COMBINATION OF APPROXIMATION AND SIMULATION APPROACHES FOR DISTRIBUTION FUNCTIONS IN STOCHASTIC NETWORKS

S. M. T. Fatemi Ghomi and F. Jolai Ghazvini

Department of Industrial Engineering
Amirkabir University of Technology
Tehran, Iran

Abstract This paper deals with the fundamental problem of estimating the distribution function (df) of the duration of the longest path in the stochastic activity network such as PERT network. First a technique is introduced to reduce variance in Conditional Monte Carlo Sampling (CMCS). Second, based on this technique a new procedure is developed for CMCS. Third, a combined approach of simulation and approximation procedures is introduced for the networks with activity discrete distribution function to enhance the accuracy of the approximation procedures. Application of the new approach proves that the error is drastically reduced in comparison with the best existing approximation approach.

Key Words PERT Network , Completion Time, Distribution Function, Discrete, Simulation, Approximation, Combination

چکیده این مقاله به مسئله اساسی تخمین تابع توزیع مدت زمان طولانی ترین مسیر در شبکه های فعالیت از نوع پرت احتمالی می پردازد. ابتدا تکنیکی معرفی می شود تا واریانس را در نمونه گیری شرطی مونت کارلو کاهش دهد. سپس براساس این تکنیک روش جدیدی برای نمونه گیری شرطی مونت کارلو ارائه می شود. در مرحله بعدی، رویه ای حاصل از ترکیب روشهای شبیه سازی و تقریبی برای شبکه هایی که فعالیت های آنها تابع توزیع گسسته دارند معرفی می شود تا دقت روشهای تقریبی را افزایش دهد. کاربرد رویه ترکیبی ثابت می کند که خطا در مقایسه با بهترین روش تقریبی موجود به مقدار قابل ملاحظه ای کاهش می یابد.

INTRODUCTION

A stochastic network is an acyclic, connected, and directed graph consisting of N nodes and A arcs in which the duration of some or all of the arcs are random variables with known cumulative distribution functions. The network can have a single starting and a single terminating node. The nodes can be numbered such that an arc $(ij) \in A$ leads from a smaller to a larger numbered node. Therefore, we take the strating node to be node 1 and the terminating node to be nodeN.

A fundamental problem in stochastic networks is the distribution function (df) of the duration of the longest path in the network. Usually, the duration of the longest path designates the completion time of the project represented by the stochastic network. Calculating the exact df is generally very difficult; the difficulty emanates from the dependency between the paths in the network.

Many studies have been carried out to calculate completion time df in stochastic networks. In general, these studies fall into several catagories: (i) analytical procedures to approximate or bound mean completion time of the project [1-6]; (ii) analytical procedures to calculate or approximate network CDF. [7, 12]; (iii) Monte Carlo sampling procedures to approximate network CDF [13-16]; and (iv) analytical procedure to bound network CDF [17,18]. A comprehensive review of most of the above mentioned references is presented in chapter 4 of Elmaghraby [19]. The following is a short explanation on the approximation and simulation approaches.

Dodin [4] presents an approximating procedure for large networks. O'Conner [12] develops a procedure to evaluate network cumulative distribution function when the arcs have discrete df's. He develops an analytical expression for network CDF by conditioning on most of the common arcs (arcs which are constituents of more than one path), then uses complete enumeration to evaluate the simplified expression.

Van Slyke [16] develops the idea of crude simulation as a tool for finding the CDF of a PERT network's completion time. He also suggests two methods of potentially reducing simulation computational times. Antithetic variables, stratification, control variates and regression have been suggested by Burt and Garman [13] as ways of reducing the computational effort required in curde simulation. Burt and Garman [14] also developed a new simulation procedure called conditional Monte Carlo simulation in which certain activity times are fixed at their original sampled value thus reducing computational effort and variance.

Burt and Garman [13] conditioned on the set of all common arcs in the network, where they developed a procedure to identify the common arcs in the network. They concluded in [14] that the accuracy of the approximate df obtained by Conditional Monte Carlo Sampling (CMCS) increases as the number of arcs to condition on is reduced; this conclusion is based on the contention that fewer arcs to sample and more arcs in the analytical conditional probability expression leads to greater accuracy. They also recommend the use of CMCS if the ratio of the number of arcs not to condition on reaches 20% or higher.

The above conclusions led to the development of other conditional procedures. Garman [20] extended the concept of series-parallel reduction of stochastic PERT networks, developed by Marin [9], to CMCS. He first reduces the networks as far as possible by series-parallel reductions; if the network is not reduced to a single arc, then he showed that such a network possesses a node with only one arc incident into it and more than one arc emanating from it and vice-versa. By the repeated conditioning on one of the single arcs incident into (or emanating from) the node, the network can be reduced to a single activity whose distribution function can be evaluated by Monte Carlo sampling. Sigal, Pritsker, and Solberg [15] introduced and applied the concept of maximum directed cutset in the network to reduce the number of arcs to condition on. In their CMCS procedure they replace the common arcs (used by Burt and Garman) by the arcs that are not elements of the maximum directed cutset.

Dodin [21] discusses the issue of the minimum number of arcs to condition on in CMCS of stochastic network. In his paper first he shows that none of the existing CMCS procedures condition on the minimum number of arcs. Second, he introduces the concept of the path index of the arc and develops a procedure to calculate the path indices of arcs without identifying the paths. Third, he uses the path indices with the well-known series-parallel reduction to develop the new conditional procedure which conditions on no more than N-3 arcs.

In section 1 of this paper, a new technique is introduced to reduce variance in CMCS. A new procedure based on this technique is developed in section 2. Section 3 introduces a combined approach for the networks with activity discrete distribution function. This approach uses crude simulation with an approximation procedure, which has he benefits of both approximation and simulation approaches. Section 4 is a generalization of the subject in section 3, which uses conditional simulation instead of crude simulation. This section includes the application of the new approach. Finally, section 5 is devoted to summary and concluding remarks.

1. A NEW TECHNIQUE FOR VARIANCE REDUCTION IN CMCS

Burt and Garman [14] presented the algorithms of conditional simulation in which one activity is selected in each run of the simulation. But it is probable that more than one qualified activity exist to condition on. This issue which occurs especially when the network of interest has complex structure and large size, has not been discussed yet.

The new technique considers the activity which has less effect on the variations of network completion time. In other words, the activity has low correlation coefficient with project completion time. When such an activity is selected and takes a constant value, the error in the estimation of distribution function of the project completion time would be decreased. One advantage of this technique in comparison with the technique of common arcs or path index is simultaneous consideration of variance of the activity distribution and the position of the activity in the network. Other techniques behave only on the basis of network configuration and activity position. But the weakness of the proposed technique is that it needs correlation coefficient to be computed before the simulation process can be started.

As an example, consider the Wheatstone Bridge type of network Figure 1. The activities are Exponential with parameters 0.20, 0.31, 0.27, 0.50 and 0.50 respectively.

Initial simulation is made for 1000 runs to compute the correlation coefficient for activities 1 and 5 which are shown in Table 1.

Network Figure 1 can be solved in two ways by CMCS: one way is conditioning on activity 1 and the other way is conditioning on activity 5. The results of simulating the network for two cases are given in Table 2.

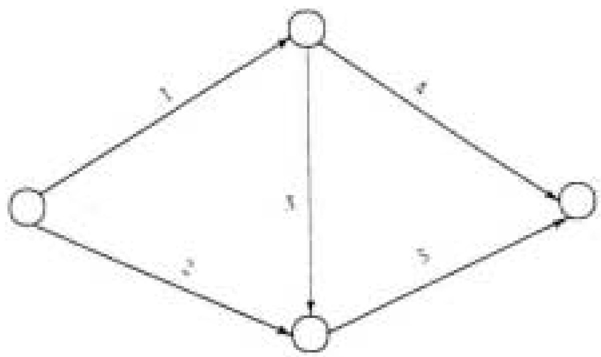

Figure 1. A typical network for study.

It is concluded that conditioning on activity 5 gives better results than those of conditioning on activity 1. Activity 5 has lower correlation coefficient, and hence the proposed technique is efficient.

2. A NEW PROCEDURE FOR CMCS

In the previous section correlation coefficient was used as a technique to find out which activity is selected for conditioning in CMCS when more than one qualified activity exist. Now, a procedure is presented based on this technique. Namely, the correlation coefficient of each activity of the network is provided and instead of providing critical index of each activity, as Dodin [6] presented they are ordered in ascending manner. In previous section it was shown that when conditioning on one activity causes the network to be solved easier, conditioning on the

TABLE 1. Correlation Coefficient for Network Figure 1.

	Exponential parameter	correlation coefficient
Activity 1	0.20	0.76
Activity 5	0.50	0.33

TABLE 2. Crude Simulation and CMCS for Network Figure 1.

t	pr (T≤t) crude simulation 10,000 runs	pr (T≤t|Activity 1) simulation 10,000 runs	pr (T≤t|Activity 5) simulation 10,000 runs
1	0.00027	0.000237449	0.0002427918
5	0.11055	0.0884261	0.09566084
10	0.48527	0.4461597	0.466216
15	0.76595	0.6495118	0.7117353
20	0.90409	0.8835261	0.8904171
25	0.96230	0.9200181	0.9562413

activity with least correlation coefficient is preferable. But using this procedure does not guarantee to solve the network with the number of activities fewer than those required by the path index procedure of Dodin [6]. If it can be shown that both procedures will solve the network with the same number of activities to condition on them, the correlation coefficient procedure will prove to be advantageous. For this purpose, some networks with different configurations and sizes are considered in Figure 2. The distribution of activities is Exponential with the parameter between 0.1 to 0.5. Series parallel operation is accomplished on the networks, once using the activity path index and once the list of activity correlation coefficient to identify the number of activities to condition on them. In all cases, not only the number of activities was the same for two procedures, but also, frequently the activities themselves were the same. This result especially is apparent in more complex networks. It appears that the path index procedure is preferable because the computation of path index is easier than the computation of correlation coefficient.

3. COMBINED APPROACH

When the size of a network is very large, even the activity distributions are approximated by discrete three-point distributions, series-parallel operations require the allocation of extensive memory of computer. This necessitates the utilization of approximation procedures which does not need to perform the operations. But the approximation procedures operate on the basis of unreasonable assumptions, like independency of paths, to reach the solution quickly. Hence, the high error will result when the network size is large and consequently many dependent paths exist.

A new approach introduced here, combines the simulation with an approximation procedure. This procedure acts in this way that random values are given for a reasonable number of activities in each run of the simulation, and the network is solved with an approximation procedure. So, without any need to perform the series-parallel operation, the dependency between a number of paths is demolished; consequently the approximation procedure presents more accurate solution.

The efficiency of proposed approach can be enhanced with the caution in the selection of activities to condition on them. However, the selection of activities depends on the behaviour of the operation of the approximation procedure when it is utilized in combination with the simulation.

4. GENERALIZATION AND APPLICATION OF COMBINED APPROACH

By conditioning, it is possible to use the actual distributions of some activities to reduce the error due

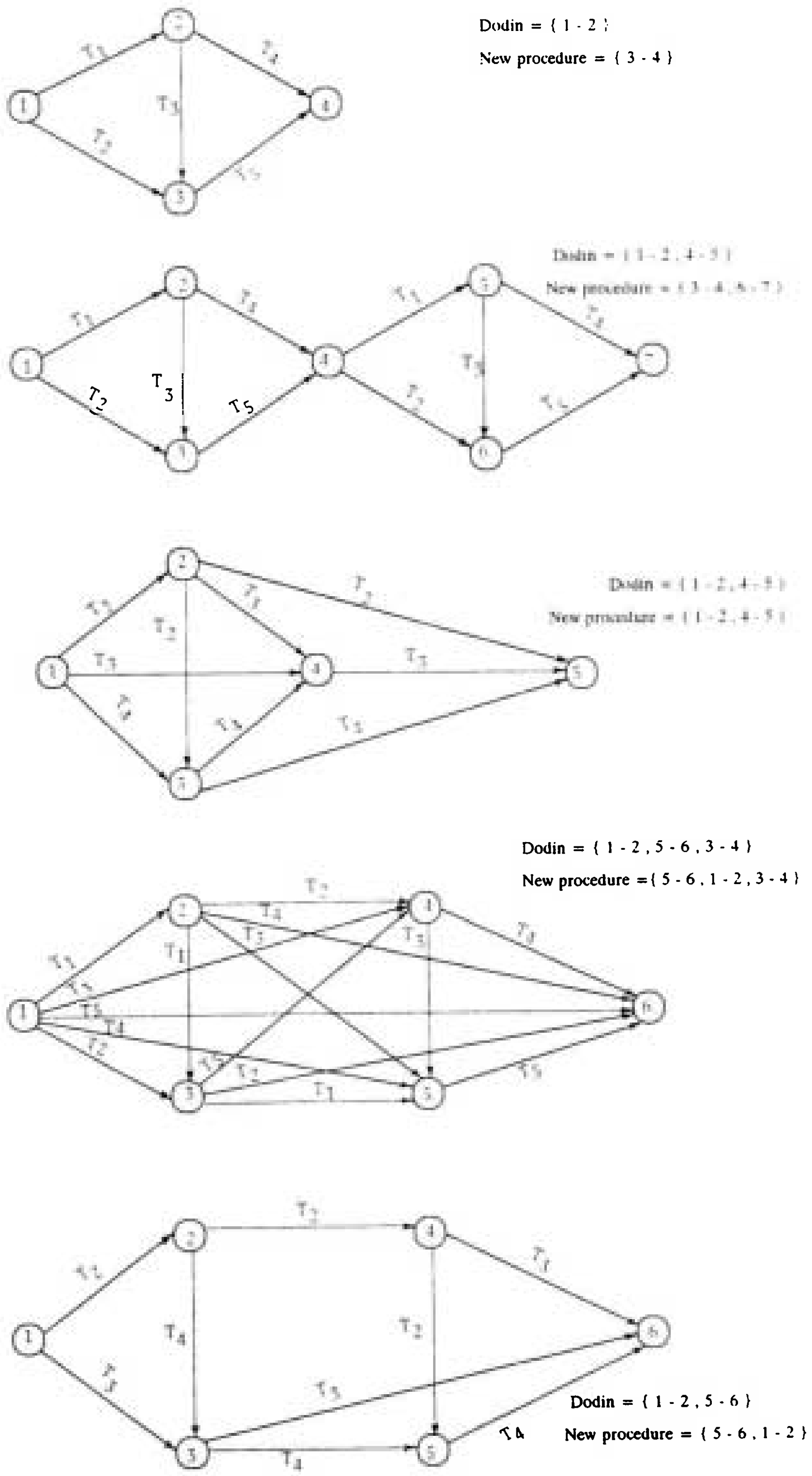

Figure 2. Different network configurations to identify the number of activities for conditioning (Dodin's and new procedure).

to the approximation. But in the network with activity discrete distribution, there is not a closed formula like the case where the activities have continuous distributions. Here there are discrete points, which is a restriction. For example, suppose a netwrok has nine activitis and each activity has 3-point discrete distribution. If conditioning on six activities are made and the actual distributions of the other three activities are used, there would be 27= 3 × 3 × 3 combinations that N runs of simulation are needed for each combination. Now, if a network is complex and has many activities, at least 50 actual activity distributions may be needed which require $(3)^{50}$ combinations.

The combined approach presented in the previous section can also be used here to utilize an approximation procedure instead of N runs of simulation for each resultant combination.

To illustrate the high efficiency of the proposed procedure, different tests are designed. During the performed tests, a comparison is also made between the efficiency of activity correlation coefficient, activity path index, and activity critical index, in addition to the test of efficiency of conditional simulation in networks with activity discrete distribution. This comparison is vital because here a further problem is how the activities are identified for conditioning.

Consider Network Figure 3. Path index procedure identifies the activities 1-2 and 5-6 at the top of the list of activities. Correlation coefficient procedure identifies the activities 2-3 and 3-5 with the highest correlation coefficient. Based on these four activities, different combinations as pair of activities are made to consider their actual distributions to be used in the simulation process of the network presented in Figure 3.

Network Figure 3 is analytically tractable. The analytical solution is used as a basis of comparison. 10,000 runs of simulation is considered for each combination of activities. Simulation results and the error compared to the real solution are given in Table 3.

It is observed that the best case is for the combination of two activites 3-5 and 5-6. Repeating the same number of simulation but with only one activity, the results given in Table 4 are achieved. The results indicate even when the distribution of one activity is considered, better solutions than those of other approximation procedures are obtained.

It is necessary to mention that the best existing approximation procedure of Dodin solves the network

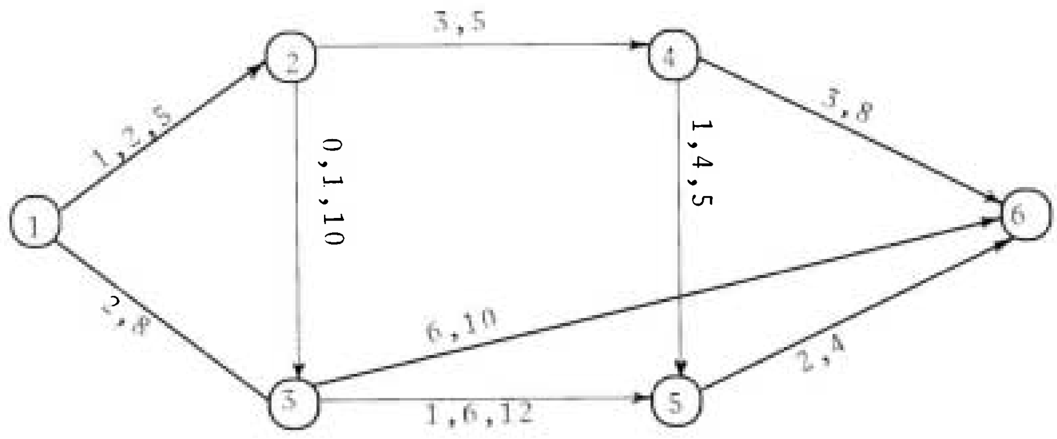

Figure 3. A network example for combined approach with activity discrete distribution.

TABLE 3. Prabability Value for Network Figure 3.

The combination of activities to consider their real values					
t	**real optimum value**	**1-2, 2-3**	**3-5, 5-6**	**1-2, 5-6**	**2-3, 3-5**
8	0.001157	0.00118	0.001083	0.00115	9.6×10^{-4}
9	0.00192902	0.00158	0.001783	0.00215	1.72×10^{-3}
10	0.005401235	0.00556	0.0051	0.005716	0.00554
11	0.005015432	0.00507	0.00473	0.0053	0.00495
12	0.040509259	0.04066	0.040283	0.04035	0.04047
13	0.038580247	0.03886	0.03856	0.037016	0.03773
14	0.0625	0.06257	0.06273	0.0623	0.06226
15	0.040123457	0.04062	0.0413	0.040083	0.04013
16	0.092592593	0.09131	0.092816	0.09206	0.09251
17	0.040123457	0.0397	0.04076	0.039183	0.4053
18	0.206018519	0.20834	0.20513	0.2086	0.2067
19	0.040123457	0.0397	0.04076	0.039183	0.0393
20	0.018518519	0.01883	0.018483	0.017783	0.01894
21	0.074074074	0.07514	0.073716	0.074883	0.07514
22	0.111111111	0.10961	0.11143	0.112083	0.11087
23	0.009259259	0.00914	0.00953	0.00915	0.00925
24	0.0555556	0.5572	0.055816	0.05635	0.05648
25	0.064814815	0.06378	0.06405	0.063183	0.06393
26	0.018518519	0.0187	0.0189	0.0187	0.01845
27	0.018518519	0.0185	0.01823	0.01896	0.01842
28	0.185185519	0.0179	0.01865	0.01816	0.01867
29	0.185185519	0.01893	0.01187	0.01865	0.01854
31	0.185185519	0.01848	0.01843	0.17616	0.0183
	percent of error	1.15%	76%	1.43%	79%

Figure 3 with 12.8% error. Also crude simulation of the same network with 100,000 runs has the error about 1%.

What is obvious is that correlation coefficient procedure is advantageous when only one activity is considered. Again for more evaluation of the proposed procedure, Network Figure 4 has been simulated of which the results are given in Table 5.

First the list of activities is obtained based on the procedures of activity path index, activity critical index, and activity correlation coefficient. Activities 1-2 and 4-5 constituted the activities of the first and second ranks in the lists of activity path index and activity cirtical index respectively. Activities 2-4 and 1-2 constituted the activities of the first and second ranks in the list of activity correlation coefficient respectively. The results of each case are then compared with the real values and the absolute total error is computed as the accuracy criterion.

As Table 5 indicates, it is observed that the correlation coefficient procedure has produced better results. It is necessary to mention that Dodin's approximation procedure has the error about 5.9% which is far from the results produced by 1000 runs of crude simulation based on activity 2-4. This simulation has the error of 4.2%.

TABLE 4. Combined Approach Results of Network Figure 3 when the Distribution of only One Activity is Used.

Activity	2-3	3-5	5-6
percent of error	0.80	1.28	1.34

5. SUMMARY AND CONCLUDING REMARKS

Procedures developed by Garman and Dodin consider one activity each time a simulation is made in CMCS. But in the case where there are several qualified activities to condition on, they have no recommendation for activity identification to condition. In this paper the correlation coefficient of the activity is introduced as a new technique and its efficiency is shown.

In another study, again the correlation coefficient concept is developed as an independent procedure to identify the activities to condition on them in CMCS. Comparisons are made between this procedure and the procedure of path index of the arc developed by Dodin. Finally it is concluded that correlation coefficient procedure is more efficient, but the path index procedure is simpler to use.

To enhance the accuracy of approximation

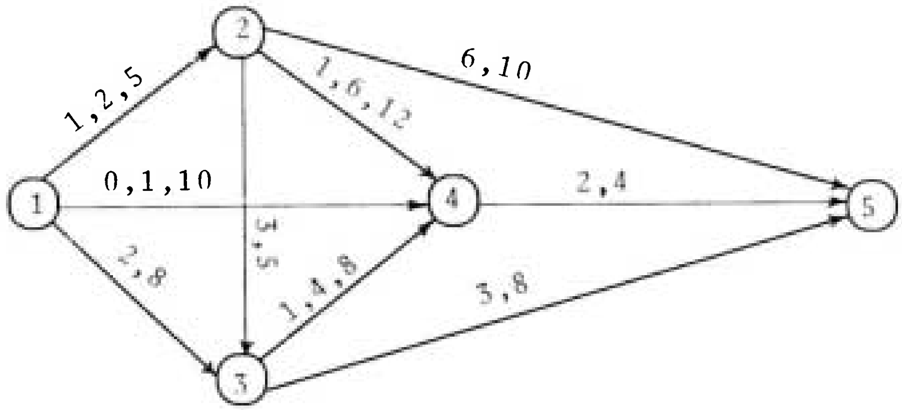

Figure 4. Another network example to apply combined approach.

TABLE 5. The Results of Combined Approach for Network Figure 4.

The activity (or activities) to consider the real distribution (or distributions)	percent error of simulation
1-2	1.44
2-4	1.29
4-5	1.53
1-2, 4-5	1.22
1-2, 2-4	1.17

procedures in the networks having activity discrete distribution, a combined approach is introduced. This new approach has the benefits of both simulation and approximation procedures and is capable of producing the solutions much more accurate than those of existing approximation procedures. This approach also permits the use of conditional simulation for the networks with activity discrete distribution.

In summary, the results of combined approach prove it is more efficient in comparison with the best existing approximation procedure developed by Dodin and can be used for different types of networks even when they are complex and have large size.

REFERENCES

1. C. T. Clingen, "A Modification of Fulkerson's PERT Algorithm," *Operations Research*, 12, (1964), 629-632.
2. L. P. Devroye, "Inequalities for the Completion Times of Stochastic PERT Network," *Mathematics and Operations Research*, 4, (1980), 441-447.
3. D. R. Fulkerson, "Expected Critical Path Lengths in PERT Networks," *Operations Research*, 10, (1962), 808-817.
4. S. E. Elmaghraby, "On the Expected Duration of PERT Type Networks," *Management Science*, 13, (1967), 299-306.
5. D. J. Malcolm, J. H. Roseboom, C. E. Clark and W. Fazar, "Applications of a Technique for Research and Development Program Evaluation," *Operations Research*, 7, (1959), 646-669.
6. P. Robillard and M. Trahan, "Expected Completion Time in PERT Networks," *Operations Research*, 24, (1976), 177-182.
7. B. Dodin, "Approximating the Distribution Functions in Stochastic Networks," *Computers and Operations Research*, 12, (1985 a), 251-264.
8. H. O. Hartley and A. W. Wortham, "A Statistical Theory for PERT Critical Path Analysis," *Management Science*, 12, (1966), 469-481.
9. J. J. Martin, "Distribution of the Time through a Directed, Acyclic Network," *Operations Research*, 13, (1965), 46-66.
10. L. J. Ringer, "Numerical Operators for Statistical PERT Critical Path Analysis," *Management Science*, 16, (1969), 136-143.
11. L. J. Ringer, "A Statistical Theory for PERT in which Completion Times of Activities are Inter-dependent," *Management Science*, 17, (1971), 717-723.
12. D. R. O'Conner, "Exact Distributions of Stochastic PERT Networks," University of College of Dublin, Belfield, Dublin 4, Ireland, (1981).
13. J. M. Burt and M. B. Garman, " Monte Carlo Techniques for Stochastic PERT Network Analysis," *INFOR*, 9, (1971), 248-262.
14. J. M. Burt and M. B. Garman, "Conditional Monte Carlo: a Simulation Technique for Stochastic Analysis," *Management Science*, 18, (1971), 207-217.
15. C. E. Sigal, A. A. B. Pritsker and J. J. Solberg, "The Use of Cutsets in Monte Carlo Analysis of Stochastic Networks, *Mathematics and Computers in Simulation*, 21, (1979), 376-384.
16. R. M. Van Slyke, "Monte Carlo Methods and the PERT Problem, " *Operation Research*, 11, (1963), 839-860.
17. G. B. Kleindorfer, "Bounding Distributions for a Stochastic Acyclic Network," *Operations Research*, 19, (1971), 1586-1601.
18. A. W. Shogan, "Bounding Distributions for a Stochastic

PERT Network, *Network*, 7, (1977), 359-381.

19. S. E. Elmaghraby, "Activity Networks: Project Planning and Control by Network Models," John Wiley & Sons Inc., New York, (1977).

20. M. B. Garman, "More on Conditional Sampling in the Simulation of Stochastic Networks," *Managament Science*, 19, (1972), 90-95.

21. B. Dodin, "Minimum Number of Arcs in Conditional Monte Carlo Sampling of Stochastic Networks, " *INFOR*, 24, (1986), 33-44.