Proceedings of the 2012 Winter Simulation Conference
C. Laroque, J. Himmelspach, R. Pasupathy, O. Rose, and A.M. Uhrmacher, eds

ESTIMATING PARAMETERS OF THE TRIANGULAR DISTRIBUTION USING NON-STANDARD INFORMATION

Seratun Jannat
Allen G. Greenwood

Mississippi State University
Department of Industrial & Systems Engineering
260 McCain Engineering Building
Mississippi State, MS 39762, USA

ABSTRACT

The triangular distribution is commonly used in simulation projects to represent probabilistic processes in absence of detailed data. The distribution can take on a variety of shapes and requires three easy to estimate basic parameters – minimum, maximum, and most likely. This paper considers two situations where different information is available than the three basic parameters. The paper provides means to use different information to estimate the remaining distribution parameters. The first situation commonly occurs in practice. For example, detail data may not be available, but the mean is known; thus, only two basic parameters need to be specified. The second situation occurs in research where controlled comparisons need to be made. For example, in order to understand the effect of variability on a system, means and general shape need to be held constant; thus, by fixing these two characteristics, only one of the basic parameters needs to be specified.

1 INTRODUCTION AND BASIC PROBLEM

In simulation studies detailed data that could be used to help select the appropriate distribution is oftentimes either not available or cannot be collected in a reasonable amount of time. The literature, e.g. (Law 2007), provides guidelines for selecting probability distributions in the absence of data. The triangular distribution is attractive since it only requires three easy-to-estimate location parameters – minimum, maximum, and most likely, typically denoted as a, b, and m, respectively. These parameters bound the values sampled during the simulation and allow the distribution to take on a variety of shapes – symmetrical, positively skewed, and negatively skewed, as well as varying degrees of variability.

Most domain experts can provide the basic or location parameter estimates for the triangular distribution. However, information other than the location parameters may be available, e.g. the mean, that can be used to specify the distribution. Oftentimes, even if detailed historical observational data are not available for distribution fitting, some summary statistics may exist, such as the mean. If the summary statistics are based on a large number of observations, then they may be considered quite reliable and thus useful and desirable for specifying the underlying distribution. In this case, it would be desirable to substitute the summary statistic(s) for the basic or location parameters used to specify a triangular distribution.

Similarly, the simulation may be used to understand the effects of shifting mean values or varying degrees of variability on system performance. In order to properly control the experiments, the mean of the assumed distribution (triangular) would change across scenarios, yet the level of variability in the distribution would remain constant. Similarly, in order to assess the affect of variability, the standard deviation could be changed across scenarios while holding the mean and skewness of the distribution constant.

978-1-4673-4781-5/12/$31.00 ©2012 IEEE

In the above examples, it is desired to substitute *non-standard information* (e.g. mean, standard deviation), when available and reliable, for *standard information* (location parameters a, b, and m) in order to specify the triangular distribution. In this paper we provide a means to use both types of information.

2 APPROACH AND RESULTS

In the case of the triangular distribution, the first three moments (mean, variance, and skewness) can be expressed in terms of the three location parameters (minimum, maximum, and most likely). Therefore, of these six values, only three need to be specified in order to define the distribution. The three remaining can be determined by solving three simultaneous nonlinear equations. Of course, this solution is not trivial. It can be done through the use of such commercial software as Wolfram's *Mathematica*.

So that modelers do not have to access special software and perform these requisite calculations each time they want to specify a triangular distribution using non-standard information, we have developed standardized tables, one for each location parameter, that are analogous to the standard normal table. These tables greatly facilitate the estimation of the required parameters. The tables provide estimates of the three basic location parameters (a, b, and m) that are needed in simulation software to sample from the triangular distribution. Each location parameter is a function of three moments, μ, σ, and S_k, the mean, standard deviation, and coefficient of skewness, respectively. The tables are standardized in terms of $\mu=1$, coefficient of variation (σ/μ), C_V ranging from 0.1 to 1.5, and positive skewness, S_k, ranging from 0 to 0.5656 (the maximum possible skewness for the triangular distribution). The tables are only generated for positively skewed triangular distributions since the extreme values, a and b, are the same for both positively and negatively skewed distributions and the most likely value (m) for negatively-skewed distribution is the complement of the positively-skewed value.

To illustrate the estimation process, consider the following examples. In the first case, assume the mean process time for an activity is well known from historical data, i.e. μ = 83, but the underlying distribution is unknown. If the triangular distribution is assumed, and if the mean is to be considered in the specification of the distribution, then only two other parameters need to specified. With the help of figures representing triangular distributions with varying skewness, domain experts believe S_k is about 0.3. They also believe the most likely value, m, is 75. Therefore, from the tables developed in this research, a and b are determined to be 47.4 and 127.3, respectively and the distribution is specified in the simulation software as *triangular(47.4, 127.3, 75)*. In the second case, researchers want to control experiments by varying distribution moments between scenarios. For example, if the assumed distribution is triangular and if μ = 180, C_v = 0.3, and S_k = 0.4, then, using our three tables, a, b, and m are determined to be 72.0, 327.5, and 140.5, respectively.

3 CONCLUSION

The proposed approach provides a quick and easy way to specify the triangular distribution by utilizing information other than the basic location parameters. Standard tables have been derived so that a triangular distribution's location parameters can be estimated based on any combination of three variables including the three distribution moments and three location parameters. This methodology is especially useful in simulation practice where non-standard information is available to support the specifying of the distribution and in research where distribution moments are specified to control experiments. The tabular approach provides a very precise approximation of the solution that would be obtained by solving the simultaneous non-linear equations directly.

REFERENCES

Law, A. M. 2007. *Simulation Modeling & Analysis.* 4th ed. New York: McGraw-Hill, Inc.