

GPH 분포에 의한 확률적 근사화[†]

윤복식* · 박광우** · 이창훈**

Probabilistic approximations based on GPH distributions[†]

B. S. Yoon*, K. W. Park** and C. H. Lie**

Abstract

The distribution of random sum of i.i.d. exponential random variables is called GPH(Generalized Phase-Type) distribution. The class of GPH distributions is large enough to include PH (Phase-Type) distributions and has several properties which can be applied conveniently for computational purposes. In this paper, we show that any distribution defined on R^+ can be approximated by the GPH distribution and demonstrate the accuracy of the approximation through various numerical examples. Also, we introduce an efficient way to compute the delay and waiting time distributions of the GPH/GPH/1 queueing system which can be used as an approximation model for the GI/G/1 system, and validate its accuracy through numerical examples. The theoretical and experimental results of this paper help us accept the usefulness of the approximations based on GPH distributions.

1. 서론

유한상태의 연속시간 마코프체인이 흡수상태(absorbing state)에 빠질때까지의 시간의 분포로 정의되는 PH(Phase-Type)분포는 Neuts

(1981)에 의해 제안된 이래 확률모형의 계산적인 접근방법에 많이 사용되어 왔다. 그러나 실용성에 있어서 몇가지 본래적인 문제점이 있다. 우선 보다 정확한 근사화를 위해서는 많은 경우에 PH분포를 표현하는 마코프체인의 상태공간이 커지게 되는데, 이에 따라 파라미터 추

[†] 본 연구는 부분적으로 홍익대학교의 지원에 의해 이루어 졌음.

* 홍익대학교 기초과학과

** 서울대학교 산업공학과

정의 문제와 계산상의 문제가 해결하기 어려운 문제로 대두된다. 또한 라플라스변환이 유리함수가 아닌 분포는 PH분포로 표현이 불가능하기 때문에 일반적인 근사화 방법으로서 제한이 있다[1,7]. 이러한 단점을 극복하는 한 방법으로 Shanthikumar(1985)는 PH(Phase-type) 분포를 일반화시킨 GPH(Generalized Phase-type)분포를 제안하였다. 동일한 지수 분포를 따르는 무작위 갯수의 확률 변수들의 합의 분포로 정의되는 GPH 분포는 PH 분포에 비해 일반적이며, 표현에 있어서 훨씬 간편하다는 잇점을 가지고 있다(2장 참조). 또한 GPH 분포는 연속함수 계산을 이산함수 계산으로 변환 시켜줌으로써, 확률분석에서 어려운 계산들을 보다 간편하게 해준다.

GPH 분포의 이러한 성질은 대기시스템을 비롯한 여러 분야의 확률적 분석에 폭넓게 응용될 잠재성을 가지고 있으나 아직까지 충분히 응용되지 못하고 있고 근사화의 효율성 및 정확성에 대한 실험적인 연구도 보고되지 않고 있다. Yoon(1988)은 임의의 분포 함수를 GPH 분포로 근사화 시킬 수 있음을 보인바 있는데 이 결과를 이용하면 보다 광범위한 모형의 확률적 분석을 GPH 분포로의 근사화를 통해 정확하고 간편하게 수행하는 과정을 확립할 수 있다. 본 논문은 우선 R^+ 에서 정의된 일반적인 분포를 GPH 분포로 근사화하는 과정을 확립하고 GPH 분포의 확률적 및 계산적 성질을 대기 시스템 분석에 효과적으로 이용하는 방법을 제시하는 것이 목적이다.

이를 위해 우선 2장에서는 GPH 분포에 대한 정의 및 확률적 계산에 유용하게 이용될 수 있는 GPH 분포의 성질을 요약하고, 3장에서는 R^+ 에서 정의된 일반적인 분포를 GPH 분포로 근사화하는 과정을 제시하고 다양한 실험을 통

해 검증한다. 4장에서는 GPH 분포를 기본으로 하는 대기시스템에서 대기시간 분포를 구하는 방법을 제시하고 정확성을 실험을 통해 확인한다. 끝으로 5장에서 결론 및 향후 연구 방향을 제시한다.

2. GPH 분포의 성질

2.1 GPH 분포의 정의 및 PH 분포와의 관계

$(E_n)_0'$ 을 평균이 $1/\lambda$ 인 지수 분포를 따르는 확률 변수열이라 하고(단, $E_0=0$ $w, p, 1$), L 을 분포함수 g 를 따르는 이산적인 확률 변수라고 할 때,

$$X = \sum_{n=0}^L E_n \quad (1)$$

의 분포 함수 F 는 다음과 같이 쓸 수 있다.

$$F_X(x) = \sum_{n=0}^L G(n) \frac{e^{-\lambda(\lambda x)^n}}{n!}, \quad x > 0 \quad (2)$$

이 분포를 GPH(λ, g)로 정의한다[6 참조]. 물론, 하나의 GPH 분포를 표시하는 λ 와 g 의 쌍이 무한히 많이 존재하게 된다.

Neuts의 PH 분포는, 흡수상태(absorbing state)를 하나 가지는 유한상태공간(finite state space)에서 정의된 연속시간 마코프 체인이 흡수상태에 빠질 때까지 경과한 시간의 분포로 정의된다[1,3,7]. 이 PH 분포는 초기상태(initial state)를 나타내는 벡터 \underline{a} 와 마코프 체인의 infinitesimal generator A 가 주어지면 확정되기 때문에, PH(\underline{a}, A)로 표현된다. 일반

적으로, 임의의 분포를 \underline{a} 와 A 를 이용하여 PH 분포로 나타낼 수 있다. 즉, PH 분포는 다음과 같은 분포 함수를 가진다.

$$F(t) = 1 - \underline{a} \exp(At)\underline{e}, \quad (3)$$

여기서, $\underline{e} = (1 \ 1 \cdots 1)'$

PH 분포는 uniformization[5]을 이용하여 GPH 분포의 형태로 변형이 가능하다. 먼저, $\lambda \geq \max_{i=1,2,\dots,n} |a_i|$ 이거나 λ 를 선택하면, uniformization을 이용하여

$$F(t) = \sum_{n=0}^{\infty} \left(\sum_{r=0}^n g(r) \right) \frac{e^{-\lambda t} (\lambda t)^n}{n!}, \quad t \geq 0 \quad (4)$$

여기서, $g(0) = a_i, g(n) = \underline{a} R^{n-1} r_n,$

$$r_n = \frac{-Ae}{\lambda},$$

$$R \text{의 component } r_{ij} = \begin{cases} a_{ij} & \text{if } i \neq j \\ 1 - \lambda_i/\lambda & \text{if } i = j \end{cases}$$

와 같이 F 를 표현할 수 있다. 결국, 모든 PH 분포는 uniformization에 의해 GPH 분포로 표현될 수 있으므로 GPH 분포가 좀더 일반적인 분포라 할 수 있다. 더우기, (2)과 (3)의 표현을 비교하면, (2)가 훨씬 간단하므로 GPH 분포가 PH 분포보다 편리하게 사용될 수 있다.

2.2 GPH 분포의 성질

GPH 분포는 3장에서 설명되듯이, 연속 변수의 연산을 이산 변수의 연산으로 변환시켜주기 때문에 근사적 접근법에 편리하게 응용될 수 있다. [6]에서 종합적으로 파악된 GPH 분포의 성질중 우리가 이용할 계산상의 편리성에 관한 성질을 요약해 보자.

1. GPH 분포는 convolution에 닫혀 있다.

— 이 사실은 쉽게 파악할 수 있는데 X_1 을

$GPH(\lambda, g_1), X_2$ 를 $GPH(\lambda, g_2)$ 분포를 따르는 독립인 두개의 확률 변수라고 하고,

$L_1 \sim g_1, L_2 \sim g_2$ 라 하면, $X_1 = \sum_{n=0}^{L_1} E_n$ 이고, $X_2 \sim$

$\sum_{n=0}^{L_2} E_n$ 이기 때문에, $X = X_1 + X_2$ 라 하면, $X \sim$

$\sum_{n=0}^{L_1+L_2} E_n$ 가 되어, X 는 $GPH(\lambda, g)$ 분포를 따르

게 된다.(단, $g(n) = (g_1 * g_2)(n), (g_1 * g_2)(n) =$

$$\sum_{k=0}^n g_1(k) g_2(n-k), \sim : \text{같은 분포를 따르}$$

는) 여기서, 보듯이 연속함수의 변수인 X_1

과 X_2 을 convolution할 때, 연속함수 적분

식이 이산함수의 합식 식으로 바뀌게 된다.

이 성질은 GPH 분포의 계산적인 응용을

가능하게 하는 중요한 성질이다. 물론, 이

와 같은 연산이 가능하기 위해서는, 두

GPH 분포의 λ 값이 같아야 한다. 2.1절에

서도 언급했듯이 하나의 GPH 분포는 λ 값

에 따라 많은 표현 쌍들이 존재하는데, 원

하는 정확도에 따라 λ 를 적절히 고정시켜

야 이러한 성질을 이용할 수 있다(3장 참

조).

2. X_1 과 X_2 는 각각 $GPH(\lambda, g_1)$ 와 $GPH(\lambda,$

$g_2)$ 분포를 따르는 확률 변수라고 하자.

이때, $X = X_1 - X_2$ 는 (λ, g) 의 BPH(Bi-

lateral Phase Type)([부록 A]) 확률 변

수가 된다. 단, $(g(n))'$ 는 다음과 같이

주어진다.

$$g(n) = \begin{cases} \sum_{k=0}^n g_1(k+n) g_2(k), & n=1, 2, \dots \\ g_1(0) g_2(0), & n=0 \\ \sum_{k=0}^n g_2(k-n) g_1(k), & n=-1, -2, \dots \end{cases} \quad (5)$$

, (g_1, g_2) 는 [부록 B] 참조)

3. X_1 과 X_2 는 각각 $GPH(\lambda, g_1)$, $GPH(\lambda, g_2)$ 분포를 따르는 확률 변수라고 하자. 그러면, $X = \max(X_1 - X_2, 0)$ 도 $GPH(\lambda, g)$ 분포를 따르는 확률 변수가 된다. 여기서,

$$g(n) = \begin{cases} \sum_{k=0}^{\lambda} g_1(k) \sum_{r=k}^{\lambda} g_2(r) & n=0 \\ \sum_{k=0}^{\lambda} g_1(k+n) g_2(k) & n=1,2,\dots \end{cases} \quad (6)$$

(g' 는 [부록 B] 참조)

3. GPH 분포로의 근사화

3.1 근사화 방법

다음의 사실로부터 $[0, \infty]$ 에서 정의된 임의의 분포는 GPH 분포로 근사화 할 수 있음을 알 수 있다.

[정리 1] F 를 $[0, \infty]$ 에서 정의된 유계함수 (bounded function)이라 할 때,

$$F_\lambda(t) = \sum_{n=0}^{\lambda} F\left(\frac{n}{\lambda}\right) \frac{e^{-\lambda t} (\lambda t)^n}{n!}, \quad t \geq 0, \lambda > 0 \quad (7)$$

라고 정의하면, $F_\lambda(t) \rightarrow F(t)$ as $\lambda \rightarrow \infty$.

[증명] 식 (8)은 $F_\lambda(t) = E[F(\frac{N}{\lambda})]$ (단, N 은 평균이 λt 인 포아송 분포를 따른다.)으로 다시 표현될 수 있다. 그런데, $\frac{N}{\lambda} \rightarrow t$ w. p. 1을 쉽게 보일 수 있으므로 결과를 얻는다. (QED)

이와 함께, 경험적인 데이터에 의한 근사화

도 가능하다. 즉, 샘플 규모가 m 인 데이터 U_i ($i=1,2,\dots,m$)를, 알려지지 않은 분포 $F(x)$, $x \geq 0$ 으로 부터 얻었다고 하자. 이 때, F 의 경험적인 분포

$$F_m(x) = \frac{\sum_{i=1}^m I(U_i \leq x)}{m} \quad (8)$$

여기서, $I(A) = \begin{cases} 1 & \text{if } A \text{ is true} \\ 0 & \text{else} \end{cases}$

를 얻을 수 있고 강대수의 법칙(strong law of large number)을 이용하면 $F_m(x) \rightarrow F(x)$ w.p. 1 as $m \rightarrow \infty$ 을 알 수 있다. 이제 $F_m(x)$ 의 GPH 형태를 얻기위해 (7)을 이용하여 $F_m(x)$ 를

$$F_m(x) = \sum_{n=0}^{\lambda} F_m\left(\frac{n}{\lambda}\right) \frac{e^{-\lambda x} (\lambda x)^n}{n!} \quad (9)$$

로 놓으면 [정리 1]에 의해 모든 고정된 m 에 대해

$$F_m(x) \rightarrow F_m(x) \text{ as } \lambda \rightarrow \infty \quad (10)$$

의 결과를 얻는다.

이상에서 두 가지의 근사화가 가능함을 알 수 있다. 즉, 상황에 따라 모형의 구체적인 분포 함수가 주어지면, (7)의 F 가 주어지는 경우이므로 [정리 1]의 근사화 방법을 이용한다. 만약, F 가 구체적으로 주어지지 않는 경우라도 데이터가 주어지기만 하면 (8), (9)를 이용하여 GPH로 근사화 시킬 수 있다. 이때 (7)과 (9)는 무한급수의 형태인데, 실제 계산에서는 오차의 한계를 설정하여, 이것에 맞게 계산을 수행한다. 오차의 한계를 ϵ 이라 하면, $\sum_{n=0}^K \frac{e^{-\lambda t} (\lambda t)^n}{n!} > 1 - \epsilon$ 인 최소의 K 를 계산하여 항의 갯수를 구할 수 있는데, 그러면 $\sum_{n=K+1}^{\infty} \frac{e^{-\lambda t} (\lambda t)^n}{n!} < \epsilon$ 이 되고,

$$\sum_{n=K+1}^{\infty} F\left(\frac{\lambda}{n}\right) \frac{e^{-\lambda}(\lambda t)^n}{n!} < \sum_{n=K+1}^{\infty} \frac{e^{-\lambda}(\lambda t)^n}{n!}$$

이므로,

$$\sum_{n=K+1}^{\infty} F\left(\frac{\lambda}{n}\right) \frac{e^{-\lambda}(\lambda t)^n}{n!} < \varepsilon \text{ 이 되기 때문이다.}$$

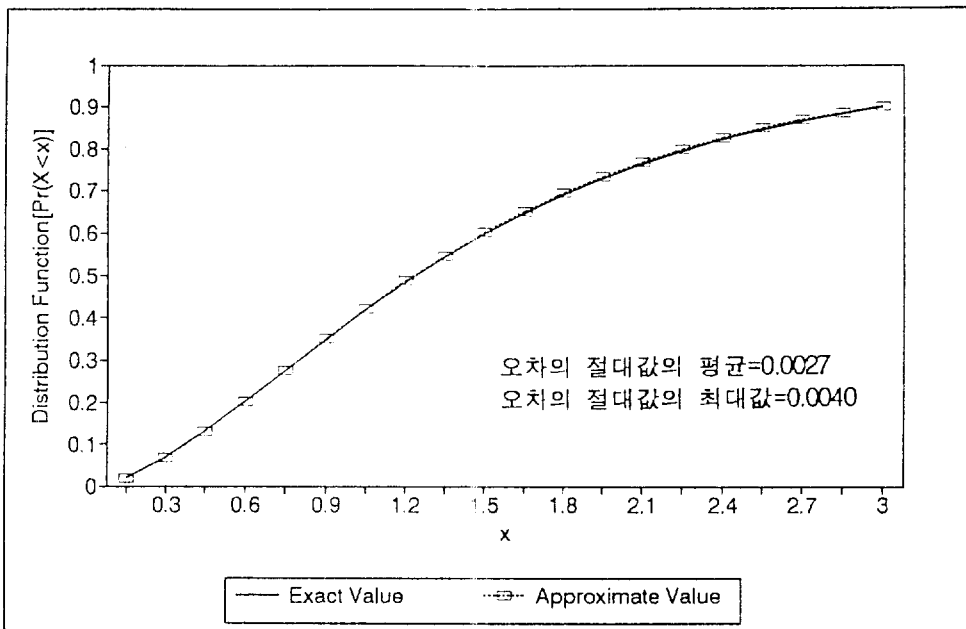
3.2 근사화 예제

3.2.1 알려진 분포로부터의 근사화

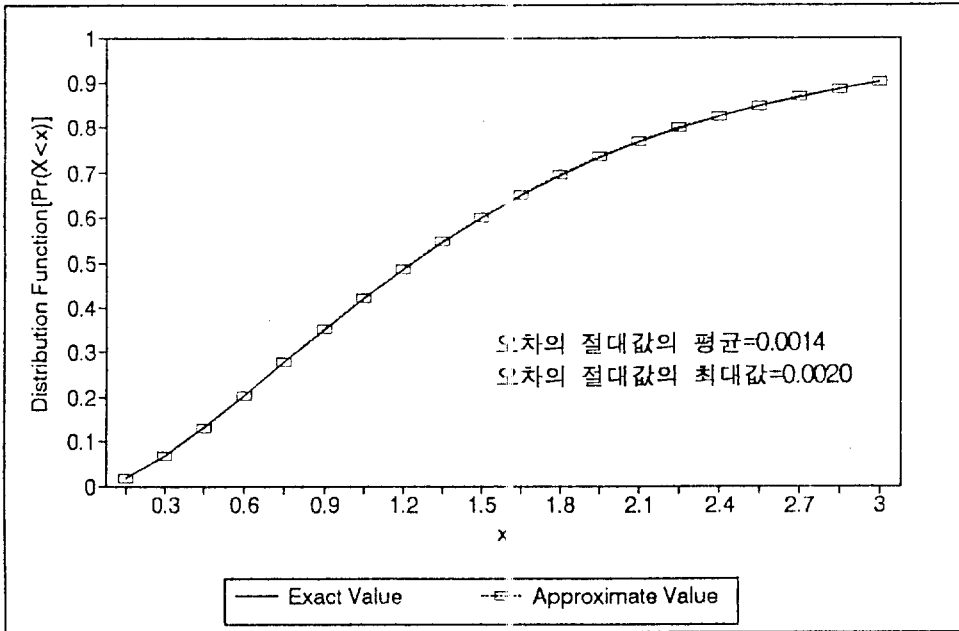
GPH로의 근사화의 정확도를 실험하기 위한 함수로서 Generalized Erlang-2(GE-2) ($\lambda_1=1, \lambda_2=2$) 함수와 Hyper-Exponential(HE) ($\lambda_1=1, \lambda_2=2, p_1=0.6$) 함수를 선택한다. 식(7)에서 λ 값이 근사화의 정확성에 영향을 미치는 정도를 알아보기 위해 λ 값을 변화시킨다음 근사화의 추이를 살펴본다. [그림 1] [그림 2]는

GE-2($\lambda_1=1, \lambda_2=2$) 함수를 근사화시키면서 각각 $\lambda=50, 100$ 으로 변화시켜, 근사화의 정확성의 정도를 비교해본 것이다. 결과를 보면, [정리 1]의 결과처럼 λ 값이 클수록 정확한 근사화가 이루어지는 것을 확인할 수 있다.

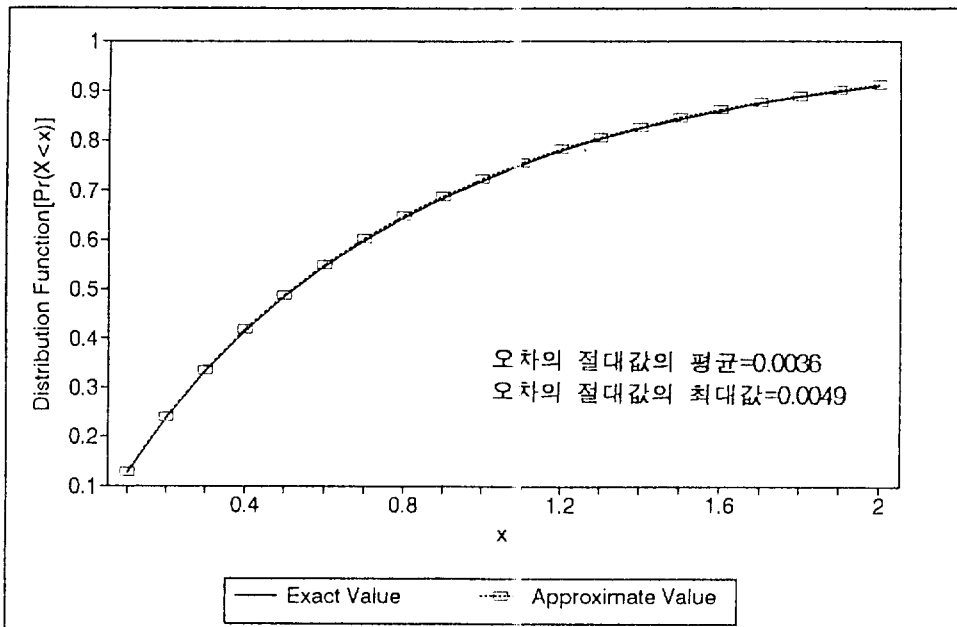
[그림 3]과 [그림 4]는 HE($\lambda_1=1, \lambda_2=2, p=0.6$) 분포함수를 GPH로 근사시키면서, 역시 λ 의 변화에 따른 근사화의 정확성을 비교한 것이다. 앞의 예제와 같이 λ 값의 크기가 근사화에 영향을 미치고 있음을 알 수 있다. 이 결과들을 종합해보면, λ 값을 충분히 크게 주변 거의 오차없이 임의의 분포를 GPH 분포로 근사화 될 수 있음을 확인할 수 있고 비교적 적은 값(50)에서도 매우 정확한 근사화가 이루어짐을 알 수 있다.



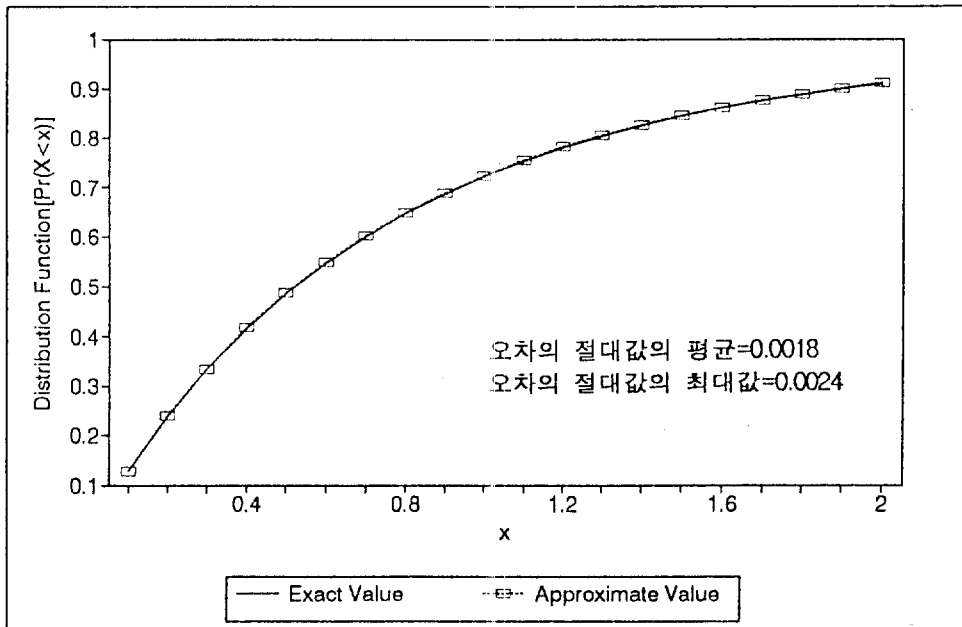
[그림 1] GE-2 분포의 GPH로의 근사화
: GE-2($\lambda_1=1, \lambda_2=2$) GPH($\lambda=50$)



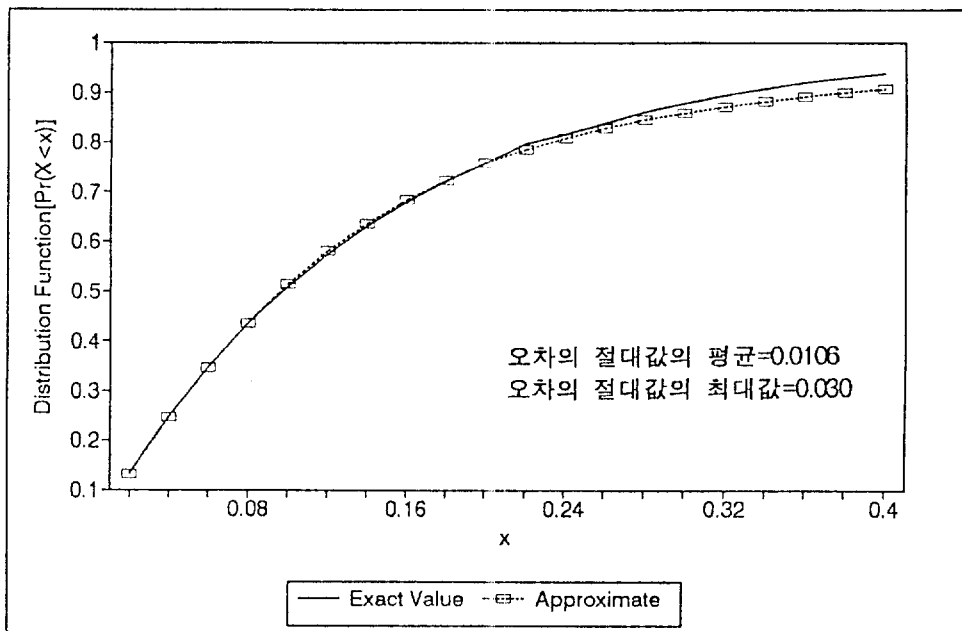
[그림 2] GE-2 분포의 GPH로의 근사화
: GE-2($\lambda_1=1, \lambda_2=2$) GPH($\lambda=100$)



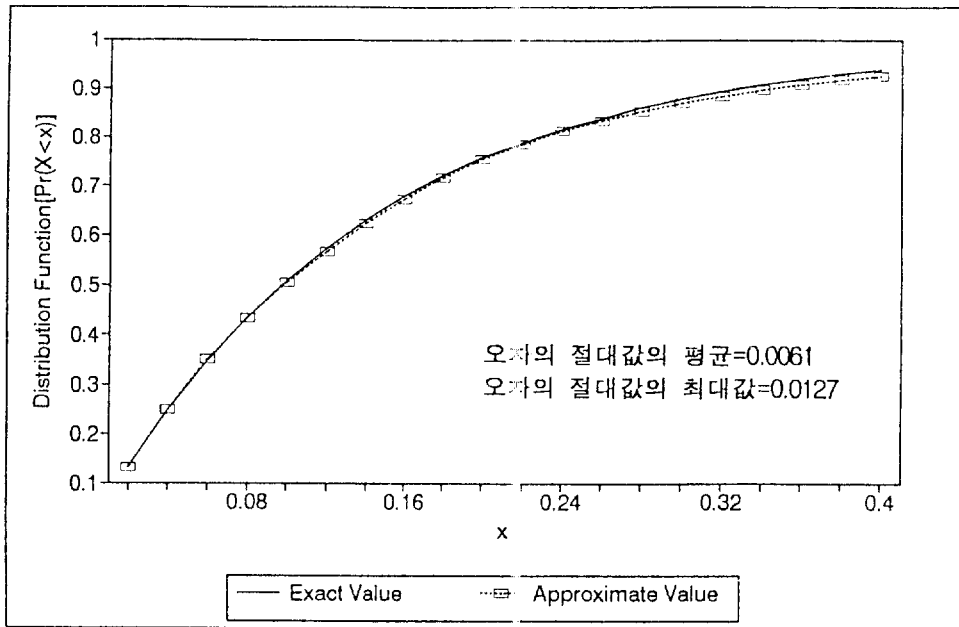
[그림 3] HE 분포의 GPH로의 근사화
: HE($\lambda_1=1, \lambda_2=2, p=0.6$) GPH($\lambda=50$)



[그림 4] HE 분포의 GPH로의 근사화
: HE($\lambda_1=1, \lambda_2=2, p=0.6$) GPH($\lambda=100$)



[그림 5] HE 분포로부터 데이터 추출에 의한 GPH로의 근사화
: 데이터 갯수 = 100, HE($\lambda_1=6, \lambda_2=8, p_1=0.4$) GPH($\lambda=100$)



[그림 6] HE 분포로부터 데이터 추출에 의한 GPH로의 근사화
 : 데이터 갯수=1000, HE($\lambda_1=6, \lambda_2=8, p=0.4$) GPH($\lambda=100$)

3.2.2 데이터 추출에 의한 근사화

근사화 실험 예제 함수로서 HE($\lambda_1=1, \lambda_2=2, p_1=0.6$) 함수를 선택한다. 주어진 데이터의 갯수에 따른 근사화의 정확도를 실험하기 위해 HE 분포를 따르는 데이터를 시뮬레이션을 통해 각각 100개, 1000개씩 추출한다. 그 다음, 이 데이터를 가지고 (9)와 (10)의 방법으로 GPH 분포로 근사화 시켰을 때의 함수와 원래의 함수를 비교하면, [그림 5], [그림 6]과 같은 결과가 나온다. 아래의 그림에서 보듯이 데이터의 갯수가 커질수록 좀더 정확한 근사화가 이루어진다. 결국, 데이터의 갯수만 충분히 크면, 미지의 분포도 GPH 분포로 근사화 시킬 수 있음을 알 수 있다.

4. 대기시스템에 있어서 GPH 분포의 응용

4.1 M / GPH / 1 Queue

도착 분포가 포아송(λ)이고, 서비스 분포가 F 인 단일봉사자 대기모형에서 FCFS 규칙하에서 안정상태의 시스템에서의 대기시간 분포함수 F_w 는 Pollaczek - Khinchin식에 의해 다음과 같이 계산된다.

$$F_w(x) = \sum_{k=0}^{\infty} (1-\rho)^k \rho^k F_k^{*k}(x) \quad (11)$$

(단, $F_k(x) = \frac{1}{E(x)} \int_0^x (1-F(\tau))^k d\tau, x \geq 0$ 이고 F_k^*

는 F_R 의 k -fold convolution)

단, $\rho \triangleq \lambda \{ \sum_{n=0}^{\infty} \bar{G}(n) \} / \mu$ 는 1보다 작아야 한다(여기서 $\bar{G}(x) = 1 - G(x)$). (11)의 식은 다중 convolution 때문에 일부 특수한 경우를 제외하고는 사실상 계산이 불가능하다. 그러나 F 가 GPH라면 convolution을 이산화할 수 있는데 이 과정을 통해 Shanthikumar(1985)는 다음과 같은 관찰을 한 바 있다.

[정리 2] F 가 $GPH(\mu, g)$ 이면 F_n 는 $GPH(\mu, g_n)$ 이다.

$$\text{단, } g_n(0) = 1 - \rho, \quad g_n(n+1) = \left(\frac{\lambda}{\mu}\right) \sum_{r=0}^n \bar{G}(n-r) \\ g_n(r), \quad n=0,1,\dots \quad (12)$$

[정리 2]을 보면 알 수 있듯이 M/GPH/1 Queue에서는 Queue 대기시간의 계산에 있어서, 적분식과 convolution의 복잡한 연산 대신에 간단한 이산적인 반복식 하나로 모든 계산이 완료된다. 따라서, 도착이 포아송 분포를 따를 때는, 서비스 분포를 GPH 로 근사화 시켜서, 대기시간을 쉽게 계산할 수 있다.

4.2 GPH / GPH / 1 Queue

A_n 을 n 과 $n+1$ 번째 고객의 도착 시간 간격, B_n 을 n 번째 고객의 서비스 시간, $n=0,1,2,\dots$ 을 나타내는 i.i.d. 확률변수 열이라고 하자. 그리고, W^n 과 W_n^* 을 각각 n 번째 고객의 시스템과 대기열에서의 대기시간이라고 하자. 이 시스템에서 안정 상태의 대기시간 분포는 Lindley의 적분방정식으로 표현될 수 있는데 [1,2], 이 식으로부터 실제 분포함수를 구하는 것은 매우 힘들다. 그러나, 위와 같은 GI/G/1 대기 시

스템을 근사화한 GPH/GPH/1 대기 시스템에서는 다음과 같은 사실이 알려져 있다 [6].

[정리 3] A_n 과 B_n 이 각각 $GPH(\lambda, g_n)$, $GPH(\lambda, g_n)$ 분포를 따른다고 할 때, W_n^* 가 $GPH(\lambda, g_n^*)$ 분포를 따르면, n 번째 고객의 시스템 대기시간 W^n 과 n 번째 고객의 대기행렬 대기시간 W_n^* 은 각각 (λ, g_n^*) (λ, g_n^*) 로 표현되는 GPH 분포를 따른다.

$$\text{여기서, } g_n^* = g_n^* * g, \quad n=0,1,\dots \quad (13)$$

$$g_n^* = g_n^* \oplus g_n^*, \quad n=1,2,\dots \quad (14)$$

(단, \oplus 는 nonnegative condition을 갖는 negative convolution, g^* 는 [부록 B]에 정의되어 있다.)

$$\text{[증명]} \quad W^n = W_n^* + B_n, \quad n=0,1,\dots \quad (15)$$

$$W_n^* = [W^{n-1} - A_n]^+, \quad n=1,2,\dots \quad (16)$$

의 관계이므로 2.2절의 GPH 성질 2, 3을 각각 적용하여 결과를 얻는다. (상세한 과정은 Shanthikumar(1985) 참조.)

[정리 3]의 결과를 이용하여 GPH/GPH/1 시스템에서의 대기시간의 안정상태 확률 분포를 계산하는 효과적인 방법을 찾기 위해 다음과 같은 과정을 따른다.

(1) (13), (14)에 의해 g_n^* 와 g_n 를 계산하고 이들을 각각 도착간격 시간의 확률분포와 서비스 시간의 확률분포로 갖는 GI/G/1 이산 대기 시스템을 고려한다.

(2) GI/G/1 이산 대기시스템의 정상상태에서의 대기열과 시스템에서의 대기시간 분포를 (15), (16)을 적용하여 이산함수 convolution 계산에 의해 계산한다. 먼저,

$$c(n) = \begin{cases} \sum_{k=0}^r g_s(k+n)g'_s(k) & n=1,2,\dots \\ \sum_{k=1}^r g_s(k)g'_s(k) & n=0 \\ \sum_{k=0}^r g_s(k)g'_s(k-n) & n=-1,-2,\dots \end{cases} \quad x > 0 \quad (21)$$

을 계산하고, $(c(n))$ 은 임의의 고객의 대기행렬에서의 대기시간이 주어져 있을 때, 그 대기시간에 $(c(n))$ 의 확률 분포를 가지는 시간만큼 더함으로써, 바로 그 다음 고객의 대기행렬에서의 대기시간을 구하는데 사용된다.) 이것을

$$g_q(n) = \begin{cases} \sum_{k=0}^r g_q(k) \sum_{r=0}^k c(r) & n=0 \\ \sum_{k=r}^{n-1} g_q(n-k)c(k) & n=1,2,\dots \end{cases} \quad (18)$$

와 같이 다시 convolution함으로써, 그 다음 고객의 대기열에서의 대기시간의 분포를 구한다. 이때 FFT(Fast Fourier Transform) 알고리즘 [4]을 사용하고 초기의 $g_q(n)$ 은

$$g_q(n) = \begin{cases} 1 & n=0 \\ 0 & n=1,2,\dots \end{cases} \quad (19)$$

로 놓는다. (17), (18)의 계산을 반복적으로 수행하면서, $g_q(n)$ 의 변화가 일정한 값보다 작으면, 정상상태로 간주한다. g_q 과 g_s 을 다시 한번 convolution하여, 이산시스템에서의 정상상태의 대기시간의 확률분포 g_n 를 구한다.

(3) GPH/GPH/1 에서의 고객의 시스템 대기시간과, 대기 행렬 대기시간 정상분포를 (정리 3)를 적용하여 다음과 같이 구한다.

$$\lim_{n \rightarrow \infty} P \{W_q^n \leq x\} = \sum_{k=0}^r G_q(k) \frac{e^{-\lambda x} (\lambda x)^k}{k!}, \quad x > 0 \quad (20)$$

$$\lim_{n \rightarrow \infty} P \{W^n \leq x\} = \sum_{k=0}^r G^w(k) \frac{e^{-\lambda x} (\lambda x)^k}{k!},$$

위의 과정을 보면, 대기시간을 계산하기 위한 모든 연산은 이산함수의 단순연산들 뿐이다. 이런 잇점 때문에 대기 시스템에서 도착과 서비스가 GPH 분포로 표현만 되면, 대기시간의 분포의 계산이 가능하게 된다.

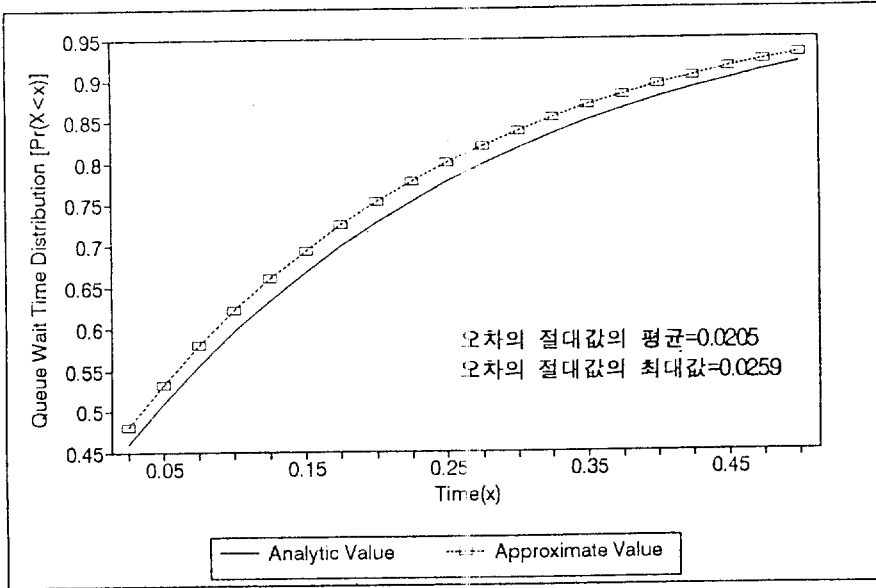
4.3 근사화 예제

본절에서는 4.2절의 결과를 이용하여 GI/M/1 시스템을 GPH/GPH/1 시스템으로 근사화시켜 대기시간의 분포를 구해본다. 여기서, GI/G/1 대신에 GI/M/1의 예제를 사용한것은, 근사화의 정확성을 알아보기 위해서는 대기시간의 정확한 해와 비교가 필요한데, GI/M/1 시스템에서는 대기시간의 정확한 분포의 계산이 가능하기 때문이다. 도착분포의 예제 함수를 각각 GE-2($\lambda_1=15, \lambda_2=12$)와 HE($\lambda_1=6, \lambda_2=8, p_1=0.4$)로 하고, λ 값이 정확성에 영향을 미치는 정도를 실험하기위해 GPH 함수에서 λ 값을 두가지 값으로 변화시켰다. GPH 분포로의 근사화를 이용해서 구한 대기시간의 분포와 이론적으로 구한 대기시간의 분포를 비교해 본 결과는 [그림 7]~[그림 10]와 같다.

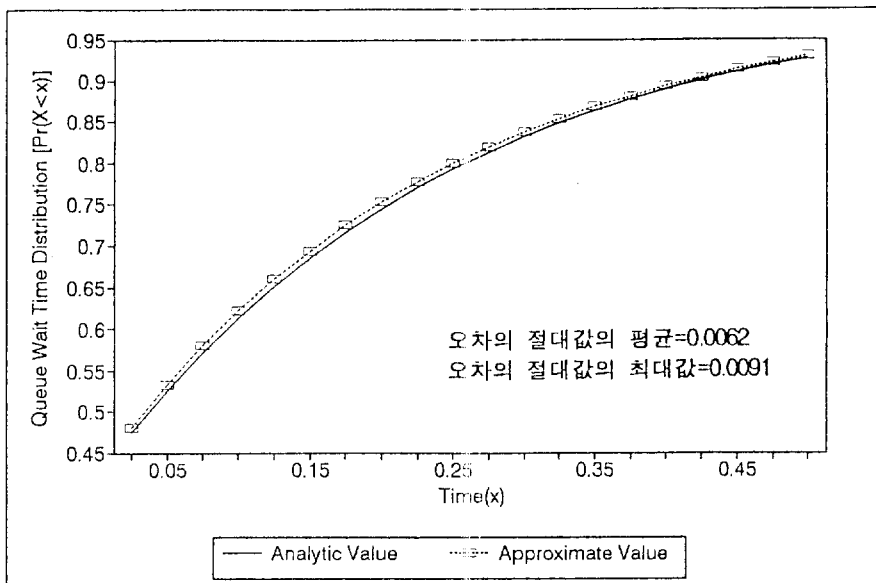
결과를 보면, 역시 λ 값이 근사화에 큰 영향을 끼친다는 사실을 알 수 있다. 여기서, λ 값을 크게 하는데는 제약이 발생한다. (14)에서 G_q 를 계산할 때 이산확률 변수의 확률 값을 저장하는 변수의 갯수가 λ 값이 커질수록 증가하기 때문인데 이는 또한 (7)의 근사화 식중에서 $F(\frac{n}{\lambda})$ 항의 값을 나타내는 변수의 갯수가 λ 값이 커질수록 증가하기 때문이다. 물론 이 변수

의 개수가 증가하면, convolution 연산을 수행할 때 계산시간도 더 많이 필요할 것이다. 따라서, 계산시간과 저장능력, 근사화에 있어서

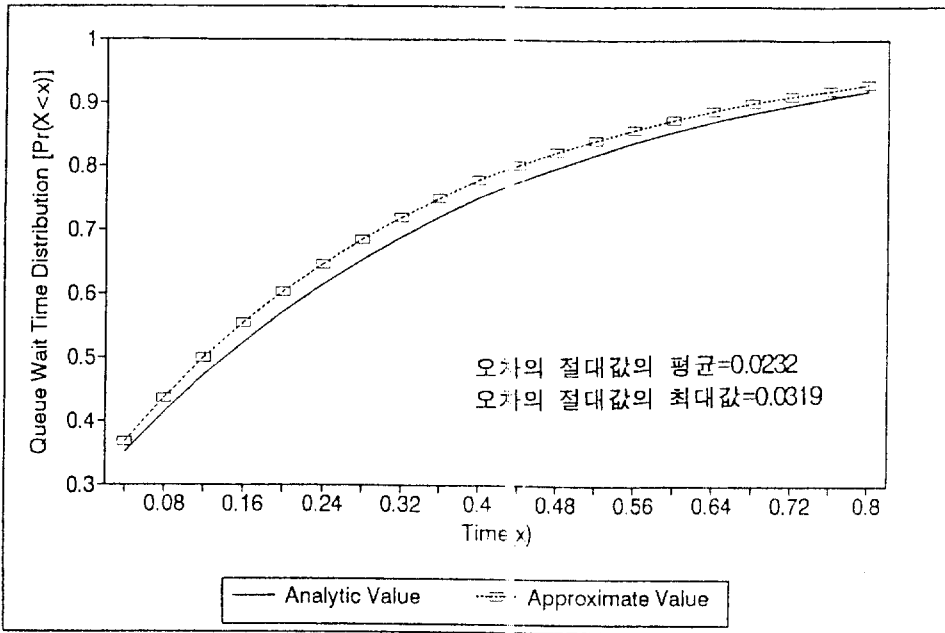
요구되는 정확도의 정도를 감안하여 λ 값을 정해야 한다.



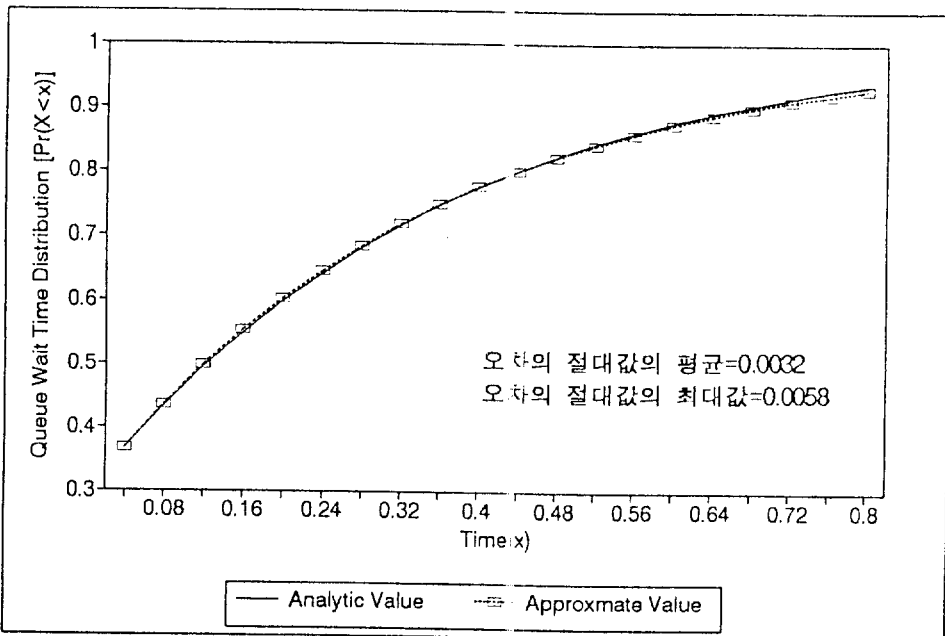
[그림 7] GE-2/ M/ 1 Queue의 GPH/ GPH/ 1 Queue로의 근사화
: GE-2($\lambda_1=15, \lambda_2=12$), $\exp(\mu=10)$, GPH($\lambda=80$)



[그림 8] GE-2/ M/ 1 Queue의 GPH/ GPH/ 1 Queue로의 근사화
: GE-2($\lambda_1=15, \lambda_2=12$), $\exp(\mu=10)$, GPH($\lambda=150$)



[그림 9] HE/ M/ 1 Queue의 GFH/ GPH/ 1 queue로의 근사화
 : GE-2($\lambda_1=6, \lambda_2=8, \rho_1=0.4$), $\exp(\mu=10)$, GPH($\lambda=40$)



[그림 10] HE/ M/ 1 Queue의 GPH/ GPH/ 1 Queue로의 근사화
 : GE-2($\lambda_1=6, \lambda_2=8, \rho_1=0.4$), $\exp(\mu=10)$, GPH($\lambda=80$)

5. 결론 및 토의

[부 록]

본 연구에서는 $[0, \infty)$ 에서 정의된 임의의 분포함수를 GPH 형태로 매우 정확하게 근사화시킬 수 있고, 또한 GPH 분포는 확률적인 계산에 편리한 성질을 가지고 있어서 확률적인 시스템 분석에 매우 유용하게 사용될 수 있음을 이론적, 실험적으로 보였다. 본 연구를 통해 GPH 분포를 기초로 하는 근사화의 정확성과 유용성이 어느 정도 확인되었으므로 앞으로 다양한 확률 분석에서의 GPH의 사용을 기대할 수 있다. 향후 연구에서는 보다 복잡한 시스템에서의 GPH의 효과적인 적용 방법과 주어진 정확도에 따라 적절한 λ 를 설정하는 규칙 등이 검토되어야 할 것이다.

(A) Bilateral Phase-type distribution (BPH)

L^+ 과 L^- 를 각각 $(g^+(n))_0'$ 과 $(g^-(n))_0'$ 의 확률 분포를 가지는 임의의 확률 변수라고 하자. 그리고, $E_n=0$ w.p 1, $(E_n, n=\pm 1, \pm 2, \dots)$ 을 L^+ 과 L^- 와는 독립이고, $1/\lambda$ 의 평균을 가지는 지수분포를 따르는 확률변수들이라고 하자. 이때 다음의 확률변수의 분포를 BPH라고 한다.

$$X = \begin{cases} \sum_{n=1}^{L^+} E_n, & w.p. p^+ \\ -\sum_{n=1}^{L^-} E_n, & w.p. p^- \end{cases}$$

(단, $p^+ + p^- = 1$)

$$(B) \begin{aligned} g^*(n) &= \sum_{k=0}^n \binom{n+k}{k} \left(\frac{1}{2}\right)^{n+k+1} g(k+1), \\ & \quad n=1, 2, \dots \\ g^*(0) &= \sum_{k=0}^{\infty} \left(\frac{1}{2}\right)^{k+1} g(k+1) + g(0) \end{aligned}$$

참 고 문 헌

1. Asmussen, S. *Applied Probability and Queues*, Wiley, New York, 1987.
2. Kleinrock L., *Queueing System Vol.2: Computer Applications*, Wiley, New York, 1976.
3. Neuts M. F., *Matrix-Geometric Solution in Stochastic Models: An Algorithmic Approach*, The Johns Hopkins University Press, Baltimore, 1981.
4. Press W. H., Flannery B. P., Teukolsky S. A., and Vetterling W. T., *Numerical Recipes in C: The Art of Scientific Computing*, Cambridge University Press, 1988.
5. Ross S. M., *Stochastic Process*, John and Sons, Inc., 1983.
6. Shanthikumar J. G., "Bilateral Phase-Type Distributions," *Naval Research Logistics Quarterly*, Vol.32, 1985, pp. 119-136.
7. Shaked, M and Shanthikumar, J. G., Phase Type Distributions, *Encyclopedia of Statistical Science*, N.L. Johnson and S. Kotz Eds., Wiley, New York, 1984, 709-715.
8. Yoon B. S., *Approximations for the Transient Behavior of Stochastic Processes: Discretization and Uniformization*, Ph. D. Dissertation, U. C. Berkely, 1988.