Monte Carlo

A Light Bulb for Illuminating Uncertainty

By Paul D. Kaplan, PhD, CFA®, and Sam Savage, PhD

In 1946, Stanislaw Ulam, a Polish-born mathematician who was a member of the Manhattan atomic bomb project, lay convalescing after an illness. To while away the time he played solitaire. Of course some shuffles of the deck resulted in games that could not be won, and he began to wonder about the likelihood of success. So he stopped playing with the cards and tried calculating the percentage of successful games out of all possible shuffles. This turned out to be harder than he thought. In case you were wondering, there are

80,658,175,170,943,900,000,000,000,000,000,000,000,
000,000,000,000,000,000,000,000,000,000,000,000

distinct ways to shuffle a deck of 52 cards (larger than the number of all the grains of sand on all the world's beaches), so checking them all clearly wasn't an option. Then in his own words,

> *I wondered whether a more practical method than "abstract thinking" might not be to lay it out say one hundred times and simply observe and count the number of successful plays. This was already possible to envisage with the beginning of the new era of fast computers, and I immediately thought of problems of neutron diffusion and other questions of mathematical physics.* (Eckhardt 1987)

He took the idea to his Manhattan Project colleague John von Neumann, the greatest mathematician of the 20th century, inventor of game theory, and father of the modern computer. Thus was born the computational technique code named "Monte Carlo." And the basic building block of Monte Carlo simulation, as it is called today, is none other than a computerized version of a roulette wheel with many billions of numbers around the edge. Von Neumann saw the merit in Ulam's approach, but he correctly predicted that truly random number generators might be difficult to implement. He said at the time, "Anyone who considers arithmetical methods of producing random numbers is, of course, in a state of sin." And in fact it took decades to work out all the kinks in the computerized roulette wheel. Today Monte Carlo has become a standard tool of risk management, and it is about to be taken to new levels in its latest incarnations.

"Too bad Ulam didn't stop with Monte Carlo simulation, because his next invention was a real bomb—the hydrogen bomb to be exact, which he cooked up with physicist Edward Teller."

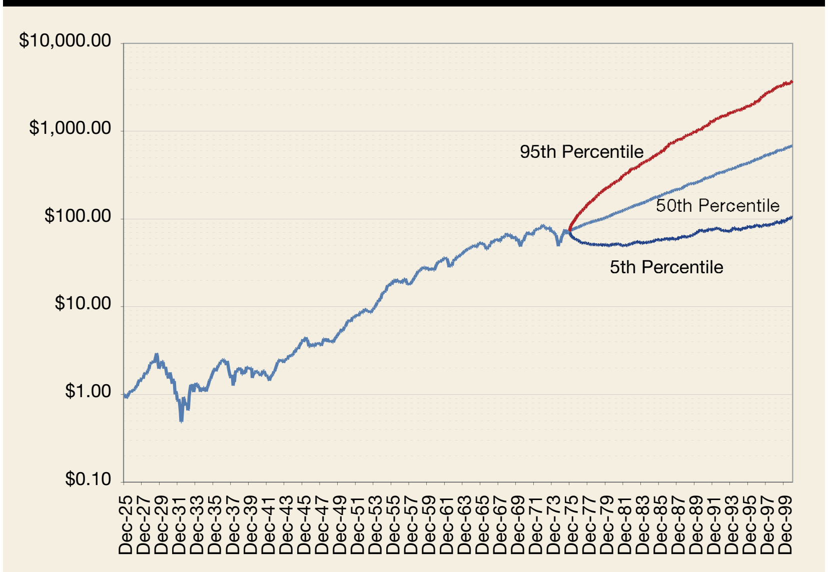
Too bad Ulam didn't stop with Monte Carlo simulation, because his next invention was a real bomb—the hydrogen bomb to be exact, which he cooked up with physicist Edward Teller.

Thirty years later in 1976, Roger Ibbotson, then an assistant professor at The University of Chicago, and Rex Sinquefield published a paper in the *Journal of Business* titled "Stocks, Bonds, Bills, and Inflation: Simulations of the Future (1976–2000)." It was a companion piece to their historical study of asset class returns. In "Simulations," they used the Monte Carlo method developed by Ulam to make probabilistic predictions of the form "there is an X percent chance that $1 invested in the portfolio will grow to $Y or more in Z years." Putting together history with the forecasts, they generated "tulip" or "fan" charts similar to figure 1.

Like Harry Markowitz's 1952 mean-variance model, the Ibbotson-Sinquefield simulation model was an early attempt to cure what Sam Savage calls the "flaw of averages" (Savage 2009). In general, the flaw of averages is a set of systematic errors that occurs when people use single numbers (usually averages) to describe uncertain future quantities. For example, if you plan to rob a bank of $10 million and have one chance in 100 of getting away with it, your average take is $100,000. If you described your activity beforehand as "making $100,000," you would be correct, on average. But this is a terrible characterization of a bank heist. Yet, as Savage writes, this very mistake is made all the time in business practice. It helps explain why everything is behind schedule, beyond budget, and below projections, and it was an accessory to the economic catastrophe that culminated in 2008.

Ibbotson and Sinquefield simulated each future month's return on a portfolio from historical monthly returns over the period 1926–1974, a period of 588 months. Like Ulam, Ibbotson

FIGURE 1: IBBOTSON-SINQUEFIELD SIMULATION CHART



and Sinquefield used a computer program to spin a roulette wheel with 588 spots 300 times for each simulated future. By running only a few thousand possible futures, they were able to complete the calculations on a mainframe computer of the era in time for publication.

While there was interest in the Ibbotson-Sinquefield simulation model at the time of its publication, technology for running Monte Carlo simulations was not readily available to many in the investment community. But four years later in 1980, four researchers published a paper in the *Journal of Business* that showed that to a large degree the results of the Ibbotson-Sinquefield simulations could be replicated without Monte Carlo simulation (Lewis et al. 1980). Titled "The Ibbotson-Sinquefield Simulation Made Easy," this paper showed that by making a number of simplifying assumptions and applying the central limit theorem,[1] probabilistic forecasts of cumulative wealth can be made using mathematical formulas.[2] The "made easy" model became the standard method for probabilistic forecasting and is in wide use today.

However, as powerful as the "made easy" model is, it is not up to the task of forecasting problems other than simple wealth accumulation with no inflows or outflows. Consider the problem of forecasting how long a retiree can make a given amount of wealth last before going broke, assuming that she invests her unspent wealth in a portfolio of risky assets. If we were to assume a fixed rate of return on investments during retirement and solve for the year in which the retiree runs out of money, we would run afoul of the flaw of averages because there are many plausible scenarios in which poor returns in the early years cause the retiree to go broke well before the time forecasted. Except under highly simplified assumptions, the only practical way to approach this problem is Monte Carlo simulation.[3] Hence, the Monte Carlo approach has become the most common method for modeling the drawing down of wealth during retirement.

Furthermore, the capital markets do not always behave in the way that the simplified models assume. As Paul Kaplan discusses, history is replete with fat-tail events that are not captured by models based on the bell curve (as are all of the simplified models).[4] This is another reason why Monte Carlo simulation is usually the most practical approach to investment forecasting.

This is not to say that Monte Carlo simulation is a silver bullet. As with any quantitative model, the inputs are key. When simulating asset class returns, it is important that the assumptions reflect views about the future, not merely restatements of the past. In developing the inputs, care needs to be taken in selecting the indexes to represent the asset classes and the time period used. For asset classes that lack highly liquid markets, the volatility of the simulated returns may need to be significantly higher than that of the observable data.[5]

In addition, there are a number of practical issues when implementing the Monte Carlo model that must be taken into consideration. Gambera (2002) summarizes a number of these issues; namely:

1. The accuracy of the results is limited by the number of simulated histories. Hence there is a trade-off between the accuracy of the model and the time it takes to run it.
2. The amount of time needed to run enough simulated histories might be too long to be practical to obtain enough accuracy to make the model useful.
3. The amount of computer storage needed to run a model might be impractically large. For example, to store 1,000 simulated histories over a 25-year period of monthly returns requires storing 300,000 numbers per asset class.

Fortunately, 21^{st} century technology addresses these issues, making Monte Carlo simulation practical, interactive, and flexible. This is due to three computer technologies that recently have come together: interactive simulation, the DIST standard distribution string, and cloud computing.

Interactive simulation. The central processing unit (CPU) in today's iPhone is hundreds of times more powerful than the machine used by Ibbotson and Sinquefield, and many times faster than computers of 2002, the date of Gambera's publication. Furthermore, several recent software breakthroughs have focused specifically on the speed of Monte Carlo simulation. Risk Solver Platform, for example, from Nevada-based Frontline Systems, can simulate 100,000 spins of the roulette wheel in Microsoft Excel before the user's finger has left the enter key of the computer. The result is a new interactive mode of simulation that provides an unprecedented level of intuition into uncertainty. And more speed is on the way. Not only are CPUs getting faster but machines are being fitted with parallel processors. Many applications cannot be programmed to take advantage of multiple processors, but Monte Carlo simulation is a notable exception; it's known in the trade as "embarrassingly parallel." It may not be long before the high speed graphical processing units (GPUs) currently used to drive computer screens are harnessed for the purpose of running simulations.

The DIST™ standard distribution string. Vastly increased simulation speeds and interactive simulation both address the first two issues raised by Gambera. In effect they are a new light bulb for illuminating uncertainty. Continuing the analogy, the distribution string is the AC current that lights the bulb and addresses Gambera's third issue. The DIST standard distribution string is an open format for encapsulating thousands of Monte Carlo scenarios into a single compact XML string (figure 2). Thus the 300,000 data elements required to store a 25-year simulation is reduced to 300 elements. When people say that size does not matter, this does not apply to factors of 1,000.

Cloud computing. The DIST standard distribution string is so compact that large data sets may be downloaded over the web in seconds. Thus a "cloud" of widely accessible distribution strings will provide the basis for a collaborative network in which specialists in financial statistics can produce probability distributions, for immediate consumption by investors, worldwide. Hence it may unleash an industry in the distribution of probability distributions. Dare we call it distribution distribution?

The above advances allow for a probability power grid, which can drive asset allocation, retirement models, and valuations, on everything from laptop computers to BlackBerries and iPads.

Monte Carlo models built with DISTs also are highly flexible. In the wake of the recent global financial crisis there is much debate about how to best model the probability distributions of asset class returns. Some researchers are proposing that we replace models based on the bell curve or normal distribution (which are tractable from a theoretical perspective) with fat-tail models in which extreme events occur (which require simulation to analyze). Others argue that the models based on the normal distribution are adequate. Distribution strings do not favor any side in these

FIGURE 2: THE DIST STANDARD DISTRIBUTION STRING ENCAPSULATES THOUSANDS OF SCENARIOS

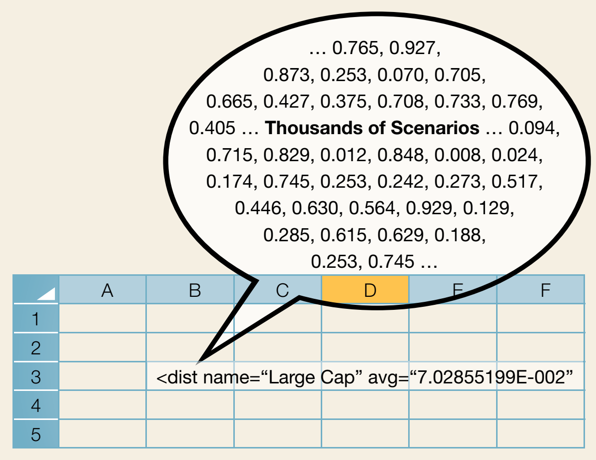


FIGURE 3: SCATTER PLOT OF "UNCORRELATED" HAP AND PY

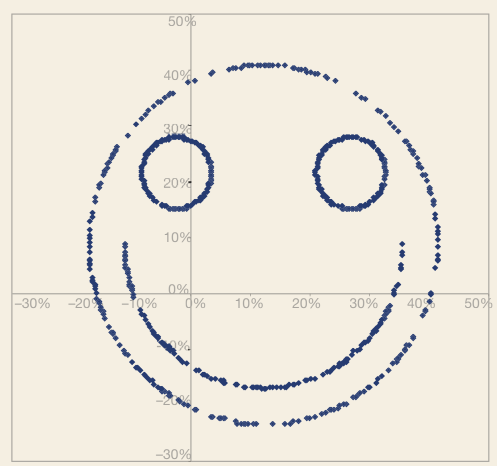
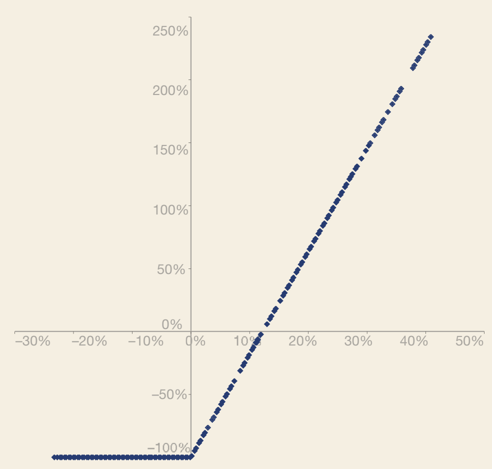


FIGURE 4: SCATTER PLOT OF ASSET VS. CALL OPTION



Software Supporting the DIST Standard

The power of this technology is beginning to appear in several programming packages for the computer savvy investment professional. Frontline Systems, the firm that developed the mathematical "solver" within Microsoft Excel, was the first to support the DIST standard with Risk Solver. This is an extremely fast interactive simulation package for Excel that can be coupled optionally to stochastic optimization or run in other environments from a server (see Solver.com). XLSim from VectorEconomics.com is a smaller DIST-compliant simulation package for Excel. In addition, XLSim enables DIST support in the popular @RISK and Crystal Ball simulation packages, and the JMP statistical exploration program. Analytica from Lumina.com is a powerful multi-dimensional modeling tool that can support DISTs of up to millions of trials.

For those who want ready-to-use interactive asset allocation software with Monte Carlo models, Morningstar is creating new tools based on DIST technology. In the near future, it will be possible to include many types of distributions, including those that model the occasional financial crisis, in an interactive environment on the desktop or laptop.

The DIST standard is open and simple to implement in such environments as Javascript, the R Project, Matlab, etc. Contact probabilitymanagement.org to receive pseudocode and other assistance in implementation.

debates because the possible outcomes stored in a DIST can come from almost any underlying probability model.

Similarly, there is debate about the usefulness of correlation matrixes to represent the interrelationships between asset class returns, with many arguing that during down markets asset classes become more correlated. Again, the DIST approach allows any pattern of co-movements to be modeled. As an extreme, the scatter plot in figure 3 of asset classes HAP and PY is a "happy face," which is certainly a type of relationship, even though the correlation coefficient between HAP and PY is almost zero. Compressing the underlying data into a pair of DISTs preserves the relationship in its entirety.

The ability to model nonlinear interrelationships between return distributions has important real-life applications. Consider figure 4, which is a scatter plot of returns on a stock index and a call option on the index. The DIST approach allows us to preserve the exact "hockey stick" relationship between the returns of these two assets, which, like the happy face, cannot be captured by a correlation coefficient. This is important if options are being considered as part of a portfolio.

These examples illustrate the importance of preserving underlying relationships between assets when creating a Monte Carlo model out of DISTs. Sets of DISTs that preserve such relationships are said to be "coherent." The creation of coherent DIST libraries is one of the most important functions of probability management, a field devoted to managing databases of probability distributions (see www.probabilitymanagement.org).

Appendix: The DIST Standard

The DIST 1.1 standard stores Monte Carlo trials with a numerical accuracy of one part in 256 (single precision) or one part in 65,536 (double precision) depending on the degree of compression and precision desired. In the single precision case, each trial falls into one of 256 buckets or bins (or 65,536 for double precision). The corresponding XML string has two parts: a header containing human readable information (figure A.1) and a body containing a sequential list of the bin numbers of each trial, encoded as a character string. Figure A.2 contains the first 20 and last 20 body characters of the DIST whose header is shown Figure A.1. For further details and downloadable interactive simulations in Excel, visit www.probabilitymanagement.org.

FIGURE A.1: THE DIST HEADER

<dist name="Large Cap Stock Return" avg="7.02855199E-002" min="-1.80592557E-001" max="3.89851560E-001" count="1000" type="Double" ver="1.1">

FIGURE A.2: THE BEGINNING AND ENDING PORTIONS OF THE DIST BODY

aWBOQxGZZo8xOLUak9hm …
hSckT4kr4ihoiUYAAAAA</dist>

The Distribution string was developed by Sam Savage, in collaboration with Oracle Corp., SAS Institute, and Frontline Systems. It has been applied at Royal Dutch Shell, Merck & Co, and Morningstar.

Paul D. Kaplan, PhD, CFA®, is quantitative research director at Morningstar Europe, Ltd. in London, UK. He earned a BA in mathematics, economics, and computer science from New York University and MA and PhD degrees in economics from Northwestern University. Contact him at paul.kaplan@morningstar.com.

Sam Savage, PhD, is chairman of Vector Economics Inc., Consulting Professor at Stanford University's Engineering School, and Fellow of Cambridge University's Judge Business School. He earned a PhD *in computational complexity from Yale University. Contact him at savage@stanford.edu.*

References

Chen, Peng, Gary T. Baierl, and Paul D. Kaplan. 2002. Venture Capital and its Role in Strategic Asset Allocation. *Journal of Portfolio Management* 28, no. 2 (winter): 83–89.

Eckhardt, Roger. 1987. Stan Ulam, John von Neumann, and the Monte Carlo Method. *Los Alamos Science* (special issue) 15: 131–137, available at http://www.fas.org/sgp/othergov/doe/lanl/pubs/00326867.pdf.

Gambera, Michele. 2002. It's a Long Way to Monte Carlo: Simulated Solutions of Commonly Used Financial Models Do Not Converge Very Quickly. *The Free Library* (July 1), http://www.thefreelibrary.com.

Ibbotson, Roger, and Rex Sinquefield. 1976. Stocks, Bonds, Bills, and Inflation: Simulations of the Future (1976–2000). *Journal of Business* 49, no. 3 (July): 313–338.

Kaplan, Paul D. 2009. Déjà Vu All over Again, *Morningstar Advisor* (February/March), available at http://corporate.morningstar.com/us/documents/MethodologyDocuments/ResearchPapers/DejaVuAllOverAgain.pdf.

Kaplan, Paul D. 1995. Reverse Mean-Variance Optimization for Real Estate Asset Allocation Parameters. *Real Estate Investing in the 1990s* (Charlottesville, VA: Association for Investment Management and Research).

Lewis, Alan L., Sheen T. Kassouf, R. Dennis Brehm, and Jack Johnston. 1980. The Ibbotson-Sinquefield Simulation Made Easy. *Journal of Business* 53, no. 2 (April): 205–214.

Milevsky, Moshe, and Chris Robertson. 2005. A Sustainable Spending Rate without Simulation. *Financial Analysts Journal* 61, no. 6 (November): 89–100.

Morningstar, Inc. 2010. *Ibbotson Stocks, Bonds, Bills and Inflation*. Classic Yearbook (Chicago, IL: Morningstar).

Savage, Sam. 2009. *The Flaw of Averages: Why We Underestimate Risk in the Face of Uncertainty*. (Hoboken, NJ: John Wiley & Sons, Inc.).

Endnotes

1 The central limit theorem states that when a large number of statistically independent variables with the same distribution are added, the result is a normal (bell-shaped) distribution, for almost any distribution of the underlying variables. An exception occurs when the underlying variables are drawn from fat-tail distributions discussed in Kaplan (2009).

2 For a description of the model and the formulas, see Morningstar (2010, 113–118).

3 Milevsky and Robertson (2005) presents a probabilistic formula for the sustainable spending rate that the authors derive under a set of assumptions, one of which is that the spending rate is constant. However, for other spending patterns, such as including occasional lump-sum amounts to finance a child's wedding, a grandchild's education, or a retirement home in addition to regular spending, Monte Carlo simulation remains the only practical option.

4 See note 1.

5 In the context of forming inputs for mean-variance analysis, see Kaplan (1995) for the case of direct real estate; see Chen et al. (2002) for the case of venture capital.