ON RANDOM NUMBER GENERATION

R. R. Coveyou

Union Carbide Corporation
Oak Ridge, Tennessee 37830, U.S.A.

ON RANDOM NUMBER GENERATION

Robert R. Coveyou

Union Carbide Corporation Nuclear Division
Oak Ridge, Tennessee 37830, U.S.A.

ABSTRACT

The subject of random number generation is currently controversial. Differing opinions on this subject seem to stem from implicit or explicit differences in philosophy; in particular, from differing ideas concerning the role of probability in the real world of physical processes, electronic computers, and Monte Carlo calculations.

An attempt is made here to reconcile these views. We propose not to discuss the role of stochastic ideas in the real world, but rather to discuss their role in our mathematical models. In illustration of these ideas, we construct a mathematical model of the use of random number generators in Monte Carlo calculations, and use the analysis of this model to set up criteria for the comparison and evaluation of random number generators.

INTRODUCTION

It is not easy to understand why the design, used in Monte Carlo calculations, and analysis of random number generators (RNG) is so hard to understand. Or, put another way, why it is so easy to understand in several different, and mutually inconsistent, ways.

Nevertheless, the set of all expressed understandings of the subject slightly outnumbers the set of all workers who have devoted a significant amount of attention to it (people sometimes change their minds).

It is my belief that some of the current variant opinions concerning the design and use, in Monte Carlo calculations, of RNG stem from equally variant opinions, explicit and implicit, concerning the wider question of the logic of the application of stochastic (indeed, any mathematical) methods to the analysis of physical (and other scientific) problems.

This is neither the time nor the place for an extended discussion of these wider questions. We will attempt to deal with them here only insofar as is necessary for the partial understanding of the subject at hand.

The philosophical questions referred to here are those revolving around the relations between physical reality and our mathematical comprehension of this reality. Just what these relations really are, I do not pretend to know. And, so far as I can tell, the physicists don't really pretend to know either.

Here, we can and must restrict ourselves to narrower issues. These issues are those which revolve around the relations between the physical realities which we seek to model in Monte Carlo calculations, and our mathematical models of these realities.

Suppose that we wish to analyze some physical (or other) problem. Let us say that the problem is that of the effective prediction of the future behavior of some physical system. For whatever reason, we desire to use a stochastic model of the system in our analysis. It may well be that our choice is motivated by experience; that we know that stochastic models of systems similar to the system of interest have been effective in the past. Or it may be motivated by reasons of personal preference for stochastic models. The point is that the reasons for the choice of a stochastic model are not so much logical as aesthetic.

So, we construct a mathematical model, of stochastic character, for the physical system which gives rise to our problem, and for its operation. This model takes the form of a probability space, each member of which is interpreted as (a possible record of) a possible outcome of the operation of the system. There is a distinguished single member of the probability space, whose identity we may or may not know, which is interpreted as (the record of) the actual performance of the system. We may also interpret observations of the actual performance of the system as an effort to identify this distinguished element, as precisely as is feasible.

Having constructed our model, we decide, for whatever reason, to use the Monte Carlo method in our analysis (perhaps we wish to present our results at this meeting).

Now, an actual Monte Carlo calculation is, itself, a physical process, involving the physical (and, sometimes, mental and emotional) behavior of computers; human, electronic, and other. Hence, discussion of the (stochastic or deterministic) character of the calculation is subject to much the same difficulties as before. These difficulties can be bypassed, if not resolved, by the same (model theoretic) technique.

So, we here discuss, not actual Monte Carlo calculations, but mathematical models of such calculations. And, in particular, a specific model for such calculations, chosen so that the properties of the RNG, used in Monte Carlo calculations, are emphasized in the analysis of the model.

THE MODEL

We will base our analysis of RNG on consideration of their performance in the evaluation of definite integrals of integrable functions defined on the real unit interval and on higher-dimensional analogues of the unit interval. All of the results here can be and will be extended to evaluation of such integrals defined on compact separable abelian topological groups, which are the most general manageable by the methods here used.

Let

$$U = [u \in R: \ 0 \leq u < 1]$$

denote the real unit interval, considered as a compact, separable, abelian topological group, with group operation ordinary addition (mod 1), furnished with Haar measure equivalent to ordinary Lebesque measure.

For the positive integral n, let

$$U^n = [\vec{U}_n = (U_1,\ldots,U_n): \ U_j \in U; \ 1 \leq j \leq n]$$

denote the n-dimensional real unit cube, furnished with appropriate algebraic, topological, and measure theoretic paraphenalia. These too are as described above.

Let Ω denote the space of all real (or complex) valued integralbe functions

$$\omega: \ U^n \to R: \ \omega(\vec{u})$$

defined on U^n.

We will make Ω into a probability space, in accordance with rules set down later. Then each $\omega \in \Omega$ will be a <u>random function</u>.

Definition: a random number generator (RNG) is (a procedure designed to produce) a sequence

$$\Gamma\{\gamma_0,\gamma_1,\cdots,\gamma_{p-1},\gamma_p,\gamma_{p+1}\cdots\},$$

with each $\gamma_j \in U$.

For specific ω, Γ, and positive integers N, let

$$I(\omega) = \int_{U^n} \omega(\vec{u})\, d\vec{u}$$

denote the integral of ω over U^n, and let

$$\hat{I}(\omega,\Gamma,N) = \frac{1}{N}\sum_{j=0}^{N-1} \omega(\vec{v}_j)$$

$$= \frac{1}{N}\sum_{j=0}^{N-1} \omega(v_j, v_{j+1}, \ldots, v_{j+n-1})$$

denote the familiar Monte Carlo estimate of the integral of ω of sample size N firnished by the RNG Γ.

Our interpretation of this model for the use of RNG in Monte Carlo calculations is that, being armed with a specific RNG Γ, we contract, given any function $\omega \in \Omega$, to provide an estimate $\hat{I}$ of the integral I of ω.

We also assume that the functions ω, presented to us for estimate of their integrals, are chosen at random from the probability distribution we have imposed on Ω

Of course, the specification of the model is not yet complete. We have not yet specified the required probability distribution on the space Ω of <u>problems</u>.

For a really detailed analysis, it would be necessary to specify this distribution exactly. But it would not be wise to follow this course, since it would inevitably lead to the suspicion that the resulting judgments concerning the evaluation of specific RNG, or specific classes of RNG, would be quite sensitive to the choice of problem distribution, and would, possibly be quite different for different distributions.

So, we will <u>not</u> specify the problem distribution in complete detail here. In fact, we will specify the problem distribution to that extent here, and to that extent only, that allows us to make some little progress with the analysis.

So, we ask just what we need to know about the problem distribution before we can say anything meaningful about the comparison of RNG in the light of our chosen criterium of mean square error.

Since we have allowed <u>any</u> sequence Γ to call itself a RNG, we must also set forth criteria which will serve to separate "good" RNG from "bad." It seems to me that a very natural figure of merit for specific RNG is furnished by the mean square error $Q_N(\Gamma,\Omega)$

$$= \left\langle \omega \in \Omega: \ |\hat{I}(\omega,\Gamma,N) - I(\omega)|^2 \right\rangle .$$

An easy calculation yields that

$$Q_N(\Gamma,\Omega)$$

$$= \frac{1}{N^2} \sum_{j=0}^{N-1} \sum_{k=0}^{N-1} \left\langle \omega \in \Omega: \ \omega(\vec{v}_j)\omega(\vec{u}) \right\rangle d\vec{u}$$

$$+ \int_{U^n} \int_{U^n} \left\langle \omega \in \Omega: \ \omega(\vec{u})\ \omega(\vec{v}) \right\rangle d\vec{u}\ d\vec{v}$$

$$= \frac{1}{N^2} \sum_{j=0}^{N-1} \sum_{k=0}^{N-1} A(\vec{v}_j,\vec{v}_k)$$

$$- \frac{2}{N} \sum_{j=0}^{N-1} \int_{U^n} A(\vec{v}_j\vec{u})\ d\vec{u}$$

$$+ \int_{U^n} \int_{U^n} A(\vec{u},\vec{v})\ d\vec{u}\ d\vec{v},$$

if we assume the existence of the correlation function

$$A(u,v) = \left\langle \omega \in \Omega: \ \omega(\vec{u})\ \omega(\vec{v}) \right\rangle .$$

It is quite natural to assume, as we do, that

$$A(\vec{u},\vec{v}) = A(\vec{u} - \vec{v});$$

that is, that A is translation invariant. Then

$$Q_N(\Gamma,\Omega) = \frac{1}{N^2} \sum_{j=0}^{N-1} \sum_{k=0}^{N-1} A(\vec{v}_j - \vec{v}_k) - \int_{U^n} A(\vec{u})\ d\vec{u}.$$

Therefore, in order to make some progress in our analysis, we explicitly make the following assumptions concerning the problem distribution, and for the moment, these only:

For each $\vec{u} \in U^n$, (0)

$$\langle \omega \in \Omega: \ \omega(\vec{u}) \rangle = 0.$$

These exists a function (1)

$$A: \ U^n \to R: \ A(\vec{u})$$

such that, for each $\vec{u}$ and $\vec{v}$ in U^n,

$$\langle \omega \in \Omega: \ \omega(\vec{u}) \ \omega(\vec{u} + \vec{v}) \rangle = A(\vec{v}).$$

also,

$$\langle \omega \in \Omega: \ |\omega(\vec{u})|^2 \rangle = A(\vec{0}) = 1. \tag{2}$$

It is known that, given an otherwise appropriate function A defined on U^n, the formal necessary and sufficient condition that A be the correlation (or covariance) function of a distribution of random functions defined on U^n is that, for each set of N points u_j; $1 \leq j \leq N$; of U^n, and each set of N real constant λ_k; $1 \leq k \leq N$, we have

$$\sum_{j=1}^{N} \sum_{k=1}^{N} \lambda_j A(u_j - u_k) \lambda_k \geq 0.$$

That is, A must be a real, positive definite, function defined on U^n.

The analysis thus far furnishes criteria for the comparison of RNG; it is natural to say that, of the two RNG Γ, and Γ_1 is preferable to Γ_2 iff $Q(\Omega,\Gamma_1) \leq Q(\Omega,\Gamma_2)$. In this case, of course, all we can yet say is that Γ_2 with respect to the problem distribution Ω.

But another natural question, and one for which we must, if we can, provide some sort of answer, is that of determining whether or not a specific RNG, of those RNG of u specific class, or the best RNG of u specific class may be regarded as satisfactory at all.

Now, it might seem natural to assume that the best RNG Γ are those which most clearly resemble a sequence formed by successive random independent choices from U^n with probability distribution that specified by Haar (or Lebesque) measure. We will not find this to be true. But it does seem natural to define a RNG to be satisfactory if it is at least as good as the average such random sequence.

Now this average is quite easily calculated to be

$$\frac{1-\int A(\vec{u})\, d\vec{u}}{N}$$

and we can say that the RNG Γ is satisfactory (at least for sample size N, starting with $\vec{u}_j$) iff

$$Q(\Omega,\Gamma) \leq \frac{1-\int A(\vec{u})\, d\vec{u}}{N} .$$

DISCUSSION

I have carried this analysis much further and, in particular, have carried out the indicated generalization to compact separable abelian groups, and have used the theory of harmonic analysis on such groups to get much more extensive results.

In particular, I have come interesting results about periodic RNG. I have found quite strong evidence that periodic RNG whose output vectors fill a lattice (finite subgroup) of the unit cube may be very good RNG and, in particular, may be quite a bit better than the "random" RNG briefly discussed above. I should also say that *some are much worse!* This has a bearing on the current, widespread, and, I think, unmotivated, distrust of such lattice filling periodic RNG.

DISCUSSION

Golbard: Is it not true that the first objection to conventional random number generators came from Marsaglia, who was playing poker on the computer and found that he was losing consistently? Doesn't this suggest a significant defect in the generator?

Kalos: It was dice!

Coveyou: In Marsaglia's original paper, to which my paper on the Fourier analysis was a partial reply, he said that the conventional method was bad. Actually what he had found out was that this choice $\lambda = \sqrt{p}$ is a very bad choice. It's bad because, while it makes two random consecutive numbers practically independent, it makes three consecutive numbers almost completely dependent, in a very simple way. The fact is that each one of Marsaglia's examples (the first time he wrote about this subject) was a random number generator in which the choice of parameters was faulty and, the fact that it was faulty was known in the business at the time. What I did in my paper was to apply my test to each of the generators he had analyzed, and I found that my test predicted that they would be faulty. On the other hand, Marsaglia's test, a particular form of an χ-square test that he used, showed nothing at all to be wrong with his generator. I said at the time that he had simply chosen random number generators that were not typical of the performance of the best congruential generators. If you chose the multiplier at random, your chance of getting a random number generator as bad as the ones that Marsaglia analyzed would be practically zero.

Caldwell: There does seem to be something funny about Marsaglia's tests, because Whitesides said that with our generator he could not duplicate Marsaglia's results at all. He got good results.

Kalos: Maybe he just had a bad run.

Coveyou: Marsaglia had a bad congruential random number generator, and he simply assumed that he had a typical congruential random generator. He did not — he had a bad one. It was one which, at the time, people in the field knew was not good. In fact, it was not one — it was three or four like this that he tried. All of them had this characteristic.

Gast: In his talk Coveyou discussed a figure of merit, q, which he used to test random number generators. Why isn't it sufficient to use the serial coefficient of various lags, directly, to test the quality of a random number generator?

Coveyou: Because, as I pointed out yesterday, the serial correlation coefficient is not invariant under translations of the addend. Now, what that means is that effectively identical random number generators can have different correlation coefficients. This is a characteristic of correlation coefficients of lag 1, and all other lags for that matter. This fact was first pointed out in Berger Johnson's book. He didn't stress it, but he did point out that the serial correlation coefficient is distinctly not a good test. It is a good test in the sense that, if a random number generator has a high serial correlation coefficient, then it is not a good generator. However, a random number generator that is not good can also have a low serial correlation coefficient.

Indeed it has been suggested (and I was guilty of making the suggestion) that you can reduce the serial correlation coefficient of an additive, linear, congruential random number generator simply by changing the addend. In fact, there is a choice for the addend that actually makes the serial correlation coefficient very small and Berger Johnson pointed out that changing the addend does not change the distribution of the pairs or triplets at all. Without actually going through the mathematics it is a little difficult to see how this can happen: but it does happen, and therefore the serial correlation coefficient is just not a complete description of the correlation.

Kalos: But it must be true that the amount by which the serial correlation coefficient can be changed by this translation is bounded.

Caswell: Yes.

Kalos: If you know that they are bounded, you know that serial correlation coefficients below some value are all equally good. So if you use this information and try to make as many serial correlation coefficients as small as possible, that sounds like a sensible test.

Coveyou: Perhaps it could be. Maybe you can actually exploit this fact I talked about by essentially adjusting the addend not to make the serial correlation coefficient as small as possible, but to make it as large as possible. The maximized correlation coefficients might be used as a test for the original random number generator. You wouldn't actually use the random number generator that you constructed to have as high a correlation as possible. You just use it as an example of this whole class of generators which really all have the same distribution of pairs and triplets.

Kalos: But what you are telling us is that minimum serial correlation coefficient is not the right criterion.

Coveyou: Right.

Kalos: But at the same time there is a bound which you ought to attain, if the generator is any good.

Coveyou: Yes, that is true.

Kalos: So I think that you dismissed Gast's suggestion a little bit too quickly.

Borgwaldt: I have heard the argument, in connection with application of random number generators, that the main point is the character of the problem that you are treating. In other words, that a well-behaved problem will get along with a bad random number generator, if one could specify what a bad random number generator is. Now, I would say that if this is true, and if one can say whether a problem is well behaved or not, that this should enter into your definition of a probability space. Badly behaved problems should have very low measure in your probability space. Could you comment on the definition of the probability space, R, which you introduced. I think you skipped some important points in your argument at that time.

Coveyou: Well, what I have to say about such an approach is this. I am thinking about doing large Monte Carlo calculations on computers, and I believe that the only reasonable choice for my purpose is a random number generator that is, generally, essentially as good as you can make it. In point of fact, I do not agree with the argument. It is true that, in certain problems, you can afford a sloppy random number generator, but, I do not think that should be relevant. The point is that one thing that we all do not want to do is to wonder, everytime we do a Monte Carlo calculation, what random number generator we can get away with. The Monte Carlo user wants a generator he can depend upon.

Kalos: I would like to make two comments: An example, an amusing example, of a situation in which you can get away with a sloppy random number generator is a linear transport problem in which importance sampling has been carried out to the ultimate so that, in fact, you use a zero variance estimation procedure. Then the answer is the same, independent of what random number you use, independent of all properties of the random generator. Thus, in some sense you can get away with sloppy random number generation. The second remark is that I assume that everybody does soluble test problems, from time to time, of a general character. One such problem, for example, is the linear straight-ahead model Boltzman equation. One solves such a problem and looks for the right answer. And one does a few integrals here and there to make sure he gets the right answer for those integrals. You don't tell your friends about this, but you want to make sure that, as installed and as you use it, your random number generator can be relied upon.

Coveyou: I think that is probably correct, the right way to do business. But I suspect that, when people do this, they are not particularly thinking about the random number generator. They ary thinking about the logic of their Monte Carlo code.

Gelbard: Yes, I think that you are agreeing that you would like a random number generator that you can rely on even when you are doing rather peculiar, over-simplified problems, in which the systematics of the generator might become important; I think that was Kalos' point.

Coveyou: I should have mentioned further, the magnitudes of the errors we are talking about. Let me be more specific. If you choose the multiplier at random, then the largest magnitude you expect in the mean square error is proportional to $1/\sqrt{p}$ for a randomly chosen random number generator, and that means that the errors introduced by the random number generator, unless it is very bad, are quite small. In fact they are, in almost every case, far smaller than the statistical error of the calculation.

Borgwaldt: Well, I must come back again to the problem of specification. A random number generator which creates random numbers between zero and one may have a period of p or p over four. But, if I use this random number generator in a specific problem, and call it up in different situations, I will not get this period, but a period which is considerably larger. At one time, for example, we used the following system. We used one random number generator to drive four secondary random generators for exponential distribution, isotropic angular distribution and so on. Each was initialized by a separate initial random number. But then we found that there was a big advantage in having all of these random number generators driven by one primary generator,

and this should, according to my understanding, give a period which is far higher than that of the fundamental random number generator with a period of p over four. Is that right?

Coveyou: But my one comment is that the fact that the random number generator is periodic, and the magnitude of the period, are just simply not relevant to Monte Carlo problems. Perhaps I should make an exception of very extensive Monte Carlo calculations done on the IBM-360 with a single precision random number generator. There the period is a billion. For almost every other machine it is far higher than that. It is just not relevant what that period is; it is just too big already.

Kalos: As I understand the comment, there is a conjecture that when you go through the period, and come back to the same random numbers, they are being used for a different purpose, so that the period does not matter.

Coveyou: I agree with that.

Kalos: However, that strikes me as being something that I would not want to rely on. It is perfectly possible, but I don't want to rely upon it. I suggest, for everybody who is engaged in this sort of work, the following calculation that I make. Consider the period of the random number generator you have, and suppose that your machine is generating nothing but random numbers as fast as it can generate random numbers. On the 6600 in assembly language, since there are two multipliers and there is an instruction which generates one random number, you can generate random numbers at the rate of two per microsecond. If you write a little program, an assembly language program, to generate the whole sequence of random numbers, it will repeat itself after a year. So you run your 6600 for a year doing nothing but generating random numbers before you exhaust the period. I was somewhat reassured by that. On the 7600 the calculation was a little bit harder because of the fact that the multiplier is pipelined; exactly what the cycle time is, I do not know, but it is the order of months. Now, you should do that calculation with your IBM-360 using, I would hope, a double precision generator.

Coveyou: A double precision generator on the IBM-360 would last forever.