IMPLEMENTING BETA-DISTRIBUTION IN PROJECT MANAGEMENT

Doron GREENBERG

PhD, Department of Economics and Business Administration, Faculty of Social Science,
Ariel University Center (AUC) of Samaria, Ariel, Israel

E-mail: dorongreen2@gmail.com

Dimitri GOLENKO-GINZBURG

Prof, Department of Industrial Engineering and Management (Emeritus),
Ben-Gurion University of the Negev, Beer-Sheva, Israel
& Department of Industrial Engineering and Management,
Ariel University Center (AUC) of Samaria, Ariel, Israel

E-mail: dimitri@bgu.ac.il

Abstract: *A research is undertaken to justify the use of beta-distribution p.d.f. for man-machine type activities under random disturbances. The case of using one processor, i.e., a single resource unit, is examined. It can be proven theoretically that under certain realistic assumptions the random activity – time distribution satisfies the beta p.d.f.*

Changing more or less the implemented assumptions, we may alter to a certain extent the structure of the p.d.f. At the same time, its essential features (e.g. asymmetry, unimodality, etc.) remain unchanged.

The outlined above research can be applied to semi-automated activities, where the presence of man-machine influence under random disturbances is, indeed, very essential. Those activities are likely to be considered in organization systems (e.g. in project management), but not in fully automated plants.

Key words: *random activity duration; time – activity beta-distribution; operating by means of a single processor; convergence to a beta-distribution "family"*

1. Introduction

In PERT analysis [1-24, etc.] the activity-time distribution is assumed to be a beta-distribution, and the mean value and variance of the activity time are estimated on the basis of the "optimistic", "most likely" and "pessimistic" completion times, which are subjectively determined by an analyst. The creators of PERT [3, 17] worked out the basic concepts of PERT analysis, and suggested the estimates of the mean and variance values

$$\mu = \frac{1}{6}(a + 4m + b), \tag{1}$$

$$\sigma^2 = \frac{1}{36}(b - a)^2, \tag{2}$$

subject to the assumption that the probability density function (p.d.f.) of the activity time is

$$f_y(y) = \frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)} \frac{(y-a)^{\alpha-1}(b-y)^{\beta-1}}{(b-a)^{\alpha+\beta-1}}, \quad a < y < b, \quad \alpha,\beta > 0. \tag{3}$$

Here a is the optimistic time, b - the pessimistic time, and m stands for the most likely (modal) time.

Since in PERT applications parameters a and b of p.d.f. (3) are either known or subjectively determined, we can always transform the density function to a standard form,

$$f(x) = \frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)} x^{\alpha-1}(1-x)^{\beta-1}, \quad 0 < x < 1, \quad \alpha,\beta > 0, \tag{4}$$

where $x = \frac{y-a}{b-a}$ has the following parameters:

$$\mu_x = \frac{\mu_y - a}{b-a}, \quad \sigma_x = \frac{\sigma_y}{b-a}, \quad m_x = \frac{m_y - a}{b-a}. \tag{5}$$

Let $\alpha - 1 = p$, $\beta - 1 = q$. Then p.d.f. (4) becomes

$$f(x) = \frac{\Gamma(p+q+2)}{\Gamma(p+1)\Gamma(q+1)} x^p(1-x)^q, \quad 0 < x < 1, \quad p,q > -1, \tag{6}$$

with the mean, variance and mode as follows:

$$\mu_x = \frac{p+1}{p+q+2}, \tag{7}$$

$$\sigma_x^2 = \frac{(p+1)(q+1)}{p+q+2}, \tag{8}$$

$$m_x = \frac{p}{p+q}. \tag{9}$$

From (6) and (9) it can be obtained

$$f(x) = \frac{\Gamma(p+q+2)}{\Gamma(p+1)\Gamma(q+1)} x^p(1-x)^{p(1/m_x-1)}. \tag{10}$$

Thus, value m_x, being obtained from the analyst's subjective knowledge, indicates the density function. On the basis of statistical analysis and some other intuitive arguments, the creators of PERT assumed that $p+q \cong 4$. It is from that assertion that estimates (1) and (2) were finally obtained, according to (6-9).

Although the basic concepts of PERT analysis have been worked out many years ago [3, 17], they are open till now to considerable criticism. Numerous attempts have been made to improve the main PERT assumptions for calculating the mean μ_x and variance

σ_x^2 of the activity-time on the basis of the analyst's subjective estimates. In recent years, a very sharp discussion [7, 10, 14, and 21] has taken place in order to raise the level of theoretical justifications for estimates (1) and (2).

Grubbs [12] pointed out the lack of theoretical justification and the unavoidable defects of the PERT statements, since estimates (1) and (2) are, indeed, "rough" and cannot be obtained from (3) on the basis of values a, m and b determined by the analyst. Moder [18-19] noted that there is a tendency to choose the most likely activity – time m much closer to the optimistic value a than to the pessimistic one, b, since the latter is usually difficult to determine and thus is taken conservatively large. Moreover, it is shown [8] that value m, being subjectively determined, has approximately one and the same relative location point in $[a,b]$ for different activities. This provides an opportunity to simplify the PERT analysis at the expense of some additional assumptions. McCrimmon and Ryavec [16], Lukaszewicz [15] and Welsh [22] examined various errors introduced by the PERT assumptions, and came to the conclusion that these errors may be as great as 33%. Murray [20] and Donaldson [4] suggested some modifications of the PERT analysis, but the main contradictions nevertheless remained. Farnum and Stanton [6] presented an interesting improvement of estimates (1) and (2) for cases when the modal value m is close to the upper or lower limits of the distribution. This modification, however, makes the distribution law rather uncertain, and causes substantial difficulties to simulate the activity network.

In this paper, a research will be undertaken to develop some theoretical justifications for using the beta-distribution p.d.f.

2. The Operation's Description

We will consider a man-machine operation which is carried out by one processor, i.e., by one resource unit. The processor may be a machine, a proving ground, a department in a design office, etc.

Assume that the operation starts to be processed at a pregiven moment T_0. The completion moment F of the operation is a random value with distribution range $[T_1, T_2]$. Moment T_1 is the operation's completion moment on condition that the operation will be processed without breaks and without delays, i.e., value T_1 is a pregiven deterministic value. Assume, further, that the interval $[T_0, T_1]$ is subdivided into n equal elementary periods with length $(T_1 - T_0)/n$. If within the first elementary period $[T_0, T_0 + (T_1 - T_0)/n]$ a break occurs, it causes a delay of length $\Delta = (T_2 - T_1)/n$. The operation stops to be processed within the period of delay in order to undertake necessary refinements, and later on proceeds functioning with the finishing time of the first elementary period
$T_0 + (T_1 - T_0)/n + (T_2 - T_1)/n = T_0 + (T_2 - T_0)/n$.

It is assumed that there cannot be more than one break in each elementary period. The probability of a break at the very beginning of the operation is set to be p. However, in the course of carrying out the operation, the latter possesses certain features of self-adaptivity, as follows:

- the occurrence of a break within a certain elementary period results in increasing the probability of a new break at the next period by value η, and
- on the contrary, the absence of a break within a certain period decreases the probability of a new break within the next period, practically by the same value.

3. The Concept of Self-Adaptivity

The probabilistic self-adaptivity can be formalized as follows:

Denote A_i^k the event of occurrence of a break within the $(i+1)$-th elementary period, on condition, that within the i preceding elementary periods k breaks occurred, $1 \le k \le i \le n$. It is assumed that relation

$$P\left(A_i^k\right) = \frac{p + k \cdot \eta}{1 + i \cdot \eta} \qquad (11)$$

holds. Note that (11) is, indeed, a realistic assumption.

Relation (11) enables obtaining an important assertion. Let $P\left(A_i^0\right)$ be the probability of the occurrence of a break within the $(i+1)$-th period on condition, that there have been no breaks at all as yet. Since

$$P\left(A_i^0\right) = \frac{p}{1 + i \cdot p}, \qquad (12)$$

it can be well-recognized that relation

$$\frac{P\left(A_i^{k+1}\right) - P\left(A_i^k\right)}{P\left(A_i^0\right)} = \frac{\eta}{p} \qquad (13)$$

holds. Thus, an assertion can be formulated as follows:

Assertion. *Self-adaptivity (11) results in a probability law for delays with a constant ratio (13) for a single delay.*

4. Calculating the Activity-Time Distribution

Let us calculate the probability $P_{m,n}$ of obtaining m delays within n elementary periods, i.e., the probability of completing the operation at the moment

$$F = T_1 + m \cdot \Delta = T_1 + \frac{m}{n}\left(T_2 - T_1\right).$$

The number of sequences of n elements with m delays within the period $\left[T_0, F\right]$ is equal C_n^m, while the probability of each such sequence equals

$$\frac{\left[\prod_{i=0}^{m-1}\left(p + i\eta\right)\right]\left[\prod_{i=0}^{n-m-1}\left(1 - \eta + i\eta\right)\right]}{\prod_{i=0}^{n-1}\left(1 + i\eta\right)}. \qquad (14)$$

Relation (14) stems from the fact that if breaks occurred within h periods and did not occur within k periods, the probability of the occurrence of the delay at the next period is equal

$$\frac{p+h\eta}{1+(k+h)\eta}, \tag{15}$$

while the probability of the delay's non-appearance at the next period satisfies

$$\frac{1-\eta+k\eta}{1+(k+h)\eta}. \tag{16}$$

Using (14-16), we finally obtain

$$P_{m,n} = C_n^m \frac{\left[\prod_{i=0}^{m-1}(p+i\eta)\right]\left[\prod_{i=0}^{n-m-1}(1-\eta+k\eta)\right]}{\prod_{i=0}^{n-1}(1+i\eta)}. \tag{17}$$

Note that $\eta=0$, i.e., the absence of self-adaptivity, results in a regular binomial distribution.

Let us now obtain the limit value $P_{m,n}$ on condition that $n\to\infty$. From relation (17) we obtain

$$\frac{P_{m+1,n}}{P_{m,n}} = \frac{n-m}{m+1}\,\frac{p+m\eta}{1-p+(n-m-1)\eta}. \tag{18}$$

Denoting $\frac{p}{\eta}=\alpha$, $\frac{p}{\eta}\left(\frac{1}{p}-1\right)=\beta$, we obtain

$$\frac{P_{m+1,n}-P_{m,n}}{P_{m,n}} = \frac{(\alpha-1)n+(2-\alpha-\beta)m-\beta+1}{(m+1)(\beta+n-m-1)} = \frac{(\alpha-1)+(2-\alpha-\beta)\frac{m}{n}+\frac{1-\beta}{n}}{n\frac{m+1}{n}\left(1-\frac{m+1}{n}+\frac{\beta}{n}\right)}.$$

Denoting $m/n=x$, $(m+1)/n=x+\Delta x$, $P_{m,n}=y$, $P_{m+1,n}=y+\Delta y$, via convergence $n\to\infty$ or $\Delta x\to 0$ and, later on, by means of integration, we finally obtain

$$y = C x^{\alpha-1}(1-x)^{\beta-1}. \tag{19}$$

It can be well-recognized that the p.d.f. of random value $\xi=\lim_{n\to\infty}\frac{m}{n}$ satisfies

$$p_\xi(x) = \frac{1}{B(\alpha,\beta)}x^{\alpha-1}(1-x)^{\beta-1}, \tag{20}$$

where $B(\alpha,\beta)$ represents the Euler's function. Thus, relation (20) practically coincides with (10).

Thus, ξ is a random value with the beta-distribution activity–time p.d.f. By transforming $x=(y-a)/(b-a)$, we obtain the well-known p.d.f. (3).

5. Conclusions

The following conclusions can be drawn from the study:

1. Under certain realistic assumptions we have proven theoretically that the activity-time distribution satisfies the beta-distribution with p.d.f. (3) being used in PERT analysis.
2. Changing more or less the implemented assumptions, we may alter to a certain extent the structure of the p.d.f. At the same time, its essential features (e.g. asymmetry, unimodality, etc.) remain unchanged.
3. The outlined above research can be applied to semi-automated activities, where the presence of man-machine influence under random disturbances is, indeed, very essential. Those activities are likely to be considered in organization systems (e.g., in project management), but not in fully automated plants.

References

1. Battersby, A. **Network Analysis for Planning and Scheduling**, 3rd edition, MacMillan: London, 1970
2. Berny, J. **A new distribution function for risk analysis**, J. Oper. Res. Soc., 40(12), 1989, pp. 1121-1127
3. Clark, C.E. **The PERT model for the distribution of an activity**, Opns. Res., 10, 1962, pp. 405-406
4. Donaldson, W.A. **The estimation of the mean and variance of PERT activity time**, Opns. Res., 13, 1965, pp. 382-385
5. Elmaghraby, S.E. **Activity Networks: Project Planning and Control by Network Models**, Wiley: New-York, 1977
6. Farnum, N.R. and Stanton, L.W. **Some results concerning the estimation of beta distribution parameters in PERT**, J. Oper. Res. Soc., 38, 1987, pp. 287-290
7. Gallagher, C. **A note on PERT assumptions**, Mgmt. Sci., 33, 1987, pp. 1360-1362
8. Golenko-Ginzburg, D. **Statistical Models in Network Planning and Control**, Nauka: Moscow (in Russian), 1966
9. Golenko-Ginzburg, D. **On the distribution of activity time in PERT**, J. Oper. Res. Soc., 39(8), 1988, pp. 767-771
10. Golenko-Ginzburg, D. **PERT assumptions revisited**, Omega, 17(4), 1989, pp. 393-396
11. Gonik, A. **Planning and controlling multilevel stochastic projects**, Ph.D. Thesis, Ben-Gurion University of the Negev, Beer-Sheva, Israel, 1995
12. Grubbs, F.E. **Attempts to validate certain PERT statistics or "picking on PERT"**, Opns. Res., 10, 1962, pp. 912-915
13. Kelley, J.E. Jr. **Critical path planning and scheduling: Mathematical basis**, Opns. Res., 9(3), 1961, pp. 296-320
14. Littlefield, T.K. Jr. and Randolph, P.H. **Another note on PERT times**, Mgmt. Sci., 33, 1987, pp. 1357-1359
15. Lukaszevicz, J. **On the estimation of errors introduced by standard assumptions concerning the distribution of activity duration PERT calculations**, Opns. Res., 13, 1965, pp. 326-327
16. MacCrimmon, K.R. and Ryaveck, C.A. **An analytical study of the PERT assumptions**, Opns. Res., 12, 1964, pp. 16-37
17. Malcolm, D., Roseboom, J., Clark, C. and Fazar, W. **Application of a technique for research and development program evaluation**, Opns. Res., 7, 1959, pp. 646-669

18. Moder, J.J., Phillips, C.R. and Davis, E.W. **Project Management with CPM and PERT and Precedence Diagramming**, Van-Nostrand Reinhold Co., Inc.: New-York, 1983
19. Moder, J.J. and Cecil, R.P. **Project Management with CPM and PERT**, Van-Nostrand Reinhold Co., Inc.: New-York, 1970
20. Murray, J.E. **Consideration of PERT Assumptions**, Conduction Corporation, Ann. Arbor: Michigan, 1962
21. Sasieni, M.W. **A note on PERT times**, Mgmt. Sci., 32, 1986, pp. 1652-1653
22. Welsh, D. **Errors introduced by a PERT assumption**, Opns. Res., 13, 1965, pp. 141-143
23. Williams, T.M. **Practical use of distributions in network analysis**, J. Oper. Res. Soc., 43(3), 1992, pp. 265-270
24. Williams, T.M. **What are PERT estimates?** J. Oper. Res. Soc., 46(12), 1995, pp. 1498-1504