Analytic Uncertainty Modeling Versus Discrete Event Simulation

Henry Neimeier
The MITRE Corporation
7525 Colshire Drive
McLean, Virginia, 22102, USA

The analytic uncertainty modeling technique is useful whenever sensitivity analysis is important. It provides the entire resulting probability distribution instead of a single uncertain point estimate of the mean. Both analytic development costs, and computer execution costs are far less than in discrete event simulation. The price paid is some lack in modeling flexibility.

Discrete simulation requires multiple long simulation runs to obtain a statistically significant point estimate. The different result values from multiple runs with identical parameter values but different random number seeds, are averaged to obtain the point estimate of the mean result value. Conversely, the analytic solution gives the entire resulting probability distribution with minimal calculation. The analytic solution also considerably simplifies sensitivity analysis. A single analytic run is done for each input parameter setting. Discrete event simulation requires multiple runs for each input parameter, to obtain a statistically significant mean result.

Analytic Uncertainty Modeling Versus Discrete Event Simulation

Overview

The beta distribution analytic technique outlined in this paper is useful whenever sensitivity analysis is important. It provides the entire resulting probability distribution vice a single uncertain point estimate of the mean. Both analytic development costs, and computer execution costs are far less than in discrete event simulation. The price paid is some lack in modeling flexibility.

With the appropriate choice of parameter values, the beta distribution closely fits all the classical probability distributions. The sums and products of beta variates are also approximately beta distributed (see Neimeier reference for error bounds). The beta distribution can be fit based on the minimum, mean, maximum, and standard deviation statistics. In a complex results calculation all that is required is to keep track of these statistics as the calculation proceeds. At any point in a calculation, the probability distribution of the result, can be derived by fitting a beta distribution based on the four statistics.

Problems With Discrete Event Simulation

Discrete simulation requires multiple long simulation runs to obtain a statistically significant point estimate. The different result values from multiple runs with identical parameter values but different random number seeds, are averaged to obtain the point estimate of the mean result value. Conversely, the analytic solution gives the entire resulting probability distribution with minimal calculation. The analytic solution also considerably simplifies sensitivity analysis. A single analytic run is done for each parameter setting vice multiple runs for a statistically significant result. The simulation time (T) required to be 95 percent confident in a relative error (ε) is approximated by the following equation for open GI/G/1 (general independent inter arrival times, general service times, 1 server) queuing networks:

$$T = 8\,\tau\,(C_a^2 + C_s^2)\,Z^2/(\rho^2 (1-\rho)^2\,\varepsilon^2)$$

Where:

- T = simulation time for a specified relative error
- τ = service time
- C_a^2 = square coefficient of variation in inter arrival time (variance in inter arrival time divided by the mean inter arrival time squared)
- C_s^2 = square coefficient of variation in service times
- Z = unit normal deviate (Z=2 for 95 percent confidence
- ρ = utilization (service time divided by inter arrival time)
- ε = tolerated relative error

Figure 1 is a semi-log plot of simulation time required for 95 percent confidence in a specified relative error as a function of utilization. It represents the exponential inter arrival and service time case (Ca=Cs=1). Note that at high utilization and low relative errors extremely long simulation times are required. To achieve 5 percent relative error in the mean on an 80 percent utilized queuing network requires one million service times.

In functional economic analysis we are interested in the relative future costs of alternative systems. There are uncertainties in process performance, resource requirements, cost estimates, investment required, workload, interest and inflation rates. There is also uncertainty in the future projection of these elements. Thus there is uncertainty in the discounted present value cost distribution for each alternative system. A plot of cumulative probability versus cost, aids the decision process. The entire range in cost distribution is of interest. Figure 2 shows the expected number of simulation events required to obtain an event in the tail of the result distribution when using discrete event simulation. The equation plotted is:

$$E = 1 / P^{C}$$ Where:

E = expected number of simulation events
P = distribution tail probability
C = uncertain model components

The lower the tail probability, and the more components in the model, the more events are required. For example, an average of one million simulation events are required in a six component model to simultaneously be in the 10 percent tail of all component distributions. To simultaneously be in the 1 percent tail requires an average of one trillion simulation events. In the limit it requires an infinite number of simulation events to capture the entire range of results. Thus except for very small models with few components, discrete event simulation is not a practical method for generating the entire result distribution. If the minimum and maximum distribution values are not needed then discrete event simulation is practical. However, even in this case the model development, execution, and sensitivity analysis costs are higher.

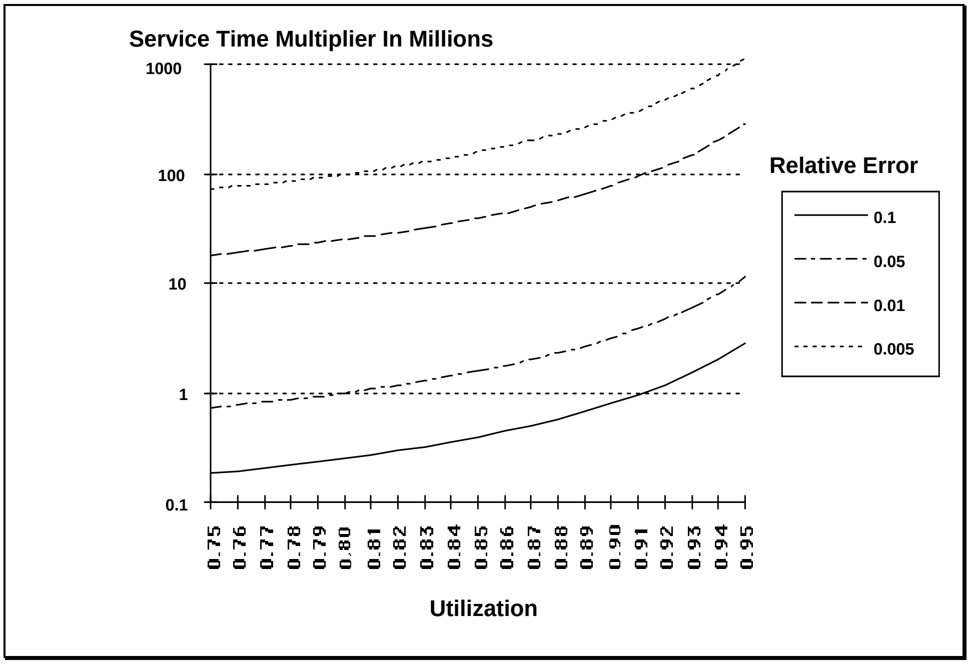

Figure 1. Simulation Time For Specified Relative Error

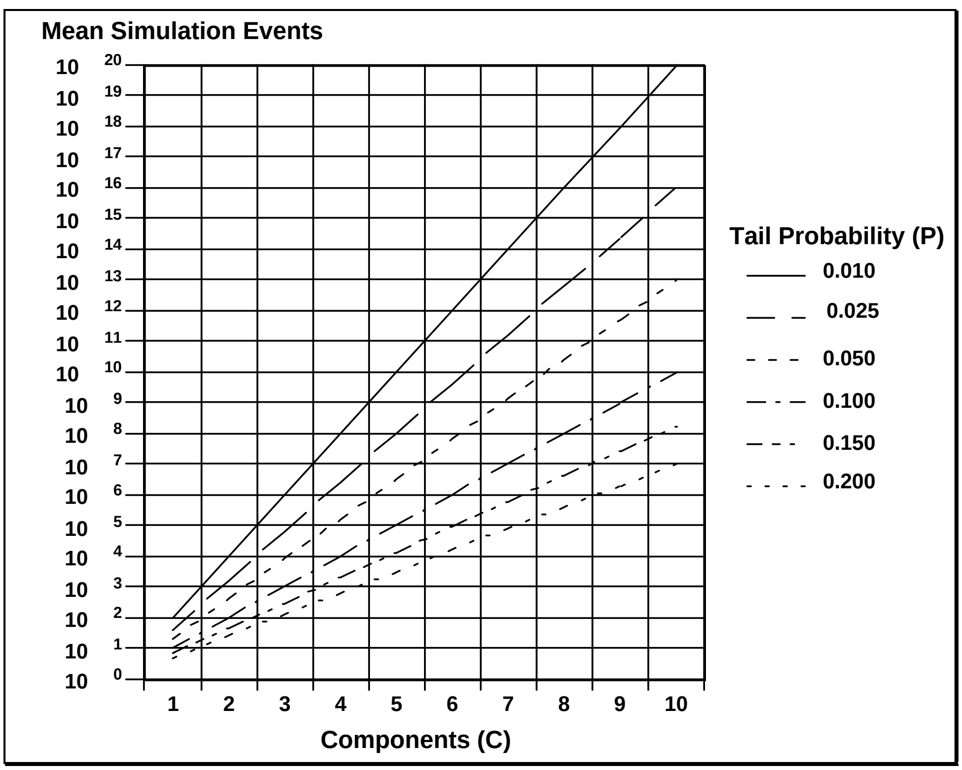

Figure 2. Simulation Events For A Specified Tail Probability Versus Model Components ($1/P^C$)

Beta distribution

The beta density distribution is bounded by high (max) and low (min) values.

$$P = C \{(x-min)/(max-min)\}^{a-1} \{1-(x-min)/(max-min)\}^{b-1}$$

Where:

P is the probability density

C normalizes the area under the distribution to unity

$$(C = 1/\int_0^1 x^{a-1} (1-x)^{b-1} dx)$$

min is the minimum variable value

max is the maximum variable value

If a random variable is bounded then its distribution function is uniquely determined by its moments (Wilks, p 126). The beta distribution can be fit based on the minimum, maximum, mean and variance. The fit parameters a and b are based on a range variable r and a skewness variable s, and are defined in standard terms as:

$$r = \text{variance} / (max - min)^2$$

$$s = (\text{mean} - min) / (max - min)$$

$$a = (s^2 (1 - s) / r) - s$$

$$b = (s (1 - s) / r) - 1 - a$$

If (a-1)(b-1)>0 then the beta distribution has a mode (a>1, b>1) or a anti-mode (a<1,b<1) at

min + (max - min) (a - 1) / (a + b - 2)

Figure 3 shows examples of different beta distribution shapes. For the mean midway between the minimum and maximum values, the distribution is symmetric with equal a and b parameter values. Unity a and b parameters yield a uniform distribution (1,1). Values less than 1 lead to a "U" shaped distribution (.5,.5). With a and b equal 2 (2,2) the distribution has a parabolic shape. At higher a and b values (6,6; 12,12) the distribution has a shape similar to a normal distribution. The lower the standard deviation relative to the range, the higher the a and b parameter values, and the more peaked the distribution shape. If a is greater than b then the distribution has a negative skew (9,2). Conversely, if a is less than b the distribution has a positive skew. Triangular distributions (1,3) and left (.5,3) and right "J" shaped distribution shapes are also possible.

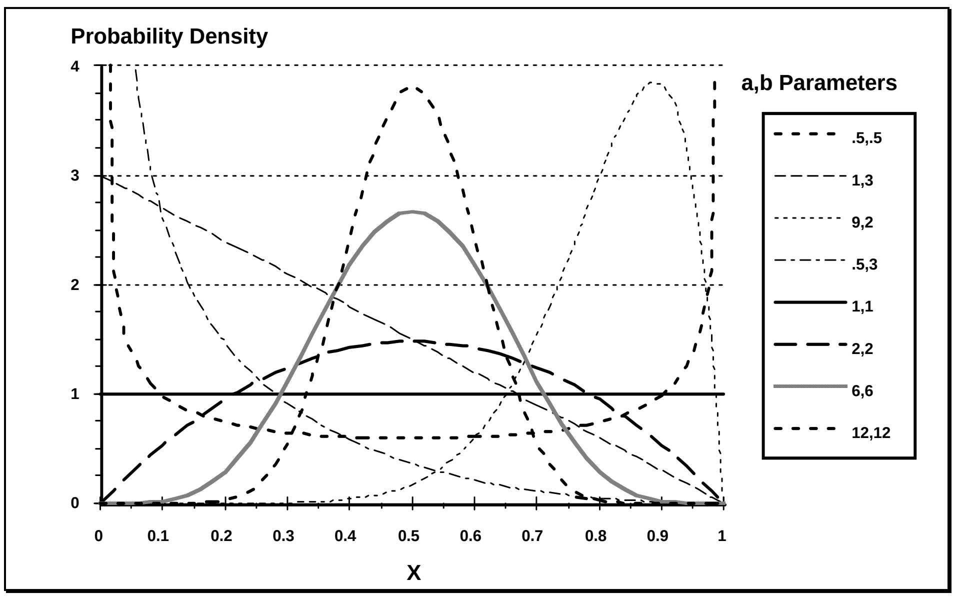

Figure 3. Beta Distribution Shapes

Spanning the classical distribution space

Figure 4 is a skew ($\sqrt{\beta 1}$) kurtosis ($\beta 2$) plot of the popular classical continuous statistical distributions. The beta distribution covers most of the classical distribution area except the log normal line. In the case of the log normal line a log transformation is performed before the beta distribution is fit. The beta distribution shapes are shown on the chart regions. Beta distributions can fit uniform, triangular, J shaped, U shaped, exponential, Erlang, hyper-exponential, gamma, and normal distributions.

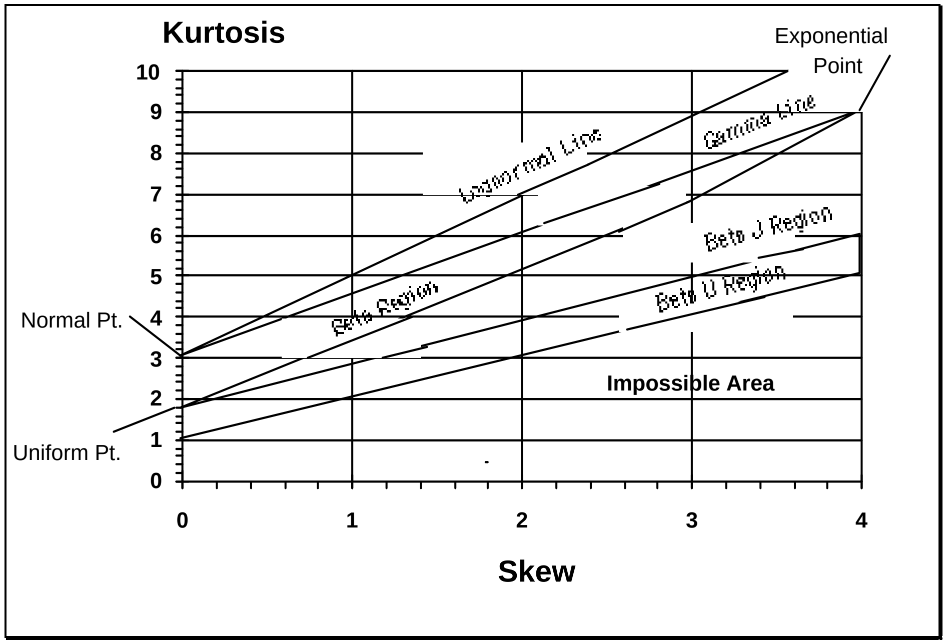

Figure 4. Skew Kurtosis Plot of Classical Distributions

Mean and variance statistics

In practice any calculation with beta variates usually yields approximate beta variates. Thus if one encodes parameter uncertainty with a beta distribution, the results of a calculation with many uncertain parameters will also be approximately beta distributed. This allows analytic solution for the resulting distribution without resorting to discrete simulation. This greatly reduces the calculation requirement and simplifies parametric sensitivity analysis.

In a process cost or performance calculation, operations must be performed on uncertain parameter values to determine expected results and uncertainty in results. Operations include: addition, subtraction, multiplication, power, polynomial function, general function. Using the Beta distribution we must keep track of the calculation minimum, maximum, mean (u) and variance (σ^2) values. The variance of the sum or difference of two independent parameters is just the sum of their component variances. If the parameters are correlated the following equation is used:

$$\sigma^2_{1+2} = \sigma^2_1 + \sigma^2_2 + 2\,\sigma_{12}\;; \quad \sigma^2_{1-2} = \sigma^2_1 + \sigma^2_2 - 2\sigma_{12}$$

Where σ_{12} is the covariance of parameters 1 and 2. The variance of the product of two uncertain independent parameters 1 and 2 is given by the following formula:

$$\sigma_{12}^2 = \sigma_1^2\sigma_2^2 + \mu_1^2\,\sigma_2^2 + \sigma_2^2\,\sigma_1^2$$

The expected value of the product for correlated variables is:

$$\mu_1\,\mu_2 + \rho\,\sigma_1\,\sigma_2$$

If parameter a is a constant with value a (minimum = mean = maximum = a, and variance = 0) then the variance of the product of parameters 1 and 2 is:

$$\sigma^2_{12} = a^2\,\sigma^2_2$$

In the case of division, the maximum result is the maximum quotient divided by the minimum divisor. Conversely the minimum result is the minimum quotient divided by the maximum divisor. The variance of the result and the corrected mean are obtained from:

$$\sigma_{1/2}{}^2 = (\mu_1/\mu_2)^2 \{(\sigma_1/\mu_1)^2 + (\sigma_2/\mu_2)^2 - 2\rho(\sigma_1/\mu_1)(\sigma_2/\mu_2)\}$$

$$\rho = \sigma_{12}/(\sigma_1\sigma_2); \qquad \mu_{1/2} = (\mu_1/\mu_2)\,(1+(\sigma_2/\mu_2)^2 + \rho(\sigma_1/\mu_1)(\sigma_2/\mu_2)$$

If y is a linear function of n input variables ($y=a_0+a_1x_1+a_2x_2+...+a_nx_n$) then the variance of y is given by:

$$\sum_{i=1}^{n} (\partial f/\partial x_i)^2 \sigma_i^2 + 2 \sum_{i=1}^{n} \sum_{j=i+1}^{n} (\partial f/\partial x_i)(\partial f/\partial x_j)\, \sigma_{ij}{}^2$$

If the input variables are independent the second term is ignored. If y is not a linear function of the input variables the above equation is only an approximation. In the case of a general single variable function ($y=f(x)$) with known derivatives, the variance of the result is approximated by the following:

$$(\partial y/\partial x)^2 \sigma_x{}^2 + 1/2(\partial^2 y/\partial x_2)^2 \sigma_x{}^4 + (\partial^3 y/\partial x^3)(\partial y/\partial x)\, \sigma_x{}^4 + ...$$

Usually only the first term provides a reasonable approximation. Additional terms can be found in the Tukey or Seiler. Approximations for functions of several variables ($z=f(x,y,...)$) are derived based on the multidimensional Taylor series. Note that the mean y ($y=f(x)$) is not equal to the result of substituting the mean x value into the function, if the function is non-linear. The shift in mean value is given by:

$$1/2\,(\partial^2 y/\partial x^2)\, \sigma_x{}^2 + 1/8\,(\partial^4 y/\partial x^4)\, \sigma_x{}^4 + ...$$

Conclusions

In Cost and Operational Effectiveness Analysis COEA we are interested in the future performance and costs of alternative systems. There are uncertainties in process parameters, cost estimates, and system workload. The analytic uncertainty analysis technique provides a simple way of calculating cost and performance uncertainty distributions from the component uncertainties. It considerably simplifies sensitivity analysis. It should be considered when the probability distribution of a result is desired rather than a single point estimate of the mean. Both analytic development costs, and computer execution costs are far less than in discrete event simulation. The price paid is some lack in modeling flexibility.

References

Whitt, W. 1989.*Planning Queueing Simulations,* Management Science, Vol 35, No.11, November 1989.

Johnson, N., Kotz,S. 1970. *Continuous Univariate Distributions-2,*

John Wiley & Sons, New York.
Wilks,S. 1962 *Mathematical Statistics*, John Wiley & Sons, New York.
Tukey,J. *The Propagation of Errors, Fluctuations, and Tolerances Basic Generalized Formulas,* AD155082, Princeton University, Princeton, New Jersey.
Springer,M.1979. *The Algebra of Random Variables,* John Wiley & Sons, New York.
Seiler,F. 1987. *Error Propagation for Large Errors,* Risk Analysis, Vol.7, No.4,pp 509-518.
Goldman, L. On the Exact Variance of Products, Journal of the American Statistical Association, vol.55, 1960, pp 708-713