UniFit II: TOTAL SUPPORT FOR SIMULATION INPUT MODELING

Stephen Vincent
Averill M. Law

Averill M. Law & Associates
P.O. Box 40996
Tucson, Arizona 85717

ABSTRACT

In this paper, we explain the important role of simulation input modeling in a successful simulation study. Two pitfalls in simulation input modeling are then presented and we explain how *any* analyst, regardless of their knowledge of statistics, can easily avoid these pitfalls through the use of UniFit II. We use a set of real-world system data to demonstrate how the package automatically specifies, evaluates, and ranks candidate probability distributions, and then assists an analyst in deciding whether the "best" candidate probability distribution provides an adequate representation of the data. If no candidate probability distribution provides an adequate fit, then UniFit II can define an empirical distribution function. In either case, the probability model can be automatically expressed in the analyst's simulation software. We then consider the general case of selecting a probability distribution in the absence of data. As an example, we show how UniFit II can be used to create busy-time and downtime models for machines that are subject to random breakdowns.

1 THE ROLE OF SIMULATION INPUT MODELING IN A SUCCESSFUL SIMULATION STUDY

In this section we will describe simulation input modeling and show consequences that might result if this important, but sometimes neglected, activity is performed improperly. We then suggest that with the use of UniFit II any simulation analyst can perform simulation input modeling more quickly and with greater accuracy than would otherwise be possible.

1.1 The Nature of Simulation Input Modeling

One of the most important activities in a successful simulation study is that of representing each source of system randomness by a probability distribution. For example in a manufacturing system, processing times, operating times before failure, and repair times for a machine should usually be modeled by probability distributions.

In this paper, we use the phrase "simulation input modeling" to mean the process of choosing a probability distribution for each random component of the system under study and expressing this representation in a form that can be used with the analyst's choice of simulation software. In Sections 2 and 3 we will demonstrate how an analyst can easily and accurately choose an appropriate probabilistic representation using the UniFit II software.

1.2 Two Pitfalls In Simulation Input Modeling

The authors have identified a number of pitfalls that can undermine the success of simulation studies (Law, McComas, and Vincent 1994, Law and Kelton 1991, Law and McComas 1989). Two of these pitfalls relate directly to simulation input modeling and are summarized in this section.

1.2.1 Pitfall Number 1: Replacing a Distribution by its Mean

Simulation analysts have sometimes replaced an input probability distribution by its mean in their simulation models. This practice may be caused by a lack of understanding on the part of the analyst or by lack of information on the actual form of the distribution (e.g., only an estimate of the mean of the distribution is available). Such a practice may produce completely erroneous results, as is shown by the following example.

Consider a manufacturing system consisting of a single machine tool at which jobs arrive to be processed. Suppose that the mean interarrival time of jobs is one minute and the mean processing time is 0.99 minute. Suppose further that the interarrival times and processing times actually have an exponential distribution. Then it can be shown that the long-run mean number of jobs waiting in the queue is *approximately 98*. On the other hand, suppose we were to follow the dangerous practice of replacing a source of randomness with a constant value. If we assume that each interarrival time is *exactly* one minute and each processing time is *exactly* 0.99

minute, *then each job is finished before the next arrives and no job ever waits in the queue!* The variability of the probability distributions, rather than just their means, has a significant impact on the congestion level in most queueing-type (e.g., manufacturing) systems. In Section 2 we shall show how use of UniFit II makes choosing an appropriate probability distribution a simple and easy process.

1.2.2 Pitfall Number 2: Incorrect Modeling of Random Machine Downtimes

The largest source of randomness for many manufacturing systems is that associated with random machine downtimes. An analyst is often faced with representing in a simulation model the random machine downtimes of a machine that has not yet been purchased. Data concerning the actual downtime behavior of machine tools is, thus, unavailable and the analyst must rely on estimates of reliability provided by vendors and engineers. Suppose, for example, that a vendor claims that a machine tool will be down 10 percent of the time, but is unwilling or unable to provide more information on its operating time before breakdown and its repair time. Given the limited available information, some simulation analysts account for downtimes by simply reducing the machine processing rate by 10 percent. Law and McComas (1989) compare through the use of simulation the described practice to a more accurate model that we shall demonstrate in Section 3. Although the two modeling approaches led to similar results for an average throughput measure of performance, the use of the reduced-production-rate model led to large errors with regard to measures such as average time in system and maximum number of jobs in queue. Accurate estimation of the latter performance measures is essential in many simulation studies. Thus, serious errors can result if an incorrect, simplified approach is taken. We will show in Section 3 how easy it is to obtain a more accurate model of random machine downtimes using UniFit II.

1.3 Advantages of Using UniFit II

With the assistance of UniFit II *any* analyst, regardless of prior knowledge of statistics, can avoid the two pitfalls introduced above. When system data are available, a complete analysis with the package takes just minutes. The package identifies the "best" of the candidate probability distributions, and assists the analyst in deciding whether the fit is good. If none of the candidate distributions provides an adequate fit, then an empirical distribution function can be created

by UniFit II. In either case, the representation of system randomness can be automatically expressed in the analyst's choice of simulation software. Appropriate probability distributions can also be selected when no system data are available. For the important case of machine breakdowns, UniFit II will determine appropriate busy-time and downtime probability distributions that match the system's behavior, *even if the machine is subject to blocking or starving.*

2 USING UniFit II WHEN SYSTEM DATA ARE AVAILABLE

We now consider the general case where an analyst has data corresponding to the source of randomness to be represented in the simulation model. Our intention is to highlight the capabilities of UniFit II. Detailed examples of program operation are available in the documentation that accompanies the free demonstration version of the software, which is available from the second author.

Three modes of operation are available for selecting probability models. The guided model selection mode and manual model selection mode options are used when data are available; otherwise, the no-data model selection mode option is used (see Section 3). The manual model selection mode option makes available to the experienced analyst a full set of tools for analyzing data sets. We have designed the guided model selection mode option to make accurate and thorough analysis of data sets easy for *any* analyst, regardless of prior knowledge of statistics. It is this mode that we will discuss.

Upon entering guided mode, an analyst uses a menu selection to read a data sample from a disk file. The data set we have chosen for this example consists of part processing times provided to us by a major automobile manufacturer. Immediately after reading the data file, UniFit II provides us with a sample summary, which is not shown here.

The basic information required by UniFit II to begin the fitting and evaluation process is a specification of the range of the underlying random variable. At two menus we selected options that characterized the range of the random variable as being greater than zero with no definite upper bound. UniFit II responded to our choices by fitting distributions with ranges starting at zero, distributions whose lower endpoint was estimated from the data itself, and distributions with no definite lower bound. These candidate models were then automatically evaluated. After a few seconds the result screen shown in Figure 1 was displayed.

UniFit II fit and ranked 20 candidate models, with

```
Guided Selection Model Rankings For Sample: Part Processing Times
```

Range of Random Variable

   During the fitting process UniFit considers distributions having any
   reasonable range (not just the specified range), provided they produce
   values in the specified range at least 99.99% of the time.

   Specified random variable range   At least 0.

Relative Evaluation of Candidate Models

	Relative Score (0-100)	Random Variable Range (if different from that specified)
Models		
1-Inverted Weibull	100.0	
2-Gamma (E)	89.5	At least 24.7981
3-Log-logistic (E)	88.2	At least 24.7981
4-Extreme Value Type B	85.5	Unrestricted
5-Weibull (E)	82.9	At least 24.7981

In addition, 15 other models were considered having scores from   .0 to 72.4.

Figure 1: Evaluation of Candidate Models for the Processing-Time Data

the five best-fitting models listed on the screen along with their scores. The displayed scores are calculated by a proprietary evaluation scheme that is based on our 16 years of research in this area, including the analysis of 35,000 computer-generated data sets. Results from the heuristics that we have found to be the best indicators of good model fit are combined and the resulting numerical evaluation is normalized

so that 100 indicates the best possible model and 0 indicates the worst possible model. These scores are *comparative* in nature and do not reflect an absolute assessment of the quality of fit. UniFit provides a separate evaluation of the adequacy of fit provided by the best-ranked model. In Figure 1 we see that the inverted Weibull distribution (range starts at zero) is the best model for the processing-time data. It should
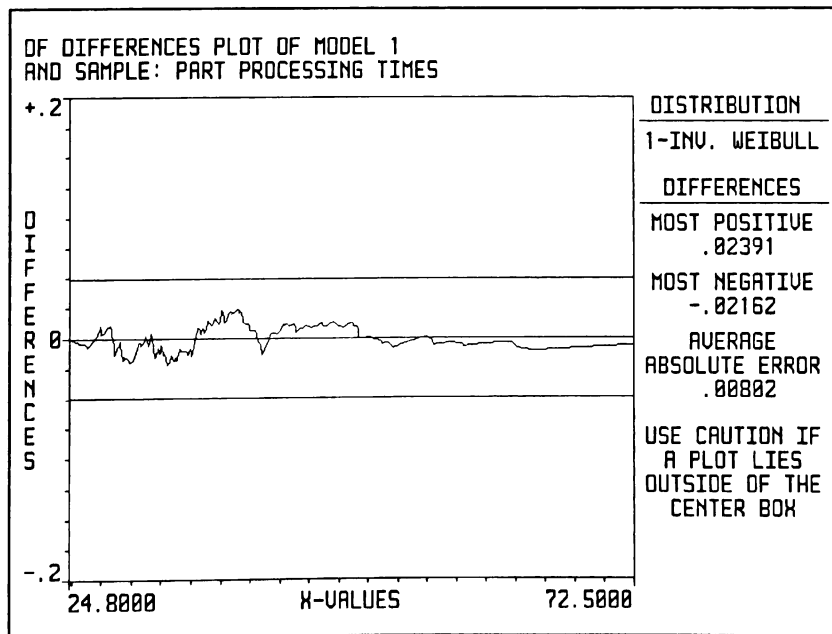

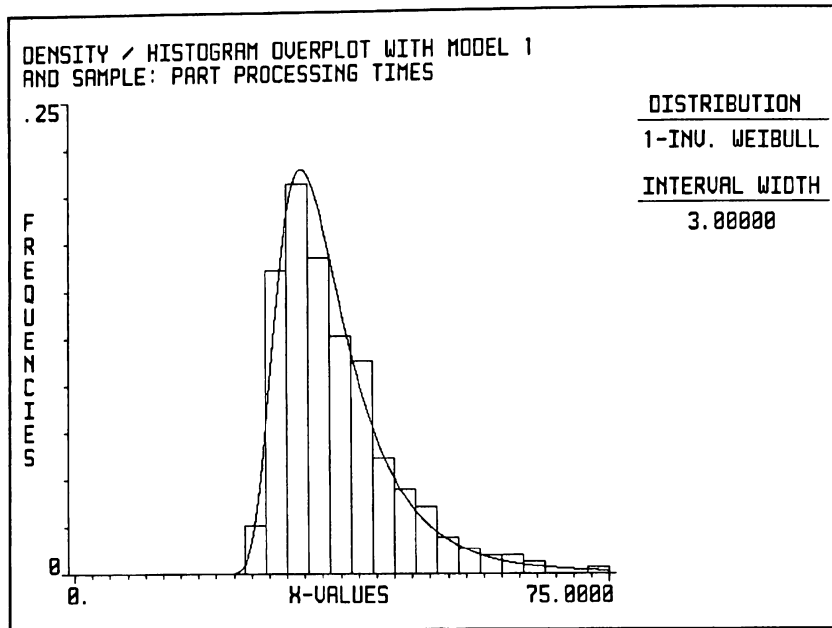Figure 2: DF Differences Plot for the Processing-Time Data

Figure 3: Density/Histogram Overplot for the Processing-Time Data

be noted that although the inverted Weibull distribution may be unfamiliar to you, it is supported by most simulation packages since it can be generated as the inverse of a Weibull random variable. It should also be noted that UniFit II completed the entire analysis without further input from the analyst; only the range had to be specified.

Immediately after presenting the summary screen seen in Figure 1, UniFit displays the DF (distribution function) differences plot shown in Figure 2. This heuristic plots the differences between the sample cumulative probabilities (distribution function) and the cumulative probabilities for the inverted Weibull distribution. Since these vertical differences are small (i.e., within the two horizontal error bounds), this suggests that the inverted Weibull distribution is a good model for the data. After viewing the plot automatically provided by UniFit II, a number of graphical and tabular confirmation options are available. We show as Figure 3 a histogram-based com-

parison. Here the density for the inverted Weibull has been plotted over the histogram (an estimate of the true density function). This plot also indicates that the inverted Weibull distribution is a good model for the observed data. UniFit II includes an option that allows us to display the representation of the inverted Weibull using different software packages. We show in Figure 4 the representations for four of the software packages supported by UniFit II.

With some data samples, no candidate model provides an adequate representation. In this case we recommend the use of an empirical distribution function. One useful feature of UniFit II is that in addition to using all of the sample values in the simulation software representation, it is possible to reduce the amount of information required through the use of a histogram-based empirical distribution function. We show a histogram-based representation, with 20 intervals, for two simulation software packages in Figure 5.

Simulation Software	Representation
Arena	1./WEIB(.03046, 6.27206, <stream>)
ProModel	InvWeibull(6.27206, 32.8341, <stream>, 0.)
TAYLOR II	1./weibull[.02832,6.27206]
Witness	1./WEIBULL(6.27206, .03046, <stream>)

Figure 4: Simulation Software Representation of the Inverted Weibull Distribution

Simulation Software	Representation
AutoMod II	continuous(.0000:24.8,.0322:27.185,.1576:29.57,.3183:31.955, .4791:34.34,.5981:36.725,.6945:39.11,.7942:41.495,.8457:43.88, .8778:46.265,.9068:48.65,.9421:51.035,.9550:53.42,.9711:55.805, .9807:58.19,.9839:60.575,.9904:62.96,.9968:65.345,.9968:67.73, .9968:70.115,1.0000:72.5)
GPSS/H	<name> FUNCTION RN<stream>,C21 .000000,24.8/.032154,27.185/.157556,29.57/.318327,31.955/ .479099,34.34/.598070,36.725/.694532,39.11/.794211,41.495/ .845657,43.88/.877812,46.265/.906751,48.65/.942120,51.035/ .954982,53.42/.971059,55.805/.980705,58.19/.983921,60.575/ .990352,62.96/.996783,65.345/.996784,67.73/.996785,70.115/ 1.000000,72.5

Figure 5: Simulation Software Representation of the Empirical Distribution Function

3 USING UniFit II WHEN NO DATA ARE AVAILABLE

Quite often a simulation analyst must model a source of randomness for which no data are available. In this section we show how UniFit II can be used to assist in the case of modeling random machine downtimes. UniFit II supports accurate modeling of systems with or without significant blocking or starving. For the example in this section we will assume that the machine of interest is never blocked or starved.

Consider a machine that has an efficiency of 0.9; that is, it is actually producing parts 90 percent of the time. When the machine goes down, the average downtime is 60 minutes. However the minimum downtime is 10 minutes. This information is specified to UniFit II through a sequence of easy-to-use menus. After all of the required information has been specified, the average number of downs (actually the average number of busy-time/downtime cycles) per 8-hour shift is calculated by the package to be 0.8. This makes sense since the average length of a busy-time/downtime cycle is 10 hours. A menu then allows

various characteristics of the busy-time and downtime distributions to be displayed. We show the simulation software representations for two packages in Figure 6.

4 CONCLUSIONS

UniFit II can help you develop more valid simulation models than if you use a standard statistical program, an input processor built into a simulation package (language or simulator), or hand calculations to determine input probability distributions. UniFit II uses a sophisticated algorithm to determine the best-fitting distribution and, furthermore, has 27 built-in distributions. On the other hand, a typical simulation package contains roughly 10 distributions.

UniFit II can represent most of its 27 distributions in approximately 20 different simulation packages such as AutoMod II, FACTOR/AIM, GPSS/H, MedModel, Micro Saint, MODSIM II, PACKAGING, ProModel for Windows, SIMAN V, SIMFACTORY II.5, SIMSCRIPT II.5, SLAMSYSTEM, TAYLOR II, and WITNESS, even though the distribution may not be available in the simulation package itself.

Simulation Software	Busy-Time and Down-Time Representations
FACTOR/AIM	GAMMA(771.429, .70000, <stream>) 10.0000 + GAMMA(35.7143, 1.40000, <stream>)
SIMSCRIPT II.5	GAMMA.F(540.000, .70000, <stream>) 10.0000 + GAMMA.F(50.0000, 1.40000, <stream>)

Figure 6: Simulation Software Representations of Busy-Time and Downtime Models

REFERENCES

Law, A.M. and W.D. Kelton. 1991. *Simulation Modeling and Analysis*, Second Edition. New York: McGraw-Hill.

Law, A.M. and M.G. McComas. 1989. Secrets of successful simulation studies. *Industrial Engineering* 21: 28-31 (May 1989).

Law, A.M., M.G. McComas, and S.G. Vincent. 1994. The Crucial Role of Input Modeling in Successsful Simulation Studies. *Industrial Engineering* 26: 55-59 (July 1994).

AUTHOR BIOGRAPHIES

STEPHEN VINCENT is Vice President for Software Development of Averill M. Law & Associates and is the codeveloper of the UniFit software package. He received his Ph.D. in Management Information Systems from the University of Arizona and has B.S. and M.S. degrees in Industrial Engineering from the University of Wisconsin-Madison. He was an Assistant Professor and taught courses in simulation modeling and software engineering at the University of Wisconsin-Milwaukee.

AVERILL M. LAW is President of Averill M. Law & Associates (Tucson, Arizona), a company specializing in simulation model building, training, and software. He has been a simulation consultant to more than 80 organizations, including General Motors, IBM, AT&T, ALCOA, General Electric, Nabisco, Xerox, NASA, the Air Force, the Army, and the Navy. He has presented more than 230 simulation seminars in 12 countries, and delivered more than 100 talks on simulation modeling at technical conferences.

He is the author (or coauthor) of three books and more than 35 papers on simulation, manufacturing, communications, operations research, and statistics, including the textbook *Simulation Modeling and Analysis* that is used by more than 40,000 people worldwide. His series of papers on the simulation of manufacturing systems won the 1988 Institute of Industrial Engineers' best publication award. He is the co-developer of the UniFit II software package for selecting simulation input probability distributions, and he developed several simulation videotapes. Dr. Law previously wrote a regular column on simulation for *Industrial Engineering* magazine from 1990 through 1991.

He has been a tenured faculty member and has taught simulation at the University of Wisconsin and the University of Arizona. Dr. Law has a Ph.D. in Industrial Engineering and Operations Research from the University of California at Berkeley.