FORMULATION OF A βETA PROBABILITY DISTRIBUTION FUNCTION FOR CONSTRUCTION PROCESS CONTROL

Kraig Knutson,[1] Cliff Schexnayder,[2] and Javier Fente[3]

[1]Assistant Professor, [2]Eminent Scholar, and [3]Graduate Student, Del E. Webb School of Construction, Box 870204, Ariz. State Univ., Tempe, AZ 85287-0204, USA. cliff.s@asu.edu

Abstract: The lack of confidence in the selection of a Probability Distribution Function (PDF) to depict an activity duration has limited the use of simulation as a tool for constructors. The PDF used for a construction simulation should be continuous, limited between two positive time intercepts, and with a unique mode in its defined range. The beta PDF satisfies these conditions. It is recommended that the subjective information used to define beta parameters include the minimum and maximum activity durations, as well as the mode and the 75^{th} or 25^{th} percentiles. This paper presents a formulation for developing a Beta PDF that can be used in construction process control, automation or for simulation modeling.

Keywords: Simulation, Modeling, Probability Distribution Function, Beta Distribution, Fitting Beta Distributions, Beta Shape Parameters, Density Curves, Construction Activity Duration.

1 INTRODUCTION

It has been proposed to use simulation models for construction process control and as an estimating tool to assess process productivity. At the beginning of a process or when preparing an estimate these types of applications inherently present the modeler with a situation where there is only a limited amount of data available for choosing an appropriate Probability Distribution Function (PDF) to use in the simulation. In the absence of data, when attempting to utilize simulation models, an expert will use subjective estimates about a given activity duration to derive a PDF that presumably represents the distribution of the process under study.

Based on research in experimental psychology, it is recommended that the subjective information used to define beta parameters include the minimum and maximum activity durations, as well as the mode and the 75^{th} or 25^{th} percentiles. This paper presents a formulation for developing a Beta PDF for use in construction process control, automation or in the development of simulation models.

2 ASCERTAINING A CONSTRUCTION PROCESS PDF

Weiler [1] concluded that many of the errors in the output of a simulation model could be traced to assigning the wrong values to the parameters of a distribution. However, if the form of the PDF is unknown, a further error may be introduced by assuming one distribution function when in fact some other distribution would have been appropriate. But this type of error is likely to be small compared with errors in assigning the wrong values to the distribution parameters. This fact has been confirmed in many studies.

Wilson et al. [2] compared the use of beta PDFs with triangular PDFs in a simulation of space shuttle ground operations. The comparison was based on matching the mean and variance of both PDFs. That study concluded that there were no significant differences in the simulation outputs. Klein and Baris [3] tested the effect of using various distributions to model the time required to perform numerous tasks in a health care facility. The INSIGHT simulation language was used to compare the results of using the Johnson, triangular, and beta distributions. Again, it was concluded that there was no statistically significant difference in the output estimates.

Touran [4] conducted a sensitivity analysis on the simulation of a tunnel-boring schedule. The SLAM II simulation language was employed for the study. When inputting either a normal or a lognormal PDF for the duration of the boring, the results showed no statistically significant difference in the predicted mean completion time. Maio [5] studied the effect of PDF choice in a haul cycle simulation model. The comparisons included a trace of actual cycle data, a beta distribution whose parameters were estimated using the actual cycle 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 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.

All of these studies found that the outputs of their respective simulation models were not affected by the choice of PDF.

In most construction applications, the underlying PDF is generally unknown. Consequently, an expert will have to select a PDF with the assumption that the one that is selected represents the shape of the underlying distribution. 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 PDF's capable of attaining a wide variety of shapes should be used. Farid and Koning [6], using data from O'Shea et al. [7], proposed the beta PDF as the closest fit for modeling truck load and travel time distributions. A modified β_1-β_2 plane analysis by AbouRizk and Halpin [8] revealed that most earthmoving construction operations can be described by the beta PDF. Their work demonstrated that 64 out of 71 construction operations from an earthmoving project fell in the beta region of the graph.

According to MacCrimmon and Rayvec [9], the PDF used for construction simulations should be continuous, limited between two positive time intercepts, and with a unique mode in its defined range. The beta PDF satisfies all of these conditions.

3 POSTULATING STATISTICAL CHARACTERISTICS

The beta distribution is a bounded continuous PDF. Therefore the end points or minimum and maximum durations of the distribution are defined at exact locations. According to Wilson et al. [2], 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.

The minimum duration for each construction activity can be calculated deterministically using the equipment manufacturer's machine data and a set of physical site constraints. The maximum duration is a function of the quality of management on the project, i.e. how long must a problem persist before it comes to management's attention and corrective action is taken. Additionally for construction applications, the mode (or duration occurring most frequently) of a construction activity duration can be elicited fairly accurately 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.

3.1 Selecting a Fourth βeta Probability Distribution Charactistic

Experiments in psychology, summarized by Peterson and Miller [10] and Peterson and Beach [11], indicated that in the case of estimating the mean of a skewed distribution there is a tendency to bias the estimate of the mean toward the median. Alpert and Raiffa [12] in their 1969 progress report on the training of probability assessors, recommended direct fractile assessments as the most reliable method. Direct fractile assessments are also less prone to bias. This led Apostolakis and Mosleh [13] 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."

Alpert and Raiffa, [12], and Lichtenstein et al. [14] have indicated that subjective estimates for certain percentiles of a population can be reasonably accurate, especially for the 25^{th}, 50^{th}, or 75^{th} percentiles; these are also known as the lower, median, and upper quartiles. With regards to their research, the results of an estimate using the quartile approach were more accurate than those obtained through the tertile method, i.e. the fractiles 33^{rd} and 67^{th}, and the lower and upper fractiles procedure, i.e. the fractiles 10^{th} and 90^{th}.

Based on the work of Wilson et al. [2], and Lichtenstein et al [14] the 75^{th} percentile is advocated in this research. Input combinations using the mean and/or the standard deviations were rejected on the basis of the study presented by Peterson and Miller [10] which explains the difficulty of assessing the mean and standard deviation of skewed distributions.

3.2 Statistical Characteristics Mythology

A typical job efficiency for construction work is a 50 min/hr (a day shift), or an efficiency factor of 0.83. Therefore, in this study the most likely duration, or mode, is predicted by applying a job efficiency factor of 0.83 to the activity minimum duration calculated by deterministic evaluation of the physical project factors. It is realized however, that appropriate efficiency factors depend on project specific conditions and these should always be selected on a case by case basis.

An expert can reasonably assess the maximum duration of an activity by considering the character of the project's management. This time span should correspond to the duration that with certainty management would notice a laggard truck or a slow loading situation and take action. It must allow for reasonable delays and unfavorable site conditions. In this study the maximum duration was calculated by multiplying the mode by a factor. The factors used with each of the four activities included in the simulation model are: for loading a factor of 4; for hauling a factor of 2; for dumping a factor of 4; and for return travel a factor of 2.

These multipliers yield reasonable maximum durations for the specified activities. This formulation assumes that the foreman in charge of the activity would consider durations beyond these values as obvious irregularities, and would take action to correct the problem. Consequently, larger durations are not accounted for in the PDF.

The 75th percentile value corresponds to that duration for which there is a 25% chance it will be exceeded. This value is determined by multiplying the mode by a factor. Analyzing a total of 54,300 haul cycles from an actual heavy construction project provided the basis for these factors in the case of loading, hauling, dumping and return travel.

The haul cycles analyzed were document during the construction of the Eastside Reservoir Project in Southern California. Data was accumulated from the project's fleet of Caterpillar 785B off-highway trucks through the use of Caterpillar's Vital Information Management System (VIMS). The VIMS onboard data recording system archived machine production and payload data. The fleet accumulated more than 14,000 operating hours, and transported close to 8 million tons of material during the study period.

4 APPLICATION OF BETA PDF METHODOLOGY

The following example from an earthmoving operation illustrates the use of subjective information to define a beta distribution representing a construction process. The mode and the 75th percentile, along with the minimum and maximum durations are the designate inputs for defining a unique beta distribution.

4.1 Deterministic Time Calculations

The minimum duration activity times for the process are calculated using the equipment manufacturer's machine data for the given set of project specific physical characteristics and for the required travel distance. A set of calculations is shown in Figure 1. The mode and the maximum duration were found by applying the factors as previously discussed. In the case of the mode this is an efficiency factor based of the expected work conditions. The maximum duration factor reflects the quality of the activity supervision and possibly to a minor extent the mechanical condition of the equipment.

The minimum travel time is found by dividing the hauling distance by the speed obtained for a specified total effective grade and truck gross weight. The gross weight used corresponds to that of the loaded truck for the haul and the empty truck for the return. The total effective grade is the result of adding the rolling resistance, expressed as a percent, and the slope of the analyzed segment. For the project under study, the total effective grade of the roadway for the haul and return travel are estimated as –4% and 8% respectively. The truck's speed for these conditions based on Caterpillar Inc. data is 58.0 kmph (36 mph) for the haul and 38.6 kmph (24 mph) for return. These speeds are for a steady state condition, therefore, to account for acceleration and deceleration effects there is an adjustment at the beginning and end of the travel distance.

4.2 Defining the Beta PDF

The beta PDF of the return travel time for an empty CAT 785 off-highway truck moving in a distance range of 3.8 kilometers (2.35 miles) can be estimated using the following subjective information:

- The *minimum possible time a truck can complete the travel* is 7.67 minutes considering the physical characteristics of the project site and manufacturer's truck data.
- The *most likely duration for the travel* (mode) time will be 9.21 minutes, allowing for acceleration and deceleration, and applying an efficiency factor.
- The *maximum possible time a truck should take to complete the travel* is estimated to be 18.42 minutes, two times the mode. It is assumed that management would notice such a laggard truck and take action.

The minimum, mode, and maximum durations of a construction activity do not define a unique beta PDF. One more activity time characteristic is needed to define a unique beta curve. For the purpose of this methodology, the fourth input is the 75th percentile.

Scheduling personnel at Kennedy Space Center use a standard procedure for modeling the duration of space shuttle ground operations when data is scarce. Their method of estimating the distribution's parameters is to apply multipliers to the estimated minimum duration of an activity [2]. At the present time this is a subject for further research in relation to multipliers that relate the mode of a construction activity to the 75th or some other percentile. However, based on the relationships propounded by Wilson [2], a factor of 1.2 was used to relate the mode of the truck return time to the 75th percentile.

- There is at least a *75% chance that the truck travel time* will not exceed 11.05 minutes, which is 1.20 times greater than the mode.

Using this information and the microcomputer based software Visual Interactive Beta Estimation System (VIBES) developed by AbouRizk et al. [15] a set of beta distribution parameters are calculated. VIBES uses a combination of four activity-time characteristics to determine the parameters of the unique beta PDF. Two of the input data points must be the end points.

LOAD ***(CAT 5230 Excavator)***			
Minimum time:			
• First pass (dump time)			0.05 minutes
• Cycle time			0.45 minutes
Minimum load (4 passes ≅ 118 tonne (130 tons))			**1.40 minutes**
Mode:			
• 5 passes (≅ 136 tonne (150 tons))			1.85 minutes
Given a 50 min/hr			**2.22 minutes**
MAXIMUM TIME:			
• 4 x Mode			**8.88 minutes**
END DUMP AND MANEUVER ***(CAT 785B off-highway truck)***			
Minimum time:			**1.00 minutes**
Mode (50 min/hr):			**1.20 minutes**
Maximum time (4 x Mode):			**4.80 minutes**
HAUL ***(CAT 785B Truck – Downhill)***			
Grade:		-6%	
Rolling Resistance:		2%	
Total effective grade:		-4%	
Distance:	3.8 km	(2.35 miles = 12,408 ft.)	
Minimum time:			
• Acceleration time	8.1 kmph	(5 mph for 500 ft) =	1.136 minutes
• Deceleration time	8.1 kmph	(5 mph for 500 ft) =	1.136 minutes
• Constant speed travel time	58.0 kmph	(36 mph for 11,408 ft) =	3.601 minutes
			5.87 minutes
Mode:			
• 50 min/hr			**7.05 minutes**
Maximum time:			
• 2 x Mode			**14.10 minutes**
RETURN ***(CAT 785B Truck – Uphill)***			
Grade:		6%	
Rolling Resistance:		2%	
Total effective grade:		8%	
Distance:		2.35 miles = 12,408 ft.	
Minimum time:			
• Acceleration time	8.1 kmph	(5 mph for 500 ft) =	1.136 minutes
• Deceleration time	8.1 kmph	(5 mph for 500 ft) =	1.136 minutes
• Constant speed travel time	38.1 kmph	(24 mph for 11,408 ft) =	5.401 minutes
			7.67 minutes
Mode:			
• 50 min/hr			**9.21 minutes**
Maximum time:			
• 2 x Mode			**18.42 minutes**

Figure 1. Deterministic Calculation of Process Minimum, Mode, and Maximum Durations.

4.3 Comparison of the Created Beta PDF to Beta PDF's Computed from Actual Data

The shape parameters calculated by VIBES were compared to the parameters from an actual data set of 1000 travel durations for CAT 785 returning empty and traversing a 3.7 to 3.9 kilometer, (2.35 mile) travel distance. The data generated parameters were calculated using the Maximum Likelihood (MLE), Methods of Moments (MM), and the Ordinary Least Squares (OLS) procedures implemented in the Betafit software developed at Purdue University [16]. The software computes the statistics of the sample and of the fitted PDF.

The parameters calculated by VIBES based on methodology proposed here and those calculated using the Betafit program and actual project data are shown in Table 1.

Table 1. Beta PDF Shape Parameters.

Beta Parameter (1)	VIBES (2)	MLE (3)	MM (4)	OLS (5)
a	1.898	4.345	0.652	3.646
b	6.372	10.561	4.741	9.135

The resultant distributions given by the parameters of Table 1 are compared in Figure 2. The figure illustrates the minor differences between the PDFs derived from VIBES and those given by the MLE and OLS methods with Betafit. For this case, the difference in the output of a simulation model derived from using any of the three beta PDFs would be relatively insignificant. The Beta PDF rendered by the MM method varies considerably from the data histogram in Figure 2.

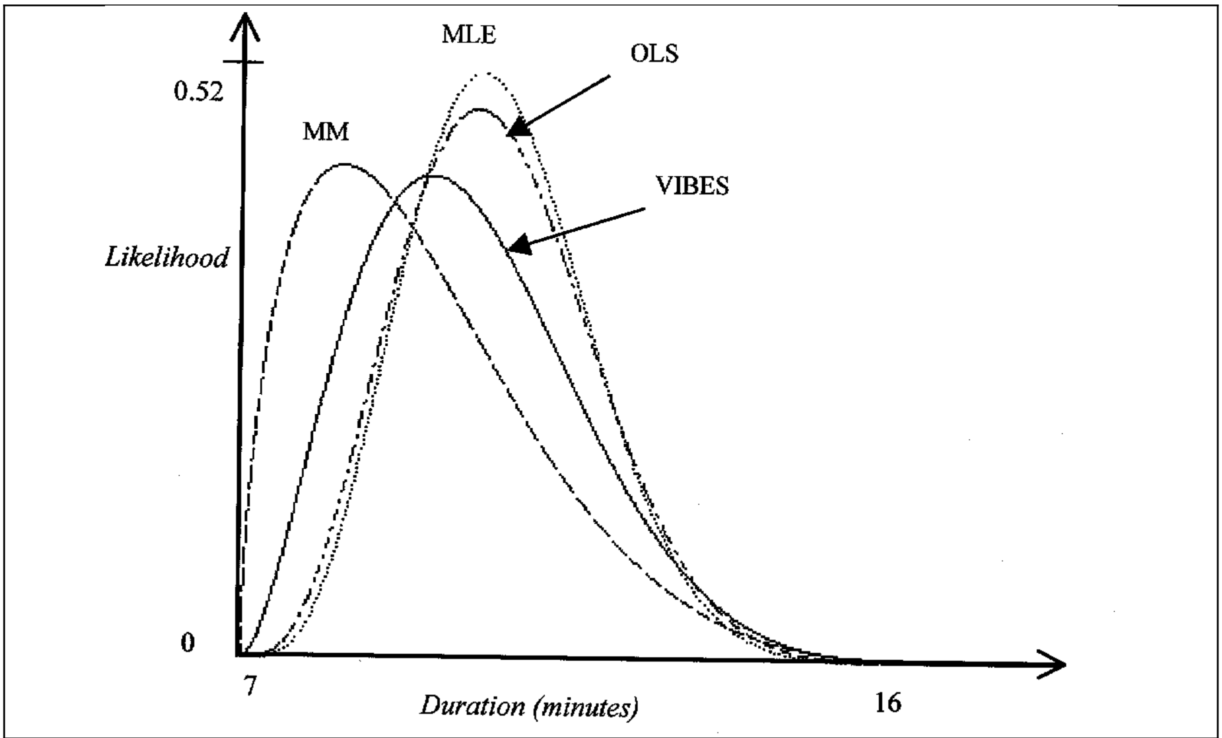

Figure 2. Methodology and VIBES Calculated vs. Actual Data and Betafit Calculated Beta PDF Shape.

5 CONCLUSION

Simulation applications often use 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 [8] lay in the beta region. Their research agreed with the findings of several other researchers. Thus, it was concluded that the beta distribution is suitable for modeling durations of construction processes.

Due to the natural range and mode of construction activity durations, the parameters of the beta distribution must be positive. Consequently, a subset of the beta curves is used to represent the duration of construction activities. The restricted subset of beta curves corresponds to the unimodal polynomial graphs.

Data quality and availability control how estimation of PDF parameters is approached. When data is available, an expert can use a number of numerical methods to compute the parameters of the underlying PDF. In the absence of data, an expert uses subjective activity-time information to estimate the parameters of the PDF expected to describe the construction process. Based on the research in experimental psychology, it is recommended that the subjective information used to define beta parameters include the minimum and maximum activity durations, as well as the mode and the 75th or 25th percentiles.

AbouRizk et al. [15] developed VIBES to compute the shape parameters of the beta PDF using the subjective information estimated by an expert. VIBES requires four activity time characteristics to define a unique beta PDF for an activity. These time characteristics can be the minimum, the mode, the maximum and the 75th percentile duration.

The proposed method for estimating the beta PDF relates the minimum activity duration, to the mode, 75th percentile, and maximum durations by factors or ratios. The minimum duration of the activity was calculated deterministically using the equipment manufacturer's machine data and a set of physical site constraints.

The mode was calculated by multiplying the minimum duration of the activity by a job efficiency factor. In the example studied 0.83, the factor used, represents a 50-minute work hour. It is realized that every job is different and the estimator must consider project specific factors when selecting an efficiency factor.

The maximum duration for a haul cycle operation was considered as the time at which management would notice a disruption and take action. In this study the duration was calculated by multiplying the mode of an activity by a factor.

The fourth activity time characteristic, the 75th percentile duration, was related to the mode of the activity by a ratio of 1.2 following the work by Wilson et al. [2].

It is shown in this research that using a ratio methodology as suggested here to develop activity duration inputs to VIBES yields a beta PDF which closely represents the underlying probability distribution of the construction process. Finding a relationship among the activity-time characteristics that define the PDF of an activity's duration reduces the degree of subjectivity in formulating simulation models for process control.

6 REFERENCES

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

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

[3] Klein R.W., and Baris, P.M. (1990). "Selecting and generating variates for modeling service times." *Research Memorandum No. 90-2*, Purdue Univ., West Lafayette, IN.

[4] Touran A. (1997). "Probabilistic model for tunneling project using Markov chain." *Journal of Construction Engineering and Management, ASCE, 123:* 444-449

[5] Maio, C. (1998). *Investigation of probability distribution functions for haul cycle simulation*. MS Thesis, Arizona State Univ., Tempe, AZ.

[6] Farid, F., and Koning, T.L. (1994). "Simulation verifies queueing program for selecting loader-truck fleets." *Journal of Construction Engineering and Management, ASCE, 120(2):* 386-404

[7] O'Shea, J.F., Slutkin, G.N., and Shaffer, L.R. (1964). "An application of the theory of queues to the forecasting of shovel-truck fleet production." *Construction Research Series No.3,* Dept. of Civil Engr, Univ. of Illinois, Urbana, IL.

[8] AbouRizk, S.M., and Halpin, D. W. (1992). "Statistical properties of construction duration data." *Journal of Construction Engineering and Management, ASCE, 118(3):* 525-543

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

[10] Peterson, C., and Miller, A. (1964). "Mode, median, and mean as optimal strategies." *Journal of Experimental Psychology, 68(4):* 363-367

[11] Peterson, C., and Beach, L. R. (1967). "Man as an intuitive statistician." *Psychology Bulletin, 68:* 29-46

[12] Alpert, M., and Raiffa, H. (1969). "A progress report on the training of probability assessors." *Judgement under uncertainty: Heuristics and biases.* Cambridge Univ. Press, Cambridge, England: 294-305

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

[14] Lichtenstein, S., Fischhoff, B., and Phillips, L.D. (1977). "Calibration of

probabilities: The state of the art to 1980." *Judgement under uncertainty: Heuristics and biases.* Cambridge Univ. Press, Cambridge, England. 306-334

[15] AbouRizk, S.M., Halpin, D. W., and Wilson J. R. (1991). "Visual interactive fitting of beta distributions." *Journal of Construction Engineering and Management*, ASCE, 117(4): 589-605

[16] AbouRizk, S.M., Halpin, D. W., and Wilson J. R. (1995). *BETAFIT Application version 1.0,* Univ. of Alberta, Edmonton, Alberta, Canada.