



Institute of Internet and Intelligent Technologies
Vilnius Gediminas Technical University
Saulėtekio al. 11, 10223 Vilnius, Lithuania
<http://www.isarc2008.vgtu.lt/>

**The 25th International Symposium
on Automation and Robotics in Construction**

June 26–29, 2008

ISARC-2008

MONTE CARLO SIMULATION WITH EXACT ANALYSIS FOR STOCHASTIC PERT NETWORKS

Zdzisław Milian

Cracow University of Technology
31-155 Kraków, ul. Warszawska 24
milian@usk.pk.edu.pl

ABSTRACT

This paper deals with one of methods for estimating the probability distribution of the project completion time in stochastic PERT network. This method combines Monte Carlo simulation with exact analysis. The exact probability distribution for a subnet is used to improve an estimator of probability distribution for larger networks. In the paper computational results for different distributions of task duration are presented. More complicated than unimodal distributions are considered also. Both symbolic and simulation calculation were made with MatLab and own procedures.

KEYWORDS

PERT network, control variates, exact distribution

1. INTRODUCTION

Many planning problems can be represented by PERT networks. A graph in PERT networks composed of nodes and arcs is directed, acyclic and connected. In a PERT network for a project, an activity (task, job) can be represented by an arc, and an event by a node. The structure of the network represents the order in which these activities may be performed, and the project duration is simply the longest path through the network. In stochastic PERT networks, it is assumed that the random variables (RV) describing the task duration are associated with some or all the arcs. If the duration of each activity is assumed to have an associated probability distribution, then the problem of determining the project completion time (PCT) immediately presents itself. This ancient problem

was formulated by Malcolm et al. who proposed a simple approximation method for evaluating the mean and the variance and using the Gaussian distribution as a distribution of the project duration (PERT method). The approximate methods were the subject of a great number of papers. They deal with various ways of estimating the cumulative density function (c.d.f.) of PCT in the networks and/or the moments of this distribution in order to avoid difficulties arising in the exact method. Monte Carlo simulation methods provides a powerful methodology to obtain desired statistics for any network with specified distribution of task durations. These methods have been discussed in literature [1,2,6,7]. However, the straightforward Monte Carlo method requires large amount of computational expense (see Table 1).

Table 1

Type of estimation	Size of sample
The expected duration of PCT, it should be within $\sigma/50$ of the true mean, where σ is known and denotes standard deviation	10 000
The variance of a project duration, it should be correct within 5%	3 000
Probability within the limit of 0.01, that a task will be lie on the critical path	10 000
The c.d.f. of PCT (the greatest absolute difference between the estimation of c.d.f. and the true c.d.f. is less then 0.01)	18 500

In this paper we want to indicate an improved technique of simulation that uses analytical solutions concerning subnet to estimate the unknown c.d.f. of lager networks (Control Variates technique). Although one can find many papers considering the PERT problems, there are few concerning the exact methods and giving analytical solutions.

Since analytical solutions can be more available [4,5], we will show that the Control Variate (CV) technique can find wider applications, for not unimodal distributions also. First, we recall some information on the simplest simulation method that is used in PERT networks.

“Straightforward” Sampling.

By a realization of a network we mean the network with a fixed value for each of its arc duration. This value is obtained through a random sample from all possible values from each arc following the c.d.f. of the arc duration. For a particular realization of network we apply the critical path algorithm to evaluate PCT. In this way we obtain a sample value of PCT. Next, by repeating this procedure to a long series of realizations, one may apply standard statistical methods to estimate the c.d.f. of PCT and /or its parameters of interest.

It is clear that an amount of computing effort to obtain estimates depends on the size of samples. In Table 1 we present examples of the sample size [7].

All sizes of sample in Table 1 were calculated to ensure the accuracy of estimation with probability 0.95. One can show that if 0.99 confidence is desired, size sample should be taken much higher.

2. CONTROL VARIATES

The basic idea underlying Control Variates (CV) technique is the generation of a twin RV that is *positively correlated* with the original RV. This new RV is called the *control variate*.

Assume we wish to estimate some parameter θ of the RV X , say $\theta = g(X)$ where the function g may be expectation, variance or probability. Let $\Phi(X)$ denote the CV and we can write the identity $\theta = [g(X) - \Phi(X)] + \Phi(X)$. Since Φ is positively correlated with g , it “mimics” it in some fashion and the difference, $g(X) - \Phi(X)$, should be small. The latter term is estimated by Monte Carlo sampling with a great deal of precision. Thus θ can be estimated precisely because we assume $\Phi(X)$ can be done analytically. To illustrate power of this technique consider the project depicted by network of Fig. 1a.

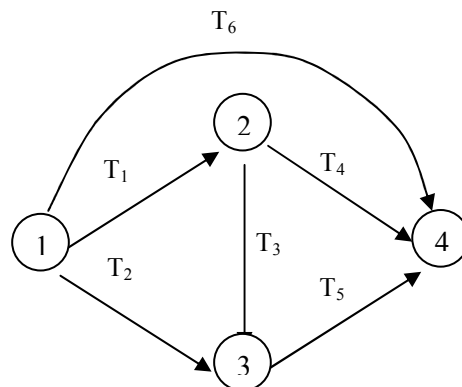


Figure 1a. Original network

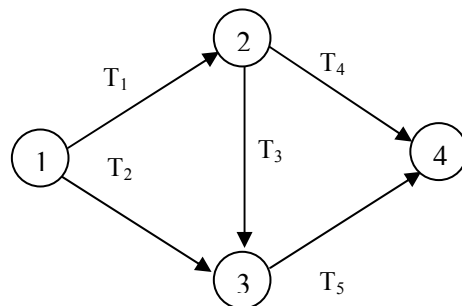


Figure 1b. CV network 1 (subnet b)

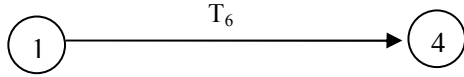


Figure 1c. CV network 2 (subnet c)

First, we estimate the expected value of PCT for the original network. We take n realizations of T_a^i (Fig. 1a) and T_b^i (Fig. 1b), using the same set of random numbers for each. To estimate $E(T_a)$ we use the following estimator

$$\bar{T} = \frac{1}{n} \sum_{i=1}^n (T_a^i - T_b^i) - E(T_b), \quad (1)$$

where $E(T_b)$ is evaluated analytically. It is easy to show that the estimator (1) is unbiased and often has a much smaller variance than does the straightforward estimate $\bar{T}_a = \frac{1}{n} \sum_{i=1}^n T_a^i$.

In a similar manner, we use the known c.d.f. of the CV. If $\bar{F}_a(t)$ and $\bar{F}_b(t)$ are empirical estimates of $F_a(t)$ and $F_b(t)$, then an improved estimator would be

$$\bar{F}(t) = \bar{F}_a(t) - \bar{F}_b(t) + F_b(t), \quad (2)$$

where $F_b(t)$ is evaluated analytically.

The success of the CV technique depends on how closely the CV, equals T_b , mimics the realization of $RV=T_a$ in the given network. To choose the CV network properly we can use the concept of “criticality” [7]. In brief, the criticality of any task of a given network is the probability that it will be lie on the critical path. If an activity has a high criticality, it should be included in a control network. This suggestion will be examine in next section.

3. VERIFICATION OF ESTIMATORS

We consider the original network (Fig. 1a) and two CV networks (Fig. 1b and 1c). The tasks are distributed as follows: T_1, T_3, T_5 – uniformly on the

segment $[0,1]$, T_2, T_4 – uniformly on the segment $[0,2]$, and T_6 is exponentially distributed with a parameter λ . Let $f_1(t)$, $f_2(t)$ and $f_3(t)$ denote the density functions of those distributions. They are illustrated in Fig. 2 and Fig. 3.

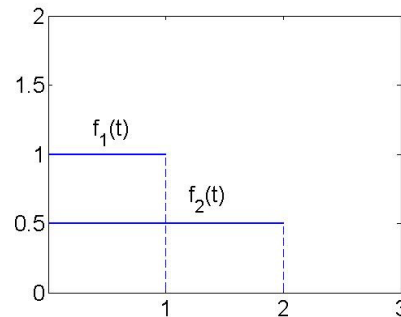


Figure 2. Distribution of T_1, T_2, T_3, T_4 and T_5

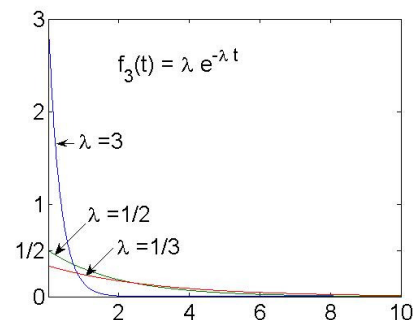


Figure 3. Distribution of T_6

Under such assumptions the c.d.f. of PCT, $F_b(t)$, related to the subnet b there is as follows

$$F_b(t) = \begin{cases} 0, & t < 0 \\ \frac{11}{480}t^5, & t \in [0,1] \\ -\frac{11}{240}t^5 + \frac{7}{32}t^4 - \frac{11}{48}t^3 + \frac{1}{16}t^2 + \frac{1}{32}t - \frac{7}{480}, & t \in [1,2] \\ \frac{11}{480}t^5 - \frac{5}{32}t^4 - \frac{1}{48}t^3 + \frac{31}{16}t^2 - \frac{111}{32}t + \frac{259}{160}, & t \in [2,3] \\ 1, & t > 3 \end{cases} \quad (3)$$

and is illustrated in Fig. 4. The probability density function (p.d.f) of PCT, $f_b(t)$, is shown in this figure also. The distribution for the CV network 2 is, obviously, identical to the distribution of T6, so, $F_c(t) = 1 - \exp(-\lambda t)$.

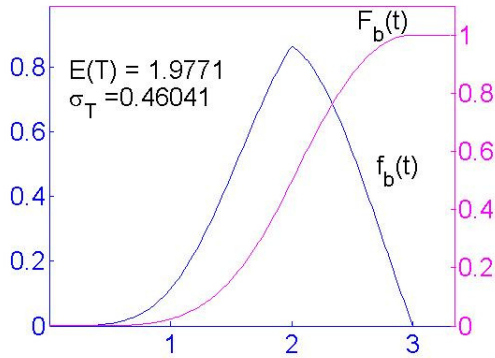


Figure 4. The p.d.f and c.d.f. of PCT

The way of calculation (3) is presented in [4,5]. Let us notice that subnet b (called as a bridge) is an irreducible network (not a series-parallel network). In the case of series-parallel network calculations are much simpler. The special program for calculating analytically the c.d.f. of PCT for any PERT network was written.

The statistical experiment was carried out to compare the quality of estimators based on the subnet b and subnet c. As mentioned earlier, the choice of the CV network has a great influence on the quality of estimators. To verify this problem,

simulation was carried out for $\lambda = 3, \frac{1}{3}$ and $\frac{1}{2}$, because it implies three cases: $E(T_b) > E(T_c)$, $E(T_b) < E(T_c)$ and $E(T_b) \approx E(T_c)$ respectively, since $E(T_b) = 1.9771$ and $E(T_c) = 1/\lambda$. It should be intuitively clear that criticality is a function of activity time's variance, as well as its mean. We can expect that if $E(T_b) > E(T_c)$ then the criticality of tasks of subnet c is greater than for subnet b. The c.d.f. $F_a(t)$ was approximated by four estimators

$\bar{F}_a(t)$, $\bar{F}_{a1}(t)$, $\bar{F}_{a2}(t)$, $\bar{F}_{a3}(t)$ as follows

$$\begin{aligned} \bar{F}_a(t) &= v(t) / n \\ \bar{F}_{a1}(t) &= \bar{F}_a(t) - \bar{F}_b(t) + F_b(t) \\ \bar{F}_{a2}(t) &= \bar{F}_a(t) - \bar{F}_c(t) + F_c(t) \\ \bar{F}_{a3}(t) &= \frac{1}{2} \bar{F}_{a1}(t) + \frac{1}{2} \bar{F}_{a2}(t) \end{aligned} \quad (4)$$

where: $v(t)$ - denotes the number of realizations of T_a that do not exceed t , n - the number of realizations, $\bar{F}_b(t)$ i $\bar{F}_c(t)$ calculated similarly to $\bar{F}_a(t)$ for the subnets b and c.

The size of samples n was 200. The calculation results are presented in Fig.5a-e.

The quality of approximation is evaluated by the root mean squared error RMSE and $RMSE_{0.5}$ (the root mean squared error for $F_a(t) > 0.5$). Since

$$F_a(t) = F_b(t)F_c(t) = F_b(t)(1 - \exp(-\lambda t)), \quad (5)$$

we can calculate the error exactly.

From comparison of graphs, RMSE and $RMSE_{0.5}$

follows that $\bar{F}_a(t)$ (straightforward simulation) is the worst estimator. The estimators based on the CV networks are much better. It was observed that the best estimators are: $\bar{F}_{a1}(t)$ - if $E(T_b) > E(T_c)$, $\bar{F}_{a2}(t)$ - if $E(T_b) < E(T_c)$ and $\bar{F}_{a3}(t)$ - if $E(T_b) \approx E(T_c)$.

3. APPLICATION NOTES

We want to point that the sample size was 200, this means 15 to 100 times decrease compared to Table 1. However, we have to include a price has to be paid in the form of additional computing. This is because, analytical integration over more and more complicated regions is needed and the formulae became longer and longer as number of tasks increases. Referring to computational complexity it is shown [3], that the problem of computing the c.d.f. of PCT within a given deadline is NP-hard even for discrete distributions of task durations, and simple series-parallel network structures. So it is

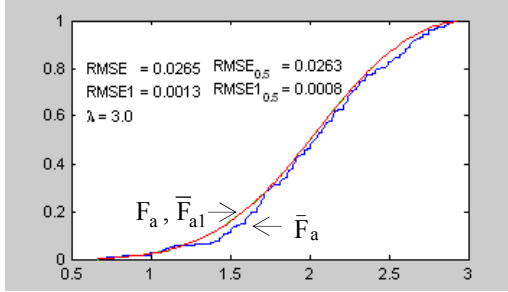


Figure 5a. Graphs $F_a, \bar{F}_a, \bar{F}_{a1}$ for $\lambda = 3$ - (F_a and \bar{F}_{a1} overlaps).

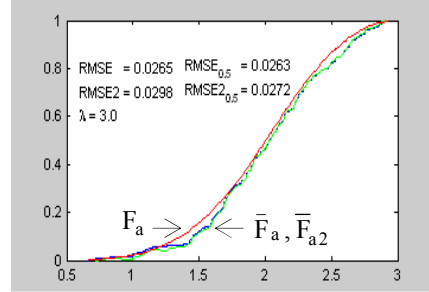


Figure 5b. Graphs $F_a, \bar{F}_a, \bar{F}_{a2}$ for $\lambda = 3$ - (\bar{F}_a and \bar{F}_{a2} overlaps).

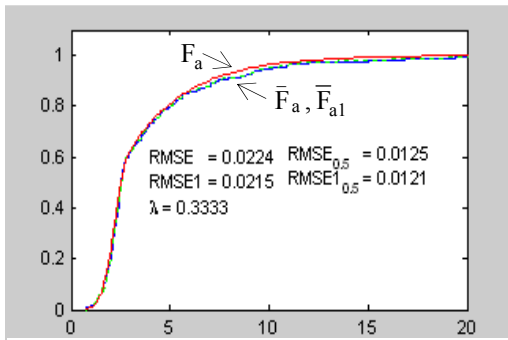


Figure 5c. Graphs $F_a, \bar{F}_a, \bar{F}_{a1}$ for $\lambda = 1/3$ - (\bar{F}_a and \bar{F}_{a1} overlaps).

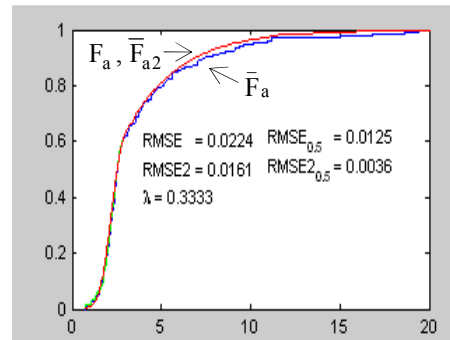


Figure 5d. Graphs $F_a, \bar{F}_a, \bar{F}_{a2}$ for $\lambda = 3$ - (F_a i , \bar{F}_{a2} overlaps).

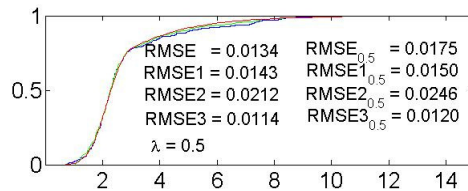


Figure 5e. . Comparison of $\bar{F}_a(t), \bar{F}_{a3}(t)$ and $F_a(t)$, $\lambda=0.5$

therefore unlikely that a polynomial time solution can be found with the more general functions. This remark refers to proper choice of the CV subnet rather than to the considered method.

If the c.d.f. of PCT is more complicated then applications of the CV are more interesting. For example, if we assume that the random variable T_1, T_2, T_4 and T_5 in subnet 1b are distributed according to f_3 presented in Fig. 6, and T_3 is

distributed uniformly between 2 and 3 then the c.d.f of PCT for subnet 1a, Fig.7, is far a way from Normal distribution.

From practical point of view such distribution should be examined also but in this paper there are no room for study such cases. More computational results concerning such cases will be presented during the conference.

4. CONCLUSIONS

Since analytical solutions of PERT problems can be more available, one can suggest the control variate technique to improve Monte Carlo simulation for estimating the probability distribution of the project duration, especially if high accuracy is required or distribution is not unimodal. This suggestion does not exclude other techniques which can be apply in combinations.

REFERENCES

- [1] Burt, J.M., Jr., Garman M.B. (1971). "Conditional Monte Carlo: A Simulation Technique For Stochastic Network Analysis", Management Science, Vol. 18, 207-217.
- [2] Elmaghraby S.E. (1977). Activity Networks: Project Planning And Control By Network Models, J.Wiley & Sons, New York.
- [3] Hagstrom J.N. (1988). Computational Of A Pert Problems. Networks, 18, 139-147.
- [4] Milian Z., (2006). Metody Określania Rozkładu Czasu Realizacji Przedsięwzięć Budowlanych W Acyklicznych Sieciach Stochastycznych, Mon. 327, Politechnika Krakowska (In Polish).
- [5] Milian Z., A Recursive Method For Computing The Exact Probability Distribution Of The Project Completion Time In A Stochastic Pert Network (To Be Published).
- [6] Sigal, C.E., Pritsker A.A.B., Solberg J.J. (1980). "The Stochastic Shortest Route Problem", Operations Research, Vol. 28, No. 5, Sep-Oct 1980, 1122-1129.
- [7] Van Slyke, R.M., (1963). "Monte Carlo Methods And The Pert Problem", Operations Research, Vol. 11, 1963, 839-86.

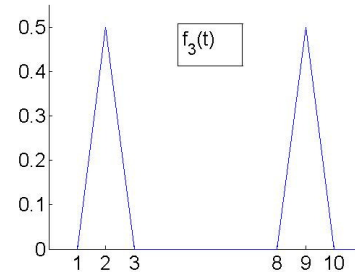


Figure 6. P.d.f. of a task duration

$$f_3(t) = \begin{cases} 1/2(t-1) & t \in [1,2] \\ 1/2(3-t) & t \in [2,3] \\ 1/2(t-8) & t \in [8,9] \\ 1/2(10-t) & t \in [9,10] \\ 0 & \text{otherwise} \end{cases} \quad (6)$$

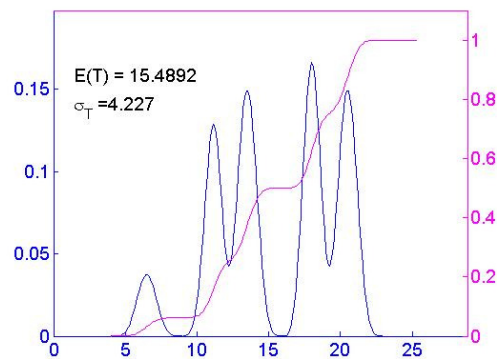


Figure 7. P.d.f. and c.d.f. of PCT