Proceedings of the 1987 Winter Simulation Conference
A. Thesen, H. Grant, W. David Kelton (eds.)

GENERALIZED ZERO-VARIANCE SOLUTIONS AND INTELLIGENT RANDOM NUMBERS

by

Thomas E. Booth
Los Alamos National Laboratory
Los Alamos, NM 87545

ABSTRACT

It is shown that a zero-variance solution exists for any linear Monte Carlo problem and that this solution can be obtained by sampling the random numbers proportional to the expected score subsequently produced by using these random numbers in the random walk.

I. INTRODUCTION

The existence of zero-variance solutions to certain classes of Monte Carlo problems is well known (at least in the particle transport field.) For example, Lux (1987?) writes

> "Zero-variance schemes were first derived through a special importance sampling procedure and these schemes involve last event (absorption) estimators (Kahn 1954, Goertzel and Kalos (1958), Kalos, Nakache, and Celnik (1968)). Zero-variance biasing schemes with collision estimator were introduced by Ermakov (1975) and Hoogenboom (1979). Schemes with arbitrary partially unbiased estimators were derived from the moment equations by Dwivedi (1982) and were generalized by Gupta (1983). Both derivations concern non-multiplying games."

Lux generalized the zero-variance schemes to include multiplying games, e.g. fission.

The present paper shows that <u>any</u> linear Monte Carlo problem (transport or non-transport) has a zero-variance solution if the random numbers used in the calculation are sampled from a score-weighted random number density rather than a uniform density on (0,1]. Because the score-weighted random number density produces a zero-variance solution these random numbers are called "intelligent random numbers." Furthermore, intelligent random numbers exist no matter what other variance reducing schemes the calculation uses. Thus the intelligent random number technique can be used together with all variance reduction techniques.

II. INTRODUCTION TO THE INTELLIGENT RANDOM NUMBER TECHNIQUE

In order to get the reader in the right frame of mind, a simple example will be given before jumping into the more general case.

Consider a particle in an infinite homogeneous medium. The particle moves a distance x sampled from the probability density function (p.d.f.)

$$p_o(x) = \sigma_o e^{-\sigma_o x} \quad , \tag{1}$$

(where σ_o is a material constant). The particle collides at x and it is absorbed at x. If a standard random number generator uniform on (0,1] is used, x is sampled by solving

$$r = \int_0^x p_o(y)dy \tag{2}$$

that is

$$x = -\log(1-r)/\sigma_o \quad . \tag{3}$$

If x is sampled by Eq. (3), this is called an "analog" sampling because x is sampled with the p.d.f. that nature uses. Suppose Monte Carlo is used to estimate the average distance to collision $\bar{x}$.

That is, sample r from a uniform p.d.f., calculate x using Eq. (3), and assign a score

$$s = x \tag{4}$$

to the sample. The empirical average $\hat{s}$ is then an

estimate of $\bar{x}$ where

$$\bar{x} = \int p_o(x)x\,dx \quad . \tag{4.1}$$

A common variance reduction technique is to sample x from a fictitious density

$$p(x) = \sigma e^{-\sigma x} \tag{5}$$

so that (see steps between Eqs. (1) and (3))

$$x = -\log(1-r)/\sigma \tag{5.1}$$

with a weight correction

$$w_p(x) = p_o(x)/p(x) \tag{6}$$

because a fictitious density is sampled. Thus the score becomes

$$s(x) = w_p(x)x \tag{7}$$

The average score $\bar{x}_p$ is

$$\bar{x}_p = \int p(x)s(x)dx = \int p(x)\left[\frac{p_o(x)}{p(x)}\right]x\,dx$$

$$= \int p_o(x)x\,dx = \bar{x} \tag{8}$$

and thus the mean is preserved as desired.

Note from Eq. 3 that x is a function of r and thus the score can be viewed as a function of r rather than x; that is

$$s(r) = w_p(x(r))x(r) \tag{9}$$

Thus

$$\bar{x} = \int p(x)s(x)dx = \int_0^1 s(r)dr \tag{10}$$

is estimated (with N samples) as

$$\hat{x} = \frac{1}{N}\sum_{i=1}^{N} s(r_i) \quad . \tag{11}$$

This is an "analog calculation in the random number space" because the random numbers (r_i) are sampled from the correct (uniform) density. In other words, one can view the calculation as a "black box" with the random number r_i as input and the score $s(r_i)$ as output. The FORTRAN program in Fig. 1 will serve as the black box for this example. The program will produce a calculation that is:

1) analog in the physical density (Eq. 1) and analog in the random number density, if sigma1=sigma0, crit=0, and q(r)=1 for $0 < r \leq 1$.

```
      program blackbx(tty,output=tty,input=tty)
c  n=total number of samples
c  s=score
c  st=sum of scores
c  w=particle weight
c  crit=user supplied critical weight
c  a=0 random numbers sampled from analog density
c  a=non-zero, random no.'s sampled from non-analog pdf
c  ranf()=random number generator uniform on (0,1]
c  sig0=constant for true density
c  sig1=user supplied constant for fictitious density 1
c  sig2=user supplied constant for fictitious density 2
      write(*,*)'input n,a,sig0,sig1,sig2,crit='
      read(*,*)n,a,sig0,sig1,sig2,crit
      st=0
      st2=0
      i=0
   20 i=i+1
      w=1
      if(i.gt.n)go to 110
c  random number selected from q(r)=(1+a*r)/(1+.5*a)
c   with weight multiplier wr=1 if r sampled
c   from uniform density (a=0)
c   with weight mult. wr=1/q(r) if r sampled from q(r)
      eta=ranf()
      if(a.ne.0)r=(-1+sqrt(1+2*a*(1+.5*a)*eta))/a
      if(a.eq.0)r=eta
      q=(1+a*r)/(1+.5*a)
      wr=1./q
      w=w*wr
      if(w.lt.crit)go to 50
c  sample distance to collision using fictitious sig1
c   rather than true sig0 and apply appropriate weight
c   correction
      dist=-log(1-r)/sig1
      wp=sig0*exp(-sig0*dist)/( sig1*exp(-sig1*dist) )
      w=w*wp
      s=w*dist
      st=st+s
      st2=st2+s**2
      go to 100
   50 continue
c  sample distance to collision using fictitious sig2
c   rather than true sig0 and apply appropriate weight
c   correction
      dist=-log(1-r)/sig2
      wp=sig0*exp(-sig0*dist)/( sig2*exp(-sig2*dist) )
      w=w*wp
      s=w*dist
      st=st+s
      st2=st2+s**2
  100 continue
      go to 20
  110 continue
c  compute estimated mean distance edist
      edist=st/n
      relerr=sqrt( (st2/n-edist**2)/n )/edist
      write(*,*)'edist,relerr=',edist,relerr
      end
```

2) non-analog in the physical density (Eq. 5) and analog in the random number density, if sigma1 $\neq$ sigma0, sigma2 $\neq$ sigma0, and q(r)=1 for $0 < r \leq 1$.

3) analog in the physical density (Eq. 1) and non-analog in the random number density, if sigma1 = sigma0, sigma2 = sigma0, and q(r) $\neq$ 1.

4) non-analog in the physical density (Eq. (5)) and non-analog in the random number density, if sigma1 $\neq$ sigma0, sigma2 $\neq$ sigma0, and q(r) $\neq$ 1.

Cases 1 and 2 have already been treated and the mean scores were shown to be the same in both cases. For cases 3 and 4 the random number r is <u>not</u> selected from a uniform p.d.f. but rather from a p.d.f. q(r).

(Note that case 3 is a limiting case of case 4 so that only case 4 need be discussed).

For case 4 r_i is selected from $q(r_i)$, by solving

$$\eta_i = \int_0^{r_i} q(\zeta)d\zeta \qquad (12)$$

for r_i as a function of the uniformly distributed random number η_i. The particle's weight is multiplied by

$$w_q(r) = 1./q(r) \qquad (13)$$

because the random number r was not selected from the correct (uniform density). Eq. (5) is then sampled for x by solving

$$r = \int_0^x p(y)dy \qquad (14)$$

that is,

$$x = -\log(1-r)/\sigma \quad . \qquad (15)$$

If s(r) is the score produced by sampling x from the fictitious density in Eq. (5) with an unbiased random number r, then

$$s_q(r) = w_q(r)s(r) \qquad (16)$$

is the score produced by sampling the fictitious density in Eq. (5) with a biased random number sampled from q(r).
The expected score is

$$\bar{x}_{pq} = \int_0^1 q(r)s_q(r)dr \quad . \qquad (17)$$

Note in Fig. 1 that depending on $w_q(r)$ the σ in Eq. (5) might be either $\sigma=\sigma_1$ or $\sigma=\sigma_2$ in the computer code. If the variable crit=0.9 then an analog sampling of r will always use $\sigma=\sigma_1$ to sample the distance to collision because $w_q(r) = 1 > 0.9$ for an analog sampling of r. However, $w_q(r)$ might be less than 0.9 if r is sampled from q(r) and there exists a range of r for which $q(r) > 1/0.9$. Thus biasing in the random number space can affect how the program calculates the distance to collision. If σ_1 is used, then $x_1 = -\log(1-r)/\sigma_1$ whereas in the biased random number case then σ_2 might be used and $x_2 = -\log(1-r)/\sigma_2 \neq x_1$. In this case not only has the particle's weight been altered by the biasing of r, the particle's random walk has also been altered ($x_1 \neq x_2$).

Rewriting Eq. (17) to explicitly indicate whether σ_1 or σ_2 is used yields (w_c = crit):

$$\bar{x}_{pq} = \int_0^1 q(r)\left[H(w_q(r)-w_c)s_{q_1}(r) + \left(1-H(w_q(r)-w_c)\right)\right.$$

$$\left. s_{q_2}(r)\right]dr \qquad (18)$$

where $s_{q_i}(r)$ uses x_i from Eq. (5.1) with $\sigma=\sigma_i$, substitutes x_i into Eq. (9) to obtain $s_i(r)$, and substitutes $s_i(r)$ into Eq. (16) to yield

$$s_{q_i}(r) = w_q(r)w_p(x_i(r))x_i(r) \quad . \qquad (19)$$

The q(r) in Eq. (18) will cancel the $w_q(r)$ of Eq. (19). Next change variables in Eq. (18) from r to x_i (use Eq. (5.1)) to yield

$$\bar{x}_{pq} = \int_0^\infty w_p(x_1)x_1\sigma_1 e^{-\sigma_1 x_1} H(w_q(r)-w_c)dx_1 \qquad (20)$$

$$+ \int_0^\infty w_p(x_2)x_2\sigma_2 e^{-\sigma_2 x_2}\left(1-H(w_q(r)-w_c)\right)dx_2 \quad .$$

Using Eqs. (1), (5), and (6) in Eq. (20) yields

$$\bar{x}_{pq} = \int_0^\infty p_o(x_1)\, x_1\, H(w_q(r)-w_c)dx_1$$

$$+ \int_0^\infty p_o(x_2)x_2\left(1-H(w_q(r)-w_c)\right)dx_2$$

$$= \int_0^\infty p_o(x)xdx = \bar{x} \qquad (21)$$

as desired.

Thus far it has been shown that selecting r from a biased random number density q(r) and multiplying the sample weight by 1/q(r) results in the same mean as without the random number biasing. The "intelligent random number technique" provides a zero-variance solution to this problem by score-weighting the random number density. That is,

$$q(r) = s(r)/\bar{x}$$

so that from Eq. (13) and (16),

$$s_q(r) = w_q(r)s(r) = \frac{\bar{x}}{s(r)}\, s(r) = \bar{x} \quad .$$

Thus every sample contributes the mean score and a zero-variance solution results. In the next section this score-weighting is generalized to provide zero-variance solutions to any linear Monte Carlo problem.

III. THE GENERAL CASE

Typical Monte Carlo calculations require more than one random number to obtain a sample. Let $\vec{r} = (r_1, r_2, \ldots)$. That is, each sample takes as many random numbers from a random number generator as the sample requires. For each random number vector simulated an estimate $s(\vec{r})$ is produced. Following particle transport terminology $s(\vec{r})$ is called the sample's score. Each time a random number is selected an event occurs. (The event being the particular value of the random number). There is an event score (often zero) associated with each event. The sample score is the sum over all the event scores in the sample.

A typical event tree is illustrated in Fig. 2a. The branches indicate the dependence/independence of the games and scores. For instance, the knowledge of r_1 and r_4 (along with the densities they were sampled from) determines the game r_{10} will be used to sample. Every game depends only on the events above it. For example, in particle transport calculations r_4 might have selected a fission process with three emerging fission neutrons whose random walks correspond to the three independent branches emerging from node 4. These branches are shown independent because the fission neutrons' subsequent (after node 4) random walks are sampled independently. An s_i is the event score that would be generated by the selection of r_i from a uniform density on (0,1].

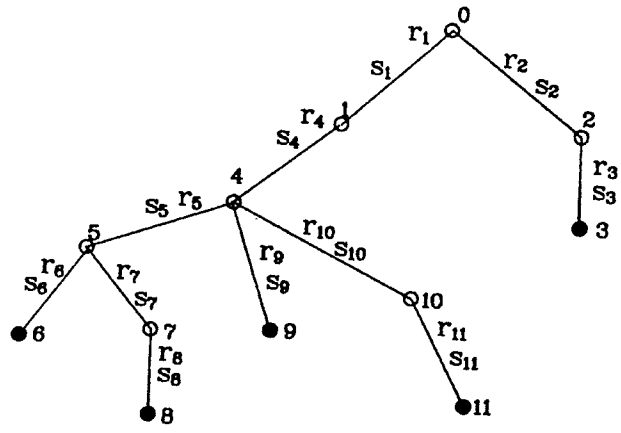


Fig. 2a. Typical random number tree for general Monte Carlo simulation.

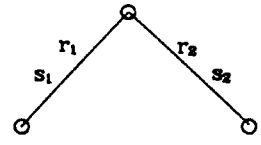


Fig. 2b. Random number representation of two independent games.

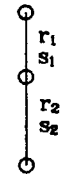


Fig. 2c. Random number representation of dependent games.

Two things are worthy of note here. First, all linear Monte Carlo calculations have a tree structure like Fig. 2a whether or not variance reduction techniques are used. Second, the random numbers need not be selected from a uniform density. There will of course be a weight correction.

$$w_i = 1./q_i(r_i) \quad ,$$

if r_i is sampled from $q_i(r_i)$ instead of $q(r_i) = 1$. That is, if a random number is made $q_i(r_i)$ times as likely to occur, then any branch containing r_i will be weighted by $q_i^{-1}(r_i)$ so that the expected score generated by r_i and all branches below i is preserved. Stated mathematically, if $s(r_i)$ is the event score generated by sampling r_i from a uniform density then $s_i(r_i)/q_i(r_i)$ is the event score if r_i

is sampled from $q_i(r_i)$. Furthermore, if $M_i(r_i)$ is the mean total score for all events below r_i (when r_i is sampled from a uniform density), then $M_i(r_i)/q_i(r_i)$ will be the mean total score for all events below r_i when r_i is sampled from $q_i(r_i)$. The mean total score produced by sampling r_i from a uniform density is

$$s_i(r_i) + M_i(r_i) \quad , \tag{22}$$

whereas the mean total score produced by sampling r_i from $q_i(r_i)$ is

$$w_i(r_i)\{s_i(r_i)/q_i(r_i) + M_i(r_i)/q_i(r_i)\} = s_i(r_i) + M_i(r_i) \tag{23}$$

as desired. (Note that the mean score for a particle of weight w has to be w times the mean score for a particle of weight one, otherwise the calculation will not produce the correct estimate).

It may help to visualize your favorite Monte Carlo code using your favorite variance reduction techniques. The only difference suggested here is that each time the Monte Carlo code calls the random number generator, a random number r_i is sampled from $q_i(r_i)$ and the branch weight is multiplied by $1/q_i(r_i)$ so that all subsequent events below r_i include the weight multiplication. (In particle transport terminology, the particle track weight is multiplied each time the random number generator is called.) All other processes continue as before. The theorem below shows the proper random number p.d.f. to produce a zero-variance solution.

Theorem

If every random number is sampled proportional to the expected score beneath it (i.e. "intelligent" random number sampling) on the event tree, then a zero-variance solution results for a finite event tree.

Proof

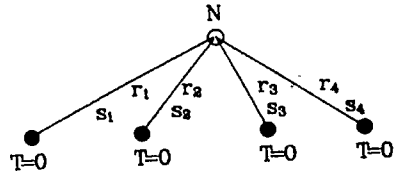
Suppose L samples are taken using "intelligent" random numbers as described above. This results in L independent event trees. It will be shown that each of the L samples produces a sample score equal to the mean sample score, independent of the structure of the sample's event tree. Thus a zero-variance solution results. Consider an arbitrary event tree, say from the k-th sample. It will be shown that the score produced below any node for the k-th event tree is equal to the mean score produced below that node (taken over all possible samplings of events below the node). Stated another way, the score below any node can be made independent of the sampling beneath that node.

Because the event trees are finite, the branches eventually terminate. For node N define:

1) T_N = total score due to all events below node N

   M_N = average score over all possible events below node N

One can evaluate T_N and M_N at branch termination points as $T_N=M_N=0$ because no score can be produced after termination. Now it is possible to go back up the tree and calculate T_N for any node that has all connected nodes below it evaluated. On the first pass, the only known nodes are termination nodes where T=M=0, so the only nodes that T_N can be evaluated at are the nodes that only have termination nodes below. For instance,



Here T_N can be evaluated because the r_i are known to have been chosen from the density:

$$g_i(r_i) = \frac{[s_i(r_i) + M_i(r_i)]}{\int[s_i(x) + M_i(x)+dx} \tag{24}$$

and the corresponding weight multiplier is

$$w_i(r_i) = g_i^{-1}(r_i) \tag{25}$$

and the unweighted score is

$$s_i(r_i) \tag{26}$$

so that the total score due to branch i is ($M_i(r_i) = 0$ upon termination)

$$w_i(r_i)[s_i(r_i) + 0] \quad . \tag{27}$$

Now for the particular r_i chosen, the branch termin-

ates and

$$w_i(r_i)s_i(r_i) = \frac{\int[s_i(x) + M_i(x)]dx}{s_i(r_i) + 0}\, s_i(r_i)$$

$$= \int[s_i(x) + M_i(x)]dx \quad . \qquad (28)$$

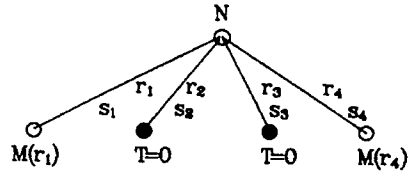
Thus the total score at node N is:

$$T_N = \sum_i \int[s_i(x) + M_i(x)]dx \quad . \qquad (29)$$

However, note that this is the average over all possible subsequent events, thus

$$T_N = M_N \quad , \qquad (30)$$

a constant, regardless of the choice of the r_i. That is, once node N is reached the score generated beneath it is always M_N with score-weighted random numbers. After all nodes that are connected (below) only to termination nodes have been evaluated one can evaluate all nodes that are connected only to previously evaluated nodes and termination nodes. The proof thus works its way back up the tree to the source node provided that each node has a constant score beneath it. Thus it needs to be shown that an arbitrary node N (like shown below) that is connected only to nodes that have a constant score below them, will ensure that node N has a constant score below it.



T_N here can be evaluated in the same way except that instead of $s(r_i)+0$ one uses $s(r_i)+M_i(r_i)$.

Thus

$$g_i(r_i) = \frac{s_i(r_i) + M_i(r_i)}{\int s_i(x) + M_i(x)dx} \qquad (31)$$

$$w_i(r_i) = \frac{\int s_i(x) + M_i(x)dx}{s_i(r_i) + M_i(r_i)} \quad . \qquad (32)$$

The total unweighted score along this branch is $s_i(r_i) + M_i(r_i)$, thus the total weighted score is

$$w_i(r_i)[s_i(r_i) + M_i(r_i)] = \int s_i(x) + M_i(x)dx \quad . \qquad (33)$$

Thus the total score to node N is

$$T_N = \sum_i \int[s_i(x) + M_i(x)]dx = M_N \quad . \qquad (34)$$

Thus every node has exactly its average score beneath it, including the source node 0. This implies that every sample contributes the average score, hence a zero variance solution.

IV. CONCLUDING REMARKS

This paper treated only finite event trees because of space considerations. The theorem is also true for infinite event trees provided only that the mean score is finite. This generalization will be submitted to Nuclear Science and Engineering along with some comments on practical experience with the intelligent random number technique for particle transport problems.

Although not shown here, it is also possible to obtain zero-variance solutions by biasing the p.d.f.'s the random numbers sample and using uniform random numbers. One simply score-weights the p.d.f.'s in the same fashion as the random number density. For example, if $\vec{P}$ represents the current state and $\vec{P}'$ represents the next state, then the zero-variance scheme for sampling $\vec{P}'$ uses the density

$$\tilde{K}(\vec{P}\to\vec{P}') = [s(\vec{P}\to\vec{P}') + M(\vec{P}')]K(\vec{P}\to\vec{P}')C \quad . \qquad (35)$$

Here K is the true density, $\tilde{K}$ is the biased density, $s(\vec{P}\to\vec{P}')$ is the score generated by the event $\vec{P}\to\vec{P}'$, $M(\vec{P}')$ is the mean score generated after arriving at $\vec{P}'$, and C is a normalization constant such that

$$\int\tilde{K}(\vec{P}\to\vec{P}')d\vec{P}' = \int K(\vec{P}\to\vec{P}')d\vec{P}' \quad . \qquad (36)$$

Note that Eq. (35) is essentially Eq. (24) except that the true random number density in Eq. (24) does not explicitly occur on the right-hand side because it is equal to one.

REFERENCES

Monte Carlo Methods for Neutron and Photon Transport Calculations, Part 5 page 200. Book to be published by I. Lux and L. Koblinger (1987?).

Applications of Monte Carlo, by Herman Kahn, April 1954 The Rand Corporation report AECU-3259.

Goertzel, G. and Kalos, M. H.: "Monte Carlo Methods in Transport Problems" in Progress in Nuclear Energy, Series I, Physics and Mathematics, Vol 2, p. 315 (1958).

Kalos, M. H., Nakache, F. R., and Celnik, J.: "Monte Carlo Methods in Reactor Computations" in Computing Methods in Reactor Physics, ed. Greenspan, H., Kelbert, G. N., Okrent, D., Gordon and Breach, New York (1968).

Ermakov, S. M.: "Monte Carlo Method and Related Problems," Nauka, Moskow (1971), (in Russian) German edition: "Die Monte Carlo Methode and Verwandte Fragen," Oldenburg Verlag, München (1975).

Hoogenboom, J. E.: "Optimum Biasing of Integral Equations in Monte Carlo Calculations," Nucl. Sci. Eng., 70, 210 (1979).

Dwivedi, S. R.: "Zero Variance Biasing Schemes for Monte Carlo Calculation of Neutron and Radiation Transport Problems," Nucl. Sci. Eng., 80, 172 (1982).

Gupta, H. C.: "A Class of Zero-Variance Biasing Schemes for Monte Carlo Reaction Rate Estimators," Nucl. Sci. Eng., 83, 187 (1983).