Proceedings of the 2015 Winter Simulation Conference
S. L. Savage, J. M. Thibault

TOWARDS A SIMULATION NETWORK
OR
THE MEDIUM IS THE MONTE CARLO
(WITH APOLOGIES TO MARSHALL MCLUHAN)

Sam L. Savage

Management Science & Engineering
Stanford University
475 Via Ortega
Stanford, CA 94305-4121, U

John Marc Thibault

Standards Committee
ProbabilityManagement.org
202-94 Sidney St.
Belleville, ON, Canada

ABSTRACT

The discipline of probability management, introduced in 2006, formalized the concept of data structures for storing arrays of simulated realizations. These are called Stochastic Information Packets or SIPs. Today the open SIPmath™ standard of 501(c)(3) non-profit ProbabilityManagement.org supports SIP libraries in XML, CSV and XLSX file formats. This article describes how such data may foster the creation of networks of simulations that bring stochastic modeling to general management. Skeptics may argue that most managers do not know how to generate the appropriate random variates. It was similarly argued that light bulbs could not be used by the general public as they would not know how to generate the appropriate electricity. In this context, probability management is devoted to the design of a power grid for probability that provides access to trusted sources of random variates. The SIP is a good candidate for the transmission standard.

1 INTRODUCTION

Sixty years ago, at the dawn of the computer age, simulation was considered a method of last resort. Today, which in the grand scale can be at most a few tenths of a second after sunrise; simulations are often the method of first resort (Lucas, Kelton, Sanchez, Sanchez & Anderson). Still, most simulations are run as standalone applications, in which uncertainty is modeled by internally generated random variates. The results are then presented as tables of summary statistics and graphs of output distributions whereupon it is game over. In principle, arrays of realizations can serve as both inputs and outputs of simulations, thereby enabling networks of stochastic models.

Coherent arrays of Monte Carlo realizations have been used since at least 1991 in stochastic optimization calculations (Dembo 1991). The discipline of probability management extends simulation through the standardized use of such arrays, which are referred to as Stochastic Information Packets (SIPs) (Savage, Scholtes, Zweidler 2006; Savage, Kirmse 2014). SIPs can serve as both inputs and outputs of simulations. A set of SIPs that maintain statistical dependence is said to be coherent, and is known as a Stochastic Library Unit with Relationships Preserved (SLURP). In Monte Carlo applications such as those created with the widely used @RISK, Crystal Ball and Risk Solver Platform packages, SIP data may be used anywhere a random variate is currently generated. Indeed, these common packages can easily import and export SIP libraries. In discrete event simulations, SIPs are currently used for communicating simulated results to other applications, but further uses are also anticipated.

2 SIP ADVANTAGES

SIPs are:
Actionable - the output SIP from one application can become the input SIP for a downstream simulation.
Additive - if coherence is maintained, the output SIPs of simulations of multiple entities may be added trial by trial to create a SIP of a consolidated entity.
Auditable - input and output distributions are treated as data with provenance supporting an audit trail.
Agnostic - SIPs comprise a simple data structure, which may be supported across many platforms.

2.1 SIPs are Actionable

Unlike the standard summary statistics and graphs of distributions that are output by most simulations, SIPs may be used as input trials to downstream simulations. In particular, Microsoft Excel has now become powerful enough to actively run thousands of trials with each keystroke using the native Data Table function (Savage 2012). Thus the output distributions of a wide variety of simulation applications may be used to drive stochastic dashboards in the hands of decision makers. Two examples are given below.

2.1.1 Forecasting

Most forecasts are communicated as single “average” numbers. This leads to several sorts of systematic errors collectively referred to as the Flaw of Averages (Savage 2009, 2012). For example, suppose we forecast the average number of surviving pathogens in chlorinated water to be 5 per gallon, but the actual number is drawn from a long tailed distribution (which it usually is). If the healthcare costs per person per year for 5 pathogens per gallon is $10, then it is tempting to believe that this is the cost to plan for. But healthcare costs are generally convex in the number of pathogens. For example, there might be no cost associated with zero pathogens, but $100 in cost associated with 10 pathogens. Therefore, the average cost may be much higher than $10. This is a manifestation of what mathematicians call Jensen’s Inequality. If the distribution of surviving pathogens is communicated as a SIP, it may be used in calculations at the local level to estimate the distribution of healthcare costs, which is necessary for effective mitigation strategies.

2.1.2 Decision Dashboards for Discrete Event Simulations

The schematic in Figure 2.1 below describes the discrete event simulation of the design study of a mobile communication system of a defense contractor. Because there are a number of engineering options, a design of experiment was run to estimate the reliability of the system for each combination of options in the design space. All told, this took many hours of computer time.

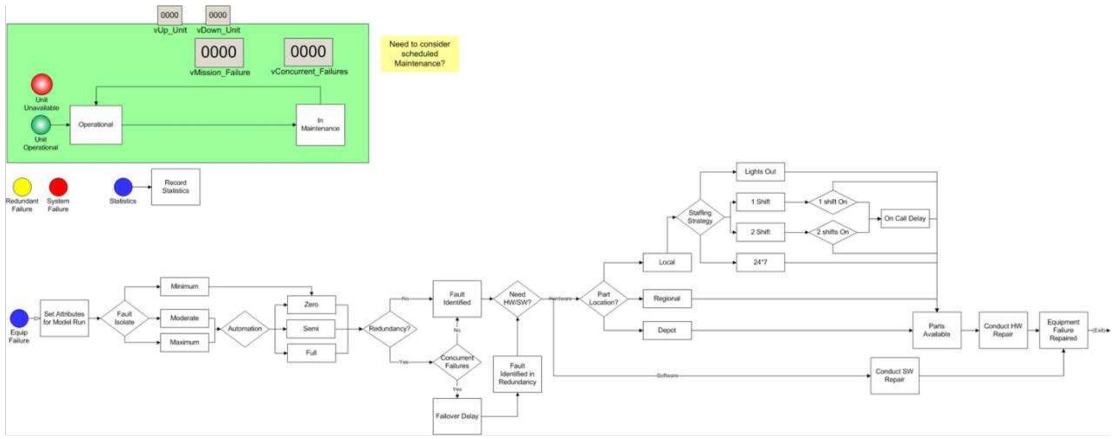

Figure 2.1 The schematic of a discrete event simulation

A separate SIP library was saved for each point in the design space. These were then loaded into an Excel model with slide bar controls for scrolling through engineering options. The decision makers were able to explore the stochastic implications of any combination of the design considerations in seconds on their own computers, even though the SIP libraries took hours to generate in a specialized simulation package.

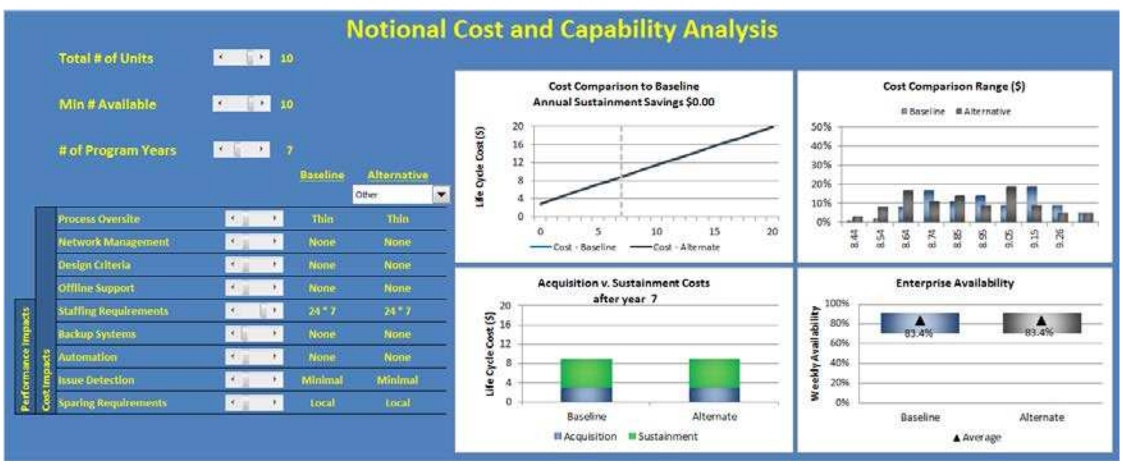

Figure 2.2 A stochastic dashboard in Excel for browsing simulation output

It should be pointed out that similar exploration dashboards may also be created using surrogate models generated from a large simulation. In theory, the SIP data standard might eventually be expanded to include surrogate models in addition to arrays of simulation realizations.

2.2 SIPs are Additive

Coherent SIPs may be operated on element by element with any algebraic operator through the common process of vectorization. That is, if *x* and *y* are random variables from a joint distribution where SIP(*x*) and SIP(*y*) are arrays of realizations that preserve the statistical dependence, then SIP(*x*+*y*) = SIP(*x*)+SIP(*y*), where addition is performed element by element over the arrays. For that matter, SIP(*x***y*) = SIP(*x*)*SIP(*y*), and SIP(*x**cos(y)) = SIP(*x*)*cos(SIP(*y*)), where the operations are taken element by element. This is just the idea behind Monte Carlo simulation in the first place except that the variables *x* and *y* are generated in advance, and stored in arrays, as are the output trials.

2.2.1 Stochastic Roll-up

Additivity allows for the stochastic roll-up of simulation results across the enterprise into a consolidated risk model.

For example, a petroleum firm was able to consolidate the simulations of numerous exploration ventures into a stochastic model of the risk and return of its portfolio as a whole (Savage, Scholtes, Zweidler 2006). In order to keep the SIPs coherent, it was necessary to separate global uncertainties such as oil price, interest rates, etc. from local uncertainties such as the volume of hydrocarbons within each venture. The simulations of all ventures were then run with the same global variable SIPs. So, if the price of oil is $60 per barrel on trial 437, then, by using common global input SIPs, the simulations for each of the ventures also have a price of $60 per barrel on trial 437. The resulting output SIPs of each venture will therefore be coherent (they comprise a SLURP), and may be added together in various combinations to model different portfolios. The final model for decision makers was again in Excel, allowing ventures to be switched in and out of the portfolio instantly while 1,000 trials were run for each change. The small demo version of this model shown below is available for download from the Models page at ProbabilityManagement.org.

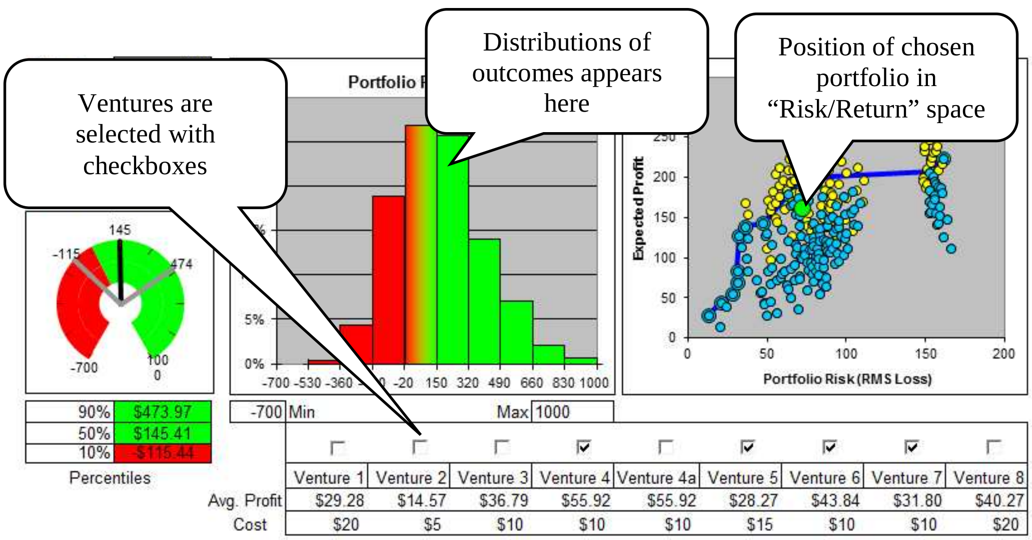

Figure 2.3 A dashboard for aggregating the simulation results of exploration projects

As another example of stochastic roll-up, consider the uncertainty in storm surge level in a certain coastal region. Each municipality in the region could independently assess the homes and businesses that would be impacted at various flood levels, in effect creating a lookup table of dollar damage by level. A common storm surge SIP could then be distributed to all local municipalities, much as the oil price SIP was delivered to each of the exploration simulations in the above example. The resulting output SIPs of damage at each municipality would thus be coherent, and could be rolled up to a SIP of total damage across the region. A demonstration SIPmath flood model is available at ProbabilityManagement.org.

2.3 SIPs are Auditable

Ironically, simulation is often viewed as suspect because of the very randomness from which it derives its power. Although one can seed the random number generator in most simulation systems to get repeatable results, it is difficult to get identical results from the same simulation across software platforms. From a computational perspective, calculations with SIPs (SIPmath) are strictly deterministic and should give the same answers on all platforms.

2.4 SIPs are Agnostic

The current data formats supported by the SIPmath 2.0 standard include XML, CSV, and XLSX. The standard specification document (Thibault 2014) is available for download from ProbabilityManagement.org.

3 THE STANDARD

The ability to easily move SIPs and relevant metadata from one system or application to another is a necessary part of the discipline of probability management. This includes the aggregation of results from separate simulations as well as accessing remote stochastic data sets - with their provenance - from trusted sources. A standard interchange format for SIPs, SLURPs, and their metadata facilitates these exchanges.

The standard was formed with the following principles in mind:

1 In general, SIP libraries will be produced by a few sources but they will be used many times by many consumers. Following the admonition of Nathaniel Hawthorne that “Easy reading is damn hard writing,” the focus of the standard is on readability rather than ease of creating the data files.

2 Options are kept to a minimum. Producer applications can choose one format option but consumer applications need to recognize them all. Again the focus is on keeping the reader simple to use.

3 Recognizing that the discipline of probability management is still evolving, the standard is relatively informal. The development is driven primarily by consensus and working code rather than extensive effort on design.

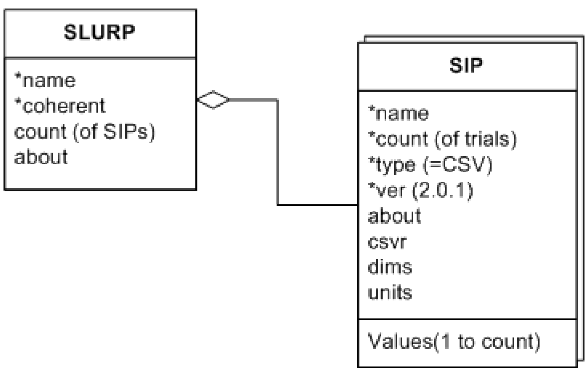

Figure 3.1 Data Architecture

Figure 3.1 presents the SIP/SLURP data architecture. It has been kept simple so that it can be easily realized using the native data structures of most, if not all, software platforms.

The SLURP is a container of SIPs with only two required attributes: a name or other unique identifier, and a flag indicating whether SIP relationships have been preserved (indicating coherence). Additional attributes may be included, and the standard includes a list of recommended attribute names.

The SIP is a simple object with four required attributes and the standard includes a list of recommended attribute names. The required attributes are: a name or other unique identifier, a count of the number of trials in the SIP, the data encoding type, and the SIP standard version.

The only data encoding currently specified is Comma-Separated Values (CSV). The *csvr* attribute is used to indicate how many digits to the right of the decimal place should be preserved.

The standard allows for multi-dimensional SIPs, using the *dims* attribute to specify the dimensions.

The standard includes three formats:

1 SIP/XML implements a universal XML format that is completely platform independent.

2 The Excel SIP Library, optimized for all-Excel applications. It supports Excel workbook-to-workbook exchanges including URL access to remote sources across the Internet. It is more complex than the other two formats, but very efficient in an all-Excel application.

3 Excel SIP/CSV defines a simple spreadsheet model based on the Excel .csv file format. Like SIP/XML it is both machine- and human-readable and simple to implement.

Being platform-independent, the SIP/XML file format is the most general-purpose and the easiest to read as-implemented.

```
<SLURP name="2016_Prices"
       coherent="true"
       count="12"
       about="2016 Month End Price Projection."
       approved="John Smith, 2015-03-15"
       copyright="Warbucks Financial Services"
       >
       <SIP …
</SLURP>
```

Figure 3.2 SLURP in XML

Figure 3.2 presents the realization of a SLURP in XML with both required and optional attributes in key="value" form. The data elements will be SIPs.

Figure 3.3 presents the realization of a SIP in XML. It has some attributes, and the data elements are sample values.

```
<SIP name="Price_2016_01"
       count="1000"
       effDate="2016-01-31"
       type="CSV" csvr="0"
       units="US$"
       ver="2.2.0"
       about="2016-01 Price Projection"
       approved="John Smith, 2015-03-15"
       copyright="Warbucks Financial Services"
       >
       12,14,51,95,35,42,42,58,91,65,43,…
</SIP>
```

Figure 3.3 SIP in XML

Translation functions for SIP/XML have been developed in Excel/VBA, the R statistical language, JavaScript, and Matlab. Export routines exist for the following widely used packages: Crystal Ball and @Risk simulation packages, and the Autobox time series analysis software.

4 ASPIRATIONS

What can be hoped for from such a standard? Ideally it will improve the way we think about uncertainty, and pave the way for a network of stochastic appliances, much in the way the power grid led to a network of electric appliances.

4.1 The Arithmetic of Uncertainty

When Fibonacci introduced Arabic numerals to Italy in 1199, it changed the way Western culture thought about numbers. It allowed people to sum up weights and measures, calculate interest payments, convert currencies, and then communicate the results using nothing but 10 simple symbols.

The quantities that can be expressed with Arabic numerals, the real numbers, are an example of what mathematicians call a field: an algebraic construct that is closed under addition, subtraction, multiplication, and division (except by zero). To best make use of numbers it helps if these operations are easy to perform. The famous sequence that bears his name was not invented by Fibonacci. Instead it was a public relations initiative to prove the superior arithmetic of his favored system over the Roman numerals and counting sticks of the time.

Similarly SIPs of a given number of trials may be thought of as a field, as any arithmetic operation may be performed on them. Given that a SIP is a computational representation of a probability distribution, it means that we can also think of probability distributions as a field, whose elements may be manipulated with the same operations we use for numbers. The examples shown make use of Sparkline graphs to demonstrate the arithmetic of uncertainty with histograms as both arguments and results in native Excel. Again, this is an active simulation in which each keystroke runs 10,000 trials through an Excel Data Table before your finger leaves the <Enter> key. Download the Sparkland file to experiment with your own calculations involving distributions.

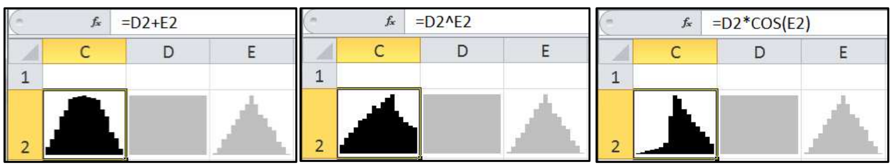

Figure 4.1 A distribution calculator

4.2 The Network Effect

The smartphone was nothing new. Computers, cell phones and touch screens had been around individually for at least a decade before the introduction of the iPhone in 2007. But with a common communication protocol the smartphone rapidly became a node in a network of 100 million other smartphones, and that *was* new. Network effects, as they are known in economics, create increasing value as the number of nodes in the network increases.

Similarly, the discipline of probability management is nothing new. Monte Carlo simulation, array arithmetic, and data bases have existed for decades. But the common communication protocol of the SIPmath standard has the potential to turn any simulation into a node in a network of many other simulations, and that *would* be new.

REFERENCES

Dembo, R.S. 1991. "Scenario Optimization". *Annals of Operations Research* vol. 30: pp 63-80.

Lucas, T.W, Kelton, W.D, Sánchez,P.J,Sanchez, S.M., Anderson, B.L. 2015. "Changing the Paradigm: Simulation, Now a Method of First Resort". *Naval Research Logistics* DOI 10.1002/nav.21628: pp 293-302.

Savage, S. L., 2009. *The Flaw of Averages.* Hoboken: John Wiley.

Savage, S. L. 2012. "Distribution Processing and the Arithmetic of Uncertainty". *Analytics Magazine* November/December: pp 28-32.

Savage, S. L., Kirmse, M. 2014. "Probability Management 2.0". *ORMS Today* vol. 41: pp 30-33.

Savage, S. L., Scholtes, S., Zweidler, D. 2006. "Probability Management". *ORMS Today* vol. 33: pp 22-28.

Thibault, J. M. 2014. "Standard Specification for Stochastic Information Packets (SIPs) and Stochastic Library Units with Relationships Preserved (SLURPs) Version 2.0" Last modified October 2014. http://probabilitymanagement.org/library/SIP-Standard-Version2.pdf.

AUTHOR BIOGRAPHIES

SAM L. SAVAGE is the author of *The Flaw of Averages: Why We Underestimate Risk in the Face of Uncertainty*. Dr. Savage holds a PhD in the area of Computational Complexity from Yale University. He is the inventor of the Stochastic Information Packet (SIP), an open data structure for performing stochastic roll-ups, a Consulting Professor at Stanford University, and a Fellow of the Judge Business School at Cambridge University. He is also Executive Director of ProbabilityManagement.org, a 501(c)(3) nonprofit that maintains the open SIPmath™ standard. His email address is savage@stanford.edu

JOHN MARC THIBAULT is an independent consultant with a twenty-year practice focused on technical analysis, design and planning. His clients have included a number of the Canadian federal government's departments and a variety of high-tech companies. His earlier experience includes over a decade of marketing and technology roles at Xerox, and senior management in two high-tech startups. He has a physics degree from Loyola College in Montreal. He is the lead architect of the open SIPmath™ 2.0 Standard for conveying realizations of stochastic processes. His email address is marc@probabilitymanagement.org