Journal of the Korean Data & Information Science Society
2011, **22**(2), 171–177

한국데이터정보과학회지

결합 로그노말-파레토 분포에서 추출된 양쪽 중도 절단된 표본을 이용한 모수추정

이광호[1]

[1]영남대학교 통계학과

접수 2011년 1월 12일, 수정 2011년 2월 9일, 게재확정 2011년 2월 14일

요 약

최근에 비약적으로 발달하는 보험 산업에 수반하여 보험금 지불 분포에 대한 연구가 활발하게 진행되고 있다. 보험금 지불금의 분포는 일반적으로 두터운 꼬리를 가지면서 좌로 치우친 왜도를 가지는 파레토 분포나 로그노말 분포로 잘 설명된다고 알려져 왔으며 Cooray와 Ananda (2005)는 이들 두 분포를 결합한 결합 로그노말-파레토분포를 제시하고 이 분포의 적합도가 높음을 보였다. 그런데 보험금 지불의 경우 보금지불 총 금액의 한도로 인하여 극단적으로 큰 보험금이나 혹은 매우 사소한 보험지불금의 경우는 옵션을 두어 예외적으로 취금하는 경우가 많다. 본 논문에서는 결합 로그노말-파레토 분포로부터 추출된 표본이 양쪽 중도 절단 되어 있는 경우에 대하여 모수를 추정하는 문제를 다루어 보았다.

주요용어: 결합 로그노말-파레토분포, 로그노말분포, 양쪽 중도절단 표본, 파레토분포.

1. 서론

금융업의 발달과 더불어 손해보험이나 생명보험과 같은 보험산업이 급속도록 발달하고 있는 것은 주지의 사실이며 또한 그에 상응하여 보험과 관련된 여러 가지 통계모형이나 분석기법이 심도 있게 연구되고 있는 것은 누구나 잘 알고 있는 사실이다. 본 논문에서는 보험금 지불 금액을 잘 설명해 줄 수 있는 통계적 분포의 특성에 대해 알아보고 현실적으로 보험금의 지불 특성으로 나타날 수 있는 양쪽으로 중도 절단된 표본을 이용하여 모수를 추정하는 문제를 다루고자 한다.

보험금의 지급 규모는 사건의 크기에 의해서 결정되며 대형사건과 소형사건의 빈도를 생각하면 사건이 크게 될수록 그 빈도는 낮을 것이고 사건의 규모가 작을 수 록 그 빈도는 높을 것으로 기대되며 이와 같은 사건의 빈도를 잘 반영하기 위한 분포의 특성은 오른쪽으로 긴 꼬리 (upper long tail)를 가지거나 혹은 왼쪽으로 치우친 왜도 (skewness)를 가지는 분포가 적합할 것이다. 이와 같은 분포의 특성을 가지는 분포함수로는 감마 (gamma), 와이블 (Weibull), 로그노말 (lognormal), 파레토 (Pareto) 등의 분포가 있으나 로그노말 분포와 파레토 분포가 보험금 지불금을 비교적 잘 설명해 주는 분포로 알려져 왔다 (Hogg와 Klugman, 1984). Resnick (1997)와 Beirlant 등 (2004)은 보험금 지불금액이 극단적으로 큰 경우가 많은 경우에는 파레토 분포 혹은 일반화 파레토분포가 유용하게 이용될 수 있다는 것을 보였다. 로그노말 분포는 Burneckit 등 (2000)에 의해 일반 재해 보험 데이터에 유용하게 적용됨을 보였다.

그러나 파레토 분포는 그 밀도함수의 모양이 아래로 볼록한 (low-humped) 단조 감소 (monotonically decreasing) 형태의 특성을 가지기 때문에 흔히 나타날 수 있는 위로 볼록한 (upper-humped) 형

[1] (712-749) 경북 경산시 대동 214-1, 영남대학교 통계학과, 교수. E-mail: khlee@yu.ac.kr

태를 보인 후 단조 감소하게 되는 데이터의 분포에 적용하기에는 다소 무리가 있다고 생각할 수 있다. 사실 이와 같은 경우의 데이터에 잘 적용되는 분포는 로그노말 분포라 할 수 있다. 그러나 로그노말 분포의 오른쪽 꼬리 (upper tail)가 파레토 분포의 그것에 비하여 상대적으로 매우 빠르게 수평축으로 가까워지는 특성을 지니기 때문에 보험금 지불금이 큰 경우가 빈번한 경우에 적용하면 과소추정 (under estimates)할 가능성이 매우 크게 될 것이라 생각할 수 있다. 그래서 Resnick (1997)과 McNeil (1997)은 지불금액이 큰 사건이 흔히 나타나는 보험금 지불 데이터에서는 상대적으로 부담이 적은 지불금액이 적은 데이터는 무시한 나머지 데이터에 대해 파레토 분포를 적용하는 것을 권장하였다. 그러나 Cooray와 Ananda (2005)은 로그노말 분포와 파레토 분포를 결합하여 적은 규모의 지불금액에 대해서는 로그노말 분포가 적용되고 지불금액이 큰 경우에는 파레토 분포가 적용될 수 있는 결합 로그노말-파레토 분포 (composite lognormal-Pareto distribution)를 정의하여 이 분포의 특성을 규명하였다. 그리고 그들은 이 분포의 적합도와 로그노말 분포의 적합도 그리고 파레토 분포의 그것을 비교하여 결합 로그노말-파레토 분포의 적합도가 더 높음을 예제 데이터를 통하여 설명하였다.

근년에 Preda와 Ciumara (2006)는 Cooray와 Ananda (2005)의 방법과 유사하게 와이블분포와 파레토분포를 결합하여 결합 와이블-파레토 분포라 명명하고 결합 로그노말-파레토 분포와 비교 연구하여 서로 비슷한 특성과 성질을 가지고 있다고 하였으며, Scollnik (2007)는 로그노말 분포와 파레토 분포의 결합을 Cooray와 Ananda (2005)들이 제시한 방법과 다른 새로운 두 가지 방법으로 결합하여 세 가지 형태의 결합 로그노말-파레토 분포를 비교 분석하였다. 그리고 Pigeon과 Denuit (2010)는 로그노말 분포와 파레토 분포를 결합할 때의 경계점 (threshold)이 랜덤 (random)인 경우까지 확장하여 연구하였으며 그는 경계점이 감마분포를 따를 때와 로그노말분포를 따를 때로 나누어서 Scollnik (2007)등의 결과와 비교하였다.

본 연구에서는 보험금 지불의 특성상 빈번하게 발생되면서 보험지불 금액이 소액인 경우와 매우 드물에 발생하면서 보험지불 금액이 매우 큰 경우는 별도로 고려하는 것이 바람직할 것으로 생각되어 보험지불 금액 모델에서의 표본이 양쪽 중도 절단된 경우에 모수를 추정하는 문제를 다룬다. 2절에서는 Cooray와 Ananda (2005)가 제시한 결합 로그노말-파레토 분포로 부터 양쪽 중도 절단된 표본을 추출하였을 때 모수의 추정량을 유도하고 3절에서는 모의실험을 통한 추정량들의 특성을 제시한다.

2. 양쪽 중도 절단된 표본을 이용한 모수추정

2.1. 결합 로그노말-파레토 분포

보험 지불금 데이터를 분석하기 위하여 Cooray와 Ananda (2005)는 아래와 같이 로그노말 분포함수와 파레토 분포함수를 결합하여 결합 로그노말-파레토 분포를 정의하고 이 분포함수의 여러 가지 특성을 규명하였다.

$$f(x) = \begin{cases} \dfrac{\alpha\theta^{\alpha}}{(1+\Phi(k))x^{\alpha+1}} \exp\left\{-\dfrac{\alpha^2}{2k^2}\ln^2(x/\theta)\right\}, & 0 < x \le \theta \\ \dfrac{\alpha\theta^{\alpha}}{(1+\Phi(k))x^{\alpha+1}}, & \theta \le x < \infty \end{cases} \tag{2.1}$$

여기서 $\Phi(.)$는 표준 정규확률변수의 누적분포함수이고, $\theta > 0$, $-\infty < \mu < \infty$, $\sigma > 0$ 그리고 $\alpha > 0$이며 k는 $\exp(-k^2) = 2\pi k^2$을 만족하는 상수이다.

위의 분포함수는 Cooray와 Ananda (2005)가 양의 실수 θ에 대해 $0 < x \le \theta$에서는 $cg_1(x)$이 되고 $\theta \le x < \infty$에서는 $cg_2(x)$가 되도록 로그노말 분포 $g_1(x)$와 파레토 분포 $g_2(x)$를 결합한 것이다. 즉,

다음과 같이 각각 두 개의 모수를 가지는 로그노말 분포 $g_1(x)$와 파레토 분포 $g_2(x)$에 대해 Cooray와 Ananda (2005)는 $0 < x \leq \theta$에서는

$$g_1(x) = \frac{(2\pi)^{-1/2}}{x\sigma} \exp\left[-\frac{1}{2}\left(\frac{lnx - \mu}{\sigma}\right)^2\right],\ x > 0.$$

$$g_2(x) = \frac{\alpha\theta^\alpha}{x^{\alpha+1}},\ x > \theta.$$

$cg_1(x)$로 표시되고 $\theta \leq x < \infty$에서는 $cg_2(x)$로 표시되는 분포함수가 되기 위해서는 상수 c는 $1/(1+\Phi(k))$가 되어야 하며 또한 모수들 사이에는 다음과 같은 관계를 만족해야 함을 보였다.

$$\frac{\ln\theta - \mu}{\sigma} = \alpha\sigma = k.$$

그들은 모수들 사이의 이러한 관계를 이용하여 $cg_1(x)$과 $cg_2(x)$를 식 (2.1)과 같이 표현하였다. 식 (2.1)에서 보는 바와 같이 결합 로그노말-파레토 분포도 로그노말 분포나 파레토 분포와 같이 단지 두 개의 모수만을 가지고 있음을 알 수 있다. 그리고 로그노말 분포와 파레토 분포의 누적분포함수가 잘 알려져 있으므로 이것을 이용하면 결합 로그노말-파레토 분포의 누적분포함수가 아래와 같이 표현할 수 있다.

$$F(x) = \begin{cases} \dfrac{1}{1+\Phi(k)}\Phi((\alpha/k)\ln(x/\theta) + k), & 0 < x \leq \theta \\ 1 - \dfrac{1}{1+\Phi(k)}(\theta/x)^\alpha, & \theta \leq x < \infty \end{cases}.$$

2.2. 모수추정

보험금 지불금 데이터는 서론에서 언급한 바와 같이 양쪽으로 중도 절단되는 경우가 많으므로 양쪽으로 중도 절단된 표본을 이용하여 최우추정법으로 모수를 추정하기로 한다. 먼저 2.1.에서 소개된 결합 로그노말-파레토 분포에서의 표본을 $X_1, X_2, \cdots, X_n$라 하고, 이 표본들의 순서 통계량의 값을 $x_1 \leq x_2 \leq \cdots \leq x_n$으로 나타내기로 한다. 그리고 왼쪽 중도 절단된 데이터는 g개, 오른쪽으로 중도 절단된 데이터는 f개라 가정한다. 그러면 실제로 관측된 데이터는 $n-f-g$개가 될 것이다. 즉, 관측된 데이터는 $x_{g+1} \leq x_{g+2} \leq \cdots \leq x_{n-f}$가 된다. 그리고 $g+1$보다 크고 $n-f-1$보다 적은 어떤 상수 m에 대해 분포함수의 경계값 (threshod) θ가 x_m과 x_{m+1}사이에 있다고 가정하면 양쪽 중도 절단된 표본에 대한 우도함수는 결합 로그노말-파레토 분포의 밀도함수와 누적분포함수로 다음과 같이 표시된다.

$$L(\alpha, \theta; x_1, x_2, \cdots, x_n) = \prod_{i=1}^{g} F(x_{g+1}) \prod_{i=g+1}^{m} f_1(x_i) \prod_{i=m+1}^{n-f} f_2(x_i) \prod_{i=n-f+1}^{n} [1 - F(x_{n-f})].$$

여기서 $f_1(x)$는 식 (2.1)에서 0< x $\leq \theta$인 경우의 밀도함수 $f(x)$이고 $f_2(x)$는 식 (2.1)에서 $\theta \leq x \leq \infty$인 경우의 밀도함수 $f(x)$이다. 위의 우도함수를 정리하면 다음과 같이 된다.

$$L = B(\prod_{i=g+1}^{n-f} x_i^{-\alpha})\alpha^{n-f-g}\theta^{(n-g)\alpha} \exp\left\{-\frac{\alpha^2}{2k^2}\sum_{i=g+1}^{m} \ln^2(\frac{x_i}{\theta})\right\} \Phi(A)^g (x_{n-f}^{-\alpha f}).$$

위의 식에서 A와 B는 다음이며 A는 모수들의 함수이고 B는 모수와 무관한 상수이다.

$$A = \left(\frac{\alpha}{k}\right) \ln\left(\frac{x_{g+1}}{\theta}\right) + k.$$

$$B = (\frac{1}{1+\Phi(k)})^n (\prod_{i=g+1}^{n-f} x_i^{-1}).$$

위의 우도함수로부터 대수 우도함수를 계산하여 모수 α와 θ에 대해 각각 편미분하여 정리하면 다음의 식을 얻을 수 있다.

$$\frac{\partial}{\partial\alpha} lnL = -\sum_{i=g+1}^{n-f} lnx_i + \frac{n-f-g}{\alpha} + (n-g)\ln\theta - \frac{\alpha}{k^2}\sum_{i=g+1}^{m} \ln^2(\frac{x_i}{\theta})$$

$$+ \frac{g \cdot \ln(\frac{x_{g+1}}{\theta}) \exp\left\{-A^2/2\right\}}{\sqrt{2\pi}k\Phi(A)} - f \cdot lnx_{n-f}$$

$$\frac{\partial}{\partial\theta} lnL = \frac{1}{\theta}\left\{(n-g)\alpha + \frac{\alpha^2}{k^2}\sum_{i=g+1}^{m} \ln(\frac{x_i}{\theta}) - \frac{g\alpha \exp\left\{-A^2/2\right\}}{\sqrt{2\pi}k\Phi(A)}\right\}.$$

대부분의 최우추정법의 적용에서 그러하듯이 위의 식에서 직접적으로 최우추정값을 얻을 수는 없지만 일반적인 수치해석법을 적용하여 모수를 추정하는 데는 큰 어려움이 없을 것이다.

3. 모의실험

이 절에서는 앞 절에서 다룬 결합 로그노말-파레토 분포에서 양쪽 중도 절단된 표본을 추출하여 모수들의 최우추정에 대한 모의실험을 수행한다. 표본의 크기가 30과 50인 경우에 왼쪽으로 중도 절단된 표본과 오른쪽으로 중도 절단표본이 동시에 5인 경우의 모의실험 결과를 정리하였다. 표본을 생성하기 위하여 모수 α는 1과 2의 값을 주었고, 경계값 θ는 5와 10 그리고 20의 값을 주었다. 표본의 크기와 모수의 참과 그리고 경계값을 다양하게 변화를 시키면서 각각 1,000번의 반복실험을 수행한 결과가 대체로 유사한 패턴을 보였으며 여기에서는 결과의 패턴을 보여 주는 최소의 결과만을 정리하였다. 계산은 c 언어를 사용하고 고정점 반복법으로 비선형 연립방정식의 해를 구하였다. 그리고 결합 로그노말-파레토 분포에서 양쪽 중도 절단된 표본에 대한 선행 연구의 결과가 없으므로 양쪽 중도 절단된 경우의 추정값과 평균제곱오차와 절단값이 전혀 없는 경우의 추정값과 평균제곱오차를 계산하여 비교하였다.

모의실험을 위한 표본의 생성은 결합 로그노말-파레토 분포의 누적분포함수를 이용하여 다음과 같이 생성하였다. 먼저 일양분포 $U(0,1)$에서 하나의 표본을 추출하여 그것을 u라 하고, 모의실험을 위해 주어진 경계값 θ에 대한 $F(\theta)$와 비교를 한다. 물론 여기에서 $F(.)$는 결합 로그노말-파레토 분포의 누적분포함수이다. 비교의 결과가 $u \le F(\theta)$이면

$$u = \frac{1}{1+\Phi(k)}\Phi((\alpha/k)\ln(x/\theta) + k)$$

를 만족하는 x를 계산하고 $u > F(\theta)$이면

$$u = 1 - \frac{1}{1+\Phi(k)}(\theta/x)^{\alpha}$$

를 만족하는 x를 계산하면 x가 결합 로그노말-파레토 분포를 따르는 모집단에서의 표본이 된다. 물론 위의 식에 포함되어 있는 k는 $\exp(-k^2) = 2\pi k^2$을 만족하는 상수이므로 여기에서는 $k = 0.372238898$을 사용할 것이다. 이와 같이 생성된 n개의 표본 중에서 작은 것 g개와 큰 것 f개는 중도 절단된 표본으로 간주하여 실험을 하였다.

사실 위의 두 식을 정리하면 생성되는 표본 x는 다음의 식과 같이 계산된다. 만약 $u \le F(\theta)$이면

$$x = \theta \cdot \exp\left\{\frac{k}{\alpha}(\Phi^{-1}(u(1+\Phi(k)) - k\right\}$$

와 같이 생성되고, $u > F(\theta)$이면

$$x = \theta \cdot \{(1-u)(1+\Phi(k))\}^{-\frac{1}{2}}$$

와 같이 생성된다.

표 3.1 n=30, α=1인 경우

		추정 값	평균제곱오차	추정 값	평균제곱오차
$\theta = 5$	양쪽 절단된 경우	$\widehat{\alpha}$ =1.098	0.165	$\widehat{\theta}$ =4.956	0.751
	완전한 경우	$\widehat{\alpha}$ =1.050	0.039	$\widehat{\theta}$ =5.079	0.417
$\theta = 10$	양쪽 절단된 경우	$\widehat{\alpha}$ =0.991	0.152	$\widehat{\theta}$ =10.195	2.721
	완전한 경우	$\widehat{\alpha}$ =1.058	0.042	$\widehat{\theta}$ =10.123	1.887
$\theta = 20$	양쪽 절단된 경우	$\widehat{\alpha}$ =0.956	0.157	$\widehat{\theta}$ =21.941	8.636
	완전한 경우	$\widehat{\alpha}$ =1.054	0.038	$\widehat{\theta}$ =20.157	6.382

표 3.2 n=50,α =1인 경우

		추정 값	평균제곱오차	추정 값	평균제곱오차
$\theta = 5$	양쪽 절단된 경우	$\widehat{\alpha}$ =1.026	0.126	$\widehat{\theta}$ =5.034	0.401
	완전한 경우	$\widehat{\alpha}$ =1.032	0.021	$\widehat{\theta}$ =5.028	0.236
$\theta = 10$	양쪽 절단된 경우	$\widehat{\alpha}$ =0.988	0.131	$\widehat{\theta}$ =10.125	1.354
	완전한 경우	$\widehat{\alpha}$ =1.034	0.020	$\widehat{\theta}$ =10.078	1.009
$\theta = 20$	양쪽 절단된 경우	$\widehat{\alpha}$ =0.993	0.129	$\widehat{\theta}$ =21.118	6.784
	완전한 경우	$\widehat{\alpha}$ =1.033	0.021	$\widehat{\theta}$ =20.137	4.062

표 3.3 n=30, α=2인 경우

		추정 값	평균제곱오차	추정 값	평균제곱오차
$\theta = 5$	양쪽 절단된 경우	$\widehat{\alpha}$ =2.086	0.232	$\widehat{\theta}$ =4.957	0.443
	완전한 경우	$\widehat{\alpha}$ =2.113	0.168	$\widehat{\theta}$ =5.020	0.109
$\theta = 10$	양쪽 절단된 경우	$\widehat{\alpha}$ =2.125	0.228	$\widehat{\theta}$ =9.751	1.532
	완전한 경우	$\widehat{\alpha}$ =2.097	0.161	$\widehat{\theta}$ =10.034	0.435
$\theta = 20$	양쪽 절단된 경우	$\widehat{\alpha}$ =2.096	0.234	$\widehat{\theta}$ =21.597	2.760
	완전한 경우	$\widehat{\alpha}$ =2.114	0.143	$\widehat{\theta}$ =20.059	1.612

위의 표로부터 당연한 결과로 볼 수 있지만 절단된 경우의 추정값의 편의와 평균제곱오차가 절단되지 않은 경우의 그것 보다 다소 크다는 것을 알 수 있다. 그러나 그 차이가 상당히 미미하여 양쪽 절단 경우라 하더라도 결합 로그노말-파레토 분포의 모수를 추정하여 활용하는 데는 크게 문제되지 않을 것으로 판단된다. 그리고 표 3.1과 표 3.2 그리고 표 3.3과 표 3.4를 비교는 표본의 크기 효과가 추정의 결과에 얼마

표 3.4 n=50, α =2인 경우

		추정 값	평균제곱오차	추정 값	평균제곱오차
$\theta = 5$	양쪽 절단된 경우	$\widehat{\alpha}$ =2.067	0.107	$\widehat{\theta}$ =4.987	0.224
	완전한 경우	$\widehat{\alpha}$ =2.062	0.083	$\widehat{\theta}$ =5.000	0.061
$\theta = 10$	양쪽 절단된 경우	$\widehat{\alpha}$ =2.082	0.093	$\widehat{\theta}$ =10.241	0.814
	완전한 경우	$\widehat{\alpha}$ =2.059	0.074	$\widehat{\theta}$ =10.038	0.231
$\theta = 20$	양쪽 절단된 경우	$\widehat{\alpha}$ =2.124	0.112	$\widehat{\theta}$ =20.574	1.325
	완전한 경우	$\widehat{\alpha}$ =2.069	0.080	$\widehat{\theta}$ =20.004	0.940

나 큰 영향을 미치는가를 알 수 있다. 당연히 표본크기가 크면 최우추정량의 일치성에 의하여 추정의 결과가 좋아지리라는 것을 예상할 수 있는데 위에서 제시된 결과에서도 예상된 결과가 나오기는 하였지만 표본의 크기가 50인 경우가 30인 경우 보다 편의와 평균제곱오차가 예상 보다 훨씬 더 적은 것을 확인할 수 있는데 이것은 결합 로그노말-파레토 분포에서 얻은 표본이 특히 양쪽 절단된 경우에는 표본크기가 추정의 결과에 상당히 크게 미치며 표본의 크기가 50이 이상이 되는 경우에는 양쪽 절단된 경우의 표본이라 하더라도 추정의 결과가 상당히 신뢰성을 가진다고 생각할 수 있다.

위의 표를 통하여 알 수 있는 결합 로그노말-파레토 분포의 또 다른 특징은 모수 α의 추정값 추정값의 평균제곱오차는 두 분포를 결합하는 경계값 θ의 영향은 별로 받지 않는 것으로 보이며 경계값 θ의 추정값 $\widehat{\theta}$은 α가 1이냐 2이냐에 따라 평균제곱오차가 큰 차이를 보인다. 즉, $\alpha = 2$인 경우의 $\widehat{\theta}$의 평균제곱오차가 $\alpha = 1$인 경우의 $\widehat{\theta}$의 그것 보다 표본의 크기가 30과 50인 경우 모두 훨씬 더 작다는 것을 알 수 있다. 또한 위의 표는 경계값 θ는 그 값이 큰 경우 보다는 작은 경우의 추정이 신뢰성이 높다는 것을 보여준다. 표에서 θ가 5, 10 그리고 20인 경우 각 모수들을 추정하였는데 모수 α의 추정과는 달리 모수 θ의 추정은 θ의 참값이 큰 경우 일수록 그 추정의 평균제곱오차가 크게 되어 추정의 신뢰성이 떨어짐을 알 수 있다.

참고문헌

Beirlant, J., Joossens, E. and Segers, J. (2004). Generalized Pareto fit to the society of actuaries' large claims database. *North-American Actuarial Journal,* **8**, 108-111.

Burnecki, K., Kukla, G. and Weron, R. (2000). Property insurance loss distribution. *Physica A: Statistical Mechanics and its Applications,* **287**, 269-278.

Cooray, K., and Ananda, M. M. A. (2005). Modeling actuarial data with a composite lognormal-Pareto model. *Scandinavian Actuarial Journal,* **2005**, 321-334.

Hogg, R. V. and Klugman, S. A. (1984). *Loss distribution,* John Wiley & Sons Inc., New York.

McNeil, A. I. (1997). Estimating the tails of loss severity distribution using extreme value theory. *ASTIN Bulletin,* **27**, 117-137.

Pigeon, M. and Denuit, M. (2010). Composite lognormal-Pareto model with random threshold. *Scandinavian Actuarial Journal,* in press.

Preda, V. and Ciumara, R. (2006). On composite models: Weibull-Pareto and lognormal-Pareto. *Romanian Journal of Economic Forecasting,* **2**, 32-46.

Resnick, S. I. (1997). Discussion of the Danish data on large fire insurance losses. *ASTIN Bulletin,* **27**, 139-151.

Scollnik, D. P. M. (2007). On composite lognormal-Pareto models. *Scandinavian Actuarial Journal,* **2007**, 20-33.

Journal of the Korean Data & Information Science Society
2011, **22**(2), 171–177

한국데이터정보과학회지

Estimation on composite lognormal-Pareto distribution based on doubly censored samples

Kwangho Lee[1]

[1]Department of Statistics, Yeungnam University

Received 12 January 2011, revised 9 February 2011, accepted 14 February 2011

Abstract

With the development of the actuarial and insurance industries, the distributions of the insurance payments data are deeply studied by many authors. It is known that theses types of distribution are very highly positively skewed and have a long thick upper tail such as Pareto or lognormal distribution. In 2005, Cooray and Ananda proposed a new model which is composed lognormal distribution and Pareto distribution. They said it as composite lognormal-Preto distribution. They showed that the proposed distribution was better fitted than lognormal or Pareto distribution. On the other hand many agreements about the insurance payment have some options for a trivially small payment or extremely large one because of the limits of total payment. Appling these cases, in this paper we consider the parameter estimation on the composite lognormal-Pareto distribution based on doubly censored samples.

Keywords: Composite lognormal-Pareto distribution, doubly censored sample, lognormal distribution, Pareto distribution.

[1] Professor, Department of Statistics, Yeungnam University, Gyeongsan 712-749, Korea.
E-mail: khlee@yu.ac.kr