RECORD, Volume 22, No. 3*

Orlando Annual Meeting
October 27–30, 1996

Session 126TS
Values and Risks of Complex Financial Instruments: Monte Carlo and Low-Discrepancy Points

Track: Investment
Key words: Credibility, Funding of Pension Plans, Investments, Mathematics, Research, Valuation of Assets, Valuation of Liabilities

Moderator: IRWIN T. VANDERHOOF
Instructors: GRAHAM LORD
ANARGYROS PAPAGEORGIOU†
LEONARD H. WISSNER‡
Recorder: FAYE ALBERT

Summary: Complex financial instruments like collateralized mortgage obligations (CMOs) and complex financial operations like asset/liability management can only be valued using Monte Carlo methods. Whenever we average a series of values that use different scenarios, we are using Monte Carlo methods. There cannot be simple, straightforward formulas that give quick answers. However, Monte Carlo may produce biased expected values and seems useless for the determination of a risk profile of the asset or the asset/liability match. Learn the background of Monte Carlo and see whether the use of low-discrepancy points will improve these impediments to it.

Dr. Irwin T. Vanderhoof: Dr. Anargyros Papageorgiou is from Columbia University where the use of low-discrepancy sequences for valuation was invented. He just returned from a conference on complexity theory in Frankfurt where icons in this

†Dr. Papageorgiou, not a member of the Society, has a Ph.D. from Columbia University in New York, NY, and is working on post-doctoral studies at Columbia University in New York, NY.

‡Mr. Wissner, not a member of the Society, is President and Chief Investment Officer of Ward & Wissner Capital Management in the Village of Hastings-on-Hudson, NY.

Note: The charts for this session are not available online. Please contact Sheree Baker at sbaker@soa.org or call 847/706-3565 for a hard copy.

area, Tezuka and Neiderreiter made presentations . Dr. Papageorgiou is currently doing postdoctoral work with Joe Traub. They are investigating the use of low-discrepancy sequences and building the Finder software for identifying such sequences. His bachelor of science degree is from the University of Athens and his Ph.D. is in computer science from Columbia University.

Graham Lord is a Fellow of the Society of Actuaries (FSA). He received his undergraduate degree from the University of Auckland and has his Ph.D. in analytic number theory. After coming to North America, he became a tenured professor of actuarial science at Lavalle University. Graham has worked for Morgan Stanley, the consulting firm Mathematica and has also taught at Wharton. He now lives in Princeton working as a teacher at Temple University and as a consultant.

Leonard Wissner is a fund manager and originally studied at City College of New York. He went on to study for his Ph.D. in operations research, but before he finished his dissertation, that branch of New York University closed. Leonard manages about half a billion dollars for pension funds and has run his business using immunization techniques for matching duration and convexity of pension liabilities.

Finally, thank you to Chalke and Tillinghast who have cooperated by allowing us to use their software so that we can show the impact of using low-discrepancy sequences in choosing scenarios to run on asset/liability problems.

This session will be broken into several sections:

- Graham is going to present an introduction to Monte Carlo, describing why it, rather than other numerical methods, is used for integration and valuation of complex formulas.

- I will discuss the paper in the current issue of *Contingencies* which presents our results using low-discrepancy points.

- Leonard Wissner will share results using low-discrepancy sequences instead of the usual Monte Carlo simulation for pension fund analysis.

- Graham will return to discuss an example applying low-discrepancy sequences to an insurance company problem.

- Finally, Dr. Papageorgiou will fill us in on the most recent and spectacular developments in speeding up the processing of these complicated problems.

All my life I have heard people saying that making things go faster does not create anything new. I disagree. When you get improvement in the speed of calculation of several orders of magnitude, all of a sudden you find you are able to do things you never thought were possible.

What we are doing with computers and personal computers (PCs) is not just a faster way of doing what we used to do by hand. We are now doing things that we never would have bothered to even think about doing by hand. Due to the increase in precision and speed of convergence available and because of the use of low-discrepancy sequences, we are going to be able to do things that we never thought were possible and that we never dreamed of doing in the past.

Dr. Graham Lord: My role is to give an overview of the fundamentals of low-discrepancy sequences. Development of this topic goes back to another area, namely to Monte Carlo simulation, which, incidentally, has not been around that long either.

Monte Carlo was a code word intended to disguise what was being attempted. The purpose of Monte Carlo simulation was to help physicists work through equations that did not have solutions which they could nicely compute. The work was in connection with developing the parameters for the atomic bomb, the Manhattan project. These methods are the process or the bag of methods used to simulate the process or model, and in that simulation, random variables are used. I am drawing a distinction here between deterministic scenarios and Monte Carlo simulation. Regulation 126, for example, has seven deterministic scenarios. Monte Carlo simulation of some annuity products, for example, processes the annuity product through a model and its behavior is determined by random variables rather than by pre-determined, pre-set interest rates such as those in the New York seven.

The key is random variables. We are trying to mimic a process which would otherwise be very difficult to understand, study its sensitivity to the input parameters, and examine the behavior of a model of some real-life process.

In most actuarial applications we tend to see an examination of the effect of increasing surrender charges or changing other product design features. If we consider the behavior of a bond portfolio, we are making some statement about future interest rates. We are not modeling the actual bond, but are determining the model which determines the interest rates, which in turn determines what the bond value is.

If we knew the closed-form solution in the first place, we wouldn't bother with simulation. This procedure of Monte Carlo simulation is undertaken when we don't have an analytical solution.

As a test, we will consider a case where an analytical solution exists. We will simulate a three-dimensional integral, and, since we know the answer, we can see how good the Monte Carlo method is. The test is not only for Monte Carlo methods, but also for quasi-random variables, that is, low-discrepancy sequences. The problem is the evaluation of integrals, not necessarily of one dimension like we learned in calculus one, but of multiple integrals. Within that framework, think in terms of the price of a collateralized mortgage obligation. Because the price is an expected value, the economic value is an expected value, and the expected value is an integral. Modeling a collateralized mortgage obligation month-by-month means evaluating 360 integrals.

Wall Street uses Monte Carlo methods to evaluate collateralized mortgage obligations, by, in essence, tossing a coin in order to evaluate the high-dimensional integral. Jim Tilley is an actuary who has been instrumental in the valuation of insurance company liabilities using Monte Carlo methods. Much of the work Jim has done at Morgan Stanley is in connection with the economic valuation of insurance liabilities, as they tie into the economic valuation of an asset portfolio, namely asset/liability management or asset/liability analysis. These are some of the techniques and topics we are thinking of when doing Monte Carlo simulation.

Let's return briefly to this application in the evaluation of an integral. A Monte Carlo simulation is equivalent to the toss of a coin, and the outcome of that coin toss will determine how the function we are evaluating is going to be estimated. We do not toss the coin just once, we toss it many times. The coin we toss is not a two-sided coin, but a multifaceted coin. A computer helps with this process, and in the simplest application, the tossing of the coin is telling us the distribution; one toss of the coin would be one point from the uniform distribution.

Some of the mathematical distributions we meet, particularly when modeling interest rates, are not uniform, but are lognormal, Brownian motion, white noise or other far more complicated probability distributions that we can approximate using something other than uniform random variables. Underlying most of the applications, we evaluate our integral via the uniform random number process. Rather than talk about how to approximate a normal by Monte Carlo methods, a Gamma, exponential, or Poisson, each of which has very special techniques for Monte Carlo simulation, I will take as my sole example simulating a uniform random variable. These other distributions, such as exponential and normal, have some desirable properties. The choice of the pseudo-random number generator with certain

desirable properties is crucial to the success of the Monte Carlo simulation method. What do we mean by desirable? If we consider uniform random numbers, there are many tests we could force our random number generator to satisfy. Here is a list of some of them. It is not an exhaustive list, but they give a sense of what we are looking for.

The one particular thing that we would like is a random number generator that does not repeat. A computer generates the random numbers, and the computer will use an algorithm to determine those numbers. In other words, there is a mathematical, deterministic formula used to come up with what we believe to be a random number. We look at the output and say, this is random. The density of the result, if we measure it, will be the uniform distribution.

The input is a deterministic sequence, and for many desirable random number generators, that deterministic sequence cycles. You start off with one value, and after a thousand or two thousand tosses or pulls of the random number, you come back to where you started. Obviously, a generator that repeats after one thousand or two thousand tosses, is just too short. If you were doing a simulation of 100 thousand runs, you would be using the same numbers over and over again. One desirable property is to have the period, or the length of the cycle of your deterministic algorithm that generates your pseudorandom numbers, be very long. Also, since we are talking about uniform random numbers, we would like the resulting sequence of numbers to be uniformly distributed between the limits of your intervals (usually zero to one).

Next, we would like statistical independence between the numbers that we pull. This can be made very precise by saying we want independence between successive ones. However, this is impossible, because we are using a mathematical formula to get the numbers. We should really put the word independence in quotes, or add, statistical almost independence. What would that accomplish? We would have to define it. You can see that some of these tests can be somewhat arbitrary or subjective.

There is another test we will speak about when we look at low-discrepancy sequence. We do not want numbers lining up in a row or regular gaps or jumps that are regular. In other words, there should not be patterns emerging in the numbers. We do not want to see a lattice structure.

When you think of usual random numbers, you are thinking of numbers between zero and one. These patterns do not emerge so clearly. Think of a 50-dimensional vector, say a set of one thousand, 50-dimensional pseudorandom vectors. Consider the 39th dimension. Sometimes you see disturbing patterns in that dimension, or

some other high dimension. One can control for this in a number of ways, and we'll touch on this briefly.

I'm not going to spend much time on nonuniform pseudorandom numbers, but a determinant of the properties of those pseudorandom numbers, say for the normal, is the process by which you generate them. Think of the Box-Mueller method or other methods which produce the normal distribution or bivariate normal distribution more easily than going through uniform random numbers. However, there are problems within. Box-Mueller fails in the tails of the normal distribution. If you are interested in an insurance application and are concerned about the probabilities of insolvency, and somewhere along the line you are using a normal distribution, then you should not be using Box-Mueller because that is where it is apt to break down. One must be careful.

This is meant to be an overview, so I am not giving you any details on how to test for the period length, other than it has to be long. The question is, how long? I have an example that will perhaps impress upon you how long is long.

For the equi-distribution properties, there are a number of very refined statistical tests, the s-dimensional Kolmogorov tests are similar to what you might have learned if you had done nonparametric statistics. There are other tests which are used.

So that we do not lose sight of where we are going, let me give you one example, perhaps the most famous example, of a pseudorandom number generator. That is the circular linear congruential method. It is a very simple one, but this is the one that is in almost every piece of software which is commercially available, whether it is a spreadsheet program like Lotus, Excel, or some of the more sophisticated software statistical packages. Invariably, they have some form of the linear congruential algorithm. You take the pseudorandom number which was just generated, multiply it by a constant A, add another constant C, divide by M, and look at the remainder; that remainder is your next pseudorandom number. This is looking at remainders after dividing by this number M.

Those who have done number theory will realize that this cycle length is going to be less than M or it is going to be at most M. If you divide a number by ten, you can get only ten remainders. What we take is very large. In fact one that is commonly taken, though it is not the only one and is not necessarily the best one, is one where the first constant is 397,204,094. The B is equal to zero and the M is 2^{31} minus one. This is a large prime number, and the question is, what is the length of this cycle? The cycle length of this is something bigger than 2.8 times 10^{13}. That is large. Suppose when using this linear congruential method, this particular

generator, we wanted to pick a thousand random numbers every second. The question is, how long would we be picking until we came back to where we started. That will be a measure of our cycle length. The time required to come back to the start, that is, to cycle, is about 800 years.

What is of interest to us is not picking a linear sequence of pseudorandom numbers, but choosing the vectors, a linear sequence of vectors, of pseudorandom numbers. Think of the collateralized mortgage obligation example. If we do a simulation of 10,000 runs, every run must have something like 360 components, the random vector must have 360 components.

Let me do a pick of one thousand pseudorandom numbers and then pick two-dimensional pseudorandom numbers, and see what they look like. More than words, Chart 1 tells why we are looking at low-discrepancy sequences rather than continuing to look at Monte Carlo methods exclusively. We have two coordinates and we have a thousand or just over a thousand pairs of random numbers.

What I see is bad. There is bunching up or points where the crosses are very close to each other. At the same time, there are areas where there are big gaps. Look at the center. Where is the equidistribution property we wanted? There may be an equal distribution in one direction, and there may be an equal distribution in the other direction, but when the two are put together, we start to get the undesirable properties. We would like a way of better filling the unit square with points so that we have better representatives when we are using the numbers, whatever the application. Keep this picture in mind because we will compare it to a picture using low-discrepancy sequences.

The researchers in this area have realized that many of these early pseudorandom number generators are flawed because of the patterns one can see, the gaps etc. The search for better pseudorandom numbers is underway, and in the future there will be even better ones. If you do go to a particular piece of statistical software, you are not going to see just one random number generator, but a whole slew of them. Each one will be using a particular method. Some others are: multiple recursive congruential, shift register (GFSR), nonlinear congruential, recursive inversive, explicit inversive, and digital inversive.

I am not going to talk about them, but one tends to think the only way that random numbers are generated is by the linear congruential method or some variant of it; but, in fact, that is one of many methods.

Some of these methods are so new that they have only been around for the last six years. Be aware that Monte Carlo methods are not dead; it is just that we may have something that is superior for many of our applications.

From the point of view of applications to finance, including actuarial applications, one of the biggest drawbacks, even with the more refined methods, is the time it takes to do them. If you consider the worst error you would get, the worst length of time, or the precision of the results, then we can show that the measure of the maximum error in a Monte Carlo run tends to be one over the square root of the number of trials. If you are doing 10,000 trials, the error is bounded by 1/100. Also $1/\sqrt{M}$ (one over the square root of M) is the bound on the error.

There have been a variety of methods which have been developed to get around this problem of how many numbers you have to pick to get a desired level of accuracy. Some of the classic methods are: antithetic variables, stratified sampling, control variate, and importance sampling.

The one that is the easiest to understand is the antithetic variables. If you pick a pseudorandom number, and the number is say a one-third, then you also use the complement of that random number, namely one minus one-third, which is two-thirds. Instead of picking 1,000 random numbers, you only choose 500 and take the complement of that 500 to get a full set of 1,000.

You can become more sophisticated about it and combine it with some other methods in order to help reduce your variance. Stratified sampling is a way of reducing variance by looking at the interval over which you are doing your simulation, chopping it up into little intervals, and doing the simulation over each interval. If you do it right, the variance over each interval added up will be less than the variance if you did not constrain it by this stratification.

Control variate uses another variable which is already known, and combines it in a linear way.

$$Y=X+c(Z-\mu)$$

The classic way is: I want to estimate the variable X. In fact, I want the expected value of X to be the estimate, and I know a random variable Z which I can estimate easily, and its mean is μ.

I take a simple linear combination of the two, let's call it Y, and then I simulate Y. What is the expected value of Y? It is the expected value of X. Depending on how X and Z are correlated, the variance of Y can be less than the variance of X. By using this control variance Z, I replace my problem of estimating the expected value

of *X* by estimating the expected value of *Y*, and I have a smaller variance to deal with. We can talk about the best choice for *c* which enables us to reduce the variance by the most.

Importance sampling, the last one I am going to mention, is a way in which you give additional weight to parts of the function you are estimating, and give less weight to the sections of the function where it has less of an impact on your overall estimate.

These methods have been used to either reduce the number of overall runs needed, such as antithetic variables, or reduce the variance in some other way. I can reduce my variance even further than these methods do. Think of the original work of calculating an integral.

Suppose we have a curve and put it in a box. Then we just fire two-dimensional random points at it, count the number of crosses underneath the curve, and divide by the total number that fall within the box. This is the so-called hit or miss method. You either hit by getting underneath the curve or you miss by getting outside it.

Next, do this method in a sneaky way by forming a grid. Then choose the pseudo-random points, say, at the points on this grid, in other words, the points of intersection. Randomness enters by how the points are ordered. With this grid process the variance is reduced from $1/\sqrt{n}$ to being no worse than 1/n, where n is the number of points. Depending upon the nature of the function, using this grid approach, I might reduce my error dramatically. I have to know how fine to make my grid, and hence, how many points to do. Although this looks nice in theory, in practice we do not know how fine to make our grid.

That leads to the question of a way other than using pseudorandom numbers to keep the grid, and choose points that might not necessarily be at the corners of each square, but somewhere inside each square. That way we may be able to preserve the bound on the error to be 1/n and better than the $1/\sqrt{n}$ of the Monte Carlo method or pseudorandom number method. Is that possible? The answer is yes, and that is what discrepancy points are.

From the Floor: When you draw the grid, do you count how many across?

Dr. Lord: Yes, you have to.

From the Floor: It works with random?

Dr. Lord: I mentioned what the randomness is. You have to write down the sequence of points that you picked, and it is the order that you write them down that is random. It is the logical extension of the hit-and-miss method that is used to evaluate very complex functions.

The problem with the grid size is that in order to have any accuracy, you are going to get a prohibitive number of points. The question is, can we still use this grid approach and do better?

Quasi-Monte Carlo methods are deterministic, but the points are no longer random. We define a quasirandom method or a quasi-Monte Carlo method as a simulation based upon what we are going to call quasirandom sequences. These are deterministic. There is going to be a formula to calculate them, and it is going to have the very nice property that we had in our hit-and-miss example; the points are going to fill up our space. It could be just the unit integral, or two dimensions (as we had in our pseudo-random number example), or multidimensional. Later on we are going to have a number of different examples.

The points we are going to talk about will have properties such that they fill in that grid in a very uniform way. We will give you an introduction to the definition of what we mean by uniform way. It does cover the unit cube or the hyper cube, but it does so in an extremely parsimonious way. No point is too close to another point, which is what I mean by, "they avoid each other," so that a point is playing the role of many points around it. You can think of little spheres working in spherical coordinates.

From the Floor: Is it really a question of the size of the grid, or do they not really follow the grid?

Dr. Lord: We disguise the grid in the algorithm that is used to construct them. The quasi-Monte Carlo points which we choose, and you will see my example, are points which are inside each cell. The measure we are going to use of how uniform these are is called discrepancy.

I will give you an introduction to the definition of discrepancy in a moment. Chart 2 shows only two-dimensional points and is one example of a sequence of low-discrepancy points. It is created by an algorithm named after Faure, the French mathematician who developed it.

When comparing Chart 2 to Chart 1 which showed 1,024 pseudorandom numbers, we see there is a far better distribution of the points within the square. I will explain in more detail what base three means when I actually give you the Faure points.

If 512 points do so well, you may ask what the corresponding 1,024 quasi-Monte Carlo points look like? How do these fill up? Chart 3 is the picture for them.

You can see how the quasi-Monte Carlo points in Chart 3 avoid each other compared to the pseudo-Monte Carlo points in Chart 1. If we are going to simulate interest rates as we do later on in our applications, then I want to make sure that when I do toss my quasi-Monte Carlo or my interest rate generator, I can be better guaranteed that I will have a better representation of interest rates.

From the Floor: If you were to complete a whole grid, you might have lower discrepancy, but the problem is that at any given time, when you are working halfway through a grid, then you are much worse off. Wouldn't that be true? If you complete a whole grid, you are going to have lower discrepancy at that particular number of points. It looks that way when you look at those charts. Are you saying that is wrong?

Dr. Vanderhoof: That's wrong. I cannot give you the proof as to why it is wrong, but I have seen the formulas. What you say is correct. For two dimensions, the grid is better. Once you go over three dimensions, then the grid falls apart. I have seen the formulas for it, but I cannot give the proof of the formulas.

Dr. Anargyros Papageorgiou: The discrepancy is a function of the number of the points, so you cannot compare two point sets that are different in size and talk about the discrepancy. If you consider the grid, even in its most trivial form, let's say a three-dimensional grid, then you have at 2^3 or eight points, one on each vertex. If you take a 360-dimensional grid, you have 2^{360} points, again with one on each vertex. This is what leads to the combinatorial explosion which does not allow you to solve these problems. You want to come up with sequences that, for a fixed number of points have as little as possible deviation from normality. If I keep on filling the grid, yes, that diminishes the discrepancy. But you are paying more because you are taking more and more points. Fix the cost. Find a point set that has a fixed number of points, and among all point sets, choose the one that has the lowest discrepancy.

From the Floor: I think what you just said was slightly different from what Irwin said. You are saying, "Yes, you could fill in the whole grid in ten or 15 dimensions." For that huge number of points you might actually do better, but there is no way you are going to do it.

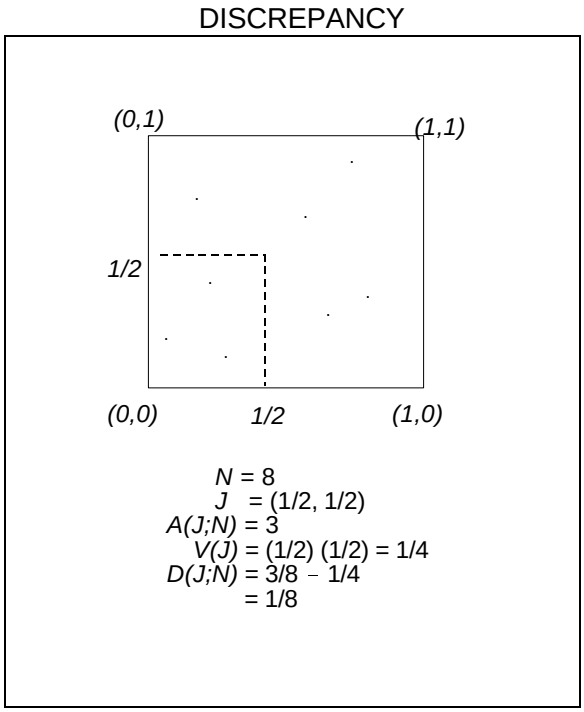
Dr. Papageorgiou: No, that is wrong. If I keep on filling, it is as if I keep on taking more points. It does not have anything to do with the grid or any other way of selecting the points. It is misleading for one to think that I can reduce the

discrepancy by increasing the number of points. You have to keep the number of points fixed and then look at the placement of those points and decide what is the deviation.

From the Floor: If you compare two sets of points which have say 100 ^{360} of points, one of which is done this way and the other is done on the points of the grid (the same number of points) which one would do better?

Dr. Papageorgiou: They are proportionately the same.

Dr. Lord: This is the measure of discrepancy, which looks more forbidding than it really is:



Here are all eight points, all together in the unit square. Define the subinterval, which is J, and count up the number of points that fall within J. In this example, there are only three. What is the portion of those points relative to the total number of points, and how does that compare with the actual area of the square? In other words, how good is it? It is like hit and miss. The area of the square, J, using the point system measure, is 3/8, compared to what the area should have been, which is a quarter. The difference between these two, the one done by counting points, 3/8, minus the true area, 1/4, is the discrepancy for that particular J.

The definition makes it a bit more formal but just think of the picture. Instead of just two dimensions, consider as many dimensions as you want.

From the Floor: Explain what the advantage of using Faure points is over using the straight Monte Carlo method.

Dr. Lord: There is a dramatic reduction in the error. There is a speed-up with which you attain your results, and in some cases that speed-up is phenomenally fast. Irwin mentioned a 100-fold increase.

To get the accuracy of an analysis using 100,000 Monte Carlo runs, you need to use only 1,000 low-discrepancy points or the quasi-Monte Carlo method. In fact, let's briefly touch on what I am going to be talking about later—the study of a single premium deferred annuity (SPDA) block of business. It took 16 hours to do the pseudo-Monte Carlo run on a computer, and it took 2 hours to do it using these quasi-Monte Carlo methods. Most of the work was not the low-discrepancy points, it was the actual computer model of the SPDA and assets that took so much time. That is a dramatic savings.

From the Floor: Every time you take a different interval you get a different discrepancy number?

Dr. Lord: Every time you take a different interval J, you get a different $D(J;N)$. What you want to look at is the worst example or the worst measure and that leads to what the discrepancy is. It is given the name D*, and it is the maximum of all the Js. There is the final definition of discrepancy. Take the maximum or the supremum over all those little discrepancies.

From the Floor: This works for Js of all sizes?

Dr. Lord: This is for Js of all sizes within the unit interval. The reason it is starred is because all those Js are anchored at the origin. They all have one vertex at the origin. There are other measures of discrepancy which are more general, but this is the one that is perhaps the easiest to use. It also leads to some interesting properties. For a uniformly distributed infinite sequence then the D* is equal to zero. This is what you were asking about? Does it actually fill up everything? The answer is, it fills it up very fast.

From the Floor: Why are they all started at the origin?

Dr. Lord: It is just mathematically convenient to do that. I could have had Js anywhere in the interval, all over the place. It is just quicker to do it this way.

Mr. Thomas N. Herzog: Do you lose generality?

Dr. Lord: You do not lose too much generality by anchoring them.

From the Floor: What if there's a problem, for whatever reason, in the upper box?

Dr. Lord: Remember this is just one *J*. One of the *J*s would be from 0–15/16. That would capture the behavior in the corner. You can see the *J* is an increasing family of little squares. That way everything is covered. You do not lose too much generality by considering only these, as compared to considering all possible little squares all over the place.

Mr. Thomas J. Mitchell: Isn't that supremum hard to calculate for general sequences?

Dr. Lord: I do not think so. You are taking areas of squares or hyper cubes.

Mr. Mitchell: You take the maximum, and then you would have to look at a large number?

Dr. Lord: Yes. I am not saying you can do it quickly.

Mr. Mitchell: By hard, I meant slow.

Dr. Lord: Yes, slow. In fact, it is so hard in that sense that we only know of special cases. The example I am going to share with you is the one-dimensional case. Take the unit interval from zero to one. Answer the question, "What is the discrepancy of the points?"

Take any bunch of points, $x(1)...x(N)$, and the discrepancy will be equal to this formula.

D^*_N Discrepancy

$$D^*_N = sup_J |D(J;N)|$$

Therefore,

$$\lim_{N\to\infty} D^*_N = 0,$$

a uniformly distributed infinite sequence.

In the one-dimensional case the D^*_N discrepancy of the sequence

$$0 \le x_1 < x_2 < \ldots < x_N \le 1$$

$$D^*_N = \frac{1}{2N} + \max_{i=1,2,\ldots,N} \left| x_i - \frac{2i-1}{2N} \right|$$

Thus to obtain the lowest discrepancy sequence, we should pick

$$x_i = \frac{2i-1}{2N}$$

This special case reduces to the midpoint rule!

Then in the true spirit of mathematicians, we ask, "What is the smallest value this thing can have?"

Consider N points $x_1, x_2, \ldots, x_N$ in the s-dimensional unit cube $I^3 = (0,1)^3$, $s \ge 1$, and a subinterval $J \in I^3$, the *local discrepancy, D(J;N)*, is defined by

$$D(J;N) = \frac{A(J;N)}{N} - V(J)$$

Where *A(J;N)* is the number of n, $1 \le n \le N$ with $x_n \in J$ and $V(J)$ is the volume of *J.*

Because this was an arbitrary sequence of N points, take the smallest or minimum of this value. You end up with all the points in the odd parts of the interval.

If there were ten points, they would be at 1/20, 3/20, 5/20, 7/20, 9/20, and 11/20. That is the mid-point for a mid-point numerical integration formula for the area underneath the curve.

This example can be misleading. If the solution to the problem is equally-spaced points between zero and one. This would imply if you are looking at a square, a two-dimensional problem, that you should be using equally-spaced points in both dimensions and putting them together. That is not the lowest discrepancy sequence. Some of the other examples which we explain do fall into the lowest discrepancy sequence.

The one I am going to show you in some detail goes by the name of the van der Corput sequence. Take a prime number, say the number three. If I take a number like 11, I can write 11 in base three. Let's use the example shown below.

CONSTRUCTION

P is a prime number

Any non-negative integer n can be expressed

$$n = \sum_{j=0} c_j \, p^j$$

(e.g., if p=3, n=7 = $1*3^0 + 2*3^1$)

Define the radical-inverse function *f* in base *p* by

$$f(n) = \sum_{j=0} c_j \, p^{-j-1}$$

(e.g., if $p = 3$, $n = 7$, $f(7) = 1/3 + 2/9$)

Note for $n > 0$, $0 < f(n) < 1$

The van der Corput sequence in base p is then:
f(0), f(1), f(2), ... , f(n), ...

The van der Corput sequence is "uniformly scattered" or "self-avoiding", and is "uniformly distributed" in the sense that its discrepancy tends toward 0 as the number of points in the sequence gets larger.

In fact, the discrepancy of the sequence is (k*log n)/n
(k is a function of the base *p*)

The best value of k is 1/(2 log 3) and occurs when $p = 3$

The constant can be improved by permuting the coefficients c_j in the representation—the resulting sequence is called the generalized van der Corput sequence.

I can write seven in base three because it is two times 3^1, plus one times 3^0. If you are going to do a base three representation of the number 7, it is going to be 21. The two and the one are the numbers that appear in the sum. They are the co-efficient in the basis expression in base three. We can write any number, a number in the millions or a number as small as seven, in base three.

Now, define the radical inverse function which takes those same coefficients, the two and the one, and now puts the base in the denominator. It says, you had two

and one next to each other, and you did a reflection after the decimal point. The digit that is in the units place becomes the digit immediately following the three-base point. The digit that is in the second place to the left of the decimal point now becomes the digit in the second place to the right of the decimal point, and so it goes on. Why are we doing this? Because we end up with a number, *f*(7), which is between zero and one. If I keep doing this, starting at zero and going on to *n*, then I will get a sequence of numbers between zero and one, and these will be my quasi-Monte Carlo points.

It is a very simple construction. You can do it even in a spreadsheet program and generate quasi-Monte Carlo points, or one-dimensional van der Corput sequences. They are uniformly distributed in the sense of our discrepancy. If I let the *n* go to infinity, the limit of the discrepancy goes to zero. I gave a slightly different definition as equivalent.

One can show that the discrepancy of van der Corput sequence is (k log n)/n. Discrepancy is the measure that is somewhat similar to the variance, in that it gives an estimate of what the error is in some applications. It is what you are missing by. It is approximately 1/n, which is much better than $1/\sqrt{n}$. The k is a constant, and it depends upon the base. This proof is for any arbitrary prime. Where do you get the best discrepancy? It is when $p = 3$, and $k = 1/(2 \log 3)$.

We can play fun games like this. This one blows Irwin's mind in that we are talking about derandomization and getting away from random points. I can improve discrepancy by commuting the digits in some random way. I leave you with that thought because I want to talk about higher dimensional quasi-Monte Carlo points.

This was an example of a quasi-Monte Carlo sequence, which has a low discrepancy, $p = 3$, and the sequence is named after its inventor, van der Corput.

From the Floor: So you have given us a different Monte Carlo method.

Dr. Lord: Yes. I gave you a way of generating numbers between zero and one.

From the Floor: If we use that, we will get a better discrepancy than if we use linear congruential modeling.

Dr. Lord: Yes. For a fixed number of points.

From the Floor: Those points are *f*?

Dr. Lord: They are $f(n)$. If I decide I want 1,000 points, then I am going to go from $f(0)$ to $f(999)$, or I could go from $f(13)$ to f(1,012).

From the Floor: Are the *n*s in sequence?

Dr. Lord: In the original way it is defined, yes. The reason is so it fills out the unit interval.

From the Floor: I could have done 1,000 points of the linear congruential method.

Dr. Lord: Yes. That is the pseudorandom number.

From the Floor: I can do it this way following the formula, and I will get 1,000 numbers, $f(0)$ to $f(n)$, suggesting that if I use $p = 3$, I get the best numbers. With those 1,000 numbers, my simulation will give me a better result.

Dr. Lord: Right.

From the Floor: It will be more evenly distributed.

Dr. Lord: What you would have to do is take your application of 1,000 Monte Carlo random numbers and repeat it say 100 times, and look at the error over those hundred. Then compare that to the corresponding thing if you did 100 replications of 1,000 using these sequences. You will find that the error in the latter case is less.

From the Floor: Why do you call this quasi-Monte Carlo?

Dr. Lord: It looks like it is random, but in fact it is deterministic. The people who invented the word called them quasi, because they look as though they are traditional Monte Carlo, but they are not.

From the Floor: You have just given us a better formula than random numbers?

Dr. Lord: In essence, yes.

From the Floor: The limitation on this is that it is one dimensional?

Dr. Lord: On this one, yes.

Mr. Herzog: Those cases are really deterministic.

Dr. Lord: It is correct that they are formulas. You can think of this as a different class of formulas, though we're looking at a slightly different measure of its effectiveness.

Mr. Mitchell: When you say the error, are you talking about the error in pricing something using these numbers?

Dr. Lord: Yes, it could be. Let's talk about the introduction to the real applications. We were not doing one dimension, because that is a bit simple. We were doing many dimensions. This algorithm was developed by Faure and is what was behind Chart 2.

Higher Dimensional Sequences

One technique -- the Faure sequence:

$$n = \sum_{j=0} c_j\,(n) p^j$$

Generate successive coefficients ${}^{i}c_j(n)$ recursively
(where ${}^{1}c_j(n) = c_j(n)$)

$${}^{i+1}c_j\,(n) = \sum_{j=0} \binom{i}{j}\ {}^{i}c_j\,(n) \pmod{p}$$

Now define the vector sequence, the Faure sequence:

$$f^k\,(n) = \sum_{j=0} {}^{k}c_j\ p^{-j-1}$$

Of 1,024 two-dimensional Faure points, base 3 could be used in comparison to two-dimensional pseudo random numbers.

Note discrepancy can be improved by permuting the coefficients as in the one-dimensional case.

We start with the same base three representation. That would generate coefficients. I have made the coefficients a function of *n*. Then add up these coefficients after multiplying them by a binomial coefficient. That *c(i,j)* is our old friend.

From the Floor: What is the summation over?

Dr. Lord: It's over *i*. That is the only thing that is moving.

From the Floor: What does *i* equal?

Dr. Lord: Wherever the binomial coefficient is not zero.

From the Floor: Zero to *j*?

Dr. Lord: The *i* has to be bigger than *j*, otherwise it is zero. It is going to stop when you get to *p*.

From the Floor: It's from *i* equal *j to p.*

Dr. Lord: Yes. I've iterated once. Then I use the result and the same formula for both. I no longer have *c* but ^{2}c , and that will give me ^{3}c. Then I use ^{3}c in this formula in place of the c and that will give me ^{4}c. By repeating this formula, that will generate *k*, a sequence, *c*, ^{2}c, ^{3}c, ^{4}c, etc. I keep getting more and more numbers. Each one of these is the next element in my vector. If I want a three-dimensional vector, then I am going to generate c^{2}c and then ^{3}c, and that will be the three components of my vector, and that will be the first Faure point. To get the second Faure point, take *n* equals another integer, and go through the same process again.

From the Floor: In this process, are the measures meant to have literally one, two, three, or a random?

Dr. Lord: Yes, one, two, and three. Anargyros will probably talk about what is the best choice for picking that consecutive sequence. You can skip over say the first thousand and then start *N* equals 1,001, for example. Then we do exactly what we did in the van der Corput sequence, which was a reflection about the decimal point, and create those numbers that are between zero and one by taking those coefficients and dividing by appropriate powers of three. What we end up with is a sequence of vectors of three elements, and that is our Faure sequence.

From the Floor: Is that *j* equals zero to p?

Dr. Lord: Yes, *j* starts at zero. The first thing is going to be one-third, or 1/*p*. It goes to the coefficients that are zero. After a while the coefficient becomes zero.

It is this algorithm that I use to generate Chart 2 and the other one that was like it in Chart 3. The *c*s are on the x-axis, and the ^{2}cs are in the y-axis, or the vertical axis. Those crosses were obtained by just doing one iteration of this thing and correlating a point, a point which has the component *c(n)* and *2c*(*n*). If I want to do a 360-dimensional Faure sequence, then I am going to choose a prime, in fact you choose a prime immediately larger than the dimension, and then do this process iteratively 359 times to get every component in the Faure vector.

The Faure sequence is only one of many such low discrepancy or quasi-Monte Carlo algorithms. Some of the early ones were mentioned because of their historic interest rather than their practicality. The equally-spaced one on the unit interval is a Hammersley sequence. LaCot is another one. The Russian mathematician Sobol extended Faure to come up with a comparable sequence. Neiderreiter did work which developed a whole theory of what Sobol was doing and came up with a very comprehensive class of quasi-Monte Carlo sequences and low-discrepancy sequences. The Japanese mathematician Tezuka came up with an extension of what Faure did, which in some sense could be considered a special case of Neiderriter, but we call it the generalized Faure sequence. The examples we will see later all use this latter algorithm. Perhaps these simple examples will show you the advantage that we have observed in using low-discrepancy points.

This first example is maybe unpleasantly mathematical, so let's imagine you have a doughnut in three space and a function that is defined on the inside of the donut. I want to evaluate that function, in other words, take the integral. Even though it looks formidable, you can get an answer. It is $1.2\Pi^2 a^2$.

The question is, can we estimate the correct answer by using pseudorandom numbers? How does that compare if we use Sobol numbers?

Pseudo versus Sobol[1]

Example 1

Integrate $f(x,y,z)=1+\cos \pi \frac{r^{2^*}}{a^2}$ where $r < a$,

inside the doughnut in 3-D; B is the major radius of the torus, and a the minor radius

Answer : $2\pi^2 \; a2B$

Example 2

Integrate $f(x,y,z) = 1$ when $r < a$, inside the same doughnut as in Example 1.

Answer : $2\pi^2 \; a2B$

Chart 4 shows the results of repeated trials of 100 using
(a) pseudorandom numbers
(b) Sobol numbers

Note the 100 fold speed up with the Sobol sequence.

[1] From *Numerical Recipes in C* by Press, et al.

If we use pseudorandom numbers, the variance is going to be 1/n. We do repeated trials of 100. Choose 100 pseudorandom numbers and use them to estimate this integral and put down the number. Then do a second one, keep doing 100, and then look at the error in those 100 trials. Chart 4 is from a book which has become almost a bible in numerical methods, *Numerical Recipes in C*, by W. Press, et. al. The other line on the graph are Sobol numbers. The other generator, the quasi Monte Carlo generator that is used here is the Sobol numbers, and we can show that the discrepancy for those is $(\log n^3)/n$.

What you should be looking at in Chart 4 is the upper dotted line and the thinner solid line. The upper dotted line is the pseudorandom number result, doing the graph against the number of points in my test. The solid thin line is the result when I use the Sobol points. The scale is logarithmic so that the curves look as though they are nicely behaved. You see the error is far smaller for the Sobol points than it is for the pseudorandom number points. It is true even if we only take 100 points. The difference between the dotted line and the solid line is still there. As you go further down and increase the number of points, that difference becomes even greater. Note the pseudorandom numbers are asymptmatic to that line, which is what we predict from the theory; the error behaves like $1/\sqrt{M}$ (in the log scale).

This line for Sobol points is 1/n, the theoretical error we claim for the low-discrepancy points. This line lies below the Sobol points because the Sobol point error is not 1/n, but $(\log n^3)/n$. That is why the curved line and the solid line do not come together.

The significance of this Chart 4 is that if, in estimating my integral, I only want an error of say 0.1%, then I will be able to use 100 fewer points generated by the Sobol method than if I use the pseudorandom number. In other words, the speed up in my estimation is 100 times faster. That is quite significant.

You see on Chart 4 that there are two other lines, the heavy dotted line and the heavy solid line. That is a second function and speaks to some of the weaknesses of the quasi-Monte Carlo method. If your function is not smooth, then the quasi-Monte Carlo methods do not give as good results as we have just talked about. Even though they do not do as well, they still do better than the pseudorandom numbers or the dotted lines. This function is the simple cliff function, that is, one in some places and zero elsewhere.

The last example was done by Phelim Boyle and some of his students. This one may be closer to our hearts than those doughnut examples. That is when we have an option. It is a European option to make it simpler, and here are some of the statistics. The current value is 100, and the exercise or strike price is 100. Looking

over a year, volatility is 30%, and the riskless discount rate is 10%. Since it is a European option, we know the answer from Black-Scholes. We plug it in and we come up with the number of $16.73. If pricing a put, we get $7.22.

The question is, what happens if we try to estimate the value of these two options, the call and the put using low-discrepancy sequence, using Faure points? We are going to get a graph for the call and a graph for the put. (See Chart 5.)

The upper graph is the call, and the little diamonds are the results, the error of doing repeated trials of the pseudorandom numbers, the crude Monte Carlo method. You can see that with a few runs, they are quite scattered around. After a while, it settles down. However, even when you get close to 10,000 simulations, the crude Monte Carlo method, the pseudorandom numbers, is still not giving reliable estimates of the value. Compare that with the value using the Faure points, the quasi-Monte Carlo one, or the solid line. Even though, at the beginning, the error is high, it drops down quickly and becomes very stable. Quite a telling example of the power and the improvement in efficiency and speed with the quasi-Monte Carlo points. It is even more dramatic in the case of the put.

How come it seems to work better for the put than for the call? The put was in the money. Current value is 100, and the exercise price is 100. From the point of view of the purchaser, the value of the put is bounded. The intrinsic value of the put will never exceed the strike price of 100. It is going to be between zero and 100. The call can go up to infinity if the price of the security goes very high.

From the Floor: It doesn't seem to improve. This one comes very near to zero and the one on top seems to come to almost 6,000, and 100,000 will still not get to zero?

Dr. Lord: It gets much closer. We created a binomial model of interest rates, and when you discretize, you are putting an additional wrench in the results. Some of that lack of convergence could be because we use a somewhat crude model to value the options. Maybe using a stochastic differential value of the security would produce a better result.

From the Floor: Is there any software available?

Dr. Lord: Yes, there is.

Dr. Vanderhoof: A researcher in Japan solved the same CMO problem. That is what IBM is saying they have done. Actually, they took the idea and the problem

from Spassimir Paskov. It is now being actively worked on around the world by many different people.

What is the intuition? This is crucial. Graham has talked about what a low-discrepancy sequence means and what low-discrepancy points are.

Figure you have a box. The low-discrepancy problem is that I have 100 points. How do I fit those 100 points in the box, so that any volume in the box that I picked has a number of points in it that is proportional to the volume? Think of the box as being a unit cube, because then the volume is always between zero and one. How do you arrange those points?

That problem was a classic problem in measure theory, and it was solved by mathematician Roth. Roth came up with low discrepancy. Discrepancy is the difference between the percentage of points in a particular volume and the volume itself, the volume in the unit cube. One of the solutions was Hammersley points, and Hammersley was a pioneer in Monte Carlo methods. Hammersley speculated that the van der Corput sequence would work better than traditional Monte Carlo methods. That was in 1957 I believe. Nothing further was done on it until Wozniakowski and Traub showed that this would also be useable for integration.

Let's go back to that unit cube. Consider that each dimension is a cumulative distribution function, that is, it is a probability. The unit cube represents the probability that everything would happen, and it is one. Each volume in that unit cube represents a probability of occurrence corresponding to the three different distribution functions for that volume. If we say certain points have a low discrepancy, then we are saying that each of those points must have about the same probability volume associated with it

I have not mentioned interest rates, prices on stocks, or anything like that. It does not matter. Once you have cumulative distribution functions, you can go from the cumulative distribution function, say of 0.4, back to whatever the function was, and get a real value. The important fact is that each of those points seems to have about the same probability volume associated with it, and that is why the whole thing works. It works in higher dimensional arrays also.

If each of these points has essentially an equal probability, because an equal probability space volume is affiliated with it. It is in that neighborhood, then the worst result of those points has the worst possible value in that number of points.

If we do 200 calculations, there is less than a 1% chance that the worst of those possible results will be worse than the worst of the 200. If I do 200 calculations of

the price of this stock, then no matter how many more I did, there is less than a 1% chance the price would be worse than the worst of those 200. That is a very strong statement. I have never heard anybody say that with a finite and reasonably small number of Monte Carlo calculations you could make any probability statements at all. This is important because value at risk is becoming a key item in the statements of financial companies. You need to be able to say something about the probability of a bad result, and nobody has set up any paradigm or demonstration that you can do it with Monte Carlo calculations.

The September/October 1996 issue of *Contingencies* has an article on "Using Low-Discrepancy Points to Value Complex Financial Instruments." The bibliography was published by New York University with the paper "Strategic Function of Life Insurance." If you would like a copy of the bibliography, or the *Contingencies* article, please contact Faye Albert at her Directory address. There was also an article, "Breaking and Tractability," in *Scientific American* on this subject in January 1994 by Wozniakowski and Traub.

I will share some results from Spassimir Paskov's dissertation. One was shown in the *Contingencies* September/October 1996 article. The question is posed, if we do a valuation of a tranche of a CMO, what kind of results do we get? (See Chart 6.)

Using traditional Monte Carlo methodology with random numbers, results differ depending on the seed. Graham discussed this with regard to the linear congruential method for generating random numbers. If you start with a different seed, you end up with a different answer. How much different? It depends. But you will end up with different answers. This does not happen with either the Halton or the Sobol sequences. These techniques give an answer which is more dependable and probably more correct.

Chart 7 shows a change in the generator. Using Ran 2 you get better convergence, closer to the Sobol sequences. Even with Ran 2 or Ran 1, results depend on the seed. The random number generators are not dependable.

The antithetic variable question was raised. In Chart 8, 20 runs were done using the antithetic variable technique. For an antithetic variable approach, use pseudo-random numbers, the traditional methodology, but with 20 different seeds. Then the average of all 20 runs, i.e., runs using different seeds. The same calculation is based upon 100,000 points for each of the Sobol and Halton sequences. Consider the number of calculations, 100,000. This is 100,000 using Sobol or Halton. In fact, it is 2,000,000 calculations, 20 times as many for the antithetic variable, since there were 20 different runs of 100,000 each. You can see that the Sobol line

shows a very slight difference on a large CMO, and that is with a very small number of runs.

Mr. Leonard H. Wissner: To many plan sponsors, the asset allocation decisions of a defined-benefit pension plan is a "no-brainer." With the stock market roaring the way it is, and with historical studies done by Ibbotson saying that stocks out-performed bonds for the last two centuries, what is the sense in going through the trouble? Why not put the whole pension plan in the stock market and then just ride it out? I have a problem with that for a couple of reasons. One reason is obvious from the introduction Irwin gave about me—I would be out of a job. The other reason is that in my experience over the last 20–25 years, I have found there is no easy way to make money in the financial markets.

If things were that easy, why couldn't people just buy the Standard and Poor's (S&P), short the bond, go home, and become rich? If it was that easy, wouldn't everyone be doing it, and consequently pricing the asset to such a rich value that the opportunity would be removed from the market? Markets generally price assets to a certain point; but then there comes a point where the market becomes over-valued. My job as an investment manager is not only to look at price, but also at value. It is the synthesis between price and value which determines investment opportunity.

To examine the allocation problem, I decided to build a simple Monte Carlo simulation to assess the stock/bond decision over a long time horizon, say 30 years. The only place I could build the model was on a spreadsheet program, and the only random numbers I knew about were the random numbers that spreadsheet program gave me. I picked 1,000 because that sounded like a round number. Then I saw a session on low-discrepancy points. I wasn't really sure what 1,000 random num-bers meant. But if 200 or 1,000 low-discrepancy points would give me more confidence in the results and be more robust, I was willing to try. Graham helped me with the simulation trials.

Before founding Ward & Wissner Capital Management, Inc. in 1981, and prior to joining the Equitable where I met Irwin, I was in the brokerage industry. The change was a big culture shock because, in the brokerage industry, the time horizon is ten seconds. What is the price of the stock market or what is the price of a long bond? In an insurance company, time horizons are considered for 20 and 30 years. Consider a long time horizon, say a simple 30-year 8% bond. What proportion of the total 8% return is coming just from the coupon stream? For a 30-year time horizon, almost 87% of the bond return is from the coupon stream, and the price of the return of principal at the end comprises only 13% of the return. Although

papers report every day the fluctuation of the prices of bonds, it is really the coupon stream that will ultimately determine the return on a bond.

A very simple valuation model used for stock is called the dividend discount model. It says that the real long-term return on stock is simple to calculate. The current dividend yield is unfortunately, at one of the lowest points in the century. There is only about a 2% dividend yield on the S&P. Add dividend growth, which has historically been only about 5% over 30-year time periods. Add the 2% and the 5% to get 7%, and then we have a little correction factor based on regression analysis. Come up with a long-term return of 7%. If I went back and used the Ibbotson data from 1926 to the present and calculated 30-year holding period returns, almost 88% of the return can be determined just by looking at the initial dividend yield and the dividend growth on the stock. The error in your estimate is only about 1%. Although everybody is talking about where the price of the S&P is, what is really going to be driving ultimate return on equity are just two factors: the initial dividend yield originally bought, and the dividend growth throughout the 30-year time period.

Chart 9 will give you an indication of how slow the bond business is right now. The dividend discount model (DDM) estimates a 30-year return. What would happen if we knew the dividend yield and dividend growth 30 years prior, for example in 1956? We do know the initial dividend yield in 1926, and say we knew with perfect hindsight what the dividend growth over the 30-year period 1926–56 was going to be. What would be the equilibrium price of the S&P in 1956 for the model to have a perfect fit? What was the percentage error of the prediction of the dividend discount model. In other words, if you predicted seven and it came out eight, that would be a 1% error.

In Chart 9, we looked at the actual price of the S&P, and compared that to what the DDM prediction would have been with perfect knowledge of the growth of the dividend stream and the initial dividend yield from 1956 to 1996. Our conjecture was that there was some type of a mean reversion in this process. In other words, there were times when the S&P and the dividend discount model were in perfect sync. There were times when the stock market was undervalued compared to what the DDM prediction was. Finally, there were times when the stock market was overvalued.

What happened when we applied this? There were some notable periods, for instance in the 1970s, when the stock market looked tremendously undervalued. Now it looks like it is tremendously overvalued. It looks like a graph of the Tokyo stock market in 1988 or 1989. There's a tremendous divergence from the model. If we carry the trend dividend growth forward for the next five years, and if there is

some type of mean reversion, somewhere in the next five-year period, the stock market could correct to a very big degree. That being said, we have looked at price and value.

What is going on in the real world as far as the asset allocation decision? Deep in everyone's heart, they want to be in stocks. As a consultant or as an advisor, one must come up with a model that will put them there. In this country, in 98% of the situations that I have been involved in, that is the state of the art. How many of you have been invited to an investment allocation meeting? In my experience, the two are distinct. In other words what the asset managers are doing and what the actuaries are doing are two separate processes.

The most popular model used is the capital asset pricing model (CAPM), which was developed by Markowitz and came into vogue in the early 1960s. It is a model that does not look at the liability side of the equation at all, and defines risk as the standard deviation of the annual return of the asset.

As you know, in a pension plan, the purpose is to pay the pension obligation as it falls due in the future, and the risk is not having enough assets to fund the pensions when they come due. Funny things start to happen when you just look at the asset side of the ledger. From the point of view of an investment person, the salary growth and interest rate assumption are critical assumptions on what the eventual liability structure of the plan would be, and these have to relate to the investment environment itself. As a result of the Financial Accounting Standards Board (FASB), liabilities are segmented. The liability that we are going to be keying in on, is the projected benefit obligation (PBO), which is the actuarial accrued liability of the plan. This tells you if the plan is sufficiently funded to date. The scheme that I am going to be employing is like a paid-up immunization scheme. In other words, if I have sufficient assets to fund the PBO, I will let the contributions of the plan fund the future service.

I was fortunate many years ago to have a copy of a book by Howard Winklevoss on pension mathematics and also a paper by Irwin which looked at the sensitivity of the Equitable liability to a fluctuation in interest rates and a fluctuation in inflation rates; in other words, it was a parallel shift of the interest/inflation yield curve structure.

Let's discuss the sensitivity of the pension liability to changes in interest rates, inflation rates, or salary growth. The example I'll use is taken from the Winklevoss book and shows that if you decrease the interest assumption and salary growth assumption by 1%, the present value of the future benefit obligation would increase by about 12%. This is a very critical number to me. Without considering the total

obligation, but just the PBO obligation, the sensitivity is near 12%. An immunization scheme which is going to always keep assets in line with the PBO needs an asset structure or a bond portfolio duration of 12 years.

The paper that Irwin wrote looked at an immunization scheme not only immunizing with respect to the PBO, but also immunizing such that the contribution as a percentage of payroll remains constant. These structures are interesting from the point of view of an asset manager. Let's say the assets in the portfolio comprise three-fifths, or 60% of the total benefit obligation (TBO). If the duration of the TBO is 12, then five-thirds times 12, or 20 years, would be the duration of the asset structure. Some of these structures were unavailable when Irwin wrote your paper. However such asset structures are very possible with cash bond instruments, strip securities, or principal only (PO) mortgages. What's even more interesting, you can create synthetics with the futures markets that will actually dominate the yield of these duration structures in the cash market and produce immunized structures as well. What's exciting here is that it is possible to immunize a pension plan with debt instruments. When one goes into the asset allocation problem as it is practiced, there are probably something like 108 assumptions which go into the analysis. The fact that you can do it with bonds in almost an assumption-free system is truly a remarkable development. This gets into an appreciation of what the bond instrument can do for the risk reduction process and for performing a funding process in an assumption-free manner.

In this particular problem, we are going to work with the PBO obligation. Our risk measure is the probability that the assets in the plan will be less than the PBO liability, given that the plan starts out in a fully funded status, and the plan is invested 100% in the stock market for 30 years. In other words, what is the probability that the funding ratio would be less than 1 after 30 years?

Funny things start happening in the traditional process when one does not look at the pension liability. Given a 1% increase in inflation and interest rates, you need a 12% increase in asset value. Also, if you are carrying an 8% assumption, you will need a 20% return to keep pace with the liability. What people do not realize is that even though stocks are behaving very well, bonds are indexed to a short duration index, such the Lehman Bond Index. At the investment committee meeting everybody is going to be happy with a falling rate environment because the stock and bond assets are performing very well. Few realize that the liability in such a setting is growing at the rate of 20%. You get a false sense of security in a disinflationary environment where the stock market is rallying. Even though in the 1980s there was a disinflationary environment, the stock market was rallying. In the 1930s in a disinflationary environment, the stock market was falling. Such an environment is a disaster for the pension plan, because the liability is growing at an

astronomical rate. Everybody is saying, "Let's load it all up with stocks," the asset side of the structure is not only unable to keep up, but it is actually going down.

The risk of many of the plan structures is precisely a disinflationary environment, where stocks are not performing well and the bond rates are going down. That is the purpose of the bond in the plan. You might have seen an article about a month ago basically saying there is no purpose for bonds in a pension plan. This is the state of where the market has gotten to. Looking at the 1980s was a very eye-opening experience. The later part of the 1980s was an environment of disinflation which followed the high inflation rates of the early 1980s, and basically good stock returns and good bond returns.

A Buck study which was recently done looked at the period 1988–94. Despite very good stock and bond performance, the percentage of companies which reported fully funded plans with respect to the accumulated benefit obligation (ABO) liability actually went down. Moreover, we constructed a pension surplus index which started back in 1981 with a plan that was 160% overfunded. We then looked at what the funding status would be at the end of 1995 and found that based on a 12-year duration liability structure, the pension surplus ratio of the typical plan which is 60% stocks, 30% bonds and 10% cash, would have gone down. Although everybody was celebrating the fact that we have had great asset markets, based on the status of the pension plan, surplus actually eroded during these good market years. You can imagine what it would be like if there were a correction in the equity markets.

Mr. Herzog: Do you think it has to do with the interest rate policy of the federal reserve?

Mr. Wissner: The strip yield went from around 14% down to 6%. One of the stated objectives of the federal reserve is price stability. Let's say that is such a setting they bring the inflation down to 2%. In other words, a disinflationary setting as an objective of the federal reserve policy is a very real possibility. If you look at the mechanism of the capital asset pricing model, one of the inputs to it is the correlation between stocks and bonds, usually a correlation coefficient of about 38%, which precludes the possibility that the bond market could go up and the stock market could go down. Yet, that is precisely what happened during the 1930s. In other words, not only does the model fail to look at the liability, it does not address the principal risk in the pension plan of the two markets decoupling as they did in 1987.

From the Floor: You mentioned surplus as being static at a point.

Mr. Wissner: Contributions would come in at the rate of 60% stock, 30% bonds and 10% cash defining the asset mix. Assets would stay in that and then be compared to a 12-year duration liability.

From the Floor: There are some new products which allow the credit to treasury bill function of the S&P 500 gross rate. Would that be a better instrument than actual stocks or bonds?

Mr. Wissner: I am not familiar with the product. The problem that I would see though is that you need the bond because that is your disinflationary hedge. It is not just good enough that you are keeping up with the S&P 500. In a disinflationary environment, the only thing that is going to save you is the long duration bond. The other problem this points out is the bonds are typically indexed too short. The duration of the Lehman index is typically five years; that is the most popular index in most of these plans. The duration of the liability is at least 12 years. The bonds are positioned about half the length of what they should be.

From the Floor: Can we use any kind of a stock index, like call or put option so somehow we can stabilize these stock market variations?

Mr. Wissner: I do not believe the problem is the stock market variation. I think the problem is that the effective duration on the stock is too short. In other words, given a 1% decline in interest rates, the stock market is not sensitive enough to produce the market appreciation of the long-duration bond. Moreover, if you get into a bad economic setting, one would expect stock prices to go down.

From the Floor: Your focus is primarily on the pension funds, U.S. insurance companies have not typically had a significant investment in equities primarily because of the regulatory environment. Outside the U.S., especially in the Far East, insurance companies often operate with significant equity positions. Your analysis is applicable to insurance companies. My question is, are you finding an ideal ratio between equities and fixed-income instruments which would be appropriate for pension funds?

Mr. Wissner: It would depend on the liability structure. What I am finding are the ones that look at the liability structure generally have more bonds, and the durations are sometimes two to three times longer than the durations of funds that do not look at the liabilities at all. This is where actuaries come in to give the sensitivities of the liabilities.

We are going to get into the actual simulation model itself, trying to value the return on stocks without making it an input. In most situations I have observed that a

model is employed which makes no mention of the liability and is preloaded with an assumption that says stocks will perform 7% better than bonds, no matter what. When this is put into the Markowitz model or this capital asset pricing model, these models will always tell you to buy stocks. Much of that is because you said that stocks will outperform bonds by 7% in the first place.

In contrast, what we try to do is come up with a way of figuring out what the total return on the stock would be, and what the return on the pension liability would be, and then come up with what the probability is that the funding ratio, which started out at 1, would be less than 1 after 30 years. We used a simple spreadsheet model with four equations. The first two equations look at the return on stocks. We use the dividend discount model, where the return on stocks is dividend yield plus dividend growth. We know the dividend yield we buy is 2%, but we do not know the dividend growth. We went back historically and measured the relationship between dividend growth and gross domestic product (GDP) growth. In the last two equations, we estimate what the bond yields will be over the next 30 years. We know that the bond yield will be related to the inflation rate, and the inflation is taken to be the nominal GDP minus the long-term trend for real growth of the economy.

The variable which is driving the model is the GDP. In other words, given a 30-year GDP path, we can derive dividend growth; and therefore we can derive the total return on stock. Given the 30-year GDP trend, we can derive what the inflation is each year and therefore the nominal yield on the bond. Once we know the nominal yield on the bond over the 30-year path, we can compute the total return of a 12-year duration liability and then compare that to the total return of the stock.

We used a distribution of GDPs which was based on the historical pattern of GDP growth from 1926 to the present. On average, that led to a GDP growth of about 6.6%. Then we wanted to put a stress on the model to see what would happen if in the future things did not behave as they did according to the pattern of GDP growth over the last 50 years. What happens if we get into a protracted period of disinflation, in other words if GDP growth is 2% less? What if GDP growth is 2% more? How would that affect the funding ratio?

How bad was the first equation, the dividend discount model? When we do regression analysis over 30-year holding periods, the model would predict a return on stock within about 1%. In other words, if the model said the return should be 7%, it would be between 6% and 8%, and that was basically the standard error of the regression. The dividend yield and dividend growth were significant. What you should realize is that although stocks returned 10% from 1926 to 1995, almost half

of that came from the initial dividend yield. This raises the red flag given that dividend yields are currently only 2%. Can you really expect that great long-term return off the stock market?

You would imagine that dividend growth would be related to the growth in the economy. We tested that particular relationship and found that dividend growth lags the growth in the economy by about 2.5%, with a standard error of the regression of about two-thirds of a percent. Given the 2% dividend yield, if dividend growth reverts back to its historical mean, in order to produce the 10% historical return on equity, you will be more than 3 standard deviations away from the mean dividend growth of 6.2%. In order to produce a 6% premium, which is one of the assumptions put into the capital asset pricing model (CAPM), dividend growth will have to actually exceed the growth in the economy by a wide margin. That does not seem very likely based on historical results.

One of the problems with equity values is they compete with an attractive bond yield above 7%. Chart 10 analyzes the nominal yield on 12-year duration immunized bonds based on a core inflation, and compares it to the rate on treasury-index-linked gilts Any time bonds have more than a 4% real yield, they look very attractive. The nominal yield on bonds less the floor rate of inflation compared to the dividend yield on stocks is at a very high margin. It means that the bond looks attractive compared to the stock. As people keep believing that stocks are a better buy, they are going to drive that relationship wider apart. That is going to make the hurdle rate for stocks greater when compared to the pension liability.

Let me get into the simulation trials. Chart 11 is the probability that the funding ratio will be less than one. If you are 100% invested in stocks over the 30-year period, given 1,000 trials and 200 low discrepancy point trials, the results are very similar. In the 1,000-trial run, the GDP is off by maybe five-basis-points compared to the 200-point sequence. All the statistics are close to one another. You come up with a significant 55% probability of the pension plan being underfunded, or the funding ratio being less than one with the all-stock strategy at this point in time. It is not that good a bet, and as more people are driven into the stocks as long as bond yields remain high, that bet becomes worse and worse. When you stress test the situation, as you might imagine, the results become even more dramatic. (See Chart 12.)

If the federal reserve is successful in producing a disinflationary environment, this would be a very poor environment for stock investment and pension plans. In this case, the probability of an underfunded plan would be almost 79%. The results again are very close to the 200 low-discrepancy point results which Graham produced for us.

The environment that will bail most people out will be the high-inflation, high-growth environment, which is the environment that most people who are currently investing in stocks today grew up in. It is characterized by rising interest rates and rising inflation. In that particular environment, the odds are overwhelming that the funding status of the plan would actually improve 73% versus 27%. Again, they agree for the 200-trial run and the 1,000-trial run (See Chart 13.).

This says to me that most people's behavior is determined largely from their experience. The baby boomers have experienced mostly inflationary type environments and that is why they are led into the stock market. What I remember about my father, who lived in the depression, was that he did not want to touch a stock with a ten-foot pole. However, if he had used the dividend discount model back in 1947, stocks were a very good buy. Right now, using the dividend discount model, the value does not seem to be in equities as long as dividend yields are low compared to the high yield on a long-duration bond portfolio.

Dr. Vanderhoof: This is actually a 32-dimensional example, 30 years and two separate variables relating the growth to the dividend and the return.

From the Floor: Why is it 32? I think all they did was GDP.

Dr. Vanderhoof: It was done by taking each year separately for 30 years; 30 values compounded to make one 30-year growth rate for that year. Then go on for that trial. The trial had 30 different values of growth on a yearly basis. They were put together to get the 30-year figure with two separate variables relating the growth to the dividend and the return.

From the Floor: When you ran 1,000 trials under the old way of calculating, how many numbers do you need?

Mr. Wissner: There were at least 30,000 for that trial, but then there were other error functions from the regression that come into it, and also from the two regression equations.

Dr. Vanderhoof: Basically, you need 32,000.

From the Floor: The purpose of this meeting is to show the value of the random numbers.

Mr. Wissner: I had no confidence in whether 1,000 would have been enough.

From the Floor: I appreciate that, but is the distribution equivalent? Does the distribution which comes from the 200, give you a sense of the highs and lows?

Dr. Vanderhoof: At this point, you know as much as we do because these are the tests that have been done. It is not that we are taking a sample of many tests that have been done previously. This is it.

If you use 1,000 randomly generated numbers you are in the position which Len has been in. You do not necessarily have the confidence that 1,000 or 5,000 is enough. By showing that 1,000 and 200 produced essentially the same answer, it gives you not only confidence in the 200, but it also gives you much more confidence that 1,000 is plenty. You can do it using random numbers and I invite you to do it.

From the Floor: What I am saying is if we ran the old way, 100,000, would the annual GDP growth be 4.62?

Dr. Vanderhoof: It would be essentially the same, unless the particular random number generator you had involved the kinds of bias which I demonstrated. All the random number generators seem to have a built-in bias based on the seed. I can say this does not seem to have a built-in bias.

Dr. Lord: In the example I am going to discuss we took a real $400 million block of business which had somewhat mature policies in it, issue years 1987–95, and an average policy size of $33,000. For this particular test, the crediting strategy was the portfolio method. The policies that are in the block of business have different guarantee periods. In some cases they are one year, and in other cases, the initial guarantee period was as long as five years. In each case, it seems as though the reset was annual thereafter. For this test, one had to choose what the competitors were doing. It was essentially spread off a medium-term rate, and an algorithm was used to determine whether the policyholders would lapse depending upon the difference between the current portfolio rate, the rate that the SPDAs were earning, and what the competitors were doing. The surrender charge was declining. In many cases it was seven, six, five, four, three, two, one. The average was not 5%, perhaps because some of these policies having been there since 1987 had no surrender charge left. Five percent is too high.

The initial yield curve was valued as of the end of the first quarter 1996. We generated interest rates using a log normal process which had certain constraints on it to make it look more reasonable than just tossing a coin. The model that this was run through was the Tillinghast Actuarial Software (TAS). Rather than look at a real

live portfolio, we selected generic assets, medium-term notes, the asset class, I think grade A, term of 5–7 years, and reinvestment in medium-term notes.

We used low-discrepancy points to come up with two sets of interest rate scenarios. Just to give you a flavor of what these are about. Chart 14 is the summary of the set of scenarios from the 200 low-discrepancy run.

What the model requires is the constant maturity yield curve, which is ten points. I did not draw a ten-dimensional yield curve, but rather I took two paths of the yield curve. Because it is an investment horizon of ten years, quarter by quarter, and I cannot do 40 quarters, I have done a rather crude statistical average by taking percentiles at every quarter. Chart 14 will give you the percentiles—the 95th, the 75th, the median, the 25th and the 5th—as we go down. These were generated by generalized Faure sequences and it produced these envelopes. There is some volatility. There is a mean reversion displayed by the stability of the median. In case of the question, "Did you have arbitrage-free scenarios?," the quick answer is "no". If you do use the switch in TAS for arbitrage-free scenarios, that tends to reduce the volatility. Since this exercise was to bring out the distinction between a run of 200 and a run of 1,000, not having the arbitrage-free scenarios would make this comparison a more stringent one.

The results are shown in the following three tables. The tests only allow 999 scenarios because there is only a three-digit field for how many scenarios you can run, and 999 is the maximum. It is not 1,000, although I keep saying it is 1,000. First is the present value of book profit (Table 1).

This helps with the question, do you reproduce the distribution with the set of 200 compared to a much larger set? Without sophisticated statistical tests, if you compare the two columns, the percentiles match up surprisingly well. This tells me the distribution of book profit for 1,000 runs is being mimicked well by the distribution with 200. Averages and standard deviations are reported, although these are not a particularly robust measure in the TAS model, because once the company goes insolvent, even if it is in the first quarter and stays insolvent, the system does not seem to handle the future behavior of the company after the period of insolvency. That affects the average.

TABLE 1
ASSET/LIABILITY STUDY

	Present Value Book Profit (BFIT) $(MM)	
Percentage	200 Scenarios	999 Scenarios
95%	60.1	59.9
90	55.7	56.1
85	53.5	53.4
75	48.5	48.9
Median	37.5	37.3
25	27.7	27.3
15	21.0	20.1
5	7.4	7.4
Average	36.3	35.9
Standard	18.3	18.5

Comparison of results based on 200 and 999 interest rate scenarios generated by low-discrepancy sequences

Table 2 shows the present value of book and market surplus. It depends on how you value your assets. Again, you compare 200 to 1,000. Again you see comparable results. The conclusion is that the 200 is mimicking the 1,000.

Depending on how you value your assets, book or market, and again comparing the corresponding columns, you get the similar conclusion. Not only are we reproducing the expected value as captured by the mean or average and also partially by the median, we are reproducing the actual probability distribution of ending surplus. That enables us to ask the question, “If I were to make a statement about the probability of insolvency, am I going to be robust in finding what that probability of ruin is if I have only 200?” Looking at the last line of Table 3, 200 scenarios is not that different from the probability of ruin under 1,000. You would expect that due to your percentile distribution higher up in the table of 1,000 is being paralleled in the 200. Table 3 is ending surplus.

TABLE 2

	Present Value Book Surplus ($MM)		Present Value Market Surplus ($MM)	
Percentage	200 Scenarios	999 Scenarios	200 Scenarios	999 Scenarios
95%	78.5	86.9	78.3	87.2
90	72.2	81.3	72.8	81.6
85	69.6	76.5	68.8	76.7
75	62.5	68.2	62.3	68.1
Med	46.4	51.5	46.1	49.2
25	32.2	28.8	32.2	30.0
15	24.0	20.9	23.4	19.4
10	19.5	15.1	17.4	12.7
5%	9.7	5.9	7.3	5.2
Average	45.7	47.9	47.9	47.6
Standard	25.6	26.0	26.0	23.1

Comparison of results based on 200 and 999 interest rate scenarios generated by low-discrepancy sequences

TABLE 3

	Ending Book Surplus ($MM)		Ending Market Surplus ($MM)	
Percentage	200 Scenarios	999 Scenarios	200 Scenarios	999 Scenarios
95%	119.2	132.3	119.1	133.2
90	113.0	124.5	113.8	125.0
85	109.3	119.6	109.7	120.6
75	103.0	111.0	102.8	111.3
Med	84.8	91.1	84.4	89.1
25	66.1	60.7	66.0	62.2
15	55.5	50.5	52.6	45.2
10	47.5	34.9	42.6	31.4
5%	24.0	16.7	20.4	14.7
Average	79.9	82.6	79.0	82.3
Standard	34.4	39.6	35.6	41.0
Probability of ruin	0.0300	0.0350	0.0360	0.0390

Comparison of results based on 200 and 999 interest rate scenarios generated by low-discrepancy sequences

Not only can we use low-discrepancy points for calculating prices, expected values, or economic values in the case of liabilities, but we actually can make some comment about the distribution of our actuarial and financial variables. Just to summarize, what we concluded from the results is that the key variables were similar in value. The distributions between the 200 scenarios and the 1,000 scenarios were similar. The probabilities of insolvency were also relatively close.

One thing I did not tell you was how long it took TAS to run the 1,000. It took between 16 and 18 hours. This is before they started printing the results. How many hours did it take to run the 200 scenarios? The answer is two hours. Here you have an enormous savings in time efficiency, and you preserve the results which you had for 1,000. The key thing is this allows you to do stress testing and product design testing in less time, and you can be assured that the results you are getting are solid. The thing that I like is whatever decision you make based upon doing 200 scenarios will be the same decision you would make if you had gone 16 hours and done the 1,000-scenario run. Where it would take you a weekend to do 1,000, now on a Friday you can do five or six tests.

We are also doing other studies on other models and other interest rate generators. It is coming out that the results are very similar to the ones that we just presented.

Dr. Vanderhoof: Denny Carr made available the services of the portfolio to do this particular test. I like this because it says something else to me. It says I have a reason to have confidence that if it shows up with seven ruin scenarios, those seven ruin scenarios are the only ones with which I can be comfortable over the range of possibilities. If the underlying model is correct, I have covered it. If I do it on Monte Carlo, I still want to check as to why those ruin scenarios existed. I never had a good feeling that there was not some other thing that was not hit by the Monte Carlo that was equally bad, because Monte Carlo doesn't really try to cover the entire probability space.

Dr. Papageorgiou: I am going to describe the work on deterministic pricing of financial instruments we are currently doing at Columbia. This is joint work with Joe Traub. I will discuss financial instruments, low-discrepancy sequences, a review of the current state of affairs, and finally some test results which show speed-up factors much larger than those you saw already. In some sense they are very surprising.

Typically with financial problems, you have to compute the expectation of a function of a random quantity. This function tends to be multivariate and the dimension is usually high. You want to do this at a high-speed despite the fact that the instruments can be complicated in terms of models, or the fact that you may

have to price a large number or a book of instruments. The accuracy is relatively low, i.e., low as compared to other engineering problems; here you are satisfied with ten to the minus two, ten to the minus four, one basis point, and so on.

For problems which involve expectations, such as the computation of an integral, the Monte Carlo method was the choice until the early 1990s. Low-discrepancy methods and deterministic methods in general were not used at all. Up to that time everybody believed that low-discrepancy methods, or deterministic methods, would be good as long as the problems had very low or moderate dimensions. However, they would lose any theoretical advantage over Monte Carlo once the dimensions would grow to 10 or 30 dimensions. On the other hand, what we have shown is that for a range of financial instruments, low-discrepancy methods beat Monte Carlo in the sense that they produce small errors using a small number of samples, and when the accuracy demand grows, the speed-up factors can be huge.

The work started in 1992 at Columbia and led to a sequence of papers. Spassimir Paskov got his Ph.D. working on a model problem. The most recent paper appeared in *Risk* magazine, and it is by Joe Traub and me. It deals with one of the test examples I am going to discuss.

Let me summarize what we do with a financial instrument.

S = underlying, asset follows model
For example, $ds = \mu s dt + \delta s dz$
G = derivative instrument, function of s, mode
$E[G(S)]$ (expected value)?

Example 1: Asian Options

$$\tilde{S}_{ik} = \frac{1}{k-i+1} \sum_{j=i}^{k} S_t \; ; \; 0 \le t_i \le t_j \le t_k \le T$$

$$G_c(S) = \max\,[S_T - \tilde{S}_{ik}, 0]\; e^{-tT}$$

average strike ↑ ↑ *discount factors*

Example 2:

$$\tilde{S}_{ik}=\max\left\{S_{t_j} : j=i,\ldots k\right\}$$

Usually you have a time frame from now to some time, *T*, in the future. Discretize the time frame into periods, and then observe the value of an asset, let's say a stock. You also have a derivative, that is a financial product, which depends on the stock. The question is, how am I going to compute the expected value of the derivative product? A typical example is an Asian option where you have arithmetic mean average strike options. This is a call option to buy the asset at its average price over a particular period. In general, the parameters may vary, and depending on the financial instruments, these quantities can be different. You can have a look-back option, where now you are looking at the maximum of the price of the stock during a certain period of time.

Let me come to low-discrepancy sequences, because we were using Monte Carlo sequences for these problems, and low-discrepancy sequences is what we are suggesting. One thing I would like to make clear is that discrepancy is a global property that we would like our point set to have. It is not a particular number which we are trying to achieve. Discrepancy says that I have a measure of uniformity for *n* points. The most characteristic of the properties, or at least the one you can visualize, is that the points that are in the d-dimensional cube do not have clusters.

Going to some examples of low-discrepancy sequences, I would like to mention the Halton sequence is a unique sequence. On the other hand, the Sobol sequence is not unique; it's a class. You have instances that differ between them. All of them obtain the same asymptotic discrepancy bounds. All of them are low-discrepancy and they have discrepancy equal to a constant times log n to the d/n. The constants however, differ between the instances, and the constants also depend on the dimension. The Faure sequence is just a single sequence; this is work of the 1980s. Later work led to the generalized Neiderreiter sequences. Finally, the generalized Faure sequence, which is also a class of sequences, not a single sequence, was obtained by Tezuka in 1995.

The clustering in Monte Carlo methods is shown in Chart 15. There is also an example of a low-discrepancy sequence in Chart 16.

This is an instance of generalized Faure sequence for you to contrast with the Faure sequence which Graham gave you before. You can see certain patterns. There is a great deal of work currently going on in algebraic curves that produce such sequences.

The following are some formulas about the low-discrepancy sequences. The (*t,d*) sequences are low-discrepancy sequences that have to satisfy even stricter uniformity properties. The generalized Neiderreiter sequences are obtained in the following way. For every natural number *n*, you expand it in some basis, by let's say *b*. For simplicity, I would assume that *b* is a prime. This expression is unique. Once you have this, you create, coordinate-wise, sums that depend on the coefficients c_{ij}. These are the important numbers that one has to consider; $a_j(n)$. The coefficients of *n* in base *b* are easy to obtain. The quality of the sequences depends on these generator matrices. The task is how to obtain these.

NEIDEREITER CONSTRUCTION (t,d) Sequences

$n = 0,1,2 \ldots$

$$n = \sum_{t-1}^{\infty} a_t(n)\, b^{t-1}$$

$$x_{(n)}^{h} = \sum_{i=1}^{\infty} x_{(ni)}^{h}\, b^{-i}, \text{dimension } h = 1,2,\ldots d$$

$$x^{(h)}\, ni = \sum_{J=1}^{\infty} c_{ij}\, a;(n)$$

$c^{(h)} = c_{ij}^{(h)}$, $h = 1,\ldots,d$ generator matrix

For example, Generalized Faure Sequence [Tezuka]

b-prime ($\geq$ d)

$c^{(h)} = a^{(h)} p^{(h-1)}$

↑

Nonsingular lower triangular

In the generalized Faure sequence, the generator matrices are given as the product of some lower triangular and nonsingular matrix multiplied by a power of the Pascal matrix. Graham showed you the Faure sequence. In that example, *a* was identity. If, however, you take various choices of *a*, then you get various instances of the generalized Faure sequence.

Let me tell you what we have done at Columbia. Spassimir Paskov began by taking a model problem and comparing the Halton sequence, the Sobol sequence, and the Monte Carlo sequence. He concluded that the low-discrepancy sequences were better than Monte Carlo and that the Sobol sequence was the method of choice. He also did some improvements on Sobol itself. He even found that Monte Carlo exhibits a tremendous sensitivity with regard to the seed. You may end up with different results using different seeds, which is important because this can put you in a very difficult situation. Later on we started using the generalized Faure sequence, and we applied it to a number of financial instruments ranging from options for

equities, to bonds, to our collateralized mortgage obligation. In all cases, we found that up to this point generalized Faure was the method of choice. All this work was put in Finder which is a software package originally built by Paskov, and it included Halton and Sobol along with some random number generators. We have recently added various instances of generalized Faure. Finder is available from Columbia University.

The test results will give you some idea. We considered a CMO based on 30-year mortgages with monthly cash flows. You end up with a 360-dimensional problem. Because it is a CMO, you have ten tranches which would leave you having to compute ten 360-dimensional integrals. Some tranches were easier to approximate than others. We have chosen to show results on the residual tranche which was the hardest to approximate, because it was influenced by all of the 360 interest rates.

The situation was quite surprising because you have a 360-dimensional problem. Chart 17 shows a simulation where the number of points is shown along the x-axis. These are 5,000 points and on the y-axis you have the relative error. What you see is that the generalized Faure sequence that achieves an accuracy of 10^{-2} and remains there, and that is at about 170 points. Monte Carlo achieves the same level at about 2,700 points. Sobol is somewhere in the middle at about 600 points. In this respect, we were very surprised.

We went on to look at the convergence rate of the method. As you already know, the error of Monte Carlo is proportional to $1/\sqrt{M}$. What you see in Chart 18 is the number of points along the x-axis came up to 100,000. I am showing that the error using the generalized Faure sequence is equal to 1/N, times a constant of moderate size. The constant is shown by the line. You see the constant does not exceed 20. On the other hand, you see that the Sobol constant is bounded by 80; Sobol was a little bit worse.

Table 4 summarizes the results for tranch R where you see that for accuracy 10^{-2}, Monte Carlo will take 2,700 points, and generalized Faure needs only 170 and we found this to be extraordinary. As the accuracy demands grow, you see that Monte Carlo may require 800,000 points versus 16,000 points for generalized Faure.

The speed-up factor is the number of points that Monte Carlo requires in order to achieve an accuracy epsilon and stay there. It is not to exit that band, relative to the number of points that generalized Faure requires for the same amount of accuracy. You see that the speed-up factor starts at 16 and goes all the way up to about 500 as the accuracy demands grow. It is not just five PCs that can do the job; but you need more than 500 PCs to do the same job.

Let me give you another example, before I summarize. This is our discount bond. This is a five-year bond, and you end up with a problem in 1,439 or almost 1,500 dimensions. The interest rate is modeled by the Vasicek model, $dr = a(b-r)dt + \sigma dz$, where a, b, and σ are given, and r_0 is also given. This was one of the test candidates that Tezuka used to test his implementation of generalized Faure. We did the same test, and the speed-up of generalized Faure relative to standard Monte Carlo can be 1,000.

TABLE 4
SUMMARY OF RESULTS FOR TRANCH R

CMO, D=360: TRANCHE R		
Rel. Error E	Method	Number of Points
10^{-2}	MC Sobol G. Faure	2,700 600 170
$\frac{1}{2}10^{-3}$	MC Sobol G. Faure	800,000 96,000 16,000
SPEEDUP		
$N_{nc}(E)/N_{GF}(E)$	$N_{SOB}(E)/N_{GF}(E)$	E
16	3.5	10^{-2}
50	~6	$\frac{1}{2}10^{-3}$
> 500	~4	10^{-4} <

In Chart 19, you have 100,000 runs and the value of the points. In the bottom, you see the error of generalized Faure and Monte Carlo. Generalized Faure is clearly much superior. I should mention that the generalized Faure method requires at most 30,000 points to give you one basis point accuracy, while Monte Carlo would require 30,000,000.

One of the important features of this method is that it is able to generate samples at a cost proportional to the cost of a linear congruential generator for those same samples. Changing and comparing it to other techniques is not always fair. The idea is that you have two methodologies: one Monte Carlo as expressed by a linear congruential generator, and then a low-discrepancy method as expressed by a

cheap-to-compute scheme. You compare the two of them. Otherwise, you could spend the extra time in generating more samples, and then you would break even.

To summarize, we have tested low-discrepancy sequences in a fairly wide range of financial instruments ranging from options on equities, to CMOs, and to bonds, and to options on swaps that depend on these bonds. We have found that the low-discrepancy methods, in particular our improvements of the low-discrepancy methods, give you small errors, using very small samples. The speed-up factors can be very huge and up to this point our improvements of the generalized Faure sequence make it the candidate of choice.

Dr. Vanderhoof: I would like to point out that the use of low-discrepancy sequences is independent of the problem. If you use the independent variable techniques or any of the other methods, such as variance reduction techniques, they tend to be specific for the problem and they must be handled for the problem. For this one, the only thing you need to hand tool is the inverse function where you go from the cumulative distribution function back to the actual value of the parameters involved. It is a one-size-fits-all-problems situation, and one set of software is used to create the low-discrepancy points.

It's very exciting because in what amounts to a research environment, it is possible to do things people had not thought were possible. It is possible to make reasonable numbers of calculations and arrive at a value at risk. That is the probability of ruin, the probability of losses beyond a certain point without an exorbitant cost, and that is something people haven't generally tried before.