

# Stochastic simulation

---

Stochastic simulation is a simulation that operates with variables that can change with certain probability. Stochastic means that particular factors (values) are variable or random.<sup>[1]</sup>

With stochastic model we create a projection which is based on a set of random values. Outputs are recorded and the projection is repeated with a new set of random (variable) values. Previous steps are repeated until reasonable amount of data is gathered (thousandfold, millionfold, ..). In the end, the distribution of the outputs shows the most probable estimates as well as a frame of expectations (fringe values dividing those we still can expect from the ones we should not).

## Stochastic

Stochastic means "pertaining to conjecture"; from Greek *stokhastikos* "able to guess, conjecturing"; from *stokhazesthai* "guess"; from *stokhos* "a guess, aim, target, mark". The sense of "randomly determined" was first recorded in 1934, from German *Stochastik*.<sup>[2]</sup>

## Discrete-event simulation

In order to determine the next event in a stochastic simulation, the rates of all possible changes to the state of the model are computed, and then ordered in an array. Next, the cumulative sum of the array is taken, and the final cell contains the number  $R$ , where  $R$  is the total event rate. This cumulative array is now a discrete cumulative distribution, and can be used to choose the next event by picking a random number  $z \sim U(0, R)$  and choosing the first event, such that  $z$  is less than the rate associated with that event.

## Probability distributions

A probability distribution is used to describe the potential outcome of a random variable.

Limits the outcomes where the variable can only take on discrete values.<sup>[3]</sup>

### Bernoulli Distribution

*Main article Bernoulli distribution*

A random variable  $X$  is Bernoulli-distributed with parameter  $p$  if it has only two possible outcomes, usually encoded as 1 (success or default) or 0 (failure or survival).

Example: Toss of coin<sup>[4]</sup>

Define

```
X = 1 if head comes up and  
X = 0 if tail comes up
```

Both realizations are equally likely:

```
P(X = 1) = P(X = 0) = 1/2
```

Of course, the two outcomes may not be equally likely (e.g. success of medical treatment).

---

## Binomial Distribution

*Main article Binomial distribution*

A binomial distributed random variable  $Y$  with parameters  $n$  and  $p$  is obtained as the sum of  $n$  independent and identically Bernoulli-distributed random variables  $X_1, X_2, \dots, X_n$

Example: A coin is tossed three times. Find the probability of getting exactly two heads. This problem can be solved by looking that the sample space. There are three ways to get two heads.

HHH, **HHT**, **HTH**, **THH**, TTH, THT, HTT, TTT

The answer is  $3/8$  ( $= 0.375$ ).<sup>[5]</sup>

## Poisson Distribution

*Main article Poisson Distribution*

The Poisson Distribution depends on only one parameter,  $\lambda$ , and can be interpreted as an approximation to the binomial distribution when the parameter  $p$  is a small number. A poisson-distributed random variable is usually used to describe the random number of events occurring over a certain time interval.

Typical exemple problem: If 3% of the electric bulbs manufactured by a company are defective find the probability that in a sample of 100 bulbs exactly 5 bulbs are defective. ( Given  $e^{-0.25} = 0.7788$  )<sup>[6]</sup>

## Methods

### Direct and first reaction methods

Published by Dan Gillespie in 1977, and is a linear search on the cumulative array. See Gillespie algorithm.

Gillespie's Stochastic Simulation Algorithm (SSA) is essentially an exact procedure for numerically simulating the time evolution of a well-stirred chemically reacting system by taking proper account of the randomness inherent in such a system.<sup>[7]</sup>

It is rigorously based on the same microphysical premise that underlies the chemical master equation and gives a more realistic representation of a system's evolution than the deterministic reaction rate equation (RRE) represented mathematically by ODEs.

As with the chemical master equation, the SSA converges, in the limit of large numbers of reactants, to the same solution as the law of mass action.

### Next reaction method

Published 2000. This is an improvement over the first reaction method where the unused reaction times are reused. To make the sampling of reactions more efficient, an indexed priority queue is used to store the reaction times. On the other hand, to make the recomputation of propensities more efficient, a dependency graph is used. This dependency graph tells which reaction propensities to update after a particular reaction has fired.

### Optimised and sorting direct methods

Published 2004 and 2005. These methods sort the cumulative array to reduce the average search depth of the algorithm. The former runs a presimulation to estimate the firing frequency of reactions, whereas the latter sorts the cumulative array on-the-fly.

### Logarithmic direct method

Published in 2006. This is a binary search on the cumulative array, thus reducing the worst-case time complexity of reaction sampling to  $O(\log M)$ .

### Partial-propensity methods

Published in 2009, 2010, and 2011 (Ramaswamy 2009, 2010, 2011). Use factored-out, partial reaction propensities to reduce the computational cost to scale with the number of species in the network, rather than the (larger) number of reactions. Four variants exist:

- PDM, the partial-propensity direct method. Has a computational cost that scales linearly with the number of different species in the reaction network, independent of the coupling class of the network (Ramaswamy 2009).
- SPDM, the sorting partial-propensity direct method. Uses dynamic bubble sort to reduce the pre-factor of the computational cost in multi-scale reaction networks where the reaction rates span several orders of magnitude (Ramaswamy 2009).
- PSSA-CR, the partial-propensity SSA with composition-rejection sampling. Reduces the computational cost to constant time (i.e., independent of network size) for weakly coupled networks (Ramaswamy 2010) using composition-rejection sampling (Slepoy 2008).
- dPDM, the delay partial-propensity direct method. Extends PDM to reaction networks that incur time delays (Ramaswamy 2011) by providing a partial-propensity variant of the delay-SSA method (Bratsun 2005, Cai 2007).

The use of partial-propensity methods is limited to elementary chemical reactions, i.e., reactions with at most two different reactants. Every non-elementary chemical reaction can be equivalently decomposed into a set of elementary ones, at the expense of a linear (in the order of the reaction) increase in network size.

## Continuous simulation

While in discrete state space it is clearly distinguished between particular states (values) in continuous space it is not possible due to certain continuity. The system usually change over time, variables of the model, then change continuously as well. Continuous simulation thereby simulates the system over time, given differential equations determining the rates of change of state variables. <sup>[8]</sup> Example of continuous system is *the predator/prey model* <sup>[9]</sup> or cart-pole balancing <sup>[10]</sup>

## Probability distributions

### The Normal Distribution

*Main article Normal Distribution*

The random variable  $X$  is said to be normally distributed with parameters  $\mu$  and  $\sigma$ , abbreviated by  $X \in N(\mu, \sigma^2)$ , if the density of the random variable is given by the formula  $f_X(x) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(x-\mu)^2}{2\sigma^2}}$ ,  $x \in \mathbb{R}$ .

Many things actually are normally distributed, or very close to it. For example, height and intelligence are approximately normally distributed; measurement errors also often have a normal distribution. <sup>[11]</sup>

## Exponential Distribution

*Main article Exponential distribution*

Exponential distribution describes the time between events in a Poisson process, i.e. a process in which events occur continuously and independently at a constant average rate.

The exponential distribution is popular, for example, in queuing theory when we want to model the time we have to wait until a certain event takes place. Examples include the time until the next client enters the store, the time until a certain company defaults or the time until some machine has a defect.

## Student's t-distribution

*Main article Student's t-distribution*

Student's t-distribution are used in finance as probabilistic models of assets returns. The density function of the

t-distribution is given by the following equation:  $f(t) = \frac{\Gamma(\frac{\nu+1}{2})}{\sqrt{\nu\pi} \Gamma(\frac{\nu}{2})} \left(1 + \frac{t^2}{\nu}\right)^{-\frac{\nu+1}{2}}$ ,

where  $\nu$  is the number of *degrees of freedom* and  $\Gamma$  is the gamma function.

For large values of  $n$ , the t-distribution doesn't significantly differ from a standard normal distribution. Usually, for values  $n > 30$ , the t-distribution is considered as equal to the standard normal distribution.

## Other distributions

- Extreme Value Distribution
- Generalized Extreme Value Distribution

## Methods

### $\tau$ leaping method

Since the SSA method keeps track of each transition, it would be impractical to implement for certain applications due to high time complexity. Gallespie proposed a solution in approximation procedure, the tau-leaping method which decreases computational time with the minimal loss of accuracy.<sup>[12]</sup> Instead of taking incremental steps in time, keeping track of  $X(t)$  at each time step as in the SSA method, the tau-leaping method leaps from one subinterval to the next, approximating how many transitions take place during a given subinterval. It is assumed that the value of the leap,  $\tau$ , is small enough that there is no significant change in the value of the transition rates along the subinterval  $[t, t + \tau]$ . This condition is known as the leap condition. The tau-leaping method thus has the advantage of simulating many transitions in one leap while not losing significant accuracy, resulting in a speed up in computational time.<sup>[13]</sup>

## Combined simulation

It is often possible to model one and the same system by use of completely different world views. Discrete event simulation of a problem as well as continuous event simulation of it (continuous simulation with the discrete events that disrupt the continuous flow) may lead eventually to the same answers. Sometimes however, the techniques can answer different questions about a system. If we necessarily need to answer all the questions, or if we don't know what purposes is the model going to be used for, it is convenient to apply combined continuous/discrete methodology.<sup>[14]</sup>

## Monte Carlo Simulation

Monte Carlo is an estimation procedure. The main idea is that if it is necessary to know the average value of some random variable and its distribution can not be stated, and if it is possible to take samples from the distribution, we can estimate it by taking the samples, independently, and averaging them. If there are sufficiently enough samples, then the law of large numbers says the average must be close to the true value. The central limit theorem says that the average has a Gaussian distribution around the true value.<sup>[15]</sup>

Simple example: We need to measure area of a shape with a complicated, irregular outline. The Monte Carlo approach is to draw a square around the shape and measure the square. Then we throw darts into the square, as uniformly as possible. The fraction of darts falling on the shape gives the ratio of the area of the shape to the area of the square. In fact, it is possible to cast almost any integral problem, or any averaging problem, into this form. It is necessary to have a good way to tell if you're inside the outline, and a good way to figure out how many darts to throw. Last but not least, we need to throw the darts uniformly, i.e., a good random number generator.

## Application

There are wide possibilities for use of Monte Carlo Method :

- Statistic experiment using generation of random variables (e.g. dice)
- sampling method
- Mathematics (e.g. numerical integration, multiple integrals)
- Reliability Engineering
- Project Management (SixSigma)
- Experimental particle physics
- Simulations
- Risk Measurement/Risk Management (e.g. Portfolio value estimation)
- Economy (e.g. finding the best fitting Demand)
- Process Simulation
- Operation Research

## Random Number Generators

*main article: Random Number Generation*

For simulation experiments (including Monte Carlo) it is necessary to generate random numbers (as values of variables). The problem is, that the computer is highly deterministic machine - basically, behind each process there is always an algorithm, deterministic computation changing inputs to outputs, therefore it is not easy to generate uniformly spread random numbers over a defined interval or set.

Random number generator is a device capable of producing a sequence of numbers which can not be "easily" identified with deterministic properties. This sequence is then called *Sequence of stochastic numbers*.<sup>[16]</sup>

The algorithms typically rely on pseudo random numbers, computer generated numbers mimicking true random numbers, to generate a realization, one possible outcome of a process.<sup>[17]</sup>

Methods for obtaining random numbers exist for a long time and are used in many different fields (like gaming). However, these numbers suffer from certain bias. Currently the best methods, expected to produce truly random sequences are natural methods that take advantage of the random nature of quantum phenomena.

## References

- [1] DLOUHÝ, M.; FÁBRY, J.; KUNCOVÁ, M. Simulace pro ekonomy. Praha : VŠE, 2005.
- [2] stochastic. (n.d.). Online Etymology Dictionary. Retrieved January 23, 2014, from Dictionary.com website: <http://dictionary.reference.com/browse/stochastic>
- [3] Rachev, Svetlozar T. Stoyanov, Stoyan V. Fabozzi, Frank J., "Chapter 1 Concepts of Probability" in *Advanced Stochastic Models, Risk Assessment, and Portfolio Optimization : The Ideal Risk, Uncertainty, and Performance Measures*, Hoboken, NJ, USA: Wiley, 2008
- [4] Bernoulli Distribution, The University of Chicago - Department of Statistics, [online] available at <http://galton.uchicago.edu/~eichler/stat22000/Handouts/112.pdf>
- [5] <http://www.elderlab.yorku.ca/~aaron/Stats2022/BinomialDistribution.htm>
- [6] <http://ncalculators.com/math-worksheets/poisson-distribution-example.htm>
- [7] Stephen Gilmore, *An Introduction to Stochastic Simulation - Stochastic Simulation Algorithms*, University of Edinburgh, [online] available at <http://www.doc.ic.ac.uk/~jb/conferences/pasta2006/slides/stochastic-simulation-introduction.pdf>
- [8] Crespo-Márquez, A., R. R. Usano and R. D. Aznar, 1993, "Continuous and Discrete Simulation in a Production Planning System. A Comparative Study"
- [9] Louis G. Birta, Gilbert Arbez (2007). *Modelling and Simulation*, p. 255. Springer.
- [10] <http://anji.sourceforge.net/polebalance.htm>
- [11] University of Notre Dame, Normal Distribution, [online] available at <http://www3.nd.edu/~rwilliam/stats1/x21.pdf>
- [12] D.T. Gillespie, A General Method for Numerically Simulating the stochastic time evolution of coupled chemical reactions, *The Journal of Computational Physics*, 22 (1976), 403–434.
- [13] H.T. Banks, Anna Broido, Brandi Canter, Kaitlyn Gayvert, Shuhua Hu, Michele Joyner, Kathryn Link, *Simulation Algorithms for Continuous Time Markov Chain Models*, [online] available at <http://www.ncsu.edu/crsc/reports/ftp/pdf/crsc-tr11-17.pdf>
- [14] Francois E. Cellier, *Combined Continuous/Discrete Simulation Applications, Techniques, and Tools*
- [15] Cosma Rohilla Shalizi, Monte Carlo, and Other Kinds of Stochastic Simulation, [online] available at <http://vserver1.cscs.lsa.umich.edu/~crshalizi/notebooks/monte-carlo.html>
- [16] Donald E. Knuth, *The Art of Computer Programming, Volume 2: Seminumerical Algorithms - chapitre 3 : Random Numbers* (Addison-Wesley, Boston, 1998).
- [17] Andreas Hellander, *Stochastic Simulation and Monte Carlo Methods*, [online] available at <http://www.it.uu.se/edu/course/homepage/bervet2/MCkompendum/mc.pdf>
- (Slepoy 2008): Slepoy, A; Thompson, AP; Plimpton, SJ (2008). "A constant-time kinetic Monte Carlo algorithm for simulation of large biochemical reaction networks". *Journal of Chemical Physics* **128** (20): 205101. doi: 10.1063/1.2919546 (<http://dx.doi.org/10.1063/1.2919546>). PMID 18513044 (<http://www.ncbi.nlm.nih.gov/pubmed/18513044>).
  - (Bratsun 2005): D. Bratsun, D. Volfson, J. Hasty, L. Tsimring, (2005). "Delay-induced stochastic oscillations in gene regulation" (<http://www.ncbi.nlm.nih.gov/pmc/articles/PMC1253555>). *PNAS* **102** (41): 14593–8. doi: 10.1073/pnas.0503858102 (<http://dx.doi.org/10.1073/pnas.0503858102>). PMC 1253555 (<http://www.ncbi.nlm.nih.gov/pmc/articles/PMC1253555>). PMID 16199522 (<http://www.ncbi.nlm.nih.gov/pubmed/16199522>).
  - (Cai 2007): X. Cai, (2007). "Exact stochastic simulation of coupled chemical reactions with delays". *J. Chem. Phys.* **126** (12): 124108. doi: 10.1063/1.2710253 (<http://dx.doi.org/10.1063/1.2710253>). PMID 17411109 (<http://www.ncbi.nlm.nih.gov/pubmed/17411109>).
  - Hartmann, A.K. (2009). *Practical Guide to Computer Simulations* (<http://www.worldscibooks.com/physics/6988.html>). World Scientific. ISBN 978-981-283-415-7.
  - Press, WH; Teukolsky, SA; Vetterling, WT; Flannery, BP (2007). "Section 17.7. Stochastic Simulation of Chemical Reaction Networks" (<http://apps.nrbook.com/empanel/index.html#pg=946>). *Numerical Recipes: The Art of Scientific Computing* (3rd ed.). New York: Cambridge University Press. ISBN 978-0-521-88068-8.
  - (Ramaswamy 2009): R. Ramaswamy, N. Gonzalez-Segredo, I. F. Sbalzarini, (2009). "A new class of highly efficient exact stochastic simulation algorithms for chemical reaction networks". *J. Chem. Phys.* **130** (24): 244104. doi: 10.1063/1.3154624 (<http://dx.doi.org/10.1063/1.3154624>). PMID 19566139 (<http://www.ncbi.nlm.nih.gov/pubmed/19566139>).
  - (Ramaswamy 2010): R. Ramaswamy, I. F. Sbalzarini, (2010). "A partial-propensity variant of the composition-rejection stochastic simulation algorithm for chemical reaction networks". *J. Chem. Phys.* **132** (4): 044102. doi: 10.1063/1.3297948 (<http://dx.doi.org/10.1063/1.3297948>). PMID 20113014 (<http://www>).

ncbi.nlm.nih.gov/pubmed/20113014).

- (Ramaswamy 2011): R. Ramaswamy, I. F. Sbalzarini, (2011). "A partial-propensity formulation of the stochastic simulation algorithm for chemical reaction networks with delays". *J. Chem. Phys.* **134** (1): 014106. doi: 10.1063/1.3521496 (<http://dx.doi.org/10.1063/1.3521496>). PMID 21218996 (<http://www.ncbi.nlm.nih.gov/pubmed/21218996>).

## External links

### Software

- ResAssure (<http://www.resassure.com>) - Stochastic reservoir simulation software - solves fully implicit, dynamic three-phase fluid flow equations for every geological realisation.
- Cain (<http://cain.sourceforge.net/>) - Stochastic simulation of chemical kinetics. Direct, next reaction, tau-leaping, hybrid, etc.
- pSSALib ([http://mosaic.mpi-cbg.de/?q=downloads/stochastic\\_chemical\\_net](http://mosaic.mpi-cbg.de/?q=downloads/stochastic_chemical_net)) - C++ implementations of all partial-propensity methods.
- StochPy (<http://stochpy.sourceforge.net>) - Stochastic modelling in Python
- STEPS (<http://steps.sourceforge.net/STEPS/Home.html>) - STochastic Engine for Pathway Simulation using swig to create Python interface to C/C++ code

# Article Sources and Contributors

**Stochastic simulation** *Source:* <https://en.wikipedia.org/w/index.php?oldid=592302514> *Contributors:* Chris the speller, Compsim, Franco3450, Ivos99, Jezmpez, Kku, Lgallindo, Mange01, Melcombe, Michael Hardy, Mv240, Ricardohz, SciCompTeacher, SeanMauch, SpaceLem, Stochasticsimulation, Tmd200, Vojtech.zrust, 9 anonymous edits

## License

---

Creative Commons Attribution-Share Alike 3.0  
[//creativecommons.org/licenses/by-sa/3.0/](https://creativecommons.org/licenses/by-sa/3.0/)

---