Proceedings of the 1999 Winter Simulation Conference
P. A. Farrington, H. B. Nembhard, D. T. Sturrock, and G. W. Evans, eds.

DEFINING A BETA DISTRIBUTION FUNCTION FOR CONSTRUCTION SIMULATION

Javier Fente
Kraig Knutson
Cliff Schexnayder

Del E. Webb School of Construction
Arizona State University
P.O. Box 870204
Tempe, AZ 85287-0204, U.S.A.

ABSTRACT

In most applications of simulation to construction, the underlying Probability Distribution Function (PDF) is generally unknown. Consequently, an expert will have to select a PDF hoping that the one that is chosen matches the shape of the underlying distribution. This research attempts to quantify, through a sensitivity analysis, the effect of subjective information in choosing parameters for a Beta distribution to be used in earthmoving simulation models.

1 INTRODUCTION

Although simulation is widely applied as a practical tool for planning and analysis in many industries, in the case of construction engineering it has not yet emerged from the research stage to practice. Modeling construction activity time elements is crucial to developing a usable simulation procedure because the time elements affect process production rates, the completion of a job, and resource utilization. The lack of confidence in the subjective selection of a Probability Distribution Function (PDF) for an activity time element is due to the unpredictable nature of construction processes. This concern has limited the use of simulation as a practical tool for constructors.

It is often recommended that, in order to model the duration parameter of a construction activity in an efficient and accurate way, a flexible family of PDFs capable of attaining a wide variety of shapes should be used. AbouRizk and Halpin (1992) demonstrated through a modified β_1-β_2 plane analysis that most earthmoving construction operations can be described by the Beta PDF. In addition, according to McCrimmon et al. (1964), the PDF used for construction simulations should be continuous and limited between two positive time intercepts, and have a unique mode in its defined range. The Beta PDF satisfies all the conditions stated.

1.1 Study Data

The data used in this research was acquired from the Atkinson-Washington-Zachry (AWZ) joint venture on the Eastside Reservoir Project in California. AWZ operates a fleet of Caterpillar (CAT) trucks equipped with a Vital Information Management System (VIMS) and a Truck Payload Monitoring System (TPMS) on the project. These two systems automatically record truck performance data (e.g., payload weight, load time, haul distance, travel time, and dump time). TPMS data from this project, representing 54,000 truck cycles, is the original data for all of the subsequent statistical research presented in this paper.

2 THE BETA PDF AND ITS PARAMETERS

The Beta PDF is a continuous distribution defined over a range. Additionally, both of its end points are fixed at exact locations and it belongs to the flexible family of distributions. Because of its extreme flexibility, the distribution appears ideally suited for the description of subjective time estimates of activity duration. The shape of the Beta PDF depends on the choice of its two parameters “a” and “b.” The parameters are any real number greater than negative one, and depending on their values the Beta PDF generated will have the “U”, the “J”, the triangle or the general bell shape of the unimodal function. Estimating these parameters is controlled by data availability.

In the absence of data, an experienced constructor uses subjective information to estimate the parameters of the Beta PDF expected to describe the construction operation. The subjective information needed to determine the two Beta parameters that will describe a unique Beta curve include four characteristics: the minimum and maximum times, as well as two of the following subjective statistics; mode, mean, variance or selected percentiles. In the context of fitting a Beta PDF for construction simulation, the percentiles represent the percent chance that the analyzed activity duration will not exceed a time "t." The natural limitations on the input data range for a construction simulation PDF were defined by McCrimmon et al. (1964) as: having a unique mode in its defined range and the parameters "a" and "b" of the Beta PDF must be positive and greater than one. Consequently a restricted subset of Beta density curves have often been applied to construction modeling.

Weiler (1965) concluded that many errors in the output of a simulation are those of assigning wrong values to the parameters of a distribution. If, on the other hand, the form of the PDF is unknown, a further error may be introduced by assuming a Beta distribution when in fact some other distribution would have been appropriate. However, this type of error is likely to be small in comparison to errors in the parameters.

Maio (1998) studied the effect of PDF input in a haul cycle simulation. The distributions compared were a Beta distribution whose parameters were estimated using the same data, and a PDF ranked, by a goodness of fit test, as the closest fit to the actual data set in the case of each operation. The simulation results using these two distributions were then compared to a simulation using a trace of the actual cycle data as input. The research concluded that the choice of a distribution function, if derived from the same original data, did not influence the output results of his SLAM II model.

3 SUBJECTIVE ASSESSMENT OF PDFS

Estimating the productivity of an earthmoving operation involves the investigation of events for which significant statistical records usually do not exist. This lack of data creates problems concerning the selection of a PDF. At that point the judgement of an experienced constructor becomes very important. In such a case, an expert will have to assume a PDF and hope that its shape matches that of the underlying distribution. For this reason the flexible Beta distribution capable of attaining a variety of shapes is often used in construction simulation applications.

Experiments in psychology summarized by Peterson and Miller (1964) and Peterson and Beach (1967) have indicated that when attempting to estimate the mean of a skewed distribution there is a bias to estimate the mean toward the median. Alpert and Raiffa in their 1982 progress report on the training of probability assessors, recommended direct fractile assessments as the most reliable method and less prone to bias. In this method, the planners express their uncertainties about an unknown quantity in terms of percentiles of a distribution representing their state of knowledge. This led Mosleh and Apostolakis (1982) to recommend that "assessors use percentiles to quantify their beliefs and to avoid the direct assessment of other measures like mean value or standard deviation." This procedure has been used in applications from nuclear energy to experimental psychology and mathematics.

The Beta distribution is a continuous PDF bounded over a range; that is, the end points, or minimum and maximum durations, are defined at exact locations. According to Wilson et al. (1982), an expert can estimate the end points of an activity duration distribution somewhat easily and accurately due to the expert's familiarity with the technological constraints on the target activity. For construction applications, the mode or best guess, of a construction activity duration can be elicited accurately from an expert using deterministic methods. The end points and a value for the mode define a number of different Beta PDFs. To fit a unique beta distribution one more characteristic of the PDF must be specified.

AbouRizk and Halpin (1992) developed a microcomputer based software system specifically designed for the subjective estimation of Beta PDFs. This software is the Visual Interactive Beta Estimation System (VIBES). VIBES uses a combination of four activity-time characteristics, two of which must be the end points, to determine the parameters of the unique Beta PDF. The possible combinations include the end points and 1) mean and standard deviation; 2) mean and a selected percentile; 3) mode and a selected percentile; and 4) two selected percentiles.

For the purpose of this study, VIBES was used as the tool to determine the Beta shape parameters "a" and "b" from given subjective information. The input combination was the mode, the 75^{th} percentile and the two end points as suggested by the investigations of Wilson et al. (1982), and Lichtenstein et al. (1982). The input combinations using the mean and/or the standard deviations were rejected on the basis of the study presented by Peterson and Miller (1964) which explains the difficulty of assessing the mean and standard deviation of skewed distributions.

The following example illustrates an application of VIBES when a combination of the mode, the 75^{th} percentile, and the end points is used to find the parameters of the underlying Beta PDF.

3.1 Application of VIBES

The following example from a hypothetical earthmoving operation illustrates the use of VIBES to fit Beta distributions to subjective information developed from deterministic calculations of truck travel time. The mode and the 75^{th} percentile, along with the minimum and maximum durations are specified as inputs.

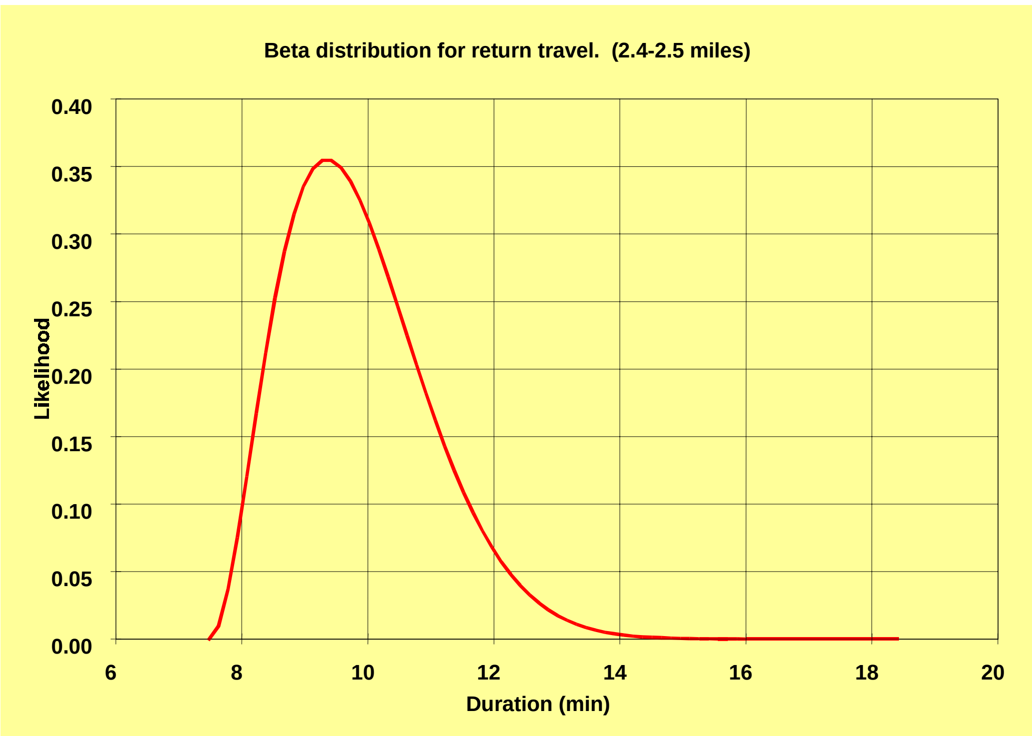
The Beta PDF of the travel time for an empty CAT 785 off-highway truck moving in a range of 2.4-2.5 miles, can be estimated using VIBES by inputting the following subjective information:

- The minimum possible time in which a truck can complete a cycle is 7.5 minutes, considering the physical characteristics of the project site and manufacturer's truck data.
- The most likely duration for the return time will be 9.3 minutes, allowing for acceleration and deceleration, and applying an efficiency factor.
- The maximum possible time will be estimated to be 18.7 minutes, two times the mode. It is assumed that management would notice such a laggard truck and take action.
- There is at least a 75% chance that the activity duration will not exceed 10.6 minutes, which is 1.13 times greater than the mode.

Using this information, VIBES yields the distribution and parameters shown in Figure 1.

The shape parameters estimated by VIBES were compared to the parameters from an actual data set of 809 travel durations for an empty CAT 785 in a 2.4-2.5 mile range. These parameters were calculated using the Maximum Likelihood Estimates (MLE), Method of Moment Matching (MM) and the Method of Ordinary Least Square Minimization (OLS) minimization procedures implemented in the Betafit software developed at Purdue University. Betafit can fit a Beta distribution to a data set based on the user's choice of fitting procedure. The software computes the statistics of the sample and of the fitted PDF. The sample and fitted data distributions graphs, Figures 2, 3 and 4, show how the methods of MLE and OLS yield a better fit to the given data histogram than the MM technique.

The parameters subjectively estimated by VIBES and those calculated by Betafit with the actual project data are shown in Table 1.



50th percentile:	9.704	Shape Parameter a:	3.156
95th percentile:	11.967	Shape Parameter b:	11.782
Mode:	9.350	Standard Deviation:	1.147
Mean:	9.851	Variance:	1.316

Figure 1: Output from VIBES

Sample Statistics		Fitted Beta distribution parameters	
Mean	8.915698	Mean	8.938812
Variance	0.904068	Variance	0.863210
Skewness	2.393677	Skewness	0.695178
Kurtosis	19.150357	Kurtosis	3.469117
Min	7.100000	Min	6.887000
Max	18.600000	Max	19.158000
		param-a	3.894371
		param-b	19.396176
		KS	11.172633

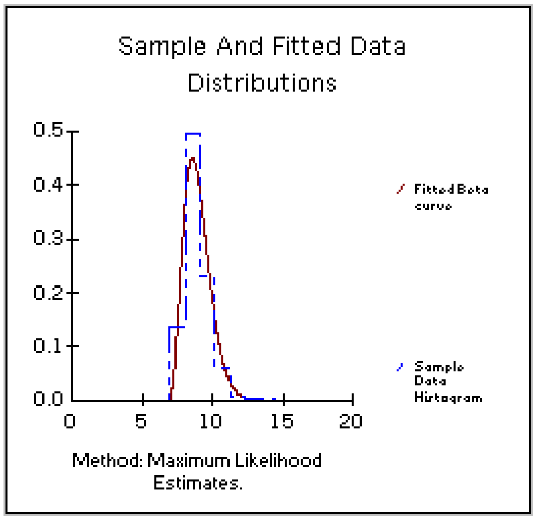


Figure 2: Maximum Likelihood Estimates (MLE)

Sample Statistics		Fitted Beta distribution parameters	
Mean	8.915698	Mean	7.519568
Variance	0.904068	Variance	0.868070
Skewness	2.393677	Skewness	3.600741
Kurtosis	19.150357	Kurtosis	18.988302
Min	7.100000	Min	7.100000
Max	18.600000	Max	18.600000
		param-a	0.158909
		param-b	4.196643
		KS	1.543546

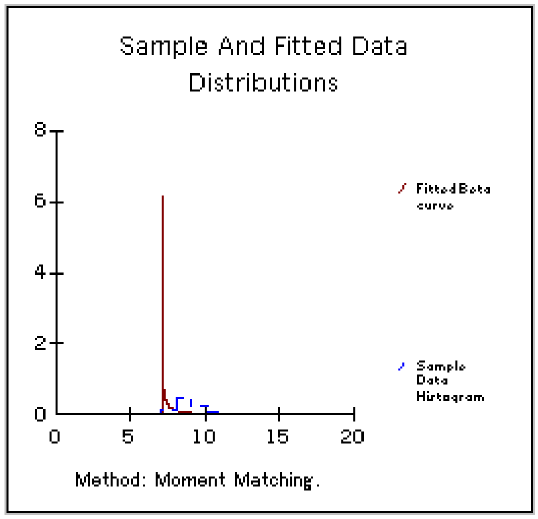


Figure 3: Method of Moment Matching (MM)

Sample Statistics		Fitted Beta distribution parameters	
Mean	8.915698	Mean	8.907456
Variance	0.904068	Variance	0.888210
Skewness	2.393677	Skewness	0.807440
Kurtosis	19.150357	Kurtosis	3.656726
Min	7.100000	Min	7.100000
Max	18.600000	Max	18.600000
		param-a	2.942815
		param-b	15.780946
		KS	8.661373

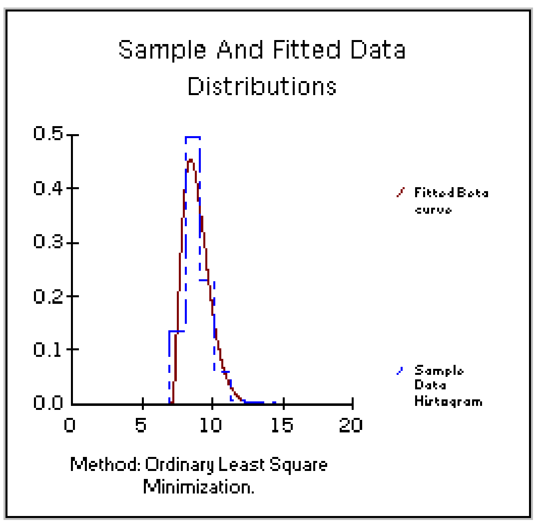


Figure 4: Method of Ordinary Least Square Minimization (OLS)

Table 1: Comparison of Calculated vs. Estimated Shape Parameters

	VIBES	MLE	MM	OLS
Parameter a	3.156	3.894	0.159	2.943
Parameter b	11.782	19.396	4.196	15.781

Figure 5 illustrates the minor differences among the PDF yielded by VIBES and those given by the MLE and OLS methods from Betafit. For this case, the difference in the output of a simulation model derived from using either of the three mentioned Beta PDFs, would be relatively insignificant. The Beta PDF rendered by the MM method varies considerably from the data histogram in Figure 5.

4 CONCLUSION

Simulation applications normally use the flexible families of probability distributions due to their capability of attaining a wide variety of shapes. Among such families is the Beta distribution. A modified β_1-β_2 plane analysis revealed that most of the construction data sets analyzed by AbouRizk and Halpin (1992) lay in the Beta region. Thus, it was concluded that the Beta distribution is suitable for modeling durations of construction activities.

Due to the natural limitations of the input data range and mode in construction activities, the parameters of the Beta distribution must be positive and greater than one. Consequently, a unique and restricted family of Beta curves is specified to the model durations of construction activities. The restricted subset of Beta curves corresponds to the unimodal polynomial graphs, (i.e. those with parameters "a" and "b" greater than one).

When sample observations are not available, a Beta distribution can be defined by using subjective information provided by experts about a given activity duration. Based on the research in experimental psychology, it is believed that a construction expert using deterministic methods can accurately estimate the minimum and maximum activity durations, as well as the mode and the 75^{th} percentiles to use in defining a Beta distribution.

The lack of accuracy in the subjective estimation of probability distributions is still a subject of research in the simulation field across various engineering disciplines. It is one of the most important factors for its limited use as a planning tool in the construction industry. This research attempted to quantify, through a sensitivity analysis, the effect of subjective information in developing earthmoving simulation models. The simulation model was developed by Maio (1998) in SLAM II (Pritsker, 1995).

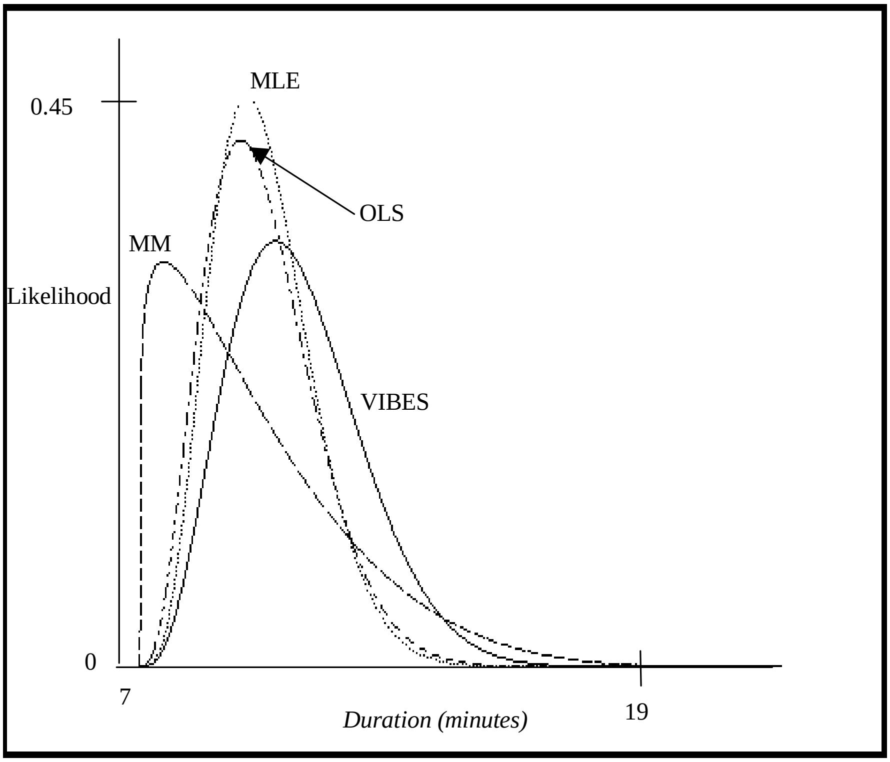


Figure 5: Comparison of Beta PDF given by VIBES and Betafit

REFERENCES

AbouRizk, S.M., and Halpin, D. W. 1992. "Statistical properties of construction duration data." *J. of Constr. Engrg. and Mgmt*., ASCE, 118(3), 525-543

Alpert, M., and Raiffa, H. 1982. "A progress report on the training of probability assessors." *Judgement under uncertainty: Heuristics and biases*, D. Kahneman, P. Slovic, and A. Tversky, eds, Cambridge University Press, Cambridge, England. 294-305

Lichtenstein, S., Fischhoff, B., and Phillips, L.D. 1982. "Calibration of probabilities: The state of the art to 1980. *Judgement under uncertainty: Heuristics and biases*, D. Kahneman, P. Slovic, and A. Tversky, eds, Cambridge University Press, Cambridge, England. 306-334

MacCrimmon, K.R., and Rayvec, C.A. 1964. "An analytical study of the PERT assumptions." *Operations Research*, 12(1), 16-37

Maio, C. 1998. *Investigation of probability distribution functions for haul cycle simulation*. MS Thesis, Arizona State University, Tempe, AZ.

Mosleh, A., and Apostolakis, G. 1982. "Some properties of distributions useful in the study of rare events." *IEEE Transactions on Reliability*, R-31(1), 87-94

Peterson, C., and Miller, A. 1964. "Mode, median, and mean as optimal strategies." *J. of Exp. Psych.* 68(4), 363-367

Peterson, C., and Beach, L.R. 1967. "Man as an intuitive statistician." *Psych. Bull.*, 68, 29-46

Pritsker, A.A.B. 1995. *Introduction to simulation and SLAM II*, John Wiley, New York.

Weiler, H. 1965. "The use of the incomplete beta functions for prior distributions in binomial sampling." *Technometrics*, 7(3), 335-347

Wilson, J.R., Vaughan, D.K., Naylor, E., and Voss, R.G. 1982. "Analysis of space shuttle ground operations." *Simulation.* 38(6), 187-203

AUTHOR BIOGRAPHIES

JAVIER FENTE is a graduate student in the Del E. Webb School of Construction at Arizona State University (ASU) in Tempe, Arizona. He holds a M.S. in Civil Engineering and a M.S. in Construction from Arizona State University as well as a B.S.C.E. from Lafayette College, Easton, Pennsylvania. Mr. Fente is a native of Madrid, Spain.

KRAIG KNUTSON, CPC, is an Assistant Professor in the Del E. Webb School of Construction at Arizona State University (ASU) in Tempe, Arizona. He holds a Ph.D. in Industrial Engineering from Arizona State University as well as bachelor's and master's degrees in construction. His research interests are related to the design, simulation and optimization of manufacturing systems and construction processes and to the decommissioning of semiconductor facilities. He is a member of AACE, ASSE, IIE, INFORMS, and SCS.

CLIFF SCHEXNAYDER is the Eminent Scholar in the Del E. Webb School of Construction at Arizona State University (ASU) in Tempe, Arizona. He holds a Ph.D. in Civil Engineering from Purdue University as well as a M.S.C.E. and B.C.E. from Georgia Institute of Technology. Prior to joining ASU, Dr. Schexnayder was the Chief Engineer for the Nello L. Teer Company of Durham, North Carolina. His research interests include construction engineering and heavy construction equipment and estimating. He is an ASCE Fellow and Chair of the Transportation Research Board's Construction Section. Dr. Schexnayder is co-author of the McGraw-Hill textbook *Construction Planning, Equipment and Methods*.