Decision Support System for Project Duration Estimation Model

(프로젝트기간예측모델을 위한 의사결정지원시스템)

조 성 빈*

Abstract

Despite their wide application of some traditional project management techniques like the Program Evaluation and Review Technique, they lack of learning, one of important factors in many disciplines today, duc to a static view for project progression. This study proposes a framework for estimation by learning based on a Linear Bayesian approach. As a project progresses, we sequentially observe the durations of completed activities. By reflecting this newly available information to update the distribution of remaining activity durations and thus project duration, we can implement a decision support system that updates e.g., the expected project completion time as well as the probabilities of completing the project within the due date and by a certain date. By implementing such customized systems, project manager can be aware of changing project status more effectively and better revise resource allocation plans.

Key word: estimation by learning, project management, decision support system

* 충북대학교 경영대학

1. INTRODUCTION

A considerable amount of effort has been devoted in the filed of project management techniques to overcome some theoretical weaknesses of traditional approaches. For example, Critical Path Method (CPM) has a deterministic view of activity durations. We can hardly estimate the duration of upcoming activities without any uncertainty. Although Program Evaluation and Review Technique (PERT) includes such uncertainty using a probability distribution for activity duration, it still has a static view of project progression. Estimating activity durations is made at the beginning of a project and they do not provide any further updating framework. Note that in the real world projects, as a project progresses, we practically update the distribution of durations of upcoming activities and as a result critical path keeps changing [Dodin and Elmaghraby, 1985].

Another limitation of traditional approaches is in their assumption such that all activity durations are independent. This significantly restricts the application of these methods or reduces the validity of methods if applied. In a project, it is not unusual that many activities share common resources such as human power, funds, material, and many other resources. In other words, many activities might be correlated with one another in terms of duration, cost, and performance/quality.

Various attempts have been made to overcome the weaknesses of existing project management models [see Chatzogou and Macaulay for a review, 1996]. For example, deterministic approaches are made by Foldes and Soumis [1993] and Babu and Suresh [1996]. Net present value maximization approaches are made by Smith-Daniels and Aquilano [1987], Elmaghraby and Herroelen [1990], Yang, et al. [1992], and Buss and Rosenblatt [1997]. However, these methods lack of learning and sequential updating processes in their models.

It is interesting to note that several researchers try to improve the accuracy of activity duration estimation by using the learning curve effect [Ayas, 1996; Badiru, 1995; Shtub, 1991; Teplitz and Amor, 1993]. However, the use of learning curve effect is extremely limited to some repetitive projects like apartment or condominium construction projects only [Urban Land Institute, 1995]. Real world projects are mostly rather small in size and possess some unique aspects [Meredith and Mantel, 1995]. Therefore, updating the distribution of durations of identical activities in more than one projects is very rare. The learning curve effect is better fitting to routine work type activities than activities in projects.

Recent studies have been made to model probabilistic dependence among activity durations in a project and sequentiality. Jenzarli [1994] simply allows probabilistic dependence between two different activity durations. van Dorp and Duffey [1999] assign the reasons of probabilistic dependence. Covaliu and Soyer [1996] [1997] and Cho [2000] make a Bayesian decision theoretical framework for learning-based estimation and sequential decision-making in project management.

2. Linear Bayesian Model

Hartigan's linear Bayes' theorem [1969] is as follows. For random variables $X, Y, Z_1, Z_2, \ldots,$ and Z_n, by defining the linear expectation of X

given $Y, Z_1, Z_2, \ldots,$ and Z_n, we can update the following quantities. The conditional precision of Y is the sum of its marginal precision and the present data precision. The conditional mean of Y can be estimated by the weighted average of its marginal mean and a term due to the present data, and the weight is the ratio of each precision over the conditional precision, respectively.

Let $\widehat{E}[X|Y,Z] = cY + d,$

where $d = a_0 + \sum_{i=1}^{n} a_i Z_i$

Then

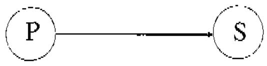
$$V^{-1}[Y|X,Z] = V^{-1}[Y|Z] + c^2 V^{-1}[X|Y,Z], \quad (1)$$

$$\widehat{E}[Y|X,Z] = \widehat{E}[Y|Z]\frac{V^{-1}[Y|Z]}{V^{-1}[Y|X,Z]} + c(X-d)\frac{V^{-1}[X|Y,Z]}{V^{-1}[Y|X,Z]}. \quad (2)$$

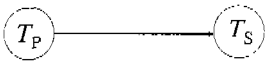
Note that 1) $\widehat{E}$ [.] represents the estimation of expected value.

2) V^{-1} [.] represents the precision (the inverse of variance).

For the sake of explanation, imagine a simple, two-activity in-series project. Under the Activity-On-Node convention, the precedence diagram (in Figure 1) depicts that succeeding Activity S is begun upon the completion of preceding Activity P while the influence diagram (in Figure 2) shows that the duration of Activity S might depend on that of Activity P. Consider a pair of activities is positively correlated with each other due to sharing human resources. If



[Figure 1] Precedence diagram for the two-activity in-series project



[Figure 2] Influence diagram for the two-activity in-series project

the preceding activity takes more time than originally expected for many reasons, e.g., insufficient labor skills, it increases the probability that the succeeding activity takes more time than originally expected. In other words, the conditional expected duration of the succeeding activity is greater than its marginal expected duration. Other types of resource sharing might be found in raw materials, utility, and equipment/facilities.

Suppose two activities have marginal means and variances of activity durations under a Normal distribution, μ_{0i} and σ_{0i}^2, for i = P, S, and correlation coefficient.

$$T_P \sim N(\mu_{0P}, \sigma_{0P}^2),$$

$$T_S \sim N(\mu_{0S}, \sigma_{0S}^2),$$

ρ_{PS}=correlation coefficient of T_P and T_S,

let, $\hat{E}[T_P | T_S] = cT_S + d$,

where c and d are chosen so that the variance of T_P given T_S is minimized.

Then, by the linear Bayes' theorem,

$$V^{-1}[T_S | T_P] = V^{-1}[T_S] + c'^2 V^{-1}[T_P | T_S]$$

i.e., $1/\sigma_{1S}^2 = 1/\sigma_{0S}^2 + c'^2 / \mathrm{Var}[T_P | T_S]$, (3)

$$\hat{E}[T_S \mid T_P] = \hat{E}[T_S]\frac{V^{-1}[T_S]}{V^{-1}[T_S \mid T_P]} + c''^2(T_P - d')\frac{V^{-1}[T_P \mid T_S]}{V^{-1}[T_S \mid T_P]}$$

i.e., $\mu_{1S} = \mu_{0S}(1/\sigma_{0S}^2)/(1/\sigma_{1S}^2) + c'^2(T_P - d'')V^{-1}[T_P \mid T_S]/(1/\sigma_{1S}^2)$. (4)

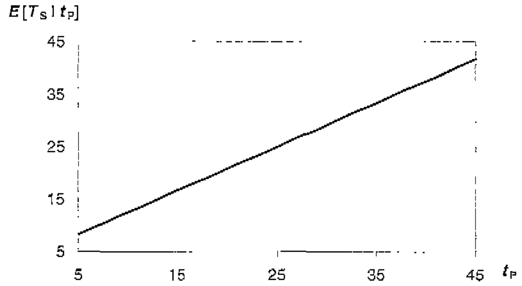
To illustrate the characteristics of the Linear Bayesian Estimation Model, consider the following data: (activity duration unit: days)

$$T_P \sim N(25, 3^2),\ T_S \sim N(25, 5^2),\ \rho_{PS} = 0.5.$$

By (3) and (4), the conditional mean and variance of the duration of succeeding activity would be:

$$T_S \mid T_P \sim N(4.17 + 0.83T_P,\ 4.33^2).$$

Figure 3 shows the learning relationship between the estimation of conditional expected duration of Activity S and the observed duration of Activity P. It is obvious that if Activity P takes 25 days, exactly same as expected, then Activity S tends to take 25 days as expected. If Activity P takes longer than expected, e.g., 29 days, the conditional expected duration of Activity S is estimated to be longer than its marginal expected duration, e.g., 28.33 days. On the other case when Activity P takes shorter than its marginal mean, e.g., 21 days, the conditional expected duration of Activity S is estimated to be shorter than its marginal expected duration, e.g., 21.67 days. Together with the update of the expected duration of succeeding activity, refer to the fact that the conditional variance given observation (4.33^2) is smaller than the marginal variance before observation (5^2) due to learning. In short, we improve the



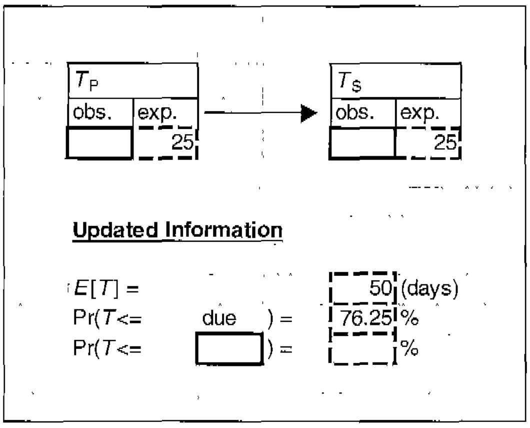
[Figure 3] The estimation of conditional expected duration of Activity S given the observed duration of Activity P

estimation of the duration of upcoming activities in terms of expectation and variance by learning from the duration of completed activities.

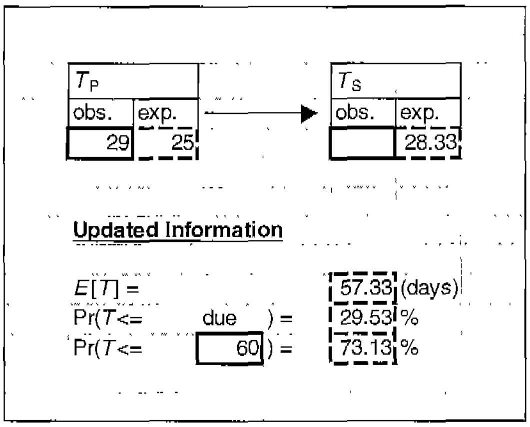
The proposed estimation model and the following decision support system can be validated by the Hartigan's linear Bayes' theorem; see the proof of this theorem [1969]. For jointly normally distributed variables, the linear Bayesian model produces the identical conditional mean and variance as those using the full Bayesian approach [Cho, 2000]. In other words, we try to measure $E[T_S \mid T_P]$. What we use to measure what we intend to measure is $\hat{E}[T_S \mid T_P]$. Under a Normal distribution, these two quantities are exactly equal to each other. The closer to a Normal or at least symmetrical distribution activity duration approaches, the better the estimation model and, of course, the decision support system fit. This paper will assume a Normal distribution for activity duration in the next section.

3. Decision Support System for Estimation Model

We can implement the above Linear Bayesian



[Figure 4] Decision support system before observing Activity P duration



[Figure 5] Decision support system after observing Activity P duration

Model using a spreadsheet modeling. In Figure 4, a box with thick solid lines (—) is used an input cell whereas a box with grid lines (---) is used an output cell. Before the project actually begins, we estimate that the expected durations (exp.) of Activities P and S are 25 days. The expected completion time of the project is 50 days. If the project delivery time is scheduled as 55 days, then the probability that the project is completed within the due time is 76.25 %.

If we observe the duration of Activity P (obs.) as 29 days, then the decision support system updates the conditional expected duration of Activity S as 28.33 days (see Figure 5). So do the expected completion time of project and the probability of completing the project within the due time as 57.33 days and 29.53%, respectively. Since the preceding activity takes more time than expected, it is less likely to finish the project within the due time of 55 days. For example, if the project manager wants to know the probability that they can complete the project by 60 days, then the system gives the answer as 73.13 %.

4. Application of the Model

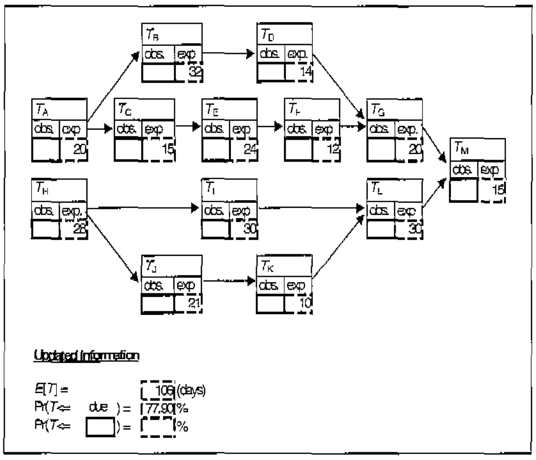
The effectiveness of model can be further demonstrated using the Fantasy Products Case [adopted from Stevenson, 1993]. A local manufacturer of high quality small appliances, Fantasy Products is working on a new kitchen appliance project consisting of 13 activities. Table 1 gives activity descriptions, precedence relationships, marginal means and variances of activity durations, and correlation coefficients.

Figure 6 shows the decision support system for the kitchen appliance project. If all correlation coefficients are known as 0.5 and the project delivery due time is 115 days, the system shows that the expected project completion time is 106 days and the probability of completing the project within the due time is 77.90 %. As the project proceeds and we sequentially observe the duration of completed activities, the system automatically updates the conditional expected durations of remaining activities and so do the expected project completion time and the

〔Table 1〕 Data about the kitchen appliance project

Activity	Description	Immediate predecessor	$E[T_i]$	$Var[T_i]$	ρ
A	Select and order equipment	-	20	5^2	ρ_{AB}
B	Receive equipment from supplier	A	32	8^2	ρ_{CD}
C	Install and set up equipment	A	15	3^2	ρ_{DG}
D	Finalize bill of materials	B	14	4^2	ρ_{HI}
E	Order component parts	C	24	7^2	ρ_{IE}
F	Receive component parts	E	12	3^2	ρ_{IL}
G	First production run	D, F	20	6^2	ρ_{JK}
H	Finalize marketing plan	-	28	6^2	ρ_{KL}
I	Produce magazine ads	H	30	6^2	ρ_{FM}
J	Script for TV ads	H	21	5^2	
K	Produce TV ads	J	30	3^2	
L	Begin ad campaign	I, K	30	6^2	
M	Ship product to customers	G, L	15	3^2	

〔duration unit: days〕



〔Figure 6〕 Decision support system for the kitchen appliance project

probability of completing the project within the due time.

For practical considerations, the elicitation of correlation coefficients between activity durations is, although not easy, not impossible. Cooke [1991] suggests some ways to combine expert opinions, i.e., project managers' and field managers' expectations, using weighted combinations. Gokhale and Press [1982] propose a method of assessing a prior distribution for the correlation coefficient in a bivariate normal distribution.

5. Discussion and Concluding Remarks

One major contribution of this paper is to allow learning in estimation of activity durations in a project, which has long been unfeasible in studies based on learning curve effects. Jenzarli [1994] fails to suggest an underlying dependence structure between activity durations. Covaliu and Soyer's model [1996] [1997] can be implemented under three specific probability density functions borrowed from reliability theories. Compared to these two studies, the proposed estimation model in this paper clearly suggests the underlying dependence structure and can be applied to any approximate probability distributions. In particular, by combining newly observable information - the observed duration of completed activities - with the existing data available from the onset of project - marginal expected durations and variances of activities as well as correlation coefficients between activity durations, we can update the distribution of the durations of upcoming activities. For example, for two positively correlated activities, our model successfully demonstrates such relationships, e.g., the succeeding activity tends to take shorter as the preceding activity takes shorter.

Another unique contribution of this paper is to illustrate a way of implementing a decision support system for the Linear Bayesian Estimation Model. For small-sized projects planned and managed by individual decision-makers, the decision support system can be built on a spreadsheet like the one in this paper. Moderately large-sized projects for organizations can be built using more advanced commercial programming languages like visual basic. As the project manager plugs the inputs in the system, e.g, the observed duration of completed activities, the system gives the updated expected durations of activities and project, and the likelihood of completing the project within the due date. In the middle of a project, the project manager might be interested in knowing the likelihood of completing the project by a certain date, possibly not the due date. Then the decision support system can easily provide such outcomes, too. More highly customized questions and answers can be programmed depending on the needs of project managers and other stakeholders.

It is often anticipated that projects will be involved with increasing uncertainty in the future as business environment changes with a shorter cycle due to accelerating speed of technology advancement. Therefore, it would be worthwhile to develop more realistic project management techniques followed by implementation tools to manage projects more rationally and preserve scarce resources.

References

[1] Ayas, Karen, (1996), Professional Project Management: A Shift towards Learning and a Knowledge Creating Structure, *International Journal of Project Management*, 14, p131-136

[2] Babu, A.J.G. and N. Suresh, (1996), Project Management with Time, Cost, and Quality Considerations, *European Journal of Operational Research*, 88, p320-327

[3] Badiru, Adedeji B., (1995), Incorporating Learning Curve Effects Into Critical Resource Diagramming, *Project Management Journal*, June, p38-45

[4] Beaver Creek Resort, Avon, Colorado, (1995), The Urban Land Institute, 15, no 5, Jan.-Mar.

[5] Buss, A.H. and M.J. Rosenblatt, (1997), Activity Delay in Stochastic Project Networks, *Operations Research*, 45, p126-139

[6] Chatzogou, Prodromos D. and Macaulay, Linda A., (1996), A Review of Existing Models for Project Planning and Estimation and the Need for a New Approach, *International Journal of Project Management*, 14

[7] Cho, Sungbin, (2000), Sequential Estimation and Decision-Making in Project Management - A Bayesian Way and Heuristic Approaches -, *Ph. D. Dissertation*, George Washington University

[8] Cooke, Roger M., (1991), Experts in Uncertainty: Opinion and Subjective Probability in Science, *Oxford University Press*

[9] Covaliu, Zvi and Refik Soyer, (1996), Bayesian Project Management, *Proceedings of the ASA section on Bayesian Statistical Science*, p208 - 213

[10] Covaliu, Zvi and Refik Soyer, (1997), Bayesian Learning in Project Management Networks, *Proceedings of the ASA section on Bayesian Statistical Science*, p257 - 260

[11] Dodin, B.M. and S.E. Elmaghraby, (1985), Approximating the Criticality Indices of the Activities in PERT Networks, *Management Science*, 31, p207-223

[12] Elmaghraby, S.E. and W.S. Herroelen, (1990), The Scheduling of Activities to Maximize the Net Present Value of Projects, *European Journal of Operational Research*, 49, p35-49

[13] Foldes, S. and F. Soumis, (1993), PERT and crashing revisited: Mathematical Generalizations, *European Journal of Operational Research*, 64, p286-294

[14] Gokhale, D.V. and S.J. Press, (1982), Assessment of a Prior Distribution for the Correlation Coefficient in a Bivariate Normal Distribution, *Royal Statistical Society*, 145, p237-249

[15] Hartigan, J.A., (1969), Linear Bayesian Methods, *Journal of the Royal Statistical Society*, Series B, 31, p446-454

[16] Jenzarli, Ali, (1994), PERT Belief Networks, Report 535, College of Business, The University of Tampa, FL

[17] Meredith, Jack R. and Samuel J. Mantel, Jr., (1995), Project Management, *John Wiley & Sons*, p536-538

[18] Shtub, Avraham, (1991), Scheduling of Programs with Repetitive *Projects, Project Management Journal*, Dec., p49-53

[19] Smith-Daniels, D.E. and N.J. Aquilano, (1987), Using a Late-Start Resource-Constrained Project Schedule to Improve Project Net Present Value, *Decision Sciences*, 18, p617-630

[20] Stevensons, William J. (1993), Production/ Operations Management, *Irwin*, p821-823

[21] Teplitz, Charles J. and Amor, Jean-Pierre, (1993), Improving CPM's Accuracy Using Learning Curves, *Project Management Journal*, Dec., p15-19

[22] van Dorp, J.R. and Duffey, Michael R., (1999), Statistical Dependence in Risk Analysis for Project Networks using Monte Carlo methods, *International Journal of Production Economics*, vol. 58

[23] Yang, K.K., F.B. Talbot, and J.H. Patterson, (1992), Scheduling a Project to Maximize its Net Present Value: An Inter Programming Approach, *European Journal of Operational Research*, 64, p188-198