Journal of the Korea Industrial Information Systems Research Vol. 22 No.1, Feb. 2017 :61-67
http://dx.doi.org./10.9723/jksiis.2017.22.1.061 ISSN:1229-3741

프로젝트 네트워크에서 사업완성시간의 적률 추정

(Estimating the Moments of the Project Completion Time in Project Networks)

조 재 균[1)*]
(Jae-Gyeun Cho)

요 약 프로젝트 네트워크 분석에서 사업완성시간의 분포를 추정하는 것은 매우 기본적이다. 본 논문에서는 활동시간이 상호 독립적이고 정규분포를 따른다는 가정 하에서 사업완성시간의 적률(평균, 분산, 왜도, 첨도)을 추정하기 위한 방법을 제안한다. 제안된 방법은 연속형의 활동시간 분포를 이산형 분포로 근사화하기 위한 이산화 기법과 난수발생을 이용한다. 제안된 방법은 대규모 네트워크에 대해서도 쉽게 적용 가능하며, 그리고 제안된 방법에 의한 결과는 몬테칼로 시뮬레이션에 의해 얻어진 결과와 비교할 때 매우 정확함을 보여준다.

핵심주제어 : 프로젝트 네트워크, 사업완성시간, 이산화 기법

Abstract For a project network analysis, a fundamental problem is to estimate the distribution function of the project completion time. In this paper, we propose a method for evaluating moments(mean, variance, skewness, kurtosis) of the project completion time under the assumption that the durations of activities are independently and normally distributed. The proposed method utilizes the technique of discretization to replace the continuous probability density function(pdf) of activity duration with its discrete pdf and a random number generation. The proposed method is easy to use for large-sized project networks, and the computational results of the proposed method indicate that the accuracy is comparable to that of direct Monte Carlo simulation.

Key Words : Project Network, Project Completion Time, Technique of Discretization

1. 서 론

사업의 효율적 통제 및 관리를 위해 사용되는 기법인 PERT(Project Evaluation and Review Technique)에서 도식적인 모형으로서 프로젝트 네트워크가 사용된다.

프로젝트 네트워크는 노드들(nodes)과 활동들(activities)로 구성되고, 1개의 시작 노드와 1개의 종료 노드를 가지며, 활동들 간의 선후행 관계가 존재하고(directed), 연결성을 가지며(connected), 순환이 허용되지 않는(acyclic) 그래프로 정의될 수 있다. 각 활동시간은 그 활동의 불확정성(uncertainty)으로 인해 독립적이고, 양의 값을 갖는 확률변수로서 나타내어지며, 이로 인해 사업완성시간(project completion time)도 확률변수가 된다. 프로젝트 네트워크 분석에서 주요 연구 분야는 사업완성시간의 분포(또는 적

* Corresponding Author : jgcho@deu.ac.kr
† 이 논문은 2010학년도 동의대학교 연구년 지원에 의하여 연구되었음.
Manuscript received Jan, 25, 2017 / revised Feb, 10, 2017 / accepted Feb, 13, 2017
1) 동의대학교 e비즈니스학과

률)을 추정하는 방법과 프로젝트의 효율적 관리를 위해 더 많은 관심을 필요로 하는 활동을 밝히기 위해 각 활동의 중요도를 평가하는 방법이다. 사업완성시간의 분포를 추정하는 것은 사업목표일까지 사업을 완성할 확률에 대한 정보 제공, 사업완성시간과 비용을 줄이기 위한 일정계획 수립 등의 의사결정을 위해 매우 기본적이고도 필수적이다. 사업완성시간의 분포를 추정하기 위해 제안된 다양한 방법들은 (a) 해석적 방법, (b) 몬테칼로 시뮬레이션(Monte Carlo simulation), (c) 근사적 방법으로 구분된다[1-7]. 해석적 방법은 정확한 결과를 얻을 수 있지만, 활동시간에 대해 매우 제한적인 가정들을 함으로 인해 일반적으로 적용하기가 어렵고, 대규모 네트워크에 대해서는 지나치게 많은 계산량으로 인해 사용하기가 거의 불가능한 문제점이 있다. 몬테칼로 시뮬레이션은 대규모 네트워크 대해서도 쉽게 적용 가능하기 때문에 가장 빈번히 사용되며, 다른 근사적 방법들에 의해 얻어진 결과들의 정확도를 입증하기 위한 비교 기준으로 종종 사용된다. 반면에, 추정치의 정확도를 높이기 위해서는 많은 계산량이 요구되는 단점이 있다. 그리고 해석적 방법과 몬테칼로 시뮬레이션이 가지는 한계점들을 보완하기 위해 많은 근사적 방법들이 제안되었다.

몬테칼로 시뮬레이션이 가장 빈번히 사용되는 방법이긴 하지만 추정치의 불안정성과 같은 근본적인 문제점을 가진다. 즉, 몬테칼로 시뮬레이션을 사용하여 사업완성시간의 적률들을 추정하고자 할 때, 활동시간의 분포가 심하게 비대칭이거나 꼬리부분이 두텁고 긴 분포(예: error factor가 큰 대수정규분포)를 가질 경우, 샘플링되는 이상점(outlier)에 의하여 큰 영향을 받기 때문에 사업완성시간의 적률 추정치들이 매 시뮬레이션마다 크게 달라짐으로 인해 안정적인 적률 추정치들을 얻는 것이 매우 어렵게 되는 문제점이 있다[8-10].

근사화 방법 중에서 널리 사용되는 방법은, 활동시간의 주어진 연속형 분포를 이산형 분포로 근사화하기 위해 이산점들(discrete points)과 그에 대응되는 확률들을 결정하는 이산화 기법(technique of discretization)이다. 제안된 이산화 기법들 중에서, 주어진 연속형 분포의 적률들과 일치하도록 이산점들과 대응되는 확률들을 결정하는 적률 매칭(moment matching) 방법이 정확도에 있어 우수한 장점을 가진다[2].

본 논문에서는 활동시간이 상호 독립적이고 정규분포를 따른다는 가정 하에서, 사업완성시간의 적률의 안정적인 추정치를 얻기 위하여 Seo and Kwak[11]에서 제안된 적률 매칭 방법에 의한 이산화 기법과 난수 발생을 이용해 사업완성시간의 적률(평균, 분산, 왜도, 첨도)를 추정하기 위한 방법을 제안한다. 2절에서는 3절에서 제안되는 평가 절차의 이론적 근거로써, 연속형의 활동시간 분포를 이산형 분포로 근사화하기 위해 필요한 이산점들과 대응되는 가중치를 결정하기 위한 이산화 기법을 설명한다. 3절에서는 사업완성시간의 적률을 추정하기 위한 절차를 제안한다. 4절에서는 제안된 절차를 예증하고, 그 결과를 몬테칼로 시뮬레이션에 의해 얻어진 결과와 비교한다. 마지막으로 5절에서는 본 연구에 대한 결론을 맺는다.

2. 연속형 분포의 이산형 분포 근사화 방법

시스템 구성요소들의 연속형 확률분포가 주어졌을 때 시스템 특성치의 확률분포를 얼마나 정확하게 추정할 수 있는가 하는 것은, 시스템 구성요소의 연속형 확률분포가 주어졌을 때 이 분포를 이산형 분포로 근사화하기 위한 이산점들과 대응되는 가중치를 어떻게 결정하는가에 달려있다.

허용차 설계(tolerance design)에 관한 연구에서는 시스템 특성치의 확률분포를 결정하기 위한 방법으로 실험계획법(experimental design)이 빈번히 사용된다. 시스템 특성치의 적률을 추정하기 위하여 Taguchi 방법[12]이 널리 사용되고 있으며, D'Errico and Zaino[13]는 Taguchi 방법을 약간 수정함으로써 고차 적률(high-order moment)을 더 정확히 추정할 수 있는 방법을 제안하고 있다. 그러나 Taguchi 방법[12]과 D'Errico and Zaino[13]에 의한 방법은 시스템 구성요소가 대칭 분포인 정규분포를 따를 경우에

시스템 구성요소의 3개의 수준값(즉, 이산점)과 가중치를 결정하는 방법을 제안하고 있다. Seo and Kwak[11]은 D'Errico and Zaino[13]의 방법을 확장하여 시스템 구성요소가 대칭 분포 뿐만 아니라 비대칭 분포를 따르는 경우에도 적용 가능한 3개의 수준값과 가중치의 결정 방법을 제안하고 있다. 여기에서 언급되는 시스템 구성요소와 시스템 특성치는 본 논문에서의 활동시간과 사업완성시간에 각각 대응된다.

본 절에서는 3절에서 제안되는 평가 절차의 이론적 근거로써, Seo and Kwak[11]에 의해 제안된 이산화 기법 즉, 연속형 분포를 이산형 분포로 근사화하기 위해 필요한 3개의 수준값과 대응되는 가중치의 결정 방법을 소개한다.

시스템이 한 개의 구성요소로 이루어져 있는 경우를 생각해보자. 시스템 특성치 함수를 $g(x)$, 시스템 구성요소 X의 확률밀도함수를 $f(x)$, X의 평균, 분산, 왜도, 첨도를 각각 μ, σ^2, $\sqrt{\beta_1}$, β_2라 할 때, 시스템 특성치의 원점에 관한 k차 적률은 절점(node)의 수가 m인 구적법식 (quadrature formula)을 이용하여 다음과 같이 근사화할 수 있다.

$$E\{g^k\}=\int_{-\infty}^{\infty}[g(x)]^k f(x)\,dx$$
$$\cong w_1[g(\mu+\alpha_1\sigma)]^k+w_2[g(\mu+\alpha_2\sigma)]^k+\cdots+w_m[g(\mu+\alpha_m\sigma)]^k \quad (1)$$

1~4차까지의 적률을 정확히 추정하기 위해서는, 식(1)에서 최소한 $m=3$이어야 한다. 이때, 시스템 구성요소 X에 대해서는 다음의 구적법식을 만족하도록 $\{w_1,w_2,w_3,\alpha_1,\alpha_2,\alpha_3\}$를 결정하면 된다.

$$M_k=\int_{-\infty}^{\infty}(x-\mu)^k f(x)\,dx$$
$$=w_1(\alpha_1\sigma)^k+w_2(\alpha_2\sigma)^k+w_3(\alpha_3\sigma)^k \quad (k=0\sim 5) \quad (2)$$

$$\{l_1,l_2,l_3\}=\left\{\frac{B-\sqrt{C}\sigma}{A},\ \mu+\Delta,\ \frac{B+\sqrt{C}\sigma}{A}\right\} \quad (4)$$

$$\{w_1,w_2,w_3\}=\begin{bmatrix}\dfrac{F-E\sqrt{C}}{C\sigma-D\sqrt{C}}\\[2ex] \dfrac{(\beta_2-\beta_1-1)\sigma^4}{\Delta^4-2\sqrt{\beta_1}\sigma\Delta^3+(\beta_2-3)\sigma^2\Delta^2+2\sqrt{\beta_1}\sigma^3\Delta+(\beta_2-\beta_1)\sigma^4}\\[2ex] \dfrac{F+E\sqrt{C}}{C\sigma+D\sqrt{C}}\end{bmatrix}^{\mathrm{T}} \quad (5)$$

단,

$$A=2\left(-\Delta^2+\sqrt{\beta_1}\sigma\Delta+\sigma^2\right),$$
$$B=\left(-\sqrt{\beta_1}\sigma-2\mu\right)\Delta^2+\left((\beta_2-1)\sigma^2+2\mu\sqrt{\beta_1}\sigma\right)\Delta+\sqrt{\beta_1}\sigma^3+2\mu\sigma^2,$$
$$C=(\beta_1+4)\Delta^4-2\sqrt{\beta_1}(\beta_2+3)\sigma\Delta^3+\left(\beta_2^2-6\beta_2+6\beta_1-3\right)\sigma^2\Delta^2+2\sqrt{\beta_1}(3\beta_2-2\beta_1+1)\sigma^3\Delta+(4\beta_2-3\beta_1)\sigma^4,$$
$$D=2\Delta^3-3\sqrt{\beta_1}\sigma\Delta^2+(\beta_2-3)\sigma^2\Delta+\sqrt{\beta_1}\sigma^3,$$
$$E=\Delta^3-\sqrt{\beta_1}\sigma\Delta^2-\sigma^2\Delta,$$
$$F=\sqrt{\beta_1}\Delta^5-(\beta_2+\beta_1-3)\sigma\Delta^4+(\beta_2-7)\sqrt{\beta_1}\sigma^2\Delta^3+(\beta_2+3\beta_1-5)\sigma^3\Delta^2+5\sqrt{\beta_1}\sigma^4\Delta+2\sigma^5$$

단, M_k는 X의 평균에 관한 k차 적률이다. 식(2)에서 $\mu+\alpha_i\sigma$를 l_i로 대체하면 식(2)는 식(3)와 같이 나타낼 수 있다.

$$
\begin{aligned}
&w_1+w_2+w_3=1\\
&w_1l_1+w_2l_2+w_3l_3=\mu\\
&(l_1-\mu)^2w_1+(l_2-\mu)^2w_2+(l_3-\mu)^2w_3=\sigma^2\\
&\frac{(l_1-\mu)^3w_1+(l_2-\mu)^3w_2+(l_3-\mu)^3w_3}{\sigma^3}=\sqrt{\beta_1}\\
&\frac{(l_1-\mu)^4w_1+(l_2-\mu)^4w_2+(l_3-\mu)^4w_3}{\sigma^4}=\beta_2\\
&(l_1-\mu)^5w_1+(l_2-\mu)^5w_2+(l_3-\mu)^5w_3=M_5
\end{aligned}
\qquad (3)
$$

식(3)의 연립방정식을 풀어 $\{w_1,w_2,w_3,l_1,l_2,l_3\}$를 구하는 것은 매우 복잡하다. 이를 해결하기 위한 대안으로서, 식(3)에서 5차 적률 조건인 마지막 등식을 삭제하고, l_2를 $\mu+\Delta$라 둔 후, 연립방정식을 풀게 되면 수준값 $\{l_1,l_2,l_3\}$과 가중치 $\{w_1,w_2,w_3\}$는 식(4) 및 식(5)와 같이 결정된다.

3. 평가 절차

2절에서의 논의를 바탕으로, 사업완성시간의 적률을 평가하기 위한 절차는 다음과 같다. 활동시간 $t_i(i=1,\cdots,n;\ n$은 활동들의 수)의 1수준값, 2수준값, 3수준값을 각각 l_{i1}, l_{i2}, l_{i3}라 하고, 대응되는 가중치를 각각 w_{i1}, w_{i2}, w_{i3}이라 할 때,

단계 0. $\Delta=0$라 둔다.
단계 1. 각 활동에 대해 3개의 수준값을 식(4)와 같이 결정한다.
단계 2. $l_{i1}\ge 0$이면, 단계 3으로 간다. 그렇지 않으면 $\Delta=\Delta+d$. 그리고 단계 1로 간다.
단계 3. 각 활동에 대해 수준값에 대응되는 가중치를 식(5)와 같이 결정한다.
단계 4. 각 활동에 대해 $[0,1)$ 사이의 난수값 u를 생성한다.

$u\le w_{i1}$이면 $t_i=l_{i1}$,
$w_{i1}<u\le w_{i1}+w_{i2}$이면 $t_i=l_{i2}$
그렇지 않으면 $t_i=l_{i3}$

단계 5. CPM(Critical Path Method)을 이용하여 사업완성시간을 계산한다.
단계 6. 단계 4와 단계 5를 K번 반복한다.
단계 7. 사업완성시간의 평균, 분산, 왜도, 첨도를 계산한다.

4. 수치 예제 및 결과 분석

Fig. 1에 있는 예제 네트워크에 대해 제안된 방법을 예증하면 다음과 같다. 이때, 활동 1, 2, 4, 5는 평균이 4이고 분산이 1인 정규분포를, 그리고 활동 3은 평균이 6이고 분산이 1인 정규분포를 따른다고 가정한다.

단계 1. 활동 1, 2, 4, 5의 3개의 수준값은 (2.268, 4.0, 5.732), 그리고 활동 3의 3개의 수준값은 (4.268, 6.0, 7.732).
단계 3. 활동 1~5에 대해 3개의 수준값에 대응되는 가중치는 동일하게 (0.1667, 0.6667, 0.1667).
단계 4. 활동1에 대해, $0\le u\le 0.1667$이면 1 수준값인 2.268, $0.1667<u\le 0.8334$이면 2 수준값인 4.0, $0.8334<u<1$이면 3 수준값인 5.732를 선택한다. 나머지 활동들에 대해서도 유사한 방법으로 수준값을 선택한다.
단계 5. 활동 1~5에 대해 선택된 수준값들을 가지고 CPM을 이용하여 사업완성시간을 계산한다.
단계 6. 단계 4와 단계 5를 100,000번 반복한다.
단계 7. 단계 6에서 구해진 100,000개의 사업완성시간을 가지고 계산된 사업완성시간의 평균, 분산, 왜도, 첨도는 Table 1의 Fig. 1 열에 보여진다.

Table 1 Moments of project completion time.

	Fig. 1		Fig. 2		Fig. 3	
	Proposed Method	MCS	Proposed Method	MCS	Proposed Method	MCS
Mean	14.001	14.001	21.021	21.023	57.109	57.268
Variance	3.000	2.997	10.808	10.771	36.883	34.491
Skewness	-0.001	0.005	0.029	0.039	0.982	1.042
Kurtosis	3.002	2.994	2.980	2.958	3.856	4.213

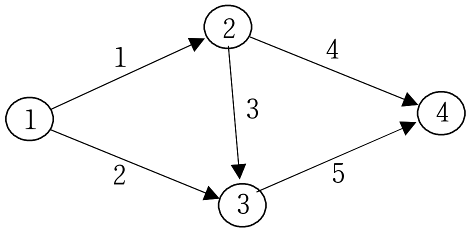

Fig. 1 Example network

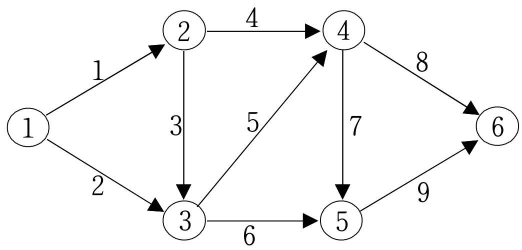

Fig. 2 Example network

제안된 방법에 의한 결과의 정확도를 입증하기 위하여, 샘플크기 100,000회의 몬테칼로 시뮬레이션을 이용하여 얻어진 사업완성시간의 평균, 분산, 왜도, 첨도는 Table 1의 Fig. 1 열에 나타낸다. 제안된 방법에 의한 결과를 몬테칼로 시뮬레이션에 의해 구해진 결과와 비교해볼 때, 제안된 방법에 의해 매우 정확한 결과를 얻을 수 있음을 알 수 있다.

제안된 방법에 의한 결과의 정확도를 추가적으로 확인하기 위하여 Fig. 2와 Fig. 3에 있는 예제 네트워크들을 채택한다.

Fig. 2의 네트워크는 Fig. 1에 있는 네트워크를 각 경로를 구성하는 활동 수 및 경로들 간의 상호의존성(dependency) 정도 관점에서 확장한 것이며, Fig. 3의 네트워크는 경로들의 수가 많고, 경로들 간에 상호의존성이 매우 높은 네트워크이다[14]. Fig. 2의 네트워크에 대해서는 활동 1, 2, 4, 5, 6, 8, 9는 평균이 3이고 분산이 1인 정규분포를, 활동 3과 7은 평균이 6이고 분산이 2인 정규분포를 따른다고 가정한다. 그리고 Fig. 3의 네트워크에 대해서는 Table 2에 주어진 평균과 분산을 갖는 정규분포를 따른다고 가정한다. 제안된 방법과 몬테칼로 시뮬레이션을 이용하여 구해

Table 2 Distributions of activities for example network in Fig. 3.

Activity	Distribution*	Activity	Distribution	Activity	Distribution	Activity	Distribution
1	(10.0, 0.167)	11	(2.0, 0.667)	21	(4.0, 0.667)	31	(1.5, 0.25)
2	(2.333, 0.389)	12	(2.0, 0.667)	22	(3.0, 0.667)	32	(10.0, 0.667)
3	(4.0, 2.028)	13	(5.333, 0.389)	23	(14.5, 0.25)	33	(4.0, 0.667)
4	(12.5, 0.25)	14	(5.333, 0.389)	24	(2.667, 0.389)	34	(1.5, 0.25)
5	(1.5, 0.25)	15	(11.0, 5.998)	25	(19.0, 24.0)	35	(7.0, 6.667)
6	(2.667, 0.389)	16	(4.0, 1.166)	26	(2.333, 0.389)	36	(13.0, 24.0)
7	(4.0, 0.167)	17	(2.333, 0.389)	27	(4.0, 0.667)	37	(13.5, 0.25)
8	(2.5, 1.25)	18	(3.0, 0.167)	28	(8.0, 8.168)	38	(10.0, 29.998)
9	(16.667, 0.389)	19	(2.0, 0.167)	29	(3.5, 1.25)		
10	(13.0, 43.997)	20	(5.5, 1.25)	30	(2.0, 0.667)		

* : (mean, variance)

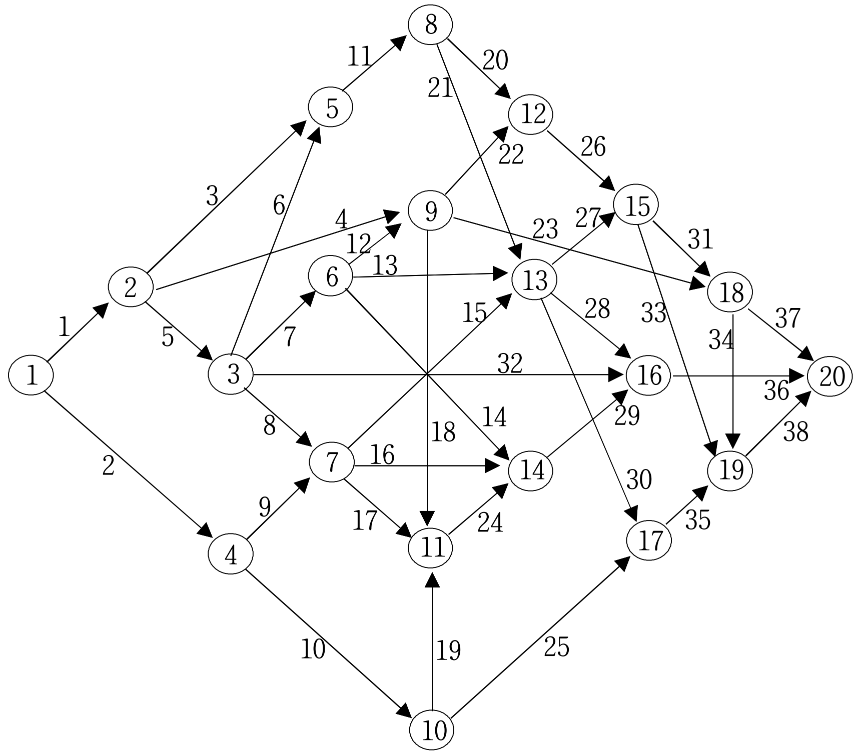

Fig. 3 Example network

진 사업완성시간의 평균, 분산, 왜도, 첨도는 Table 1의 Fig. 2 열과 Fig. 3열에 각각 보여 진다. 제안된 방법에 의한 결과는 몬테칼로 시뮬레이션에 의해 구해진 결과와 비교해볼 때 매우 정확함을 알 수 있다.

5. 결 론

본 논문에서는 활동시간이 상호 독립적이고 정규분포를 따른다는 가정 하에서 사업완성시간의 적률을 추정하기 위한 방법을 제안하였다.

제안된 방법은 사업완성시간의 분포를 추정하기 위한 방법들의 분류 중 근사적 방법에 속하며, 근사화를 위한 방법으로 이산화 기법과 난수발생을 이용한다. 그리고 제안된 방법에 의한 결과는 몬테칼로 시뮬레이션에 의해 얻어진 결과와 비교할 때 매우 정확함을 보여준다.

본 논문에서 제안된 방법이 활동시간의 분포가 정규분포가 아닌 어떠한 확률분포를 가질 경우에도 적용 가능한지의 여부를 확인하고, 다른 근사적 방법들에 의해 얻어진 결과와 정확도를 비교분석하는 것이 추후의 연구과제이다.

References

[1] Yao, M. J. and Chu, W. M., "A New Approximation Algorithm for Obtaining the Probability Distribution for Project Completion Time," Computers and Mathematics with Applications, Vol. 54, pp. 282-295, 2007.

[2] Agrawal, M. K. and Elmaghraby, S. E., "On Computing the Distribution Function of the Sum of Independent Random Variables," Computers & Operations Research, Vol. 28, pp. 473-483, 2001.

[3] Mehrotra, K., Chai, J. and Pillutla, S., "A Study of Approximating the Moments of the Job Completion Time in PERT Networks,"

Journal of Operations Management, Vol. 14, pp. 277-289, 1996.
[4] Abdelkader, Y. H., "Evaluating Project Completion Times when Activity Times are Weibull Distributed," European Journal of Operational Research, Vol. 157, pp. 704-715, 2004.
[5] Azaron, A. and Fatemi Ghomi, S. M. T., "Lower bound for the mean Project Completion Time in Dynamic PERT Networks," European Journal of Operational Research, Vol. 186, pp. 120-127, 2008.
[6] Iida, T., "Computing bounds on Project Duration Distributions for Stochastic PERT Networks," Naval Research Logistics, Vol. 47, pp. 559-580, 2000.
[7] Milian, Z., "A Note on Computing the Exact Probability Distribution of the Project Completion Time in a Stochastic PERT Network," Proceedings of the 2013 IEEE IEEM, pp. 78-82, 2013.
[8] Iman, R. L. and Hora, S. C., "A Robust Measure of Uncertainty Importance for use in Fault Tree System Analysis," Risk Analysis, Vol. 10, pp. 401-406, 1990.
[9] Cho, J. G. and Jeong, S. C., "Evaluation of Uncertainty Importance Measure in Fault Tree Analysis," The Journal of Information Systems, Vol. 17, No. 3, pp. 25-37, 2008.
[10] Cho, J. G., "Evaluation of Uncertainty Importance Measure for Monotonic Function," Journal of the Korea Industrial Information Systems Research, Vol. 15, No. 5, pp. 179-185, 2010.
[11] Seo, H. S. and Kwak, B. M., "Efficient Statistical Tolerance Analysis for General Distributions using Three-point Information," International Journal of Production Research, Vol. 40, pp. 931-944, 2002.
[12] Taguchi, G., "Performance Analysis Design," International Journal of Production Research, Vol. 16, pp. 521-530, 1978.
[13] D'Errico, J. R. and Zaino Jr., N. A., "Statistical Tolerancing using a Modification of Taguchi's Method," Technometrics, Vol. 30, No. 4, pp. 397-405, 1988.
[14] Kleindorfer, G. B., "Bounding Distributions for a Stochastic Acyclic Network," Operations Research, Vol. 19, pp. 1586-1601, 1971.

조 재 균 (Jae-Gyeun Cho)

- 정회원
- 연세대학교 응용통계학과 학사
- 한국과학기술원 산업공학과 석사
- 한국과학기술원 산업공학과 박사
- 한국전자통신연구원 선임연구원
- 현재: 동의대학교 e비즈니스학과 교수
- 관심분야: 프로젝트 관리, e비즈니스 모델 및 전략