Proceedings of the 1994 Winter Simulation Conference
ed. J. D. Tew, S. Manivannan, D. A. Sadowski, and A. F. Seila

UNKNOWN UNKNOWNS: MODELING UNANTICIPATED EVENTS

Lobna A. Okashah

Industrial Engineering and Management Systems
Department
University of Central Florida
P.O. Box 162450
Orlando, Florida 32816-2450, U.S.A.

Paul M. Goldwater

School of Accounting
University of Central Florida
P.O. Box 161400
Orlando, Florida, 32816-1400, U.S.A.

ABSTRACT

In simulations involving uncertainty, two types of unknowns must be taken into account: (1) known unknowns and (2) unknown unknowns. For known unknowns, the nature of the task is known, adequate historical data is available, and although the value of the model variable is unknown, either a theoretical or an empirical probability density function can be established to describe the variable. For unknown unknowns, the value of the variable can be zero, if the task or event does not actually occur, or may go to any amount (either negative or positive) if the event does occur. For example, there may be no definable upper limit if an in-house activity fails catastrophically or a subcontractor fails to deliver the work. These unknowns invariably result in disruptions to operations and significant cost overruns.

In industry today, we are particularly concerned with designing proactive control systems. These "unknown unknowns" therefore cannot be ignored. This paper discusses a methodology to incorporate this second type of unknown into a simulation model. Examples include modeling a forklift-pedestrian collision, a labor strike at a critical supplier, and a natural disaster at a factory.

1 INTRODUCTION

The concept of "unknown unknowns" has not been widely researched, especially in simulation. Section 2 discusses work that has been done to date in modeling and estimating unknown factors. Section 3 draws a distinction between Internal Unknown Unknowns (IUUs), which should be the product of--rather than an input to--a simulation, and External Unknown Unknowns (EUUs), which are occurrences outside the bounds of a model that nonetheless impact the model. Section 4 presents three examples of unknown unknowns, including a discussion of how they can be incorporated into simulation models. Finally, Section 5 offers conclusions.

2 LITERATURE SEARCH

Unknown unknowns in simulation are generally absent from current literature. Most references to unknown parameters come from control applications, but simulation and control are inherently different: in control, one generally desires to eliminate or at least minimize the effects of unknowns. In simulation, the goal is to preserve the integrity of the unknowns by realistically modeling them. Where references to unknown variables are found, work typically focuses on estimating the value of the unknown or on estimating limits of the value of the unknown.

2.1 Simulation Applications

Mo, Hegge and Wangensteen (1990), in a model to optimize the selection of power generation options, characterize the unknowns of their problem as fuel prices, energy demand, water inflow for hydropower, and investment costs. Simulation is presented as one possible solution, in which "different realizations of the uncertain variables" are input to simulation runs and "the final investment decisions are made by means of decision analysis."

This model has no unknown unknowns because all variables addressed by the model are defined in advance. Only the values of the variables are unknown. There is no way to predict accurately the behavior of several of the variables, specifically the price of fuel and water inflow, over the long life of a power plant. Other variables, such as energy demand, can be at least loosely modeled in accordance with demographic trends and other assumptions. The ability to know, however, that all of these factors are present and will influence the investment decision makes them "known unknowns."

Most simulations deal with known unknowns. In the best case, not only is the presence of the variable and its impact on the system known and understood, but the occurrence distribution is also known allowing an accurate distribution to be modeled. In the worst case, the presence of the variable is known but its impact on the system is based on assumptions (not observations) and its occurrence is based on an assumed distribution as well. In this case, estimation algorithms are of use, as discussed in Section 2.3.

2.2 Control Applications

There are many references in the literature to unknown factors in control problems. There are several foci of research in this field. One approach is to develop techniques to estimate the unknown, thus turning it into a known factor (although with imperfect precision). Some researchers concentrate on estimating the value of the unknown factor; others instead try to estimate its bounds. One method of estimation for control applications is Delay Time Controller (DTC), in which an estimate of uncertainty is made by evaluating the time derivative of the system state during a time delay (Chen, 1990). This technique requires that the system uncertainties vary continuously during the delay interval.

2.3 Estimation Algorithms

Assessing the probabilities associated with an unknown is a challenge when there is little or no prior data or knowledge about the random variables associated with the incident. Nonetheless, there are methods suggested in the literature by, among others, Law and Kelton (1991), Bratley and Fox (1987), and Keefer and Bodily (1983).

2.3.1 Law and Kelton

Law and Kelton define two approaches. The first, called the triangular approach, requires subjective estimates of optimistic, pessimistic and most likely estimates of the unknown phenomena under consideration. These estimates should come from experts in the field of the unknown phenomena (such as oil traders and economists for the unknown price of oil in the power plant example of Section 2.1). The distribution of the unknown is assumed to lie within the interval $[a,b]$, where $b>a$; a is the value of the pessimistic estimate and b is the value of the optimistic estimate. The estimates are used to generate a triangular random variate that can be used in simulation (such as a Monte Carlo simulation). The results obtained from running the simulation represent the results of subjecting the model to such estimate of the unknown.

The second approach assumes that the unknown has a Beta distribution on the interval $[a,b]$ with shape parameters α_1 and α_2. Since the Beta distribution assumes many shapes according to the values of shape parameters α_1 and α_2, this allows more modeling flexibility. The authors choose two methods to find appropriate values for α_1 and α_2. One method sets α_1 and α_2 to equal 1, which results in a Uniform distribution with (a,b) parameters. The other method assumes that the density function of the unknown is skewed to the right with $\alpha_2 > \alpha_1 > 1$; this method is considered to be more realistic by the authors. They provide the following formulas to calculate the mean μ and the mode c:

$$\mu = a + \alpha_1 (b - a)/ \alpha_1 + \alpha_2$$

$$c = a + (\alpha_1 - 1)(b - a)/\alpha_1 + \alpha_2 + \alpha_1 - 2$$

The authors also provided formulas to calculate estimates of α_1 and α_2 .

$$\alpha_1 = (\mu - a)(2c - a - b) / (c - \mu)(b-a)$$

and

$$\alpha_2 = (b - \mu)\alpha_1 / (\mu - a)$$

The estimated factors are used to generate a Beta random variate that is then used in the simulation.

2.3.2 Brateley and Fox

Brately and Fox (1987) offer another method that they call a "quasi-empirical distribution." This method depends on the fact that a number of observations are known, but their distribution is not known. An empirical distribution can be used to fit the data, however that distribution is probably a poor fit of the original distribution. The authors' theory is based on fitting an exponential distribution to the end tail of the empirical distribution.

2.3.3 Keefer and Bodily

The previous methods depend on estimates of the interval $[a,b]$. Keefer and Bodily (1983), on the other hand, based their methods on subjective estimates of the 0.05 and 0.95 percentiles. They also evaluate various methods done by other researchers, and document their results in a table. The basis of their evaluation is to estimate means and variances at different percentiles and to evaluate the value of error in the estimates. They then recommend the methods with the lowest errors. The method that they recommend as being the best one uses the following equations to estimate the mean and the variance:

$$\text{mean} = 0.63x(0.05) + 0.185[\, x(0.05) + x(0.95)]$$

$$\text{variance} = ([x(0.95) - x(0.05)]/3.25)^2$$

where $x(0.05)$ and $x(0.95)$ are, respectively, the fifth and ninety-fifth percentile values of the random variable x.

3 INCORPORATION OF UNKNOWN UNKNOWNS IN SIMULATION

Aggregately, unknown unknowns are quite common, but individual, specific unknown unknowns such as a strike, or a hurricane are difficult or impossible to predict. (If specific unknown unknowns were predictable their presence would be expected, making them known unknowns). Unknown unknowns can be viewed in two classes: internal and external. Internal unknown unknowns (IUUs) are those unexpected, usually unexplained things that happen within the confines of the system being modeled. Examples of this are a forklift hitting a wall or a worker, or a robot or machine tool moving wildly out of control. External unknown unknowns (EUUs) are the result of forces external to the system being modeled. A hurricane threatening or destroying a factory, or a strike at a critical supplier are both events that clearly affect a factory's operations but fall outside the factory model.

3.1 Internal Unknown Unknowns (IUUs)

IUUs should be the output of, not the input to, a simulation. This is not possible in traditional simulation, for all events must be programmed into the model, and IUUs by definition are unknown. With autonomy, however, IUUs are not only possible but expected. Because individual entities in the system are modeled, they are free to interact with each other according to their individual logic. For example, assume that two drivers each have the goal of occupying the same parking spot. In traditional simulation, they would be put in a queue. In autonomous simulation, each driver would proceed to the spot and a conflict would be identified. The conflict was allowed to develop by permitting the autonomous behavior of the entities; in the same manner, IUUs can develop. By contrast, in traditional simulation, the entire system is modeled as a coherent unit.

To promote the discovery of IUUs, uncertainty should be introduced into the model. One way of accomplishing this is with confidence factors (Widman and Loparo, 1989). To model noise, unavailable data, incorrect data or inconsistent expert knowledge, a quantifiable "confidence factor" analogous to a probability can be associated with each datum and rule. These confidence factors are combined to yield confidence factors for each intermediate conclusion and, ultimately, for each final action. While the authors note that the mathematical validity of this approach is controversial, it is one method to model uncertainty. Section 4.1 of this answer offers a demonstration of this approach.

3.2 External Unknown Unknowns (EUUs)

EUUs, like IUUs, cannot be known to the modeler. But unlike IUUs, EUUs will not result from exercising an autonomous model. The best way presently known to address the general case of an EUU is to model its effect, not the EUU itself. The effect of a tornado or a flood, for instance, may be to shut a factory down for some amount of time, render some of its equipment temporarily inoperable, and render another portion of its equipment permanently inoperable. This scenario can be modeled in both traditional and enhanced simulation as a random occurrence, with a frequency based upon researched occurrences of disasters. The length of the shutdown and the selection of disabled equipment would be a probabilistic input.

The main problem with this approach is that it is not based upon specific, causal data. There is no way of knowing whether the assumption of downtime and destruction are correct because there is insufficient data upon which to base the simulated distribution (if there were sufficient data, the phenomenon would be a "known unknown"). It is possible to model EUUs more accurately if there is sufficient reason to suspect a specific type of EUU. Sections 4.2 and 4.3 offer two examples of specific EUUs and how they could be modeled.

4 EXAMPLES

The following three examples show how one IUU and two EUUs could be incorporated into simulation models.

4.1 IUU: Discovering Forklift-Pedestrian Collisions in a Factory

An IUU that can be easily imagined is a factory collision of a forklift and a pedestrian. Suppose that the model of the forklift requires its driver, upon seeing a pedestrian obstructing his path, to honk his horn and, if the pedestrian does not move, stop the forklift. Probabilities would be assigned to each of these actions as shown in Table 1. The model of the pedestrian would have corresponding probabilities shown in Table 2.

Table 1: Probabilities Associated with Operator Model

Operator sees pedestrian	0.98
Operator remembers to honk horn	0.99
Horn functions	0.99
Operator remembers to stop	1.00
Forklift correctly stops	0.99

Table 2: Probabilities Associated with Pedestrian Model

Pedestrian sees approaching vehicle	0.70
Pedestrian hears horn	0.95
Pedestrian moves if he is aware of problem	1.00

If the uncertainty model as proposed by Widman and Loparo is used in the model, the probability P of occurrence for the IUU of a forklift striking a pedestrian is

$$P = [1 - (0.98)(0.99)(0.99)(1.00)(0.99)] * (1-0.70)(1-0.95)$$

$$P = 0.00074$$

4.2 EUU: Modeling a Labor Strike at a Critical Supplier

Suppose that when simulating a factory that there is reason to suspect that a particular supplier may experience a labor strike. This could be modeled as described in the following three steps.

1. Model normal operations

The normal simulation model consists of nodes representing arrival of parts and in-factory processes. An excerpt from a full simulation might be as shown in Figure 1.

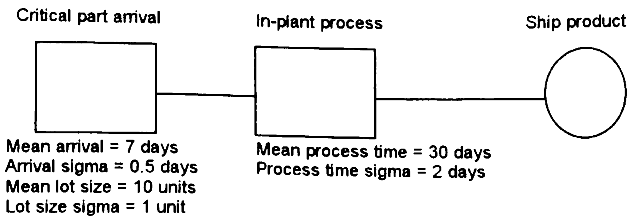

Figure 1: Example Model of Part Arrival

Ignoring capacity, and therefore queuing, for the purpose of this example, execution of the model would yield a distribution of product shipment approximating the following:

Mean ship time = 7 + 30 days
Mean ship time = 37 days

Ship sigma = $(0.5^2 + 2^2)^{1/2}$
Ship sigma = 2.062 days

2. Model a strike generator

To effectively model a strike, a separate event generator would be added to the model, based on historical or expert knowledge. The modeler must know the likelihood of a strike occurring, when it would likely begin, and how long it would likely last.

The likelihood of a strike could be ascertained by interviewing the company's contract negotiators and looking at past responses by the union when issues similar to current issues were at the heart of labor disputes. When a strike would likely begin can also be estimated by interviewing the company's negotiators, and by studying past labor disputes. A Beta-type distribution is one possibility; the exact form would probably be unknown because to gather enough data, one would need to look several decades into the past. Labor relations have changed significantly during this time. A possible distribution is shown in Figure 2.

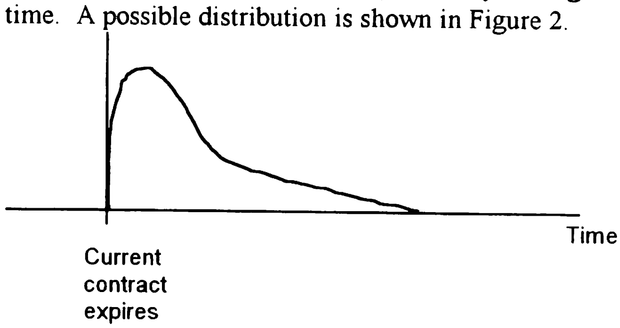

Figure 2: Representative Likelihood of a Strike at a Supplier

The possible duration of a strike can be ascertained by studying the union's history, any information known about its finances, and the company's current financial and market position. For example, public service unions (firefighters, police, trash collectors) typically strike for between a few days and a few weeks. Teacher strikes may last several weeks. Coal miner strikes, on the other hand, are known to last for many months. Companies with unique products or who depend on brand loyalty usually are more reluctant to endure a long strike than those with commodity-type products. Strike longevity would be expressed as an estimated mean and sigma, probably with a normal distribution assumed (unless research points to a different distribution).

From this information, the strike generator element would be programmed into the model. Its function is to generate a message, based upon the above parameters, that a strike will begin at time x and be of duration y.

3. Model strike operations

Under strike conditions, the node in the original (normal) model which represents the arrival of the subcontractor's product will behave in a dramatically different manner than in non-strike conditions. Upon triggering by the strike generator, the model must change the parameter values of that node to reflect strike conditions:

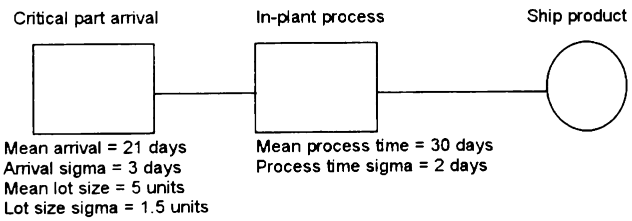

Figure 3: Part Arrival Under Strike Condition

The resulting impact on product shipment, as shown by the simulation, would be approximately as shown below:

Mean ship time = 21 + 30 days
Mean ship time = 51 days

Ship sigma = $(3^2 + 2^2)^{1/2}$
Ship sigma = 3.606 days

After an amount of simulation time equal to the strike duration y from Step 2 has elapsed, the parameters of the "normal" condition should be restored to the "Critical part arrival" node.

4.3 EUU: Modeling a Hurricane Strike of a Factory

There is ample evidence that a major hurricane hits the Southeast Coast of the United States, on average, every few years. If one were modeling a factory located near the coast in this region, the EUU of a major hurricane strike could be modeled as follows:

First, the modeler should obtain from reliable weather historians (for this example, the U.S. National Weather Service) the dates and severity of hurricanes which have hit the plant's location, and the history of hurricane predictions for the location. From this data, an element which generates messages of hurricane warnings and strikes could be modeled. Finally the modeler would include appropriate actions in the other elements of the model. Candidate actions for the supervisor might be:

IF (HURRICANE_WARNING = 2 DAYS)
  Order production control to suspend normal operations and move outdoor stock inside
  Order round-the-clock supervision and maintenance to prepare factory
  Alert employees to emergency notification procedures for plant closure
IF (HURRICANE_WARNING = 1 DAY)
  Order windows boarded
  Prepare machines for long-term shutdown (drain oil, secure moving parts, etc.)
IF (HURRICANE_WARNING = CANCELED)
  Reverse emergency actions taken

Production workers have different responsibilities than their supervisor, and may not necessarily share the supervisor's priorities when faced with an emergency such as a hurricane. Suppose that production workers are modeled as shown below:

IF (HURRICANE_WARNING = 2 DAYS)
  Reduce performance by 0.30
  Reduce attendance by 0.50
IF (HURRICANE_WARNING = 1 DAY)
  Reduce attendance by 0.95

As can be plainly seen from this example, the autonomy of the supervisor and the production worker entities are not synchronized for optimal system performance as many traditional simulation models would assume. The model incorporates the reduced performance and increased absenteeism that would result from workers urgently focusing their attention to protecting their homes and stockpiling emergency supplies, at the expense of their job performance. In this model, the supervisor will have many emergency tasks needing to be done, and few workers available to do them. Execution of the model would show failure of the plant preparation plan; plant management could then take actions to improve emergency performance. Candidate actions might be to implement plant preparation activities earlier and to encourage employees to stockpile emergency supplies at the beginning of each hurricane season.

4 CONCLUSION

It is necessary to model unknown unknowns in simulation, for they represent occurrences of the real world that, although they are rare, often have dramatic and expensive effects. In practice, few models include these phenomena. This area of research is an interesting topic that should be investigated in the future.

REFERENCES

Bratley, P., B. L. Fox, L. E. Schrage. 1987. *A Guide to Simulation.* 2d ed. Springer- Verlag Inc.

Chen, Min-Shin. 1990. Uncertainty Estimator. *1990 American Control Conference*, 2:2020-2024

Keefer, D. L., S. E. Bodily. 1983. Three Point Approximations for Continuous Random Variables. *Management Science* 29: 595 -609.

Law, A. M., W. D. Kelton. 1991. *Simulation Modeling and Analysis.* 2d ed edition. McGraw Hill,Inc.

Mo, B., J. Hegge, I. Wangensteen. 1991. Stochastic Generation Expansion Planning by Means of Stochastic Dynamic Programming. *IEEE Transactions on Power Systems* 6: 662 - 668.

Widman, L., K. A. Loparo. 1989. Artificial Intelligence, Simulation, and Modeling: A Critical Survey. *Artificial Intelligence, Simulation and Modeling..* John Wiley & Sons, Inc.

AUTHOR BIOGRAPHIES

LOBNA A. OKASHAH is pursuing her Ph.D. in Industrial Engineering at the University of Central Florida. Her research interests lie in modeling and simulation, artificial intelligence, information systems, manufacturing systems and engineering economics. She holds an M.S. in industrial engineering and a B.S. in chemical engineering. She is a member of IEEE, ORSA/TIMS, SME, and SCS.

PAUL M. GOLDWATER is an associate professor at the University of Central Florida. He is originally from New Zealand and came to the United states in 1978 to work for Arthur Young and Co. management services. His research interests include accounting education, and computerized accounting systems. He has published articles in Accounting Education, Business planning, Technological Horizons in Education Journal, Journal of accounting and Computers and Journal of Research on Computing in Education, and International Journal of Applied Expert Systems.