

5-1-2009

Estimating Task Duration in PERT using the Weibull Probability Distribution

Edward L. McCombs

The Charles Machine Works, Perry, Oklahoma, emccombs@ditchwitch.com


Matthew E. Elam

Texas A&M University, Matthew_Elam@tamucommerce.edu

David B. Pratt

Oklahoma State University, david.pratt@okstate.edu

Follow this and additional works at: <http://digitalcommons.wayne.edu/jmasm>

 Part of the [Applied Statistics Commons](#), [Social and Behavioral Sciences Commons](#), and the [Statistical Theory Commons](#)

Recommended Citation

McCombs, Edward L.; Elam, Matthew E.; and Pratt, David B. (2009) "Estimating Task Duration in PERT using the Weibull Probability Distribution," *Journal of Modern Applied Statistical Methods*: Vol. 8: Iss. 1, Article 26.

Available at: <http://digitalcommons.wayne.edu/jmasm/vol8/iss1/26>

This Regular Article is brought to you for free and open access by the Open Access Journals at DigitalCommons@WayneState. It has been accepted for inclusion in Journal of Modern Applied Statistical Methods by an authorized administrator of DigitalCommons@WayneState.

Estimating Task Duration in PERT using the Weibull Probability Distribution

Edward L. McCombs
The Charles Machine Works
Perry, Oklahoma

Matthew E. Elam
Texas A&M University
Commerce

David B. Pratt
Oklahoma State University

The Weibull probability distribution can be used as an alternative model for task time estimates in the PERT estimating methodology. It has the same advantages as the traditional beta distribution for this application. It has additional benefits, however, that make it a preferred option.

Key words: PERT; Weibull Probability Distribution; Beta Probability Distribution; Pearson Skew Plot.

Introduction

Malcolm, Roseboom, Clark, and Fazar (1959) published the project time estimating methodology that they developed for Project PERT (Program Evaluation Research Task) under the Polaris Ballistic Missile Program. The development of their methodology was motivated by the fact that there was little or no historical data available upon which to base estimates of task durations. In subsequent years, this methodology has been applied in wide variety of fields. However, various authors have identified five significant issues with PERT (e.g., Cottrell, 1999; Premachandra, 2001; Pleguezuelo et al., 2003):

1. Accurately estimating the optimistic, most likely and pessimistic durations of an activity is, in general, difficult.
2. The calculated mean and variance of the specific activity durations are estimates of the actual mean and variance.

3. The beta distribution is assumed to provide an adequate model for activity durations.
4. PERT focuses on the critical path when computing project completion time probabilities.
5. The methodology requires that multiple time estimates be developed. These estimates can be costly.

Focus on items two and three in the above list. Specifically, consider the Weibull distribution as an alternative to the traditionally used beta distribution. It is shown, among other advantages, that the Weibull distribution does not require approximations for the mean and variance, as does the beta distribution.

Beta Probability Distribution

The beta probability distribution has traditionally been used as the distribution of choice in PERT analyses based on the following advantages (Fente, Schexnayder, & Knutson, 2000; Lu & AbouRizk, 2000):

1. It is continuous.
2. It has finite endpoints.
3. It has a defined mode between its endpoints.
4. It is capable of describing both skewed and symmetric activity time distributions.

For the current discussion, consider stated advantage two. The second advantage makes sense from a practical point of view in that every activity must have a maximum completion time. The difficulty with this stated advantage,

Edward L. McCombs is the Director of Operations Support. Email him at: emccombs@ditchwitch.com. Matthew E. Elam is an Associate Professor of Industrial Engineering and is an ASQ Certified Quality Engineer. Email him at: Matthew_Elam@tamu-commerce.edu. David B. Pratt is an Associate Professor in the School of Industrial Engineering and Management. Email him at: david.pratt@okstate.edu.

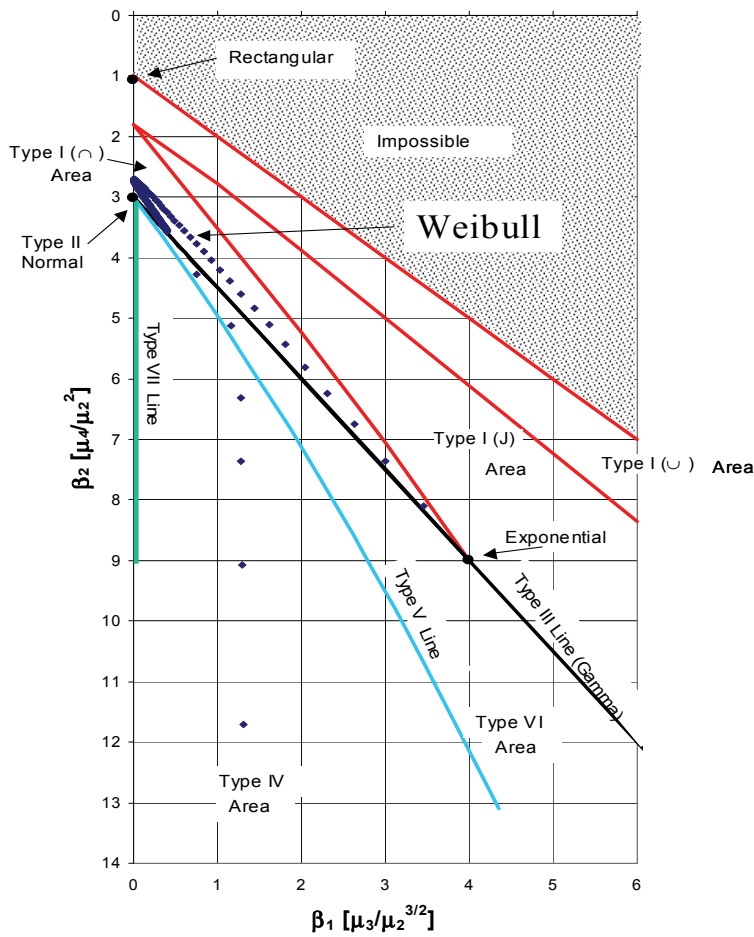
however, is determining the value of this maximum. For example, the truck travel time study described by Fente, et al. (2000). The maximum travel time is computed as two times the mode. This assumption is supported by the reasoning that management would notice the slow moving truck and take actions necessary to reduce its travel time. Undoubtedly, this type of assumption is necessary when a decision maker is constrained to using the beta probability distribution. However, it may be more reasonable to consider a distribution that can accommodate a longer tail probability than is allowed by the beta distribution.

Weibull Probability Distribution

The Weibull probability distribution can accommodate this longer right tail probability. Additionally, the Weibull distribution has advantages one, three, and four as listed above for the beta probability distribution.

Figure 1 shows a Pearson skew plot (Pearson, 1920; Pearson & Tukey, 1965) with the Weibull probability distribution plotted. The Type I areas shown in Figure 1 can be represented by the beta probability distribution. Figure 1 shows that the Weibull distribution can approximate distributions ranging from close to the normal to the exponential, can accommodate

Figure 1: Pearson's Skew Curves Plot Showing the Weibull Distribution



ESTIMATING TASK DURATION USING THE WEIBULL PROBABILITY DISTRIBUTION

distributional skewness (β_1), and can approximate activity duration models with fatter distributional tails (β_2) than can be accommodated by the beta probability distribution. Note that the Weibull probability distribution divides the triangular Type I (\cap shaped) region. It is expected that the Weibull probability distribution can satisfactorily describe those Type I (\cap shaped) models that are coincident with the beta models located in this region. Further research may also show that the Weibull probability distribution can serve as a proxy for the entire Type I (\cup shaped) region. An additional advantage of the Weibull distribution is that it should also satisfactorily model some Type III, IV, V, and VI distributions. This would be useful considering the review by Maio, et al. (2000), which shows that the beta probability distribution is not the best model for all construction operations.

Equations (1) – (5) show the Weibull probability density function, reliability, mode, variance, and mean formulas, respectively, from Ebeling (1997):

$$f(x) = (\beta/\theta)(x/\theta)^{\beta-1} \exp[-(x/\theta)^\beta] \quad (1)$$

$$R(x) = \exp[-(x/\theta)^\beta] \quad (2)$$

$$M = \text{mode} = \theta(1-1/\beta)^{1/\beta} \text{ for } \beta > 1 \quad (3)$$

$$\sigma^2 = \theta^2 \left\{ \Gamma(1+2/\beta) - [\Gamma(1+1/\beta)]^2 \right\} \quad (4)$$

$$\mu = \text{mean} = x_0 + \theta\Gamma(1+1/\beta) \quad (5)$$

where β is the shape parameter, θ is the scale parameter, Γ is the gamma function, and x_0 shifts the mean on the x-axis.

Methodology

Let x_a be the lower expert judgment percentile estimate, x_b be the upper expert judgment percentile estimate, and M be the most likely expert judgment estimate. Equation (2) can, therefore, be rewritten to solve for x_a and x_b ,

with the results as Equations (6) and (7), respectively:

$$x_a = \theta \left[\ln(1/R(x_a)) \right]^{1/\beta} \quad (6)$$

$$x_b = \theta \left[\ln(1/R(x_b)) \right]^{1/\beta} \quad (7)$$

The traditional form of the Weibull distribution has defined left and right bounds of zero and infinity, respectively. As a result, only two of the three expert opinion estimates are required to calculate the distributional parameters. If x_a and x_b , as well as their respective percentiles, are known, then Equations (6) and (7) can be used to calculate the shape parameter β in Equation (8):

$$\begin{aligned} \frac{x_a}{x_b} &= \frac{\theta \left[\ln(1/R(x_a)) \right]^{1/\beta}}{\theta \left[\ln(1/R(x_b)) \right]^{1/\beta}} \\ \Rightarrow \frac{x_a}{x_b} &= \left(\frac{\ln(R(x_a))}{\ln(R(x_b))} \right)^{1/\beta} \\ \Rightarrow \beta &= \frac{\ln[\ln(R(x_a))/\ln(R(x_b))]}{\ln(x_a/x_b)} \end{aligned} \quad (8)$$

Substituting the calculated value of β into Equation (6) or (7) allows the scale parameter to be calculated.

If x_a or x_b and also M are known, then Equation (6) or (7) and also Equation (3) can be used to calculate the shape parameter β as in Equation (9):

$$\begin{aligned} \frac{M}{x_b} &= \frac{\theta[1-1/\beta]^{1/\beta}}{\theta \left[\ln(1/R(x_b)) \right]^{1/\beta}} \\ \Rightarrow \frac{M}{x_b} &= \left(\frac{1/\beta - 1}{\ln(R(x_b))} \right)^{1/\beta} \\ \Rightarrow \left(\frac{x_b}{M} \right)^\beta &= \left(\frac{1}{\beta} - 1 \right) = \ln(R(x_b)) \end{aligned} \quad (9)$$

Finally, the scale parameter θ can be calculated using Equation (6) as in Equation (10):

$$\theta = M / (1 - 1/\beta)^{1/\beta} \quad (10)$$

An additional advantage to using the Weibull distribution exists. Specifically, a user is allowed to use whichever percentiles he/she feels are the most appropriate. Moreover, not only is a user now able to use percentiles other than the 5 and 95 percentiles with equal accuracy, the percentiles need not be symmetric; i.e., the 5 and 90 percentiles could be used.

Consider the situation in which there is a zero probability of an event occurring before a certain threshold time. For the Weibull distribution, a threshold value, x_0 , can be included as in Equations (11) – (14) from Ebeling (1997):

$$f(x - x_0) = (\beta/\theta) \left((x - x_0)/\theta \right)^{\beta-1} \times \exp \left[- \left((x - x_0)/\theta \right)^\beta \right] \quad (11)$$

$$R(x - x_0) = \exp \left[- \left((x - x_0)/\theta \right)^\beta \right] \quad (12)$$

$$M = \text{mode} = x_0 + \theta (1 - 1/\beta)^{1/\beta} \text{ for } \beta > 1 \quad (13)$$

$$x = \theta \left[\ln(1/R(x)) \right]^{1/\beta} + x_0 \quad (14)$$

The equation for the variance remains unchanged. The addition of a threshold value does not change the basic shape of the distribution, only its location on the x-axis. Because the left boundary is no longer known and there is an additional parameter, additional information needs to be incorporated.

The calculation of the ratio

$$\frac{(M - x_A)}{(x_B - x_A)}$$

in terms of its respective components from Equations (13) and (14) is shown in Equation (15):

$$\frac{M - x_A}{x_B - x_A} = \frac{x_0 + \theta (1 - 1/\beta)^{1/\beta} - \left\{ x_0 + \theta \left[\ln(1/R(x_a)) \right]^{1/\beta} \right\}}{x_0 + \theta \left[\ln(1/R(x_b)) \right]^{1/\beta} - \left\{ x_0 + \theta \left[\ln(1/R(x_a)) \right]^{1/\beta} \right\}} \quad (15)$$

The threshold value cancels, as do the scale parameters, with the result in Equation (16):

$$\frac{M - x_A}{x_B - x_A} = \frac{(1 - 1/\beta)^{1/\beta} - \left[\ln(1/R(x_a)) \right]^{1/\beta}}{\left[\ln(1/R(x_b)) \right]^{1/\beta} - \left[\ln(1/R(x_a)) \right]^{1/\beta}} \quad (16)$$

The shape parameter β can be computed using a solver program (e.g., Microsoft Excel's Solver® function). Because the threshold value is unknown, the equation for the mode cannot be used to calculate the scale parameter θ as in Equation (10). However, the variance constant K can be calculated and used to calculate the variance.

Using the calculated shape parameter β and a scale parameter θ equal to 1.0, the temporary variance is calculated as in Equation (17):

$$\sigma_{temp}^2 = \theta_{temp}^2 \left\{ \Gamma(1 + 2/\beta) - \left[\Gamma(1 + 1/\beta) \right]^2 \right\} \quad (17)$$

where $\theta_{temp} = 1$. Next, the temporary x-axis values for the required lower and upper percentiles are calculated using Equations (6) and (7). The variance constant K can now be calculated as shown in Equation (18):

ESTIMATING TASK DURATION USING THE WEIBULL PROBABILITY DISTRIBUTION

$$K = (x_{b \text{ temp}} - x_{a \text{ temp}}) / \sigma_{\text{temp}} \quad (18)$$

The variance based on the actual data can now be calculated using Equation (19):

$$\sigma^2 = ((x_b - x_a) / K)^2 \quad (19)$$

With the variance known, the actual scale parameter θ can be calculated as shown in Equation (20):

$$\begin{aligned} \sigma^2 &= \theta^2 \{ \Gamma(1+2/\beta) - [\Gamma(1+1/\beta)]^2 \} \\ \Rightarrow \theta &= \sqrt{\sigma^2 / \{ \Gamma(1+2/\beta) - [\Gamma(1+1/\beta)]^2 \}} \end{aligned} \quad (20)$$

With β and θ known, the threshold value can be calculated using the most likely value M as in Equation (21):

$$\begin{aligned} M = \text{mode} &= x_0 + \theta(1-1/\beta)^{1/\beta} \\ \Rightarrow x_0 &= M - \theta(1-1/\beta)^{1/\beta} \end{aligned} \quad (21)$$

All of the parameters for the required Weibull distribution in Equation (11) can now be calculated.

Example

As an example of these parameter calculations, consider the truck travel example as shown in Fente, et al. (2000). The travel distance is 3.7 – 3.9 km. The traditional PERT information is as follows. The minimum possible travel time is based on the physical characteristics of the project site and the truck manufacturer's specifications and is equal to 7.67 minutes. The most likely travel time is 9.21 minutes. The maximum travel time is 18.42 minutes. Additionally, it is given that the 75th percentile estimate is 11.05 minutes. Fente, et al. (2000) report that a beta probability distribution with parameters $\alpha = 1.898$ and $\beta = 6.372$ is a reasonable model for the truck travel time distribution.

To use the methodology presented in this paper for the offset Weibull probability distribution, only two percentile estimates and

the mode estimate are required. Because the 75th percentile estimate is explicitly stated, it is an obvious choice for one of the required estimates. The required second boundary estimate requires an assumption with regard to the percentile that it represents. The lower boundary of 7.67 minutes was selected because it is a finite boundary. Specifically, the lower boundary is assumed to represent the 0.01 percentile. As a result, the following parameters were calculated in Table 1. Figure 2 shows the resulting Weibull distribution plotted with the resulting beta distribution as derived in Fente, et al. (2000). The two curves converge together as the value of the lower percentile converges to zero.

Because the proposed Weibull model and the resulting beta model, as presented by Fente, et al. (2000), are both estimates of the unknown underlying distribution, it is not useful to compare the fits via a goodness-of-fit test. However, visually it seems that either model could satisfactorily model the underlying distribution. So why consider the Weibull model over the beta model? First, the Weibull model required only three estimates, while the beta model required four. Second, the Weibull model can easily be developed in a Microsoft Excel® spreadsheet. Finally, when compared to the traditional PERT methodology, the Weibull model does not require an estimate of the variance -- this value is calculated exactly (as is the mean value). Moreover, with regard to this last point, the only errors associated with the Weibull model relate to the accuracy of the original estimates and whether the Weibull model can satisfactorily describe the underlying distribution.

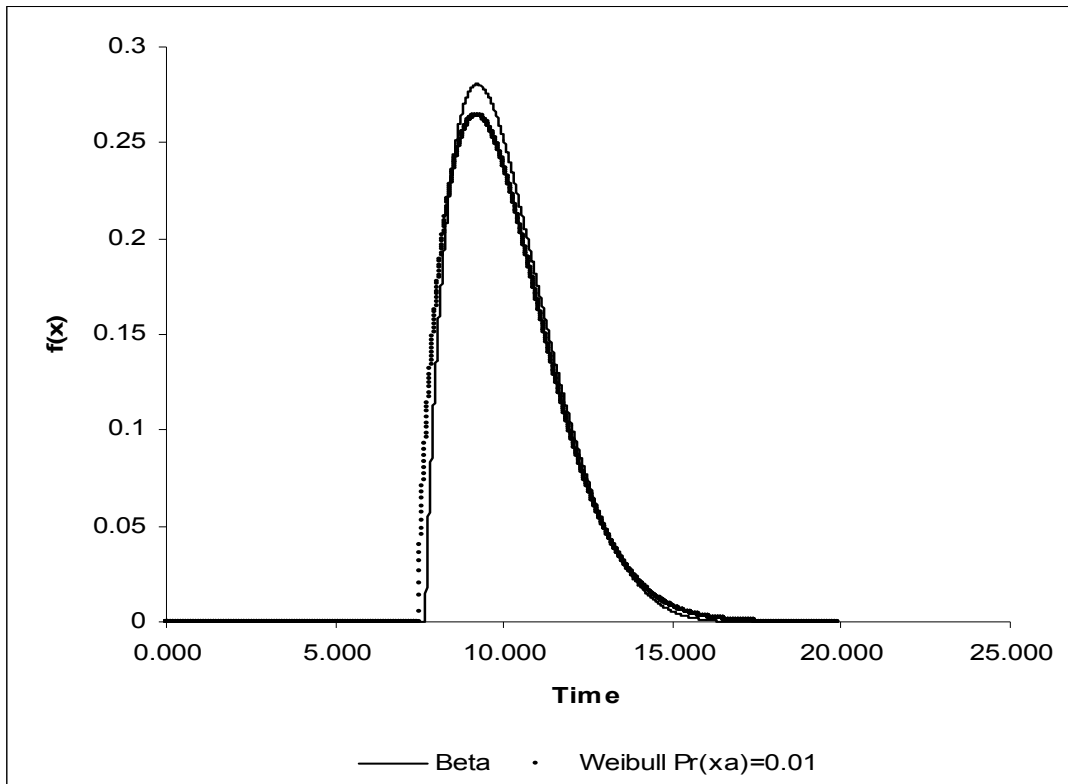
Conclusion

If an activity's duration time starts at $t=0$, and one can estimate at least two of three estimates (x_a , x_b , and M) of an unknown distribution, then one can estimate the unknown distribution with a Weibull probability distribution. This approach could be beneficial in situations where two of the three estimates (lower percentile, most likely, upper percentile) can be assumed to be known with greater certainty than the third estimate. If all three estimates are assumed known with equal certainty and/or an activity's

Table 1: Results of Fitting the Presented Weibull Model to the Data in Fente, et al. (2000)

Parameter	Eq.	Substituted Values	Result
β	(16)	$M = 9.21, x_a = 7.67$ $x_b = 11.05, R(x_a) = 1-0.01$ $R(x_b) = 1-0.75$	1.6900
$x_{a \text{ temp}}$	(6)	$\theta = 1, R(x_a) = 1-0.01, \beta = 1.6900$	0.0657
$x_{b \text{ temp}}$	(7)	$\theta = 1, R(x_b) = 1-0.75, \beta = 1.6900$	1.2132
σ^2_{temp}	(17)	$\theta = 1, \beta = 1.6900$	0.2953
K	(18)	$x_{b \text{ temp}} = 1.2132, x_{a \text{ temp}} = 0.0657$ $\sigma^2_{\text{temp}} = 0.2953$	2.1118
σ^2	(19)	$x_b = 11.05, x_a = 7.67, K = 2.1118$	2.5618
θ	(20)	$\sigma^2 = 2.5618, \beta = 1.6900$	2.9456
x_0	(21)	$M = 9.21, \theta = 2.9456, \beta = 1.6900$	7.4764
μ	(5)	$x_0 = 7.4764, \theta = 2.9456, \beta = 1.6900$	10.1056

Figure 2: Plot of the Truck Travel Time from Fente, et al. (2000) and the Weibull Model



ESTIMATING TASK DURATION USING THE WEIBULL PROBABILITY DISTRIBUTION

duration time does not begin at $t = 0$, then it is advisable to use the shifted Weibull distribution.

The objective of this article was to provide an alternative approach to the traditional Project PERT methodology using the Weibull probability distribution. It was shown that by using the Weibull probability distribution it is not necessary to estimate a future activity's mean or variance. These values are calculated exactly and have only the uncertainty inherent in the original subjective estimates and the uncertainty as to whether the Weibull probability distribution accurately models the underlying distribution of future activity times. The ease of use and the reduction in uncertainty with the proposed Weibull model will benefit both practitioners and researchers.

The beta distribution unarguably is more robust within the Pearson Type I (\cup shape) region than the Weibull distribution. However, as Lau, Lau, and Zhang (1996) have pointed out, there is a practical application for distributional models that are more robust to the third and fourth moments. The Weibull distribution satisfies this need. The true test with regard to the applicability of the Weibull distribution will be its ability to accurately model a broad range of actual problems.

References

Cottrell, W. D. (1999). Simplified program evaluation and review technique (PERT). *Journal of Construction Engineering and Management*, 125(1), 16-22.

Ebeling, C. (1997). *An introduction to reliability and maintainability engineering*. NY: McGraw-Hill.

Fente, J., Schexnayder, C., & Knutson, K. (2000). Defining a probability distribution function for construction simulation. *Journal of Construction Engineering and Management*, 123(3), 234-241.

Lau, A. H., Lau, H., & Zhang, Y. (1996). A simple and logical alternative for making PERT time estimates. *IIE Transactions*, 18, 183-192.

Lu, M. & AbouRizk, S. M. (2000). Simplified CPM/PERT simulation model. *Journal of Construction Engineering and Management*, 126(3), 219-226.

Maio, C., Schexnayder, C., Knutson, K., & Weber, S. (2000). Probability distribution functions for construction simulation. *Journal of Construction Engineering and Management*, 126(4), 285-292.

Malcolm, D. G., Roseboom, J. R., Clark, C. E., & Fazar, W. (1959). Application of a technique for research and development program evaluation. *Operations Research*, 7, 646-669.

Pearson, K. (1920). On the probable errors of frequency constants. *Biometrika*, 13(1), 113-132.

Pearson, E. S. & Tukey, J. W. (1965). Approximate means and standard deviations based on distance between percentage points of frequency curves. *Biometrika*, 52(3/4), 533-546.

Pleguezuelo, R. H., Perez, J. G., & Rambaud, S. C. (2003). A note on the reasonableness of PERT hypotheses. *Operations Research Letters*, 31, 60-62.

Premachandra, I. M. (2001). An approximation of activity duration distribution in PERT. *Computers and Operations Research*, 28, 443-452.