

PROBABILITY MANAGEMENT IN ACTION

BY SANJAY SAIGAL

HANDLING UNCERTAINTY represents the most significant need in converting quantitative insight into profitable action. Yet, explicitly incorporating uncertainty into analytics continues to present theoretical and practical challenges. Because computational results are tied so closely to model inputs, stochastic approaches require the modeler to make strong, sometime unwarranted, assumptions on key inputs such as probability distributions. When such assumptions are challenged by subject-matter experts

and end-users, it is difficult to reanalyze and regroup because stochastic methods are typically computationally intensive.

Two *OR/MS Today* articles [1] [2] by Savage, Scholtes and Zweidler introduced Probability Management, a new analytic framework for fast and robust stochastic analysis, in 2006. The first article described the key motivation for explicitly integrating uncertainty into decision-making: the outcome at the expected (or average) input is unlikely to be the expected outcome, which Savage cleverly rebranded as “the Flaw of Averages”

[3]. To address the shortcomings of average cases analysis, the authors called for “a shift in information management, from single numbers to probability distributions.” In addition, the article listed the technological, informational and managerial foundations of the field, such as interactive simulation, stochastic libraries in the form of SIPs and SLURPs, and introduced the idea of distributions certified by a “Chief Probability Officer.”

The second article expanded on interactive simulation and on the replacement of classical probability distributions by stochastic libraries. At the simplest level, stochastic libraries are pre-generated random trials that “approximate” stochastic inputs. But as likely, they could be outputs of simulation and optimization models. The magic in their simplicity is that unlike distribution formulas, uncertainties expressed as stochastic libraries are additive!

Probability Management has made a great deal of progress in the three years since the two articles were published. In addition to numerous projects that have profitably leveraged its concepts, software vendors have begun to provide native support for stochastic libraries, and quite impressively, the user community has begun to frame a best practices-based implementation methodology.

Given the rapid pace of developments, an update seems warranted. In the remainder of this article, we mention some successful implementation, list available and announced tools (software and hardware) enablers, and lay out the elements of an implementation methodology.

THE PROOF OF THE PUDDING

PROBABILITY MANAGEMENT has helped Shell Oil’s Exploration and Production (E&P) division move the energy giant from “a highly decentralized business with regional allegiances and reward systems to a single centralized organization managing a large portfolio of exploration opportunities” [1]. Critical to this reorientation, from sequential (best first) project selection to global portfolio management, is an “exploration cockpit” that incorporates simulation and optimization, as well as newly developed Probability Management tools such as “distribution strings” (DISTs), a modularized data format. (More on DISTs later.) Since it was developed in 2003, the model has undergone multiple upgrades while continuously supporting Shell’s worldwide exploration portfolio.

Probability Management posits the standardization of common uncertainties within an organization through stochastic libraries [1]. The Olin Corporation, a major

manufacturer of copper alloys, chemicals and other products, used this idea to coordinate otherwise unsynchronized functions of the company [8]. In one case, an Olin plant was frequently unable to meet customer demand, even though separate models for production and logistics had showed that capacity was more than adequate. A coupled stochastic model revealed how the interaction worked to limit the capability of the system as a whole, well below expectations.

Using interactive simulation that encoded how line personnel actually ran their respective silos, the Probability Management-savvy staff helped formulate organic policies that everyone could buy into. As a result, stochastic analysis is replacing average-based “steady state” calculations as the default at Olin.

The biotechnology giant Genentech has used Probability Management to address strategic supply chain planning questions. Genentech’s first project concerned the management of manufacturing assets [4], where interactive simulation was used to help decision-makers evaluate and derive policies that

simultaneously lower cost and delivery risk. More recent work has leveraged the company’s ongoing improvements to its product pipeline management system, and has involved using DISTs to make long-range forecasting faster and more accurate.

The effectiveness of Shell’s E&P portfolio optimizer prompted the pharmaceutical giant Merck to adopt Probability Management in its product portfolio planning, going so far as to bring on board the executive responsible for Shell’s success. Today, the use of interactive simulation allows Merck’s decision-makers to ask and answer critical questions in real-time [5]:

- How does my opportunity contribute to the portfolio?
- Is the portfolio balanced with regard to the company’s overall priorities?
- Are we making the best possible trade-offs?
- Is the overall risk of missing our aspirations tolerable?

The projects and companies mentioned above are not a complete listing of recent successes in applying Probability Management, merely ones that have been publicly documented. However they constitute a body of evidence that Probability Management is more than just a theoretical framework. Next,

let’s examine the technology artifacts that support its use.

TECHNOLOGY SUPPORT

STOCHASTIC LIBRARIES CONTAIN vectors of simulation trial data. That introduces platform-dependence: a data vector in Crystal Ball is not the same as a vector in Excel. To facilitate modularity, Sam Savage, consulting professor of management science and engineering at Stanford University and a co-author of the original Probability Management articles in *OR/MS Today*, proposed a new data type – the DIST. As mentioned earlier, a DIST is an encapsulation of a data vector (equivalently, a stochastic information packet, or SIP) in a more compact and transportable form. Any DIST conforming to the 1.0 open standard can be stored in a single spreadsheet cell, irrespective of the number of trials it encodes. Further, computations that use DISTs as inputs are faster, especially within spreadsheets. Savage has aptly described the DIST as moving simulation from the era of the six-shooter to the Gatling gun.

DIST-related functions are now available in Excel for users of Risk Solver from Frontline Systems. As of this writing, Frontline Systems distributes the only fully Probability Management-compliant

software for spreadsheets, supporting both interactive simulation and the DIST 1.0 standard. DIST support will also be available in AnalyCorp’s forthcoming upgrade of its Excel add-in, XLSim.

The traditional granddaddy of spreadsheet-based simulation, Crystal Ball, is now an Oracle product. Crystal Ball has been a pioneer in supporting libraries of probability distributions, and its leadership has actively influenced the development of the DIST 1.0 standard.

Because DISTs will soon become easily translatable with an add-in from AnalyCorp., most simulation packages should be able to support the format relatively easily. In fact an early application at Merck read DISTs into and out of @RISK, the other big player in Monte Carlo for Excel.

Specialized simulation software vendors have also begun to support the Probability Management effort, as in a real sense they were doing it before the movement began. These include Analytica from Lumina Systems, which in effect has used the stochastic library approach since the late 1970s, and Vanguard Software.

With its Enterprise focus, Vanguard has long championed collaborative, modular modeling, and was an early adopter of Probability Management [6]. That focus allows Vanguard’s software

SUBSCRIBE TO ANALYTICS

It’s fast, it’s easy and it’s FREE!
Just visit: <http://analytics.informs.org/>

to address more complex applications that desktop analysis tools cannot, for instance, estimating manufacturing cost impacts of new product designs at a \$19 billion aerospace giant. At the same time, the software can integrate spreadsheet models as calculation engines. So, for example, a simulation written in Excel by a financial analyst can be functionally embedded in a more complex Vanguard model. (Quite remarkably, a recent survey found that only one in every four simulation packages provides any Excel support whatsoever [7].) Vanguard also enables the creation and free transfer of stochastic libraries as SIPs and SLURPs, even extending the original SIP/SLURP definitions to include a time dimension. However, where Vanguard really pushes the boundary is through its grid computing capabilities. Performance speeds up to 250 times faster than spreadsheets allow Enterprise-level models to be run (and even modified) interactively over a Web browser.

The widening acceptance of simulation has further highlighted the key role of data scrubbing. Since this type of simulation modeling is largely on the desktop, Enterprise-oriented ETL tools are not well-suited for validating

and cleaning small-bore data. Unfortunately, the “garbage in, garbage out” adage is no less true for the individual analyst. The statistical analysis package JMP, from SAS Institute, has been successfully deployed for data preparation in some of the success stories mentioned in the previous section. A lightweight Excel-to-JMP DIST conversion macro, available from AnalyCorp, makes data scrubbing and preparatory analysis especially simple.

Implementation Lessons

AREAS OF APPLICATION

IT IS NOT SURPRISING that so many successful application of Probability Management have been associated with managing project portfolios and other strategic supply chain issues. Correlations between stochastic elements of a portfolio are notoriously difficult to compute, especially when the stochastics of the elements themselves are not well-understood. That is why, for instance, financial portfolio optimization uses the transparently reductive modeling assumption of linear correlation. Unlike traditional Monte Carlo, Probability Management does not impose extraneous, possibly unverifiable, relationships between

VANGUARD SOFTWARE

See What's Ahead

Advanced Forecasting
Collaborative Modeling
Interactive Web Reports
Decision Optimization

Forecasting and planning solutions for improving the quality, reliability, and speed of your business decisions

919.859.4101

www.vanguardsw.com

individual uncertainties (or classes of uncertainties). When causations or correlations are known, they can be modeled explicitly. When not, relationships encoded in historical data are still significant inputs for the simulation engine. (See [1] for a discussion of “coherent modeling”).

MODELING ENVIRONMENT

MOST OF THE REPORTED applications of Probability Management have been implemented in Microsoft Excel. However, almost by definition, many application in Analytica or the Vanguard System are examples of Probability Management. As a modeling environment, the ubiquitous Excel has a few benefits:

- Most analysts have at least an intermediate level of facility with spreadsheets. Even accounting for the simulation add-in, ramp-up by end-users is quick. In particular, since data often resides in spreadsheets, the experienced consultant can collaborate with the business subject matter expert (SME) to quickly produce “reasonable” prototypes.

- As described, many simulation environments now incorporate the key tools of Probability Management, specifically SIPs/SLURPs, and DISTs. DISTs tremendously speed up Excel-based models. Some add-ins are so efficient that tens of thousands of simulation trials are executed in the time it takes for a finger to leave the just-pressed key. There is no lag whatsoever! For the first time ever, true interactive simulation is possible on the standard business desktop.
- Excel’s charting capability enables interactivity without the need for “programming”. This is essential for engaging non-technical decision-makers, not just for getting buy-in, but also for knowledge discovery.

Of course, Excel has limitations as an analytic workbench. These include a restrictive two-dimensional data format, a non-trivial performance overhead, limited statistical capabilities, etc. Most of those issues have workarounds already implemented by add-in software vendors such as Frontline. One should also be open to the possibility of using specialized software (such as JMP) for specific needs (in this case, for data scrubbing and preparatory validation).

Excel’s limitations as a development environment are tougher to overcome.

While its embedded language – VBA – can be used to build simple custom decision-support systems (DSS), VBA’s robustness, performance and platform-dependence leaves much to be desired. Further, if the DSS is intended for use in a complex workflow, access control and data-sharing can present a challenge. If the prototype model is expected to eventually lead to a DSS, it is essential to evaluate and lay out a development strategy, possibly using a specialized simulation environment.

The beauty of the DIST data type is that eventually, distributions will be passed around between users of Excel, Vanguard, Analytica and JMP, just the way numbers are today. This leads to the best of both worlds, in which individual managers can investigate the implications of distributions generated on more robust systems in their spreadsheets.

REFERENCE DESIGN ELEMENTS

DUE TO THE OPEN and generous interaction of Probability Management practitioners, the outlines of a reference architecture are beginning to emerge. The design is driven by two common characteristics of problems most suited to the framework:

1. Many classes of basic stochastic entities (projects, products,

investments), where elements of each class are assumed to have similarly “shaped” uncertainties.

2. Sources of uncertainty can be global or class-specific. Global uncertainties, e.g., interest rates, are typically exogenous, so they do not lead to causative loops. (In other words, the results of the simulation cannot affect global factors.) Class-specific uncertainties, e.g., the availability of a certain raw material, may impact one or more entity class. There may even be inter-relationships, e.g., if the raw material is internally produced, its uncertainty may impact the price of a finished product.

Class-specific simulations are likely to be more detailed and thus, more complex. Using Probability Management, there is usually an opportunity to identify and re-use models created and knowledge collected by SMEs, say, a plant-level planner or a brand manager. Entity classes relate to each other and to global and class-specific uncertainties through a meta-model. The meta-model is typically less detailed. Cross-functional by definition, it is also less well-understood, and thus an especially fruitful area of investigation for this framework.

SUBSCRIBE TO ANALYTICS

It’s fast, it’s easy and it’s FREE!
Just visit: <http://analytics.informs.org/>

Analytics developers tend to naturally be more focused on model logic and less interested in implementing a presentation layer optimized for collaborative decision-making. Yet the end-user “dashboard” is the most important piece of an interactive application. The key design criterion is that the model should help managers make good decisions, not that it should make the best decisions. To that end, the design should rely on visual aids to communicate the impact of making different choices. In particular, data tables should be, as far as possible, hidden.

WHAT'S NEXT?

THANKS TO THE EFFORTS of a group of committed analytics practitioners, Probability Management has notched some high-visibility successes. The combined efforts of the ProbabilityManagemt.org committee members, including software vendors putting ideas within reach of implementers and academics and practitioners extending the ideas themselves, are rapidly advancing the framework. For example:

- A DIST 2.0 standard revision to incorporate full stochastic libraries including a time dimension is under discussion.
- A software-independent DIST

translator is under development.

- Many more Probability Management-inspired projects are getting underway.
- A reference design and methodology document is in preparation.

The coming year promises to be one of those interesting times for Probability Management. ■

Sanjay Saigal (ssaigal@intechne.com) is the founder of and CEO of Intechne, a provider of analytics-based solutions to ill-structured problems across multiple industries.

REFERENCES

1. S. Savage, S. Scholtes, and D. Zweidler, 2006, “Probability Management,” *OR/MS Today*, February 2006, Vol. 33, No. 1.
2. S. Savage, S. Scholtes, and D. Zweidler, 2006, “Probability Management, Part 2,” *OR/MS Today*, April 2006, Vol. 33, No. 2.
3. S. Savage, 2009, “The Flaw of Averages: Why We Underestimate Risk in the Face of Uncertainty,” Wiley.
4. D. Johnson, 2007, “Genetech Case Study: Applied Probability Management & Interactive Simulation,” *Proceedings of the 2007 INFORMS Conference on O.R. Practice*, Vancouver, Canada, April 2007: 268-270.
5. D. Zweidler, “R&D Strategy to Realization,” 2009, *Proceedings of the 2009 INFORMS Conference on O.R. Practice*, Phoenix, Ariz., April 2009: CD-ROM.
6. R. Suggs and B. Lewis, 2007, “Enterprise Simulation – A Practical Application in Business Planning,” *International Journal of Simulation*, Dec. 2007: 205-209.
7. E. M. O. Abu-Taieb and A. A. R. El Sheikh, 2007, “Commercial Simulation Packages: A Comparative Study,” *Proceedings of the 2007 Winter Simulation Conference*, Vol. 8, No. 2, July 2007: 66-76.
8. D. Cawlfeld, 2007, “Flaw of Averaged at Work at Organizational Boundaries,” *Proceedings of the 2007 INFORMS Conference on O.R. Practice*, Vancouver, Canada, April 2007: 422-426.



THE FRANZ EDELMAN AWARD
Achievement in Operations Research

2008 Presentation DVDs

AVAILABLE NOW !

Use these valuable
DVDs to see how companies
apply analytics to
increase efficiency and revenue.



**Contact INFORMS to purchase your
copy for \$45. INFORMS members receive a
complementary copy.**

Netherlands Railways,
City of Stockholm, Sweden,
Federal Aviation Administration with Metron Aviation and the
Volpe Center, U.S. Department of Transportation,
StatoilHydro and Gassco,
U.S. EPA, and
Xerox

<http://edelman.informs.org/DVD>