한국경영과학회지
제40권 제3호
2015년 8월
http://dx.doi.org/10.7737/JKORMS.2015.40.3.039

GPH 분포를 이용한 파산확률의 계산*

윤 복 식†
홍익대학교 기초과학과

Computing Ruin Probability Using the GPH Distribution

Bok Sik Yoon
Department of Science, Hongik University

■ Abstract ■

Even though ruin probability is a fundamental value to determine the insurance premium and policy, the complexity involved in computing its exact value forced us resort to an approximate method. In this paper, we first present an exact method to compute ruin probability under the assumption that the claim size has a GPH distribution, Then, for the arbitrary claim size distribution, we provide a method computing ruin probability quite accurately by approximating the distribution as a GPH. The validity of the proposed method demonstrated by a numerical example. The GPH approach seems to be valid for heavy-tailed claims as well as usual light-tailed claims.

Keywords : Ruin Probability, Insurance, GPH Distribution, Pollaczeck-Kinchine Formula

1. 서 론

파산확률(ruin probability)은 손해보험에서 보험요율을 정하는 기본적인 도구로 이용된다[2]. 그러나 파산확률의 정확한 계산은 매우 복잡하여 보통 근사적인 방법에 의존하게 된다[4, 5, 9, 10, 15 등]. Cramer-Lundberg 방법으로 대표되는 기존의 근사적인 기법은 대개 통상적인 위험 하에서는 어느 정도 정확성을 보여주지만 대재해 상황과 같은 돌발적인 상황이 고려될 때는 정확성이 떨어진다[7]. 대재해 상황에 적용되는 기법들[11, 12 등]이 제안된 바 있으나 초기유보금이 매우 클 때의 단순한 점근적(asymptotic)

논문접수일 : 2015년 05월 25일 논문게재확정일 : 2015년 07월 22일
논문수정일(1차 : 2015년 07월 03일, 2차 : 2015년 07월 17일)
* 본 연구는 홍익대학교 교내연구비의 지원을 받았음.
† 교신저자, bsyoon@hongik.ac.kr

공식만이 주어져 실용성이 떨어진다. 본 연구는 보통의 상황과 대재해 상황에서 동시에 사용할 수 있고 계산상에서도 효율적인 일반적인 기법을 개발하는 것이 목적이다.

본 연구에서는 클레임의 도착은 전통적인 푸아송 과정(Poisson process)을 가정하되, 클레임 크기의 분포는 매우 일반적인 상황을 상정한다. 즉, 일상적인 클레임의 상황이나 두꺼운 꼬리(heavy-tailed) 상황(대재해 상황)에 관계없이 일반적으로 유효성이 보장되는 파산확률 계산방법에 관심을 둔다. 물론 보다 복잡한 클레임 도착과정을 가정하고 접근할 수도 있겠으나(예를 들면 [7]), 고전적인 푸아송 도착이 보편적인 실제 상황과 충분히 부합되는 측면이 있고, 이 가정 하에서 조차 충분한 범용성과 정확성을 가진 파산확률의 계산방법을 찾기 쉽지 않다는 점 때문에, 본 연구에서는 복합푸아송 형태의 누적 클레임의 틀을 유지한다. 이 기본틀 하에서 파산확률은 제 2장에서 설명되는 대로 Pollaczeck-Kinchine 공식(다른 이름으로 Beekman 공식)으로 얻어지게 되는데, 이것을 GPH(generalized phase type) 접근방법[1, 2, 14]을 이용하면 정확하게 계산해 낼 수 있다는 것이 본 연구의 기본 착상이다.

범용성을 위해서 클레임의 크기의 분포는 되도록 일반적일 필요가 있는데, 이를 PH(phase type) 분포로 가정하고 접근하는 방법이 기존에 다양하게 존재한다([3]의 제 9장 참조), 그런데 GPH 분포는 균등화(uniformization) 기법에 의해 PH(phase type) 분포를 부분집합으로 포함한다[13]. 더욱이 비음의 실구간에서 정의되는 모든 연속 분포를 근사할 수 있고(제 3.3절의 설명 참조), 계산 과정도 PH 분포의 행렬기하(matrix geometric) 방법보다 훨씬 간편하므로 본 연구에서 제안하는 GPH의 접근법은 범용성과 효용성 양 측면에서 유용성을 갖는다고 여겨진다.

본 논문의 구성은 다음과 같다. 우선 제 2장에서 구하고자하는 보험에서의 파산확률이 명확히 정의되고 Pollaczeck-Kinchine 공식이 소개된다. 본 연구에서 제안하는 GPH 파산확률의 계산 방법은 제 3장에서 개발된다. 우선 제 3.1절에서 GPH 분포와 그것의 확률적 성질을 간단히 소개되고, 제 3.2절에서 클레임 분포가 GPH일 때의 정확한 파산확률의 공식이 유도된다. 또한 일반적인 분포를 GPH로 근사하는 방법이 제 3.3절과 제 3.4절에서 제시된다. 제 4장에서는 실험과 비교를 통해 제안된 방법의 정확성이 검증되고 제 5장에서 결론과 추후 연구 방향이 토의된 후 논문이 맺어진다.

2. 보험에서의 파산확률

일반적인 보험에서 보험회사는 주기적으로 정액의 보험료를 받고 클레임이 발생하면 사고의 규모에 따라 보험금을 지불한다. 어떤 보험의 시간 t까지의 클레임의 발생과정을 $\{N(t), t \geq 0\}$라고 하고 i번째 클레임의 규모를 $Z_i, i = 1, 2, \cdots$라고 놓자. 그러면 $(0, t]$ 사이의 누적클레임은

$$Y_t = \sum_{i=1}^{N(t)} Z_i$$

로 나타낼 수 있다. 보험료가 일정한 보험료율이 p로 연속적으로 들어온다고 가정하면 t에서의 누적 보험료와 누적클레임의 차이는

$$X_t = pt - Y_t$$

로 나타낼 수 있다. 초기($t = 0$)에 보험회사가 위험유보금을 u로 준비했다면 t에서의 위험유보금은

$$R_t = u + pt - \sum_{k=1}^{N(t)} Z_k = u + X_t \qquad (1)$$

로 정의된다. 본 연구에서는 $\{N(t), t \geq 0\}$를 발생률 α인 푸아송 과정으로 가정하고, $Z_i, i = 1, 2, \cdots$를 분포 F를 따르고 서로 독립이며 $N(t)$와도 독립인 확률변수 열로 가정한다. 우선 Z_i의 평균을 μ, 분산을 σ^2로 놓으면

$$ER_t = u + pt - \alpha\mu t = u + (p-\alpha\mu)t \qquad (2)$$

의 관계를 얻을 수 있다. 보통 보험가입자는 매우 많은 반면 개별 가입자의 사고는 발생빈도가 낮은 드문 사건이 되므로 보험회사에 청구되는 클레임 수 $N(t)$는 푸아송 과정으로 모형화 하는 것이 자연스럽다.

식 (2)에서 시간이 흐르는 동안 위험유보금이 평균적으로 양의 값을 유지하기 위해서는 보험료율과 사고청구 비용률의 차이 $p-\alpha\mu$가 양이 되어야 함을 알 수 있다. 보험학에서는 비율

$$\eta = \frac{p-\alpha\mu}{\alpha\mu}$$

를 안전 할증률(safety loading)이라 부른다.

$$\rho = \alpha\mu$$

라고 정의하면 $\eta = \frac{p-\rho}{\rho}$가 되고 특별히 $p=1$일 때는

$$\eta = \frac{1-\rho}{\rho} \text{ 또는 } \rho = \frac{1}{1+\eta}$$

의 관계가 된다. 모든 금액의 단위를 p를 기준으로 맞출 수 있으므로 앞으로 $p=1$로 가정하고, 파산이 확실히 일어나는 경우는 고려 대상에서 제외하여 $\eta > 0\ (0 < \rho < 1)$로 설정한다.

이제 파산확률은 고려 대상 시간이 무한인 경우에는

$$\Psi(u) = P\{\inf_{t \geqq 0} R_t < 0\}$$

로, 유한시간 T인 경우에는

$$\Psi(u,T) = P\{\inf_{0 \leqq t \leqq T} R_t < 0\}$$

로 정의될 수 있다. 본 논문에서는 무한 시간의 경우인 $\Psi(u)$에 초점을 맞춘다.

이제 $\overline{F}(x) = 1-F(x)$로 놓고 F의 누적꼬리분포(integrated tail distribution) 또는 평형분포(equilibrium distribution)를

$$F_e(z) \equiv \frac{1}{\mu}\int_0^z \overline{F}(x)dx$$

라고 정의하면 파산확률의 기본공식인 Pollaczeck-Kinchine 공식을 다음과 같이 얻을 수 있다([3]의 제4장).

$$\Psi(u) = \frac{\eta}{1+\eta}\sum_{n=0}^{\infty}\left(\frac{1}{1+\eta}\right)^n \overline{F_e^{(n)}}(u) \qquad (3)$$

여기서 $F_e^{(n)}(u)$는 F_e의 자신과의 n중 합성곱(n-fold convolution)을 의미하고 관례적으로 $F_e^{(0)}(u) \equiv 1, u \geq 0$으로 정의되고 $\overline{F_e^{(n)}}(u) = 1 - F_e^{(n)}(u)$이다. 식 (3)에서

$$\Psi(0) = \frac{1}{1+\eta} = \frac{\alpha\mu}{p} \quad (\eta > 0) \qquad (4)$$

임을 관찰할 수 있다. $p=1$일 경우 식 (3)을 다시 표현하면

$$\Psi(u) = (1-\rho)\sum_{n=0}^{\infty}\rho^n \overline{F_e^{(n)}}(u) \qquad (5)$$

가 되고, $\Psi(0) = \rho$이 된다.

F가 평균 μ인 지수분포이면 식 (3)은 다음과 같이 정확하게 계산된다.

$$\Psi(u) = \frac{1}{1+\eta}e^{-\eta u/\mu(1+\eta)} \qquad (6)$$

$p=1$일 경우 $\delta = \frac{1}{\mu}$라 놓으면

$$\Psi(u) = \rho e^{-(\delta-\alpha)u} \qquad (7)$$

가 얻어진다. 클레임 규모가 지수분포를 따르는 특

수한 경우에는 파산확률이 식 (6) 또는 식 (7)과 같이 정확히 계산되나 일반적으로 합성곱이 간단한 형태로 얻어지지 않기 때문에 식 (5)의 정확한 계산은 쉽지 않다.

3. GPH 파산확률의 계산

3.1 $GPH(\lambda, g)$ 분포

본 절에서는 GPH(generalized phase type) 분포와 그 성질을 간략히 요약한다. $GPH(\lambda, g)$ 분포 F는

$$F(x)=\sum_{n=0}^{\infty} G(n)\frac{e^{-\lambda x}(\lambda x)^n}{n!},\ x\geq 0 \qquad (8)$$

의 형태를 갖는다. 여기서 $G(n)=\sum_{i=0}^{n} g(i)$이고 $g=\{g(n),\ n=0, 1, \cdots\}$는 어떤 이산 확률변수의 L의 확률질량함수(pmf)이다. 만약 $g(0)>0$이면 F는 0에서 $g(0)$의 도약을 갖게 되고 $F(0)$ = 0이면 $g(0)=G(0)=0$이 된다. $x>0$에서 F는 밀도함수

$$f(x)=\sum_{n=0}^{\infty} g(n+1)\frac{e^{-\lambda x}(\lambda x)^n}{n!}\lambda,\ x>0 \quad (9)$$

를 갖는다. X가 $GPH(\lambda, g)$ 분포를 따르면 X는 L개의 발생률 λ인 iid 지수 확률변수들의 합으로 해석될 수 있는데, 균등화(uniformization) 기법을 통해 GPH가 PH(phase type) 분포를 포함함을 알 수 있다[13]. 즉, GPH 분포는 PH 분포보다 일반적이고 제 3.3절에서 소개되는 방법을 이용하면 $x\geq 0$에서 정의되는 모든 연속분포를 근사할 수 있다.

GPH 분포는 확률 분석에 유용한 성질들을 가지고 있다. 우선 X의 생존함수(survival function) $\overline{F}(x)\equiv 1-F(x)$는

$$\overline{F}(x)=\sum_{n=0}^{\infty}\overline{G}(n)\frac{e^{-\lambda x}(\lambda x)^n}{n!},\ x\geq 0 \qquad (10)$$

로 표현된다. 여기서

$$\overline{G}(n)=1-G(n),\ n=0, 1, 2,\cdots$$

즉, G의 생존함수이다.

또한 F의 평형분포를 $F_e(x)\equiv\frac{1}{E(X)}\int_0^x(1-F(t))\,dt$로 정의하면 F가 $GPH(\lambda, g)$ 분포일 때

$$\begin{aligned}
F_e(x)&=\frac{1}{E(X)}\int_0^x\sum_{k=0}^{\infty}\overline{G}(k)\frac{e^{-\lambda t}(\lambda t)^k}{k!}dt\\
&=\frac{1}{E(X)}\sum_{k=0}^{\infty}\overline{G}(k)\int_0^x\frac{e^{-\lambda t}(\lambda t)^k}{k!}dt\\
&=\frac{1}{\lambda E(X)}\sum_{k=0}^{\infty}\overline{G}(k)\sum_{n=k+1}^{\infty}\frac{e^{-\lambda x}(\lambda x)^n}{n!}\\
&=\frac{1}{E(L)}\sum_{n=1}^{\infty}\Big(\sum_{k=0}^{n-1}\overline{G}(k)\Big)\frac{e^{-\lambda x}(\lambda x)^n}{n!}\\
&=\sum_{n=1}^{\infty}G_e(n)\frac{e^{-\lambda x}(\lambda x)^n}{n!}
\end{aligned}$$

의 형태가 된다. 여기서 $G_e(n)=\frac{1}{E[L]}\sum_{k=0}^{n-1}\overline{G}(k),\ n=1, 2, \cdots$, 즉, G의 평형분포이고 $g_e(n)=\frac{\overline{G}(n-1)}{E[L]}$, $n=1, 2, \cdots$ 이다. F_e의 정의에서 $F_e(0)=0$이므로 $G_e(0)=g_e(0)=0$이고 결국 마지막 식은

$$F_e(x)=\sum_{n=0}^{\infty}G_e(n)\frac{e^{-\lambda x}(\lambda x)^n}{n!},\ x\geq 0 \quad (11)$$

과 같이 $GPH(\lambda, g_e)$ 분포로 표현된다..

GPH 분포는 유한개의 합성곱(convolution)에 대해 닫혀 있다. 이 사실은 직관적으로 쉽게 파악할 수 있는데 X_1을 $GPH(\lambda, g_1)$, X_2를 $GPH(\lambda, g_2)$ 분포를 따르는 독립인 두개의 확률 변수, $(E_n)_0^{\infty}$을 평균이 $1/\lambda$인 지수 분포를 따르는 독립인 확률변수들의 열(단, E_0 = 0 w.p.1), L_1과 L_2를 각각 g_1, g_2를 pmf로 갖는 확률변수라고 하자. 그러면 $X_1\sim\sum_{n=0}^{L_1}E_n$이

고, $X_2 \sim \sum_{n=L_1+1}^{L_1+L_2} E_n$이기 때문에, $X = X_1 + X_2$라고 정의하면, $X \sim \sum_{n=0}^{L_1+L_2} E_n$가 되어, X는 $GPH(\lambda, g_1 * g_2)$ 분포를 따르게 된다(여기서 $g_1 * g_2(n) \equiv \sum_{k=0}^{n} g(n-k)g(k)$, 이산 합성곱을 의미하고 기호 ~는 같은 분포를 따르는 관계를 의미한다). 즉, $X = X_1 + X_2$의 분포인 합성곱 $GPH(\lambda, g_1) * GPH(\lambda, g_2)$가

$$GPH(\lambda, g_1) * GPH(\lambda, g_2) = GPH(\lambda, g_1 * g_2) \quad (12)$$

과 같이 되어, 연속 변수인 X_1과 X_2의 연속 합성곱(적분)의 계산이 이산 합성곱(합)의 계산으로 바뀌어 나타남을 관찰 할 수 있다. 이 성질은 GPH을 다양한 확률적인 계산에 응용할 수 있게 해 주는 중요한 성질이다. 물론, 이와 같은 연산이 가능하기 위해서는, 두 GPH 분포의 λ값을 동일하게 설정해야 한다.

한편 $GPH(\lambda, g)$분포 F의 Laplace-Stieltjes 변환(LST)는

$$\begin{aligned} \tilde{F}(s) &\equiv \int_0^\infty e^{-sx} dF(x) \quad (13) \\ &= g(0) + \sum_{n=0}^{\infty} g(n+1) \int_0^\infty e^{-sx} \frac{\lambda e^{-\lambda x} (\lambda x)^n}{n!} dx \\ &= \hat{G}\left(\frac{\lambda}{\lambda + s}\right) \end{aligned}$$

가 된다. 여기서 $s > 0$이고 $\hat{G}$는 g에 대한 확률생성함수(pgf)이다. 즉,

$$\hat{G}(z) \equiv \sum_{n=0}^{\infty} g(n) z^n, \; |z| < 1$$

이다. 식 (13)을 이용하여 식 (12)의 양변의 LST를 구하면 $X = X_1 + X_2$의 분포의 LST는

$$\hat{G}_1\left(\frac{\lambda}{\lambda + s}\right) \hat{G}_2\left(\frac{\lambda}{\lambda + s}\right) = \hat{H}\left(\frac{\lambda}{\lambda + s}\right) \quad (14)$$

가 됨을 알 수 있다. 여기서 $h = g_1 * g_2$이고 $\hat{H}$는 h의 pgf이다.

또한 $GPH(\lambda, g_1)$와 $GPH(\lambda, g_2)$의 확률 v에 의한 혼합(mixture) 분포함수는

$$F(x) = \sum_{n=0}^{\infty} [v G_1(n) + (1-v) G_2(n)] \frac{e^{-\lambda x} (\lambda x)^n}{n!}$$

형태가 되어 역시 $GPH(\lambda, g_m)$ 분포가 된다. 여기서

$$g_m = v g_1 + (1-v) g_2 \quad (15)$$

이고 $0 \le v \le 1$이다.

3.2 GPH 파산확률의 계산

이제 위에 언급한 GPH의 성질을 종합하여 클레임 규모가 GPH 분포로 주어 질 때의 정확한 파산확률의 공식을 구해보자. 이는 다음의 정리로 요약할 수 있다.

정리 : 클레임의 분포 F를 $GPH(\lambda, g)$라고 가정하면 파산확률 $\Psi(u)$는 $GPH(\lambda, g_Q)$의 생존함수가 된다. 이때

$$g_Q = (1-\rho) \sum_{k=0}^{\infty} \rho^k g_e^{(k)} \quad (17)$$

이다.

증명 : Pollaczeck-Kinchine 식 (5)를 다시 쓰면

$$\begin{aligned} \Psi(u) &= (1-\rho) \sum_{k=0}^{\infty} \rho^k \overline{F_e^{(k)}}(u) \\ &= 1 - (1-\rho) \sum_{k=0}^{\infty} \rho^k F_e^{(k)}(u) \end{aligned}$$

또는

$$1-\Psi(u)=(1-\rho)\sum_{k=0}^{\infty}\rho^k F_e^{(k)}(u) \qquad (18)$$

가 된다. 클레임의 분포 F를 $GPH(\lambda, g)$라고 가정하면 식 (11)에서 F_e는 $GPH(\lambda, g_e)$로 표현되고 따라서 F_e의 LST는 식 (13)에서 $\widehat{G}_e(\frac{\lambda}{\lambda+s})$가 된다. 여기서 $\widehat{G}_e$는 g_e의 pgf이다. 또한 F_e의 k중 합성곱 $F_e^{(k)}$도 식 (14)에 의해 LST가 $(\widehat{G}_e(\frac{\lambda}{\lambda+s}))^k$인 $GPH(\lambda, g_e^{(k)})$가 된다. 여기서 $g_e^{(k)}$는 g_e의 k중 합성곱이다. 한편, $F_e^{(k)}$의 기하분포에 의한 혼합(mixture)인 식 (18)의 우변

$$(1-\rho)\sum_{k=0}^{\infty}\rho^k F_e^{(k)}(u)$$

도 GPH가 된다. 즉, $1-\Psi(u)$는 $GPH(\lambda, g_Q)$ 형태가 되는데 이때 g_Q는 식 (15)와 유사하게 접근하면 식 (17)과 같이 얻어진다. ■

정리에서 식 (17)이 얻어졌으나 여기에는 무한대개의 다중 합성곱의 계산이 포함되어 실제로 계산해내기가 쉽지 않다. 이제 큐잉시스템에서의 수준교차(level-crossing) 기법을 이용하여 g_Q를 보다 효율적으로 계산하는 방법을 고안해보자. 우선 $1-\Psi(u)$를 FIFO-M/G/1 큐잉시스템에서 정상상태의 대기시간(delay) 분포로 해석할 수 있음을 관찰한다[3, 15 등]. 이제 고객의 도착과정이 도착률 α인 푸아송과정이고 서비스 시간이 평균 μ인 분포 F를 따른다고 하고 $\rho=\alpha\mu<1$라고 가정하자. 이때 F가 $GPH(\lambda, g)$라고 가정하면 앞의 설명과 같이 정상상태의 대기시간의 분포 $F_Q(x)=1-\Psi(x)$도 GPH 형태가 되는데 이것을 $GPH(\lambda, g_Q)$로 표시하자. g_Q를 구하기 위해 우선 대기시간이 0일 확률은 $1-\rho$이므로

$$g_Q(0)=F_Q(0)=1-\rho \qquad (19)$$

를 얻는다. $x>0$ 대해 가상지연과정(virtual delay process)에서 수준 x를 내려가는 교차율(down-crossing rate)과 올라가는 교차율(up-crossing rate)이 같아야 하므로

$$f_Q(x)=\alpha\int_0^x(1-F(x-t))\,dF_Q(t),\quad x>0 \qquad (20)$$

을 얻는다. 식 (20)의 좌변을 GPH 형태로 나타내면

$$\sum_{n=0}^{\infty}g_Q(n+1)\frac{e^{-\lambda x}(\lambda x)^n}{n!}\lambda \qquad (21)$$

이 된다. 이제 우변에 GPH 형태를 대입해 정리해보자. 우선 $F_Q(0)=1-\rho$이므로

$$\begin{aligned}
&\int_0^x(1-F(x-t))\,dF_Q(t)\\
&\quad=(1-\rho)\overline{F}(x)+\int_0^x\sum_{m=0}^{\infty}\overline{G}(m)\\
&\qquad\times\frac{e^{-\lambda(x-t)}(\lambda(x-t))^m}{m!}\\
&\qquad\times\left[\sum_{l=0}^{\infty}g_Q(l+1)\frac{e^{-\lambda t}(\lambda t)^l}{l!}\lambda\right]dt\\
&\quad=(1-\rho)\overline{F}(x)+\sum_{m=0}^{\infty}\sum_{l=0}^{\infty}\overline{G}(m)g_Q(l+1)\\
&\qquad\times\frac{e^{-\lambda x}\lambda^{m+l+1}}{m!l!}\int_0^x(x-t)^m t^l\,dt
\end{aligned}$$

인데

$$\int_0^x(x-t)^m t^l dt=\frac{m!l!}{(m+l+1)!}x^{m+l+1}$$

이므로

$$\begin{aligned}
&\int_0^x(1-F(x-t))\,dF_Q(t)\\
&\quad=(1-\rho)\overline{F}(x)+\sum_{m=0}^{\infty}\sum_{l=0}^{\infty}\overline{G}(m)g_Q(l+1)
\end{aligned}$$

$$\times \frac{e^{-\lambda x}(\lambda x)^{m+l+1}}{(m+l+1)!}$$

$$= g_Q(0)\sum_{n=0}^{\infty}\overline{G}(n)\frac{e^{-\lambda x}(\lambda x)^n}{n!} + \sum_{n=1}^{\infty}\sum_{k=0}^{n-1}\overline{G}(k)g_Q(n-k)\frac{e^{-\lambda x}(\lambda x)^n}{n!}$$

$$= \sum_{n=0}^{\infty}\sum_{k=0}^{n}\overline{G}(k)g_Q(n-k)\frac{e^{-\lambda x}(\lambda x)^n}{n!}$$

이 되어 식 (20)의 우변은 결국

$$\alpha\sum_{n=0}^{\infty}\left[\sum_{r=0}^{n}\overline{G}(n-r)g_Q(r)\right]\frac{e^{-\lambda x}(\lambda x)^n}{n!} \qquad (22)$$

과 같이 쓸 수 있다.

이제 식 (21), 식 (22)를 등호로 연결하고 비교하면 $n=0,1,\cdots$에 대해

$$g_Q(n+1)=(\frac{\alpha}{\lambda})\sum_{r=0}^{n}\overline{G}(n-r)g_Q(r) \qquad (23)$$

과 같은 되돌림 관계를 얻을 수 있다.

이제 식 (19)와 식 (23)을 이용하여 $g_Q(n)$, $n=0$, $1,\cdots$를 구한 후

$$G_Q(n)=\sum_{k=0}^{n}g_Q(k),\ n=0,1,\cdots$$

를 구하여

$$F_Q(x)=\sum_{n=0}^{\infty}G_Q(n)\frac{e^{-\lambda x}(\lambda x)^n}{n!}$$

와 같이 대기시간 분포를 계산하여 최종적으로 파산확률을

$$\Psi(u)=1-F_Q(u)$$

로 구할 수 있다.

3.3 GPH 분포로의 근사

이제 일반적인 클레임 규모의 상황에 대해 적용하기 위래 일반 분포를 GPH로 맞추는 방법이 필요하다. 클레임 규모의 분포 F가 주어지면

$$F_\lambda(t)=\sum_{n=0}^{\infty}F(\frac{n}{\lambda})\frac{e^{-\lambda t}(\lambda t)^n}{n!},\ t\geq 0,\ \lambda>0 \qquad (24)$$

는 $\lambda\to\infty$일 때 $F(t)$로 수렴하게 된다[1]. 여기서 λ를 크게 잡을수록 GPH F_λ가 F를 더욱 정확하게 근사 할 수 있음을 관찰할 수 있다.

만약 F가 미리 알려져 있지 않다면 우선 과거의 클레임의 자료로 부터 경험분포 $F_m(x)$을

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

과 같이 구한다. 여기서

$$I(A)=\begin{cases}1, & A\text{가 참일 때}\\ 0, & \text{그 외의 경우}\end{cases}$$

이다. 이제 식 (24)에서 $F(x)$대신 $F_m(x)$를 대입하여 GPH 형태

$$F_{m,\lambda}(x)=\sum_{n=0}^{\infty}F_m(\frac{n}{\lambda})\frac{e^{-\lambda x}(\lambda x)^n}{n!} \qquad (25)$$

이 얻어진다. 식 (25)는 모든 고정된 m에 대해 $\lambda\to\infty$일 때 $F_{m,\lambda}(x)\to F_m(x)$이 되므로 m을 크게 하면 결국 확률 1로 $F(x)$로 접근함을 관찰할 수 있다.

3.4 GPH 분포의 효과적 근사계산

식 (24)(또는 식 (25))는 무한급수이므로 실제 계산을 위해서는 적절한 절단이 필요하다. 우선

$$\sum_{n=0}^{K}\frac{e^{-\lambda t}(\lambda t)^n}{n!}\geq 1-\epsilon$$

인 최소의 K를 구하면

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

이 된다. 따라서

$$F_\lambda(t) \cong \sum_{n=0}^{K} F(\frac{n}{\lambda})\frac{e^{-\lambda t}(\lambda t)^n}{n!} \tag{26}$$

로 계산하면 절단오차를 ϵ이하로 유지할 수 있다. λ가 충분히 큰 경우 $\epsilon = 10^{-7}$일 때 대략

$$K \cong \lambda t + 5\sqrt{\lambda t}$$

이면 충분하다. 이를 적용하여 식 (26)을 계산할 때 계산의 안정성과 효율을 위해 다음과 같은 방법을 사용할 수 있다.

우선 $n = \lceil \lambda t \rceil$ 일 때

$$P_n = \frac{e^{-\lambda t}(\lambda t)^n}{n!}$$

를 먼저 계산하고 관계식

$$P_{n+1} = \frac{\lambda t}{n+1} P_n$$

을 이용하여 $P_{n+1}, P_{n+2}, \cdots, P_{n+\lceil 5\sqrt{\lambda t}\rceil}$ 를 차례로 계산하고 마찬가지로

$$P_{n-1} = \frac{n}{\lambda t} P_n$$

을 이용하여 $P_{n-1}, P_{n-2}, \cdots, P_{n-\lceil 5\sqrt{\lambda t}\rceil}$ 를 차례로 구하여

$$\sum_{n=(\lceil \lambda t\rceil - \lceil 5\sqrt{\lambda t}\rceil)^+}^{\lceil \lambda t\rceil + \lceil 5\sqrt{\lambda t}\rceil} F(\frac{n}{\lambda})P_n$$

와 같이 $F_\lambda(t)$의 근사값을 계산한다. 여기서 $(x)^+ = \max\{0, x\}$이다.

4. 적용 예

실제로 제 3장에서 제시된 GPH 방법은 λ를 충분히 크게 잡으면 분포 자체를 매우 정확하게 근사할 수 있어 파산확률의 정확성이 보장되지만, 계산의 간편성과 정확성을 예시하기 위해 본 장에서 간단한 적용 예를 제시한다. [15]에 주어진 예에서 정확한 파산확률 값이 주어져 있으므로 그 논문의 예를 사용하여 비교 검증해 보자. 그 논문에서는 클레임 규모의 pdf가

$$f(x) = b(bx)^{a-1}e^{-bx}/\Gamma(a)$$

즉, 감마(a, b) 분포인 경우를 고려하였다. $a=3$, $b=3$인 경우(이때 평균은 $\mu = a/b = 1$, 분산은 $\sigma^2 = a/b^2 = 1/3$, 제곱변동계수는 $c^2 = \sigma^2/\mu^2 = 1/a = 1/3$)와 $a=0.5$, $b=0.5$(이때 $\mu = 1, c^2 = 2$)인 경우에 $\rho = 0.2, 0.5, 0.8$일 때의 파산확률의 계산결과가 각각 <표 1>과 <표 2>에 주어져 있다. 이 표들에서 GPH 방법의 정확성을 곧 바로 확인할 수 있다. 여기서는 $\lambda = 2^8 = 256$의 비교적 작은 값을 사용하였다. 또한 각 경우에 대해 GPH 방법을 이용하여 구한 파산확률 $\Psi(u)$의 그래프가 각각 [그림 1]과 [그림 2]에 주어져 있다.

5. 결 론

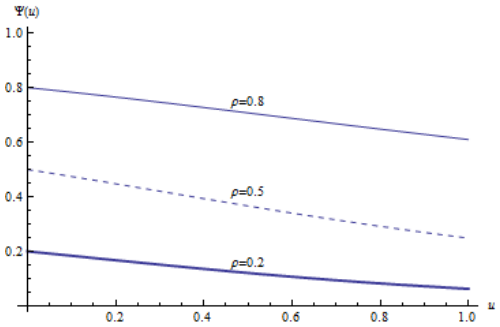
파산확률의 정확한 계산은 보험 정책수립의 기초가 되는 중요한 작업이다. 본 논문에서는 양의 실수값을 갖는 모든 분포를 근사할 수 GPH 분포를 토대로 하여 매우 정확한 파산확률을 계산해 내는 범용의 방법을 제시하였다. 또한 계산 예를 통해 이 파산확률 계산방법의 타당성과 정확성을 검증하였다. 본 논문의 방법은 다양한 상황에서 매우 정확하게 파산확률을 계산해 주므로 보험계획과 분석에서

〈표 1〉 $c^2=1/3$인 경우의 파산확률(GPH : $\lambda=256$)

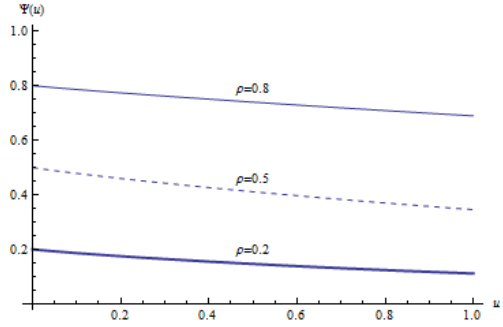
ρ	u	GPH	Willmot	참값
0.2	0.1	0.1839	0.1822	0.1839
	0.25	0.1595	0.1596	0.1594
	0.5	0.1209	0.1217	0.1209
	0.75	0.0883	0.0886	0.0882
	1	0.0627	0.0627	0.0626
0.5	0.1	0.4744	0.4728	0.4744
	0.25	0.4342	0.4344	0.4342
	0.5	0.3665	0.3672	0.3664
	0.75	0.3035	0.3037	0.3033
	1	0.2485	0.2484	0.2484
0.8	0.1	0.7834	0.7826	0.7834
	0.25	0.7562	0.7564	0.7562
	0.5	0.7075	0.7074	0.7074
	0.75	0.6578	0.6578	0.6577
	1	0.6098	0.6096	0.6097

〈표 2〉 $c^2=2$인 경우의 파산확률(GPH : $\lambda=256$)

ρ	u	GPH	Willmot	참값
0.2	0.1	0.1865	0.1849	0.1865
	0.25	0.1697	0.1685	0.1697
	0.5	0.1465	0.1461	0.1466
	0.75	0.1274	0.1275	0.1276
	1	0.1113	0.1117	0.1115
0.5	0.1	0.4786	0.4768	0.4787
	0.25	0.4511	0.4499	0.4512
	0.5	0.4113	0.4111	0.4114
	0.75	0.3765	0.3769	0.3768
	1	0.3455	0.3462	0.3458
0.8	0.1	0.7861	0.7852	0.7862
	0.25	0.7677	0.7672	0.7677
	0.5	0.7396	0.7396	0.7397
	0.75	0.7136	0.7140	0.7139
	1	0.6892	0.6896	0.6894



[그림 1] $c^2=1/3$인 경우의 파산확률 그래프 (GPH : $\lambda=256$)



[그림 2] $c^2=2$ 인 경우의 파산확률 그래프 (GPH : $\lambda=256$)

실용적으로 적용될 수 있다. 향후 분포 모형의 가정 없이 클레임 데이터로부터 직접 파산확률 얻는 단일화된 방법에 대한 연구가 계속될 것이다. 또한 계획기간이 주어진 경우의 유한시간 파산확률의 계산에 GPH 방법을 확장하는 연구가 계속될 것이다.

참 고 문 헌

[1] 윤복식, 박광우, 이창훈, “GPH 분포에 의한 확률적 근사화”, 「한국경영과학회지」, 제19권, 제1호(1994), pp.85-98.

[2] 윤복식, “일반적인 큐잉네트워크에서의 체류시간분포의 근사화”, 「한국경영과학회지」, 제19권, 제3호(1994), pp.93-109.

[3] Asmussen, S. and H. Albrecher, *Ruin Probability*(2ed.), World Scientific, Singapore, 2010.

[4] Avram, F. and M. Pistorius, “On matrix ex-

ponential approximations of ruin probability for the classic and Brownian perturbed Cramer-Lundberg processes," *Insurance : Mathematics and Economics*, Vol.59(2014), pp.57-64.

[5] Brekelmans, R. and A. De Waegenaere, *Approximating the finite-time ruin probability under interest force, Insurance : Mathematics and Economics*, Vol.29(2001), pp.217-229.

[6] Coulibaly, I. and C. Lefevre, "On a simple quasi-Monte Carlo approach for classical ultimate ruin probabilities," *Insurance : Mathematics and Economics*, Vol.42(2008), pp.935-942.

[7] Chen, Y. and K.W. Ng, "The ruin probability of the renewal model with constant interest force and negatively dependent heavy-tailed claims," *Insurance : Mathematics and Economics*, Vol.40(2007), pp.415-423.

[8] Dickson, D.C.M. and H.R. Waters, "Reinsurance and ruin," *Insurance : Mathematics and Economics*, Vol.19(1994), pp.61-80.

[9] Grandell, J., "Simple approximations of ruin probabilities," *Insurance : Mathematics and Economics*, Vol.26(2000), pp.157-173,

[10] Ignatov, Z.G., V.K. Kaishev, and R.S. Krachunov, "An improved finite-time ruin probability formula and its Mathematica implementation," *Insurance : Mathematics and Economics*, Vol.29(2001), pp.375-386.

[11] Kalashnikov, V. and D. Konstantinides, "Ruin under interest force and subexponential claims : a simple treatment," *Insurance : Mathematics and Economics*, Vol.27(2000), pp.145-149.

[12] Kalashnikov, V. and R. Norberg, "Power tailed ruin probabilities in the presence of risky investments," *Stochastic Processes and their Applications*, Vol.98(2002), pp.211-228.

[13] Shaked, M. and J.G. Shanthikumar, "Phase Type Distributions," in *Encyclopedia of Statistical Sciences*, 6, John Wiley and Sons, New York(editors S. Kotz and N.L. Johnson), (1985), pp.709-715.

[14] Shanthikumar, J.G., "Bilateral Phase-Type Distributions," *Naval Research Logistics Quarterly*, Vol.32(1985), pp.119-136.

[15] Willmot, G.E., "On a class of approximations for ruin and waiting time probabilities," *Operations Resaerch Letters*, Vol.22(1998), pp.27-32.