Journal of the Operational Research Society (2007) 58, 827–831

www.palgrave-journals.com/jors

A lognormal approximation of activity duration in PERT using two time estimates

S Mohan[1], M Gopalakrishnan[1]*, H Balasubramanian[2] and A Chandrashekar[3]

[1]Arizona State University, Phoenix, AZ, USA; [2]Arizona State University, Tempe, AZ, USA; and [3]AXP Project Center of Excellence, IBM Corporation, Phoenix, AZ, USA

The success behind effective project management lies in estimating the time for individual activities. In many cases, these activity times are non-deterministic. In such situations, the conventional method (project evaluation and review technique (PERT)) obtains three time estimates, which are then used to calculate the expected time. In practice, it is often difficult to get three accurate time estimates. A recent paper suggests using just two time estimates and an approximation of the normal distribution to obtain the expected time and variance for that activity. In this paper, we propose an alternate method that uses only two bits of information: the most-likely and either the optimistic or the pessimistic time. We use a lognormal approximation and experimental results to show that our method is not only better than the normal approximation, but also better than the conventional method when the underlying activity distributions are moderately or heavily right skewed.

Journal of the Operational Research Society (2007) **58,** 827–831. doi:10.1057/palgrave.jors.2602204

Published online 3 May 2006

Keywords: PERT; lognormal distribution; estimation; probability

Introduction

Project Evaluation and Review Technique (PERT) has long been used for managing complex projects. Real time project management and scheduling relies almost entirely on the accuracy of the activity time estimates. It has been well recognized that in a great number of situations, activity times are probabilistic, that is, they are not known with certainty. In these cases, literature suggests using a beta distribution to describe the activity duration. In order to estimate the mean and the variance of the underlying beta distribution, the scheduler should obtain three time estimates (optimistic—a, most likely—m, and pessimistic—b) for each activity. Once the scheduler has values for a, m, and b, the mean and the variance are estimated using approximation formulas. These formulas are good estimators of the true mean when the distribution is fairly symmetric. When one tail is much thicker than the other, the results could be misleading. Keefer and Verdini (1993) have shown that the average error in estimating the mean is over 40% and the average error for the variance is close to 550%. As a result, many authors have developed modifications to improve the accuracy of the original formulae. We present a summary of some of the main results in the next section.

Cottrell (1999) points out that it is often very difficult to obtain three time estimates. Callahan *et al* (1992) and Moder *et al* (1983) acknowledge that PERT is not much used in practice, one reason being the need to obtain three time estimates. As explained by one manager, estimating a and b simultaneously increases the spread and clients tend to view such estimates with skepticism. The manager added that he would be more comfortable estimating the most likely and either the optimistic or the pessimistic time estimate as the spread is narrower in these cases. Cottrell (1999) suggests using two time estimates and an approximation of the normal distribution to determine expected time and variance.

In this paper, we propose an alternate method for determining expected activity times with only two time estimates, and our method performs better than the normal approximation suggested by Cottrell (1999). Further, our method is also robust enough to work well in the conditions in which the conventional PERT method fails. The rest of the paper is organized as follows. The next section briefly discusses the various approximations with three and two time estimates. We present our new lognormal approximation with two time estimates and experimental results in the following sections. Finally, we present some insights and conclusions.

Time estimates with three parameters

Several researchers have suggested modifications of the original PERT formulae or have suggested new ones with three parameters. Farnum and Stanton (1987) have shown that the estimates fail when the modal value is near the upper

**Correspondence: M Gopalakrishnan, School of Global Management and Leadership, ASU, PO Box 37100, Phoenix, Arizona 85069-7100, USA.* E-mail: mohan@asu.edu

and lower limits of the distribution and suggest modifications to the formulae. Golenko-Ginzburg (1988) has developed improved estimates for the mean and variance when the estimated mode is located in the tails of the distribution. Other similar suggestions to handle imprecision in PERT parameter estimations have been made by Troutt (1989) and Lau *et al* (1996). Finally, Premachandra (2001) proposes new approximations for the mean and the variance that use the traditional three estimates and are free of any distributional assumptions. This approximation outperforms all other estimates proposed in the literature in terms of average percentage error. We include Premachandra's approximations in our computational experiments for comparison purposes.

Most of the approximations mentioned thus far assume that the beta distribution is used to model uncertainty in the activity times. Berny (1989) has proposed a more general distribution and claims that it is more realistic since the parameters are easier to estimate. Williams (1992) has suggested using the triangular distribution, which is simple and easy to understand. Johnson (2002) has explored the use of triangular approximations to estimate the mean and variance based on estimates of the median and 5% points. However, all of the above approximations and modifications will work well only if the scheduler is able to obtain the three time estimates with a reasonable level of confidence. This may not be possible all the time. A few researchers have addressed this issue and have developed approximations when only two time estimates are available.

Time estimation with two parameters

Golenko-Ginzberg (1988) suggests that the most likely activity time may not be required. He proposes a fixed value of $m=(2a+b)/3$ for any activity, and shows that this value is not statistically different from those proposed by values determined by analysts. Cottrell (1999) suggests using only two estimates m and b and uses a normal approximation. In this case, m becomes the mean and b a point 'z' standard deviations from the mean. In the 16 examples used, his method performs comparably with the PERT method, although the variances generated by his method are higher. Cottrell (1999) recognizes that he is imposing a symmetric distribution to approximate a non-symmetric distribution and stresses the need for developing better approximations.

Cottrell's (1999) paper does not indicate how well the method performs when some activities are highly skewed and also when there are a large number of critical activities. Given the lack of information in that paper, our primary objective is to come up with an approximation that uses only two estimates and is better than the Cottrell's normal approximation. A secondary objective is to verify the performance of the normal approximation under a variety of conditions.

Lognormal approximation

We assume that only two pieces of information are available which could be either (a and m), or (b and m). In general, if closed form expressions $a=g(\mu,\sigma)$ and $m=h(\mu,\sigma)$ can be found for some distribution, one can solve this set of equations simultaneously for μ and σ. A similar approach will work when $b=k(\mu,\sigma)$ and $m=h(\mu,\sigma)$.

A usual supposition in project management is that the distribution for most activities is right skewed, which the normal distribution is not. The lognormal distribution is right skewed, flexible and has closed form expressions for $a=g(\mu,\sigma)$, $m=h(\mu,\sigma)$, and $b=k(\mu,\sigma)$. The lognormal is a monotone transformation of a normal. If $X\sim NORM(\mu,\sigma^2)$ then $Y=\exp(X)$ is lognormal.

$$E(Y)=\exp(\mu+1/2\sigma^2) \quad (1)$$

$$\text{and } VAR(Y)=\exp(2\mu+2\sigma^2)-\exp(2\mu+\sigma^2) \quad (2)$$

The mode can be found by differentiation, which gives $\log(m)=\mu^*-z\sigma^{*2}$, where μ^* and σ^* are the mean and standard deviation of the underlying normal distribution X. Since we assume a lognormal distribution, the log of the optimistic estimate will be z standard deviations to the left of the mean. Therefore, we have the expression: $\log(a)=\mu^*-z\sigma^*$. Next, we attempt to find the mean and the standard deviation of the normal distribution that satisfies these two equations. The equations above can be combined to yield a quadratic equation in σ^*, which can be solved to get

$$\sigma^*=1/2z-\left[1/4z^2+\log(a/m)\right]^{1/2} \quad (3)$$

$$\mu^*=\log[a]+z\sigma* \quad (4)$$

A similar result using b and m yields:

$$\sigma'=-1/2z+\left[1/4z^2+\log(b/m)\right]^{1/2} \quad (5)$$

$$\mu'=\log(b)-z\sigma' \quad (6)$$

Based on pilot runs, we used $z=3$ for the *lognormal*(a,m) and *lognormal*(b,m) cases. We, therefore, have six alternate methodologies: PERT, Premachandra, *normal*(a,m), *normal*(b,m), *lognormal*(a,m), and *lognormal*(b,m). *normal*(a,m) and *normal*(b,m) are the methods using Cottrell's approximation. All the approximations use two pieces of information except the PERT and the Premachandra method.

Experimentation

In order to verify our method, we simulated critical paths with different numbers of activities and skewness. We varied the number of critical activities r, from 10 to 100. For each of these activities, four numbers were selected as follows:

- a, the lower bound for possible completion time $\sim$ UNIF(1, 4)
- b, the upper bound for possible completion time $\sim$ UNIF(8, 12)
- α, one of the two shape parameters required for the beta distribution $\sim$ UNIF(1, 5)
- k, a parameter that determined how close the mode, m, is to the lower bound, was varied over three ranges—UNIF(1, 2), UNIF(2,4), and UNIF(4, 6).

When $k=1$, the beta distribution is symmetric and as $k \rightarrow \infty$, the beta distribution becomes highly skewed. The mode, m, and the second shape parameter for the beta distribution, β, were found using the equations $k=(b–m)/(m–a)$ and $(\alpha-1)/(\alpha+\beta-2)=(m-a)/(b-a)$.

In order to compare the performance of the six methods to be studied, we needed the exact value or the 'gold standard'. By transforming a standardized (0, 1) beta distribution as $y=a+(b–a)x$, we can obtain a distribution on the interval (a, b). The mean and variance of the transformed distribution can be evaluated as

$$\mu = a + (b-a) * \alpha/(\alpha + \beta) \quad (7)$$

$$\text{and} \quad \sigma^2 = (b-a)^2) * \alpha\beta/[(\alpha + \beta + 1)(\alpha + \beta)^2] \quad (8)$$

We generated the values for each activity and then used the different methods to estimate activity mean and variance. We then calculated the true activity means and variances using formulae (7) and (8) and used this to calculate true mean project completion times and its 95th percentile. We used this to calculate the percentage error for each method as given below.

$$\text{Percentage error in mean} = [(\text{Estimated value} - \text{True value})/\text{True value}] * 100$$

If the percentage error is negative, we discard the negative sign and consider only the absolute value of the percentage error. We simulated 1000 paths for each range of k and each setting of r (number of activities on the critical path). We have three ranges for k (UNIF(1, 2), UNIF(2, 4) and UNIF(4, 6)) and six levels representing the number of activities (10, 20, 40, 60, 80 and 100). For each combination of these two parameters, the average error in the means and the 95th percentile were determined.

Performance in estimation of the mean

The average error in the means is presented in Table 1 for each of the methods. From the results, it is very clear that the PERT distribution estimates the mean very closely when the distribution is fairly symmetric (ie $1 \leqslant k \leqslant 2$) with an average error of <1%. The *lognormal*(b, m) method is also good (the maximum average error is 1.12%) and is better than the Premachandra method but not as good as the PERT method. The normal methods have an average error of around 4%. However, when the levels of skewness increase, *lognormal*(b, m) begins to perform better than the PERT method and the Premachandra method. When $2 \leqslant k \leqslant 4$ the *lognormal*(b, m) method has a maximum average error of 2%

Table 1 Average percentage error in the estimation of the means for right skewed distributions

Method→		*PERT*	*Premachandra*	*Normal (a, m)*	*Normal (b, m)*	*Lognormal (a, m)*	*Lognormal (b, m)*
Information known→		*a, m, b*	*a, m, b*	*a, m*	*a, m*	*a, m*	*b, m*
Activities	*k*						
10	(1,2)	1.70	1.90	4.24	4.24	16.30	1.13
20	(1,2)	1.50	1.94	4.31	4.31	16.31	0.93
40	(1,2)	0.36	1.92	4.27	4.27	16.16	0.87
60	(1,2)	0.31	1.91	4.26	4.26	16.31	0.87
80	(1,2)	0.28	1.92	4.28	4.28	16.28	0.86
100	(1,2)	0.26	1.92	4.27	4.27	16.39	0.86
10	(2,4)	2.81	2.15	10.19	10.19	3.02	2.01
20	(2,4)	2.36	1.86	10.26	10.26	2.35	1.62
40	(2,4)	2.28	1.65	10.23	10.23	2.00	1.38
60	(2,4)	2.28	1.56	10.16	10.16	1.85	1.25
80	(2,4)	2.26	1.57	10.20	10.20	1.70	1.24
100	(2,4)	2.21	1.60	10.20	10.20	1.78	1.25
10	(4,6)	7.06	4.68	12.33	12.33	8.52	2.59
20	(4,6)	7.05	4.42	12.34	12.34	8.52	1.80
40	(4,6)	6.96	4.24	12.39	12.39	8.55	1.33
60	(4,6)	6.91	4.18	12.43	12.43	8.58	1.11
80	(4,6)	6.88	4.15	12.47	12.47	8.63	0.94
100	(4,6)	6.87	4.14	12.42	12.42	8.61	0.79

Table 2 Average percentage error in the estimation of the 95th percentile for right skewed distributions

Method→ *Information known→*		*PERT* *a, m, b*	*Premachandra* *a, m, b*	*Normal (a, m)* *a, m*	*Normal (b, m)* *a, m*	*Lognormal (a, m)* *a, m*	*Lognormal (b, m)* *b, m*
Activities	*k*						
10	(1,2)	1.74	3.04	6.97	6.56	25.66	2.07
20	(1,2)	1.30	2.82	6.36	6.21	23.28	1.35
40	(1,2)	0.95	2.55	5.75	5.69	21.36	0.76
60	(1,2)	0.80	2.43	5.48	5.45	20.70	0.51
80	(1,2)	0.72	2.38	5.34	5.32	20.11	0.38
100	(1,2)	0.68	2.34	5.23	5.21	19.86	0.30
10	(2,4)	3.51	2.65	14.28	13.98	4.21	2.63
20	(2,4)	2.65	1.93	13.32	13.20	2.83	1.86
40	(2,4)	2.41	1.54	12.45	12.41	2.13	1.43
60	(2,4)	2.37	1.38	11.98	11.96	1.85	1.23
80	(2,4)	2.32	1.40	11.80	11.79	1.70	1.21
100	(2,4)	2.26	1.42	11.64	11.63	1.67	1.20
10	(4,6)	9.21	7.04	17.19	16.96	11.42	4.58
20	(4,6)	8.57	6.14	15.93	15.85	10.68	3.13
40	(4,6)	8.01	5.45	15.02	14.99	10.13	2.12
60	(4,6)	7.77	5.17	14.62	14.60	9.89	1.70
80	(4,6)	7.62	5.00	14.38	14.36	9.77	1.40
100	(4,6)	7.54	4.91	14.13	14.12	9.64	1.23

(the overall average is well below 2%). As we go to $4 \leqslant k \leqslant 6$, the differences between the methods are even more marked. The *lognormal*(b, m) method performs especially well (average error of around 1.5%) in this range while the PERT distribution has an error in the 6.9% range and the Premachandra method (the second best method in the range) has an average error of 4.3%. The *lognormal*(a, m) method has an average error below 9% while and the *normal*(a, m) and *normal*(b, m) methods have errors of around 12%.

Performance in estimation of the 95th percentile

We also obtained the error in the 95th percentile and present the results in Table 2. When the level of skewness is mild that is $1 \leqslant k \leqslant 2$, the *lognormal*$(b, m)$ method is the best (0.9% average error) followed by the PERT method (1.03% average error), and then the Premachandra method (2.59% average error). The *lognormal*(a, m) method on the other hand, performs quite poorly in this range. When $2 \leqslant k \leqslant 4$, the *lognormal*$(b, m)$ method again outperforms both PERT and Premachandra methods. The *lognormal*(a, m)method too performs comparably in this range. The methods based on the normal distribution have poor performances. Finally, when the times are highly positively skewed, $4 \leqslant k \leqslant 6$, the performances of all the methods deteriorate but *lognormal*(b, m) remains the most robust. When the number of activities is large (100), the error for *lognormal*(b, m) is 1.5%, which no other method achieves in this range.

Our simulations are based on $k \geqslant 1$, that is, all activities have either symmetric or right skewed distributions. A valid concern relates to the presence of a few left skewed activities in the project. We simulated the extreme and improbable case where all activities are left skewed $(0.5 \geqslant k \geqslant 1)$. The results showed that the *lognormal*(b, m) method does not perform as well as the other methods (a 5% difference from the best method), and the *lognormal*(a, m) method performs very poorly. We, therefore, propose that if only b and m are known, then the *lognormal*(b, m) method should be used, whatever the number of activities and the levels of skewness. Again, if there is a priori knowledge that many activities are highly skewed, $(k > 4)$, then the lognormal approximation may be even preferable to the PERT approximation. If a and m are given, *lognormal*(a, m) should be used for moderate and high levels of skewness and *normal*(a, m) for low levels of skewness. Also if more than a few left-skewed activities are present, *normal*(a, m) is a better choice.

Conclusion

In this paper, we proposed a lognormal approximation for estimating activity times in PERT. The proposed approximation requires only two parameter estimates (which could either be a and m or b and m). We applied this method to several projects with simulated activity times. We have shown that, over a wide range of conditions, our method is robust and works better than other two point normal approximations suggested. It works exceptionally well when the distributions are highly positively skewed.

An extension of this paper would be to study the effect of errors in estimation of the various parameters, for example, the parameter *a*. We have, in this paper, ignored the effects of near critical or multiple critical paths and this is a definite area for future research. We also propose more empirical research on projects in construction, R & D and software industries, among others to determine the true nature of the underlying distribution and level of skewness that is realistic.

References

Berny J (1989). A new distribution function for risk analysis. *J Opl Res Soc* **40**: 1121–1127.

Callahan MT, Quakenbush DG and Rowings JE (1992). *Construction Project Scheduling*, McGraw-Hill: New York.

Cottrell W (1999). Simplified program evaluation and review technique (PERT). *J Construct Eng Mgmt* **125/1**: 16–22.

Farnum NR and Stanton LW (1987). Some results concerning the estimation of beta distribution parameters in PERT. *J Opl Res Soc* **38**: 287–290.

Golenko-Ginzburg D (1988). On the distribution of activity times in PERT. *J Opl Res Soc* **39**: 767–771.

Johnson D (2002). Triangular approximations for continuous random variables in risk analysis. *J Opl Res Soc* **53**: 457–467.

Keefer DL and Verdini WA (1993). Better estimation of PERT activity time parameters. *Mgmt Sci* **39/9**: 1086–1091.

Lau AHL, Lau HS and Zhang Y (1996). A simple and logical alternative to making PERT time estimates. *IIE Transactions* **28**: 183–192.

Moder JJ, Phillips CR and Davis EW (1983). *Project Management with CPM, PERT and Precedence Diagramming*, 3rd edn. Van Nostrand Reinhold: New York.

Premachandra IM (2001). An approximation of the activity duration distribution in PERT. *Comp and Opns Res* **28**: 443–452.

Troutt MD (1989). On the generality of the PERT average time formula. *Decis Sci* **20**: 410–412.

Williams TM (1992). Practical use of distribution in risk analysis. *J Opl Res Soc* **43**: 265–270.

Received August 2005;
accepted February 2006 after one revision