STOCHASTIC SCHEDULING CONSIDERING INTERDEPENDENT ACTIVITY DURATIONS

I-Tung Yang[1]

[1] Assistant Professor, Dept. of Civil Engineering, Tamkang University, Taiwan
Correspond to ityang@mail.tku.edu.tw

ABSTRACT: A simulation model is proposed to evaluate the effect of correlations between activity durations on the overall project duration. The proposed model incorporates NORTA, a recent developed statistical method, into the simulation process to allow arbitrarily specified marginal distributions for activity durations and any desired correlation structure. The generality is of practical value when systematic data is not available and planners have to rely on arbitrary experts' estimation, which may involve a mixed situation when some activity durations are continuously distributed whereas others are discrete outcomes. The proposed model is validated by showing that the correlation coefficients of the simulation results are close to the originally specified ones. The simulation results are compared to two conventional approaches: PERT and simulation without correlation. The comparisons illustrate that the proposed model can provide important management information, which would otherwise be distorted due to the neglect of the correlations between activity durations.

Key words : Correlation, Simulation, Random Variable Generation, Simulation, Risk Analysis, Project Management

1. BACKGROUND

For decades network-oriented tools have assisted management in planning and controlling construction projects, which may be fraught with uncertainty due to unpredictable factors: site condition, crew efficiency, equipment performance, dependability of subcontractors, and the effect of weather. Since the durations of project activities are dependent upon these uncertain factors, the activity durations appear as random variables rather than deterministic numbers.

One of the earliest tools to allow probabilistic estimation of activity durations is the program evaluation and review technique (PERT), which was developed by the U.S. Navy in 1957. Each activity in PERT is assumed to be a statistical distribution of time, whose parameters are obtained based upon optimistic, most likely, and pessimistic time estimates. PERT helps estimate the probability of meeting a scheduled date for the project duration or any other milestone event.

A fundamental limitation of PERT is the assumption of independence among activity durations. In construction projects, activities share some of the same crews, equipment, supervision, site conditions, and weather. Thus the activity durations are not independent and with certain correlations [14]. Ignoring such correlation can significantly distort the realization of the actual project duration [5]. The need of recognizing correlation between activity durations in PERT analysis leads to the development of the following models.

2. PREVIOUS APPROACHES

Carr proposed MUD (model for uncertainty determination) for planners to chart the dependence among activities and to estimate the activity sensitivities to weather [3]. He indicated that the effect of weather, unlike other factors, is based on calendar date. Following this argument, Padilla and Carr treated the correlation among activity durations by using the samples drawn from a shared factor in each scheduling day [10]. The samples of activity durations are then input into simulation software to produce a statistical summary of samples of the project duration.

Other models also implemented computer simulation but took different approaches in sampling activity durations. Ahuja and Nandakumar [1] adjusted the samples of activity durations as a product of workday loss due to various uncertain factors. Levitt and Kunz [8] associated each activity with risk factors. The effect of each risk factor (favorable or adverse) is evaluated by comparing actual and planned durations of completed activities. According to the effects of risk factors, durations of associated uncompleted activities are adjusted (reduced if favorable and increase if adverse). Ranasinghe and Russell obtained the correlation coefficients among activity durations directly by expert judgment [12].

Wang and Demsetz [13] presented NETCOR to distribute the uncertainty in activity durations to factor sub-distributions (e.g., equipment, crew, and weather) that are further broken down into several condition sub-distributions (e.g., better-than-expected, expected, and worse-than-expected). NETCOR tries to capture the correlation between activity durations by sampling from the same condition sub-distribution. Ioannou and Martinez [6] developed a simulation augment, ProbSched add-on, to treat the correlation issue. They provided a formula to adjust the sample activity durations to achieve the pre-specified correlation structure.

3. GOAL

The goal of the present study is to treat practical situations when no systematic data is available and schedulers have to rely on experts' subjective estimates of activity durations and their dependences (correlations). In such circumstances the estimated distributions of activity durations may be skewed. Moreover, the distributions need not belong to the same family of continuous distributions (e.g., some are triangular while some are uniform) and may even constitute a mixed situation when some can be expressed as continuous variables whereas others have only discrete outcomes.

To obviate the difficulties, the present study proposes a more general simulation model to capture the true effect of correlations between activity durations. The proposed model starts with generating correlated random variates for activity durations based on subjective estimates of (1) marginal probability density functions (PDFs) of the duration of every activity and (2) a correlation matrix composed of pre-specified correlation coefficients between activity durations.

4. PROPOSED MODEL

The proposed model incorporates the NORTA method, which is proposed by Cario and Nelson [2] and refined by Ghosh and Henderson [4] into the process of simulating schedule networks. The proposed model is composed of the following steps:

1. Generate an IID (independent and identically distributed) unit scaled uniform random vector, $Y=(Y_1, Y_2, \ldots, Y_n)$ where n is the number of activities in the network.
2. Translate Y into a standard-normal random vector $P=(P_1, P_2, \ldots, P_n)$.
3. Transform P into a correlated standard-normal random vector $Z=(Z_1, Z_2, \ldots, Z_n)$ such that $\mathrm{Corr}(Z_i,Z_j)=\rho_{ij}$ for $i, j = 1, 2, \ldots n$.
4. Compute $U_i=\Phi(Z_i)$ for $i = 1, 2, \ldots n$, where $\Phi(.)$ denotes the standard normal CDF.
5. Compute $X_i=F^{-1}(U_i)$ for $i = 1, 2, \ldots n$, where $F^{-1}(U_i)$ represents the inverse of the marginal CDF.
6. Return X_i as the duration for activity i.
7. Perform the CPM forward pass to compute the overall project duration.
8. Repeat Steps 1 through 7 for each simulation replication, $j =1, 2, \ldots, m$.
9. Return summary statistics on all simulation replications.

Step 1 is a typical pseudo-random number generation procedure. It can be performed by many computer programs (such as MS Excel, Arena, and @Risk). Step 2 can be approximated using the following equation:

$$P_i=(Y_i^{0.135}-(1-Y_i)^{0.135})/0.1975 \qquad (1)$$

Step 3 consists of two parts. First, the correlation matrix is factorized uniquely (i.e., Cholesky decomposition) as

$$\Sigma = CC^T \qquad (2)$$

Efficient algorithms to perform the Cholesky decomposition can be found in [11]. Second, the components in C, c_{ij}, are used to produce correlated normal variates.

$$Z_i = \sum_{j=1}^{i} c_{ij} P_j \quad \text{for } i = 1, 2, \ldots, n \qquad (3)$$

In Step 4, since Z_i has a standard-normal distribution, Mood et al. [9] showed that $U_i=\Phi(Z_i)$ is uniformly distributed on (0,1). According to the inverse-transform concepts, X_i in Step 5 would have the desired marginal distributions. Note that this proof holds regardless of types of distributions.

The inverse-transform method in Step5 starts with pseudo-number uniformly distributed on (0,1). The uniform pseudo-number is input into the inverse of the desired cumulative distribution function (CDF) to obtain the desired random variates. Theoretically, the inverse-transform method can be applied in any type of distributions (both continuous and discrete) as long as the inverse of the CDF is available or can be approximated numerically. Law and Kelton [7] provided detailed survey for a wide variety of distributions, such as uniform, normal, lognormal, gamma, beta, Pearson, Bernoulli, and discrete.

5. APPLICATION

5.1 Project Introduction

A highway project in [6] will serve as an application example for using the proposed model. This project is 3.52 km (2.2 miles) long and naturally divided into two sections based on the location of the balance points for earthmoving (cut and fill). The plan and profile of the project are shown in Figure 1. The project encompasses 26 activities including equipment procurement, cut-and-fill earthmoving, the construction of one double-barrel and two single-barrel concrete box culverts, placement of subbase, paving, landscaping, and guardrail installation.

All activities and the three-point estimates (a: minimum, m: most likely, b: maximum) for the durations are listed in Table 1. These estimates are modeled as parameters of triangular distributions, not the commonly used beta distributions, because of two reasons. First, beta distributions do not have a clear-cut meaning on lower and upper bounds. Second, the four-parameter beta distribution does not have a one-to-one correspondence to the three estimates, a, m and b; it therefore requires an additional estimate for the fourth parameter, such as mean, variance, or percentiles.

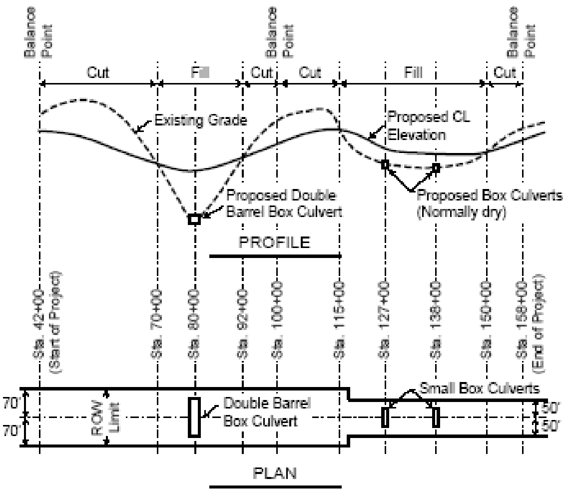

Figure 1. Plan and profile for highway project [6]

To demonstrate the capability of the proposed model in treating different families of distributions, durations of activities 24 and 25 are assumed to be uniformly distributed between 1 and 7 days while that of activity 26 has three options with corresponding probabilities as follows: (2.5 days, 0.3; 4 days, 0.6; 11 days, 0.1). It is this level of capability that distinguishes the proposed model from previous work.

Table 1. Activities duration estimates

ID	Activity	a	m	b
1	Setup batch plant	0.5	2	3.5
2	Deliver paving mesh	2	5	8
3	Rebar, double barrel culvert	2.5	5.5	11.5
4	Move in equipment	1.5	3	4.5
5	Rebar, small box culvert	1	4	25
6	Build double barrel culvert	4	10	16
7	Clear & grub Sta. 42-100	1	2.5	7
8	Clear & grub Sta. 100-158	2.5	7	11.5
9	Build box culvert at Sta. 127	1	4	13
10	Build box culvert at Sta. 138	0.5	2	9.5
11	Cure double barrel culvert	3	9	15
12	Move dirt Sta. 42-100	2.5	4	11.5
13	Sta. moving dirt Sta. 100-158	1.5	3	4.5
14	Cure box culvert at Sta. 127	1.5	6	28.5
15	Cure box culvert at Sta. 138	2	5	14
16	Stockpile paving material	0.5	2	3.5
17	Place subbase Sta. 42-100	3.6	6.1	14
18	Fin. moving dirt Sta. 100-158	1	4	13
19	Pave Sta. 42-100	4	10	16
20	Place subbase Sta. 100-158	2	4.53	21.87
21	Cure pavement Sta. 42 -100	2.5	5.5	11.5
22	Pave Sta.100-158	3	6.75	30
23	Cure pavement Sta. 100-158	2.5	5.5	11.5
24	Place shoulders Sta. 42-100	Uniform (1,7)		
25	Place shoulders Sta. 100-158	Uniform (1,7)		
26	Guardrail and landscape	Discrete {2.5,0.3;4,0.6;11,0.1}		

Table 2 lists the precedence relationships between activi ties.

Table 2. Precedence relationships

ID	Activity	Predecessor
1	Setup batch plant	-
2	Deliver paving mesh	-
3	Rebar, double barrel culvert	-
4	Move in equipment	-
5	Rebar, small box culvert	-
6	Build double barrel culvert	3
7	Clear & grub Sta. 42-100	4
8	Clear & grub Sta. 100-158	4
9	Build box culvert at Sta. 127	5
10	Build box culvert at Sta. 138	5
11	Cure double barrel culvert	6
12	Move dirt Sta. 42-100	7, 11
13	Sta. moving dirt Sta. 100-158	8
14	Cure box culvert at Sta. 127	9
15	Cure box culvert at Sta. 138	10
16	Stockpile paving material	1
17	Place subbase Sta. 42-100	12
18	Fin. moving dirt Sta. 100-158	13, 14, 15
19	Pave Sta. 42-100	2, 16, 17

20	Place subbase Sta. 100-158	18
21	Cure pavement Sta. 42 -100	19
22	Pave Sta.100-158	2, 16, 20
23	Cure pavement Sta. 100-158	22
24	Place shoulders Sta. 42-100	21
25	Place shoulders Sta. 100-158	23
26	Guardrail and landscape	24, 25

5.2 Specified Correlation

The correlations between activity pairs (13, 18), (17, 19), (20, 22), (19, 24), and (22, 25) are estimated to be strong because each pair of activities is performed by the same crew. Thus the correlation coefficient is assigned to be 0.7. Moreover, because two box culverts are close to each other (at Stations 127 and 138) and share similar geological conditions, the correlation coefficients between activity pairs (9, 10) and (14, 15) are assigned to be 0.5. The correlation coefficient between two curing activities (21, 23) is estimated to be 0.3. These different levels of correlations are intentional to show that the proposed model can treat any specified correlation structure.

Table 3. Correlation between activity durations

Activity-Activity				Correlation Coefficient	
ID	Act.	ID	Act.	Spec.	Prop.
13	Sta. moving dirt Sta. 100-158	18	Fin. moving dirt Sta. 100-158	0.7	0.7242
17	Place subbase Sta. 42-100	19	Pave Sta. 42-100	0.7	0.6960
20	Place subbase Sta. 100-158	22	Pave Sta.100-158	0.7	0.7230
19	Pave Sta. 42-100	24	Place shoulders Sta. 42-100	0.7	0.7175
22	Pave Sta.100-158	25	Place shoulders Sta. 100-158	0.7	0.6794
9	Build box culvert at Sta. 127	10	Build box culvert at Sta. 138	0.5	0.4707
14	Cure box culvert at Sta. 127	15	Cure box culvert at Sta. 138	0.5	0.5169
21	Cure pavement Sta. 42 -100	23	Cure pavement Sta. 100-158	0.3	0.2477

5.3 Simulation Result

The proposed model is implemented in Matlab code. After 1000 simulation replications, the correlation coefficients between sampled activity durations are compared to those originally specified in Table 3. While most differences are around 0.01~0.02, the largest is no more than 0.03. This validates that the proposed model can capture the correlations between activity durations with little deviations

Figure 2 analyzes the 1000 samples of project durations with sample mean and standard deviation being 77.28 and 12.06 days, respectively. The 5% probability threshold of the project duration is 60.14 days and the 95% one is 99.22 days. Palisade BestFit software is used to fit probability distributions to the samples. A beta distribution, plotting as the curve, is chosen because it ranks first in both K-S and A-D goodness-of-fit tests [7]. The fitted beta distribution has the kurtosis of 3.25 (peaked) and the skewness of 0.65 (skewed to the right).

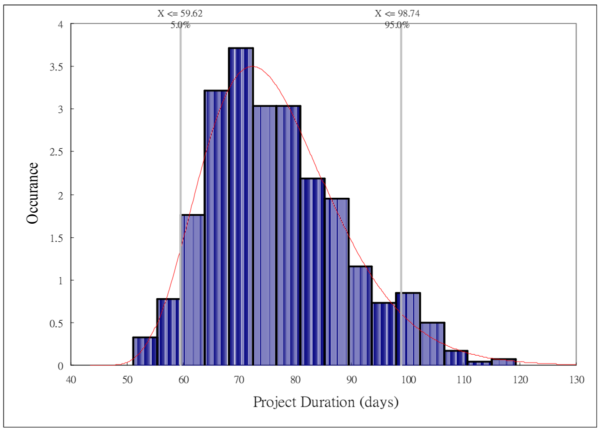

Figure 2. Distribution of project duration

5.4 Further Analysis

In this section, we compare the proposed model with PERT and conventional simulation without correlations. The PERT procedure starts with computing the mean and standard deviation of the duration of each activity. In our example project, activity durations are estimated based on triangular, uniform, and discrete distributions, thus their means and standard deviations are estimated according to the corresponding formulas. By adding up the means and variances of activities on the longest path, one can find the mean and standard deviation of the project duration to be 71.47 and 11.98 days.

The conventional simulation approach without correlations is implemented by sampling activity durations based on their estimated distributions, repeating 1000 replications, and returning summary statistics. The sample mean and standard deviation are 76.46 and 10.14 days, respectively.

All the summary statistics of three approaches are reflected in the box plot in Figure 3. The result of PERT is symmetric because of the underlying normal distribution. The conventional simulation approach generates a result closer to the proposed model but still overlooks the possibility of having the project duration over 110 days. In conclusion, the comparison raises two issues. First, the symmetry assumed by PERT cannot address the true skewness. Second, positive correlations between activity durations would increase the standard deviation of project duration and therefore increase the uncertainty of on-time completion.

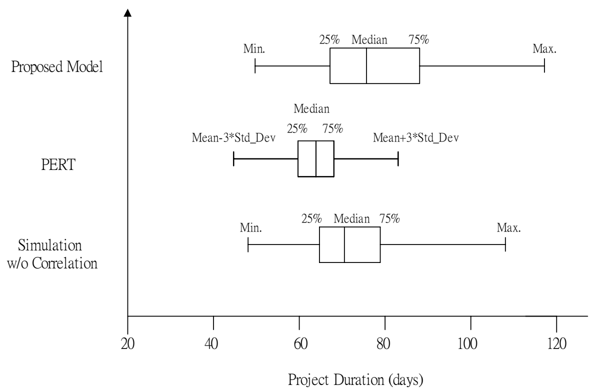

Figure 3. Comparison in box plot

6. CONCLUSIONS

In this paper, we propose a simulation model to evaluate the true effect of correlations between activity durations on the project duration. The proposed model is more flexible than previous attempts in the following aspects and hence possesses higher practicability. First, the proposed model is able to treat a mixed situation when marginal distributions of activity durations are not members of the same family and even involve both continuous and discrete distributions in the same model. Second, there is no restriction on the selection of correlation coefficients between activity durations. These flexibilities are especially important when systematic data is not available and planners have to rely on experts' subjective estimation on (1) parameters of marginal distributions, (2) types of marginal distributions, and (3) correlation coefficients between activity durations.

An actual highway project is used to demonstrate the application of the proposed model. The sample correlation coefficients between activity durations are shown to be reasonably close to the originally specified ones. This along with the proof that the activity durations generated would match the specified marginal distributions (by the inverse-transform concepts) validates the proposed model.

An ongoing research is attempting to improve the accuracy of the proposed model by minimizing the approximation error between the specified and generated correlation matrix.

REFERENCES

[1] Ahuja, H. N., Nandakumar, V. "Simulation model to forecast project completion time." *Journal of Construction Engineering and Management*, 111(4), 325–342, 1985.

[2] Cario, M. C., Nelson, B. L. "Modeling and generating random vectors with arbitrary marginal distributions and correlation matrix." Technical Report, Dept. of Industrial Engineering and Management Science, Northwestern U., Evanston, Ill., 1997.

[3] Carr, R. I. "Simulation of construction project duration." *Journal of Construction Division*, 105(CO2), 117-128, 1979.

[4] Ghosh, S., Henderson S. G. "Behavior of the NORTA method for correlated random vector generation as the dimension increases." *ACM Transactions on Modeling and Computer Simulation*, 13(3), 276–294, 2003.

[5] Hull, J. C. "Dealing with dependence in risk simulations." *Operation Research Quarterly*, 28, 201-213, 1977.

[6] Ioannou, P. G., Martinez, J. C. "Project scheduling using state-based probabilistic decision networks." *Proceedings of Winter Simulation Conference 98*, Washington D.C., December 13-16, 1287-1295, 1998.

[7] Law, A., Kelton, W. D. *Simulation modeling and analysis*, 3rd edition, McGraw-Hill, New York, 2000.

[8] Levitt, R. E., Kunz, J. C. "Using knowledge of construction and project management for automated schedule updating." *Project Management Journal*, 16(5), 57–76, 1985.

[9] Mood, A. M., Graybill, F. A., Boes, D. C. *Introduction to the theory of statistics*, 3rd edition, McGraw-Hill, New York, 1974.

[10] Padilla, E. M., Carr, R. I. "Resource strategies for dynamic project management." *Journal of Construction Engineering and Management*, 117(2), 279–293, 1991.

[11] Press, W. H., Teukolsky, S. A., Vetterling, W. T., Flannery, B. P. *Numerical recipes in C: the art of scientific computing*, 2nd ed., Cambridge University Press, Cambridge, 1992.

[12] Ranasinghe, M., Russell, A. D. "Treatment of correlation for risk analysis of engineering projects." *Civil Engineering Systems*, 9, 17–39, 1992.

[13] Wang, W. C., Demsetz, L. A. "Application example for evaluating networks considering correlation." *Journal of Construction Engineering and Management*, 126(6), 467–474, 2000.

[14] Woolery, J. C., Crandall, 'K. C. 'Stochastic network model for planning scheduling." *Journal of Construction Engineering and Management*, 109(3), 342–354, 1983.