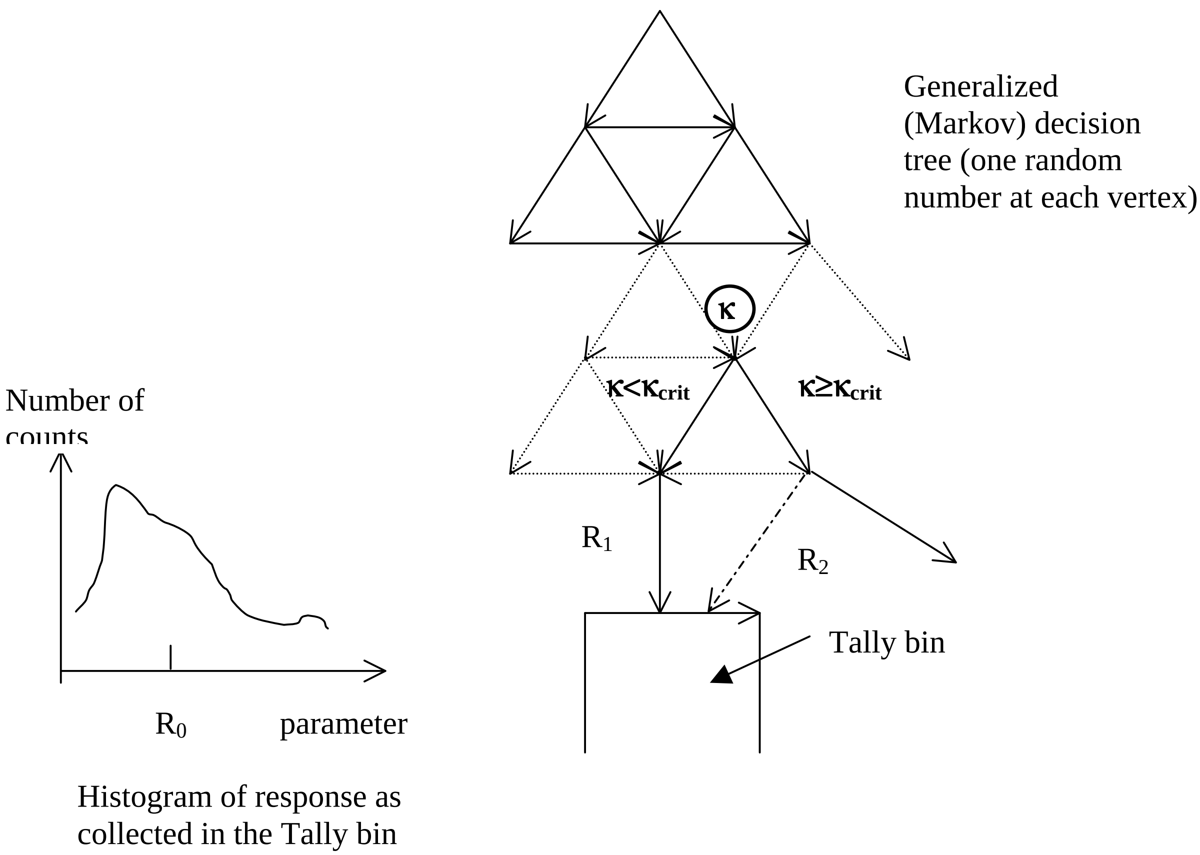

Figure 1. Model of a Monte Carlo simulation. A random number is chosen at each vertex of a decision tree, possibly leading to a score in some tally bin of interest. The collective effect of a set of histories generates a response R_0, illustrated by the histogram at the bottom left of the sketch. The pathway labeled R_2 is arbitrarily chosen to be the "weaker" or "smaller" of the two contributions to R_0 (indicated by a dashed line).

Seen another way, the random number κ controls a decision (and represents a process) that throws a random walk into, or out of, a specific sequence of interest. For example, in a source-sink problem where scattering does not play a significant role, a (source) history that does not head in the general direction of the sink, is unlikely to contribute to a bin that tallies contributions to that target. The random number κ controls the direction of such a particle in the simulation and κ_{crit} symbolizes a critical value of κ that governs the decision.

Consider the scoring as seen by a bean counter living in a tally bin of interest. Each score is the result of a sequence of random numbers chosen and corresponding decisions made along the (Markov) chain. Sometimes a history will score in the tally bin of interest, and gradually the answer, in the form of the mean R_0 and variance corresponding to that distribution, will emerge. For simplicity, consider cases where R_0 consists of two components R_1 and R_2, corresponding to two pathways governed by κ as in Figure 1. If R_1 and R_2 are very different (in magnitude or frequency), this situation becomes a candidate for treatment using antithetic variates.

2.2 Modelling a Monte Carlo Simulation – Representing the Analyst

Introduce a numerical constant (ξ) to characterize one of many possible antithetic transforms selected by an analyst (e.g. see Eq. (19) of I*). In Fig. 1, the trigger variable κ controls the pathway split, and the critical value (κ_{crit}) at which it does so represents the delineation variable representing the physics. ξ represents an analyst's approximation to κ_{crit}. If $\xi=\kappa_{crit}$, the selected transform is optimal and will reproduce any step function intrinsic to the problem, if it exists. However, κ_{crit} is usually either unknown, or inaccessible because it belongs to a continuous function that only approximates a step function. In Section 3, the properties of antithetic transforms belonging to various choices for ξ ($\xi\neq\kappa_{crit}$) are explored.

To model the analyst's simulation, choose ζ randomly from the uniform distribution and compare it to the selected value of ξ. Define response functions $\boldsymbol{R}(\varsigma)$, $\boldsymbol{R}_j(\varsigma)$ and R_j as follows:

$$\begin{aligned}\boldsymbol{R}(\varsigma) &= \boldsymbol{R}_1(\varsigma) \;\; if \;\; \varsigma < \xi \\ &= \boldsymbol{R}_2(\varsigma) \quad otherwise,\end{aligned}$$

thereby selecting one of the two possible responses (histories) of a two-part antithetic transform (e.g. Eq. 19 of I). See Figure 2. Since the reality of the physical process may not coincide with the analyst's selection ($\xi\neq\kappa_{crit}$), define a second set of response functions

$$\begin{aligned}\boldsymbol{R}_j(\varsigma) &= \mathrm{R}_2 \;\; if \;\; \varsigma > \kappa_{\mathrm{crit}} \\ &= \mathrm{R}_1 \; otherwise, \;\; with \;\; j = 1,2\end{aligned}$$

to represent the actual contributions to the tally bin corresponding to each of two paths selected by the trigger κ and accurately reproduce the physics. The common case - one of $\mathrm{R}_{1,2}$ vanishes - corresponds to a physical situation (e.g. rejection) where the trigger κ **always** yields a non-scoring tally along one pathway. Repeat this procedure (choose ζ randomly) N times, corresponding to a typical Monte Carlo simulation using N random walks.

* where ξ is symbolized by the variable α.

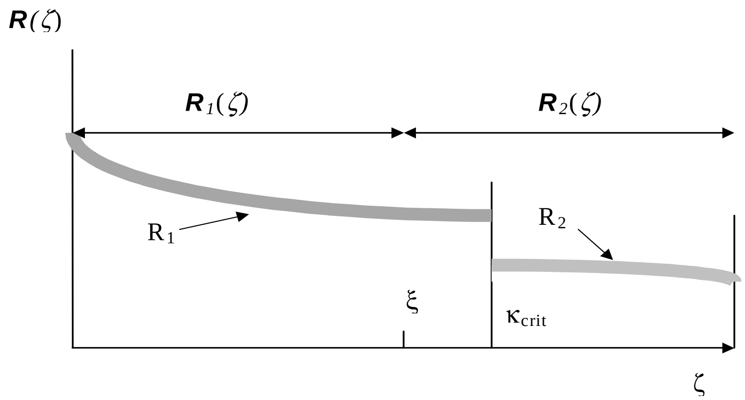

Figure 2. This sketch shows the various response functions defined in the model equations. The thicknesses of curves R_1 and R_2 represent their statistical nature.

Since the traditional estimates of mean and variance are unreliable when samples are not independent (see I), it is prudent to occasionally verify that variance is actually reduced. To achieve this, repeat each (N-history) simulation L times, using different pseudo-random number sequences, and calculate the mean of the means $\overline{\mu}$ and the variance of the resulting distribution $\mathrm{var}(\overline{\mu})$ as in I, terminating at a preset accuracy Δ=0.01. Also calculate $\overline{s_\mu^{\,2}} = \frac{1}{L}\sum_{\ell=1}^{L} s_{\mu,l}^{\,2}$ the average of all individual estimates of variance $s_{\mu,l}^{2}$. If $\mathrm{var}(\overline{\mu}) \sim \overline{s_\mu^2}$ then the estimates of variance are reliable. Otherwise, variance has been reduced (or possibly increased!), but the reduction (increase) has not been observed.

3. TWO-PART DELINEATION

3.1 Simple Models with Non-scoring Pathways (R_2=0)

Consider situations where the possibility that a history will contribute to a tally can be cleanly predetermined by the physics – e.g. a point source radiating to a spherical target through a vacuum. Choose ξ according to the geometry, since all trajectories that lie within the cone whose apex lies at the source and which subtends the target, are known to contribute, and others cannot. Choose R_1 to be the uniform, random distribution **U**. Then compare tallies as described in Section 2, according to one of the following three strategies (i=1,..,N):

1. RANDOM: Random choice of ζ_i , then sample from $\boldsymbol{R}(\zeta_i)$;

2. ANTITHETIC (mirror) (ξ=1/2): Random choice of ζ_i , first sample from $\boldsymbol{R}(\zeta_i)$, then sample from $\boldsymbol{R}(1-\zeta_i)$, and average the two estimates;

3. DIRECTED (forced collision): Random choice of $\zeta_i{*}\xi$ and sample R_1 with weight ξ.

Choose $\xi=\kappa_{crit}=1/2$ and $R_2=0$, implying a perfect choice of antithetic transform corresponding to a target that subtends one half of the emitting solid angle. The first strategy corresponds to the use of standard Monte Carlo sampling with 50% scoring efficiency; the second corresponds to the *mirror* antithetic transform discussed in I, where non-scoring histories are deliberately used to reduce variance by anti-correlating them with histories that are known to score; the third represents the conventional variance reduction strategy of directing all particles into that region of phase space in which it is predetermined that they will score, and ignoring the rest, by suitably adjusting the tally weight.

Figure 3 compares the results, verifying that the two variance reduction strategies are equivalent, as predicted in the Appendix. Either strategy reduces variance by a factor of about two relative to the same calculation performed with random sampling.

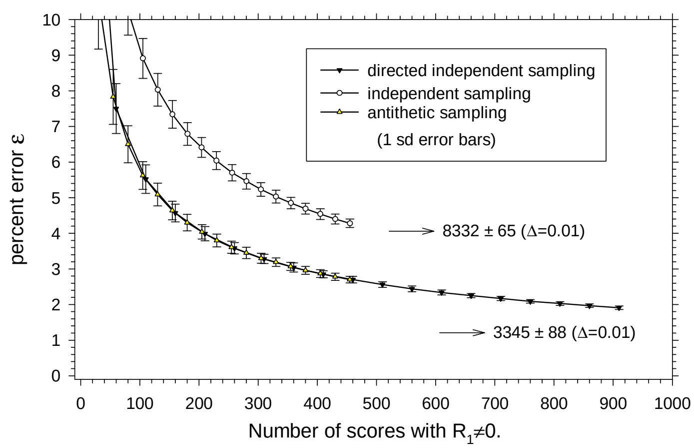

Figure 3. Convergence of the average error estimate when sampling from $R_1 \subset U$, R_2=0, and the corresponding number of scores when Δ=0.01.

It is fair to ask whether a poorly chosen antithetic transform ($\xi\neq\kappa_{crit}$) also imparts an advantage relative to random sampling. The case $\xi>\kappa_{crit}$ may be chosen by a cautious analyst when dealing with problems in which a clean upper κ_{crit} bound is known to exist, the exact value of which cannot be precisely determined by the physics or represented by a simple value of the parameter κ_{crit} - e.g. radiation through the vacuum from an extended source to an irregularly shaped target (particles initially shot "towards" the target may or may not hit; particles directed "away", will not). Tests of this case give results that lead to the same conclusion as above, verifying the utility of the antithetic transform in this situation.

The case $\xi<\kappa_{crit}$ may arise because κ_{crit} is unknown, due to incomplete ability to simply model the details of the pathway belonging to R_2 (e.g. the view factors problem, when the vacuum is replaced with a weakly scattering medium so that particles initially directed "away" from a target still have a (small) probability of hitting it). From the analyst's point of view, this case corresponds to difficult-to-model situations where it may not be possible to select an optimal antithetic transform.

The previous example (ξ=1/2) covers this situation, with case 3 no longer relevant. Sample R_1 from the normal distribution **N**(μ=10,σ=1), continue to use R_2=0 and ask if the antithetic variance reduction strategy retains an advantage relative to random sampling. Figure 4(a) gives a comparison of the convergence rate of the two cases using κ_{crit} =0.75, and Figure 4(b) shows how $\overline{N}$, the average number of evaluations required to obtain a fixed accuracy Δ=0.01, changes as a function of κ_{crit} for the two methods, demonstrating that an antithetic sampling strategy retains an advantage by anti-correlating scoring and non-scoring tallies, provided that the latter occur.

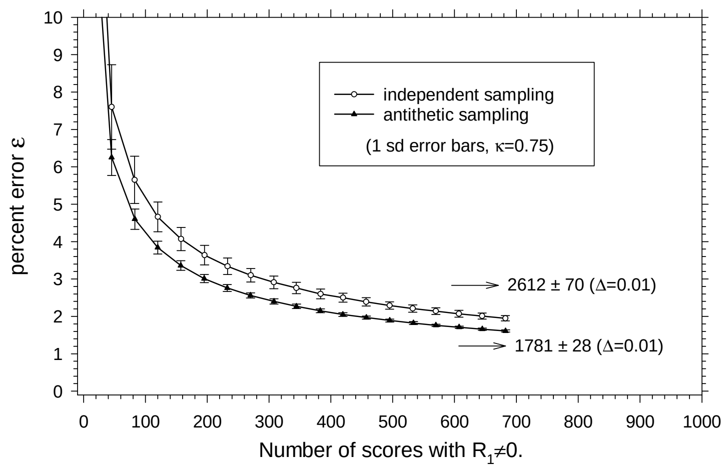

Figure 4(a). Convergence of the average error estimate when sampling from $R_1 \subset N(10,1)$ and R_2=0, as well as the corresponding number of scores when Δ=0.01

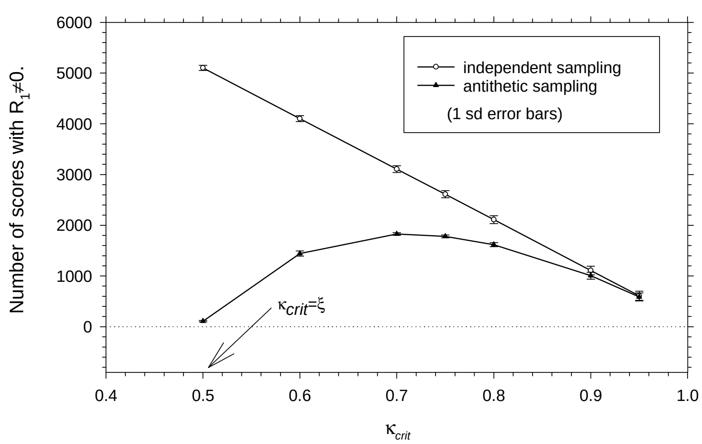

Figure 4(b). Convergence of the average number of non-zero scores with varying κ_{crit} when sampling from $R_1 \subset N(10,1)$ with ξ=1/2 and R_2=0.

3.2 Simple Model where both Pathways Score ($R_2 \neq 0$)

Consider a more realistic model, with tally bin contributions arising from a strong direct source, plus a weak secondary source covering the entire range of possibilities of the trigger. This might correspond to a source-sink problem within a scattering medium. To model this model, sample R_1 from the normal distribution $\mathbf{N}(10,1)$, representing direct collisions from the source, and R_2 from another normal distribution $\mathbf{N}(\mu_2,\sigma=1)$. The magnitude of the difference $|10-\mu_2|$ measures the "strength" of the contribution from the "scattering" medium compared to the direct source, and one choice of parameters ($\xi=0.5$, $\kappa_{crit} = 0.6$ i.e. $\kappa_{crit} > \xi$) simulates a poor choice from the universe of antithetic transforms. Figure 5 compares the use of antithetic and random sampling to model this model, as a function of the disparity between the two contributing processes, showing that the use of antithetic variates is justified in the case where the two different pathways are very different, but that the advantage disappears as the difference between the two distributions decreases, as does the anti-correlation.

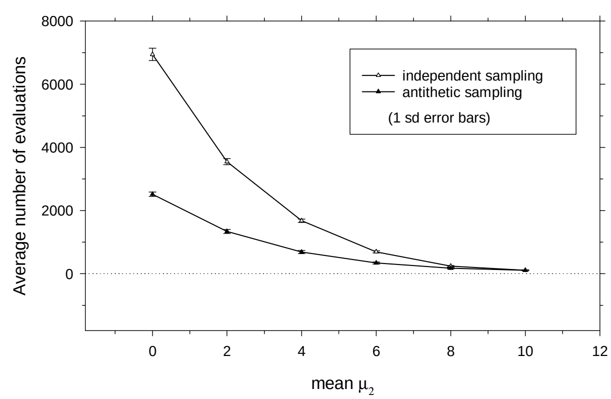

Figure 5. Comparison of the convergence rate for independent and antithetic sampling for a tally bin when $R_1 \subset N(10,1)$ and $R_2 \subset N(\mu_2,1)$ as a function of μ_2.

4. CONTINUOUS DELINEATION

There are many simulations where a step-function trigger does not exist, but a random number may still be found that strongly controls the evolution pathway of a random walk. This situation can be simulated (also see Section 5) by letting $\kappa_{crit} = 1.0$, corresponding to a single response function $\mathbf{R}(\zeta_i) = \mathbf{R}_j(\zeta_i) = R_1$. For an interesting, rather arbitrary, model, sample R_1 from the normal distribution $\mathbf{N}(\mu(\zeta_i),1)$ with a variable mean $\mu(\zeta_i) = 10 \sin(\pi \zeta_i)$. This might correspond to a scattering problem where the most likely path is the direction corresponding to $\zeta_i \sim 1/2$, and a large barrier impedes diffusion in directions corresponding to other values of ζ_i. One possible antithetic transform is[2]:

$$g(x) \to \{2[g(\frac{x}{2}) + g(\frac{1+x}{2})] - g(x)\} / 3$$

where each appearance of *g(…)* on the right represents the contribution of one random walk to the tally. For accurate antithetic sampling, use three-part grouping (and contemplate the necessity (spectre?) of introducing negative track weights to production codes). The convergence rate of antithetic and random sampling was compared, and the results, given in Fig. 6, demonstrate that the antithetic transform gives almost a factor of three advantage relative to random sampling for this model. In both cases, the correct mean (μ=2/π) was found.

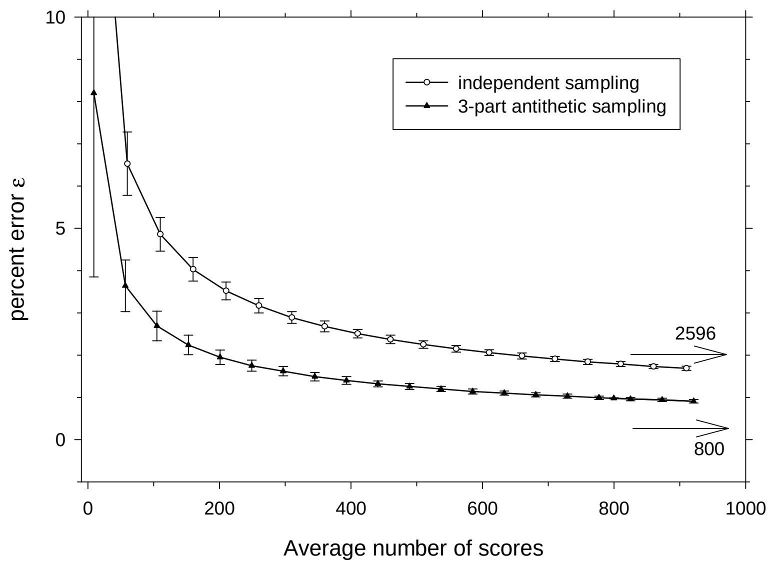

Figure 6. Convergence of the average error estimates using a normal distribution with sinusoidal mean as the response function, and the number of scores needed to achieve Δ=0.01

5. A FURTHER EXAMPLE

As demonstrated elsewhere[3], it is possible to stochastically extract a Doppler broadened cross section by sampling the outgoing energy of an incoming neutron scattered according to model dynamics, from which an (energy dependent) cross section is obtained by interpolating user-supplied cross section tables. One test of this procedure is to estimate the effective (group) cross section under a portion of a resonance by integrating over the broadened cross section sampled randomly as a function of energy. Referring to Fig. 1, accumulating cross section values corresponds to scoring in a tally bin; the energy variable corresponds to the "parameter" label on the horizontal axis of that figure.

A candidate subroutine based upon the MCNP[4] routine *"tgtvel"* was parsed according to the decision tree of Figure 1, and a *"trigger"* variable was found, corresponding to the (randomly chosen) cosine of the scattering angle[†]. The response function (corresponding to R_1 in Fig. 2, with κ_{crit} =1.0 as in Section 4) belonging to this trigger ($\zeta_7 \equiv \kappa$,) shown in Figure 7, suggests that the two-part antithetic transform referred to as *"weighted stratification"* in I (Eq. 19 with $\alpha = 0.5$) might be suitable for this problem, since any odd-numbered walk with an initially selected random value for ζ_7 will be followed by a walk with the antithetically selected value $\zeta_7 \rightarrow \zeta_7 + 0.5$. This means that small (large) values of sampled cross section will be likely paired with large (small) values of the same variable, thereby creating the desired anti-correlation. No two-part response of the type modeled in Sections 2 and 3 (κ_{crit} <1.0) was found in the decision tree, so this example corresponds to a subset of the *"continuous delineation"* model of the type explored in Section 4, with a sharply defined, but single scoring pathway.

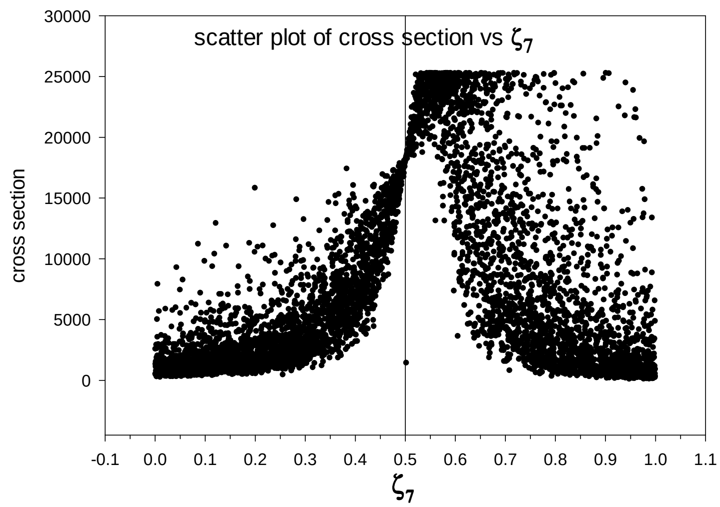

Fig. 7. Response function for random number ζ_7 in the decision tree.

Fig. 8(a) shows the realization of the histogram of the tally bin response corresponding to the sketch in the lower left hand corner of Fig. 1 using random sampling; Fig. 8(b) shows the equivalent histogram for antithetic sampling with $\alpha = 0.5$ (Eq.19 of I). In both cases, the calculated mean is the same, but antithetic sampling reduces the standard error by 30% relative to random sampling after the same number of random walks.

Figure 9 shows the convergence of the standard error estimates of the two methods of calculation to the final results cited in Figures 8.

[†] Since this was the seventh random number examined in the decision tree, it is labeled ζ_7 here. All other response functions associated with other random numbers in the decision tree were randomly scattered and are therefore of no particular interest.

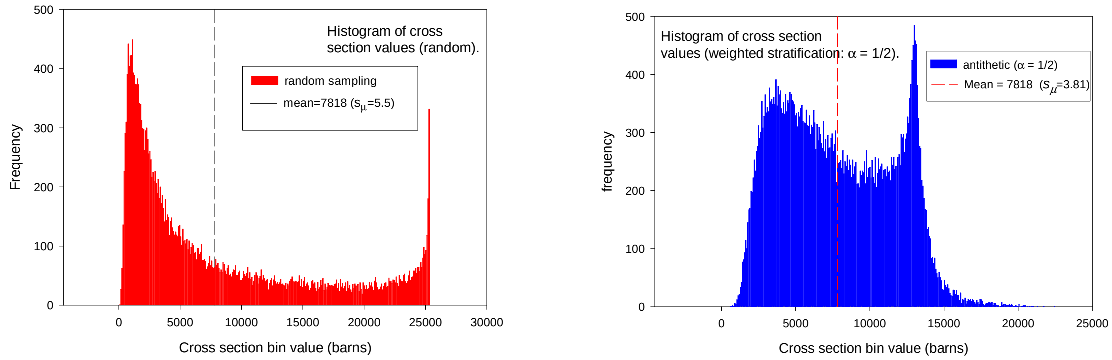

Figure 8(a) and (b). These histograms show the respective sampling distribution for random and antithetic sampling.

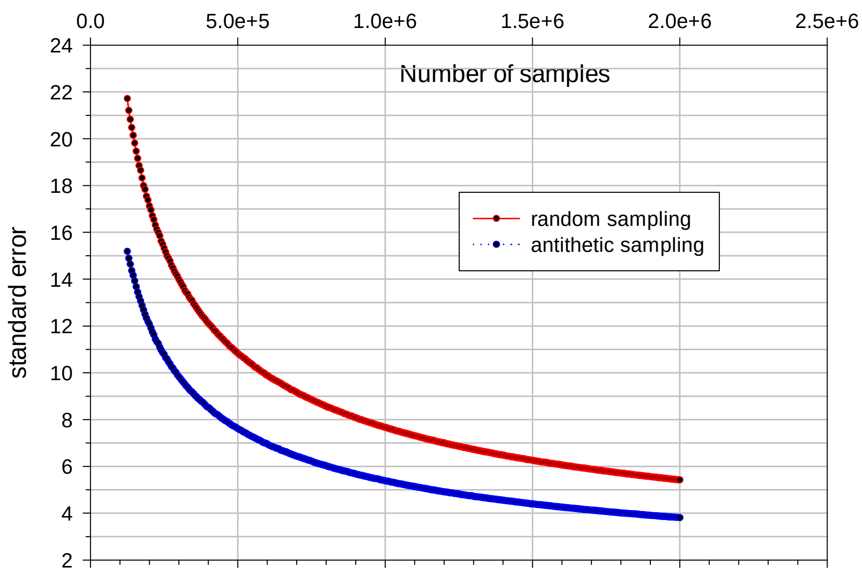

Figure 9. Convergence of the standard error estimate for random and anthetic sampling.

For verification (and for reasons discussed in I), both calculations depicted in Figures 8 were repeated 300 times with different random number sequences, and a histogram of the predicted mean was created for each. These are shown in Figure 10, along with a least squares fit to a normal distribution for each histogram. The means and standard error of each of the two fits are in accord with the estimates given in Figures 8 and 9, demonstrating that the variance reduction obtained by antithetic sampling compared to random sampling is real.

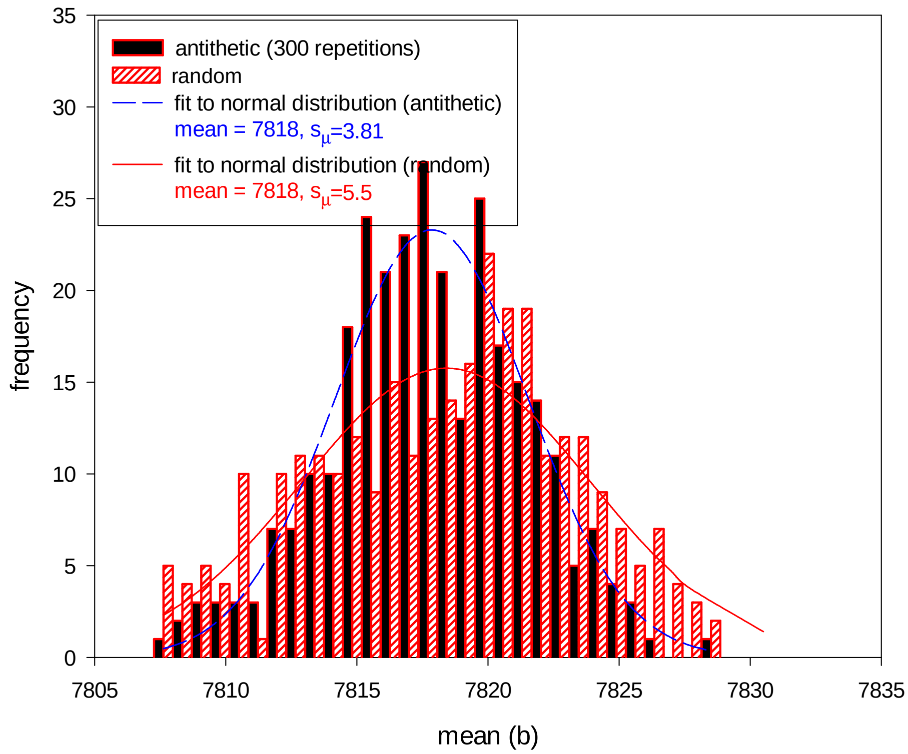

Figure 10. Histogram of the actual distribution of the mean after 300 repetitions using different random number sequences, showing agreement with the estimates of Fig. 9 and demonstrating that variance has been reduced by the use of antithetic variates.

6. SUMMARY

The use of antithetic variates has been studied by numerical experimentation and modelling. The well-known result that an antithetic transform should be carefully chosen to match the characteristics of a given problem was reproduced, and by sampling several classes of problems, the approximate advantage that can be expected by using antithetic variates was observed. It was demonstrated that the tactic of using non-scoring tallies is equivalent to another, more orthodox strategy, in cases where that strategy is available, but in cases where it isn't, the use of non-scoring tallies is a useful method of introducing anti-correlation, and thereby, variance reduction. In other instances, this effect can be obtained by introducing negative (track) weights, or by carefully weighting and ordering the choice of random numbers. The obvious application is to random walk problems where a physical "trigger" vertex exists in the (Markov) chain. It is noted that the identification of a trigger and its response function suggests the possibility of dynamically adjusting anti-correlation parameters. It is also noted that re-ordering the random walk sequence can affect the variance obtained in Monte Carlo simulations, either deliberately (antithetic) or spontaneously (see I). This could influence parallel implementations of Monte Carlo codes, where re-ordering commonly occurs because of processor and network properties.

Acknowledgments: Much of this work was performed Jan-May, 2001 while one of the authors (MSM) visited the Mechanical Engineering Faculty, Technion, Israel Institute of Technology, supported by the Lady Davis Foundation.

REFERENCES

1. Milgram, M., "On the Use of Antithetic Variates in Particle Transport Problems" *Annals of Nuclear Energy,* **28**, pp.297-332 (2001).
2. Halton, J., *Nucl. Sci. Eng.,* **98**, pp.299-316 (1988).
3. Becker, B., Dagan, R., Broeders C., Lohnert G., "An Alternative Stochastic Doppler Broadening Algorithm", *Session 29B, this conference*, *Saratoga USA* (2009).
4. X-5 Monte-Carlo Team,"MCNP: A General Monte Carlo N-Particle Transport Code, Version 5, Volume I: Overview and Theory", *Los Alamos National Laboratory report LA-UR-03-1987* (2003).

APPENDIX

Suppose there are two samples $\{t_1, t_2\}$ belonging to response function R_1, and it is known that a trigger exists. Let R_2=0, and without loss of generality, let $\kappa_{crit} = \frac{1}{2}$. According to the third strategy of Section 3 (forced collision), non-scoring tallies are removed prior to sampling by directing histories into a productive region of phase space (ζ<1/2) with weighting factor ½. In this case, the mean and variance are given by:

$$\mu = (\frac{t_1}{2} + \frac{t_2}{2}) / 2 = (t_1 + t_2) / 4 \tag{A.1}$$

and

$$\begin{aligned} s^2 &= (t_1 / 2 - \mu)^2 + (t_2 / 2 - \mu)^2 \\ &= (t_1 - t_2)^2 / 8 \quad . \end{aligned} \tag{A.2}$$

The antithetic equivalent (strategy 2) assumes that there are 4 samples $\{t_1, 0, t_2, 0\}$ from which it is possible to calculate the same parameters using 2 group batches (see I):

$$\tilde{\mu} = ([t_1 + 0] / 2 + [t_2 + 0] / 2) / 2 = (t_1 + t_2) / 4 \tag{A.3}$$

and

$$\tilde{s}^2 = ((t_1 + 0) / 2 - \tilde{\mu})^2 + ((t_2 + 0) / 2 - \tilde{\mu})^2 = (t_1 - t_2)^2 / 8 \ , \tag{A.4}$$

using $\tilde{}$ to indicate 2-group batches. (Recall[1] that, when estimating variance, batch grouping automatically includes the covariance term between elements of the batch groups.) This demonstrates that the two results are algebraically identical, and computationally competitive if the work involved in obtaining the "0" elements is negligible in comparison to the work involved in obtaining the non-zero samples t_1 and t_2. Alternatively, if non-scoring tallies are an inevitable, but unpredictable, part of the calculation, the generation of random walks should be ordered to increase the likelihood that scoring and non-scoring tallies come in pairs. To generalize the above result, let each of t_1 and t_2 represent a collection of non-correlated samples.