

## 8장. 사고 발생 금액은 어떻게 모델링할까?

### ◆ Introduction

앞선 7장에서는 빈도(frequency) 모델링에 관하여 (a,b,0) class 와 (a,b,1) class를 알아보았다. 포아송, 이항, 음이항분포 같은 이산형 분포 모델들은 주로 피보험자들의 사고건수, 즉 빈도를 표현하는데 쓰였다. 회사가 보험상품을 설계하기 위해서는 보험 계약 하나 당 손해액도 모델링 해야 한다. 즉, 보험료 책정에 사용되는 순보험료(pure premium)을 구하기 위해 빈도뿐만 아니라 심도(severity)를 알아야 한다. 그렇다면 이번 장에서는 심도를 모델링 하기 위해서 사용되는 연속형 분포들의 종류들과 특징들을 알아보자.

### 1. 정규분포(Normal Distribution), 로그정규분포(Lognormal Distribution)

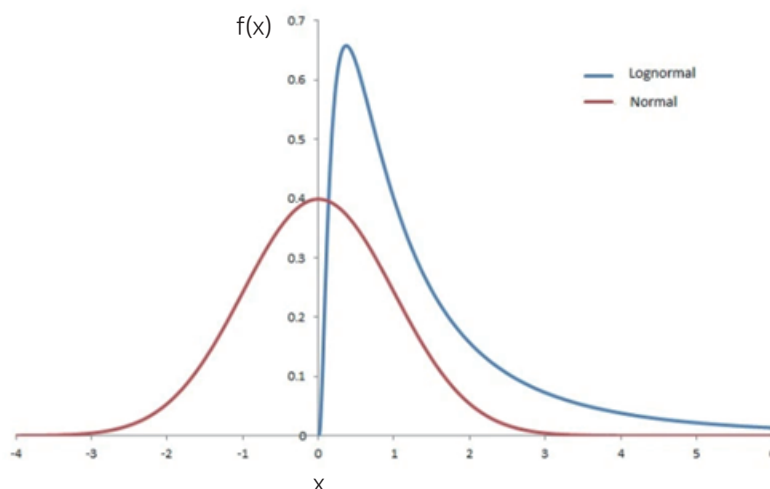
로그정규분포는 정규분포의 변수를 지수화 하여 변형시킨 분포이다. 즉 로그를 취했을 때 정규분포를 따르는 변수를 로그정규분포를 따른다고 표현한다. 다음은 정규분포와 로그정규분포의 확률밀도함수이다.

**Definition.** 정규분포와 로그정규분포의 확률밀도함수

$$\text{정규분포 : } f_{\text{Normal}}(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2} \text{ for } -\infty < x < \infty$$

$$\text{로그정규분포 : } f_{\text{Lognormal}}(x) = \frac{1}{x\sigma\sqrt{2\pi}} e^{-\frac{(\log(x)-\mu)^2}{2\sigma^2}} \text{ for } 0 < x < \infty$$

다음은 정규분포와 로그정규분포의 표준화된( $\mu=0, \sigma=1$ ) 확률밀도 그래프를 나타낸다.



[그림1] 표준정규분포, 표준로그정규분포 그래프

## 8장. 사고 발생 금액은 어떻게 모델링할까?

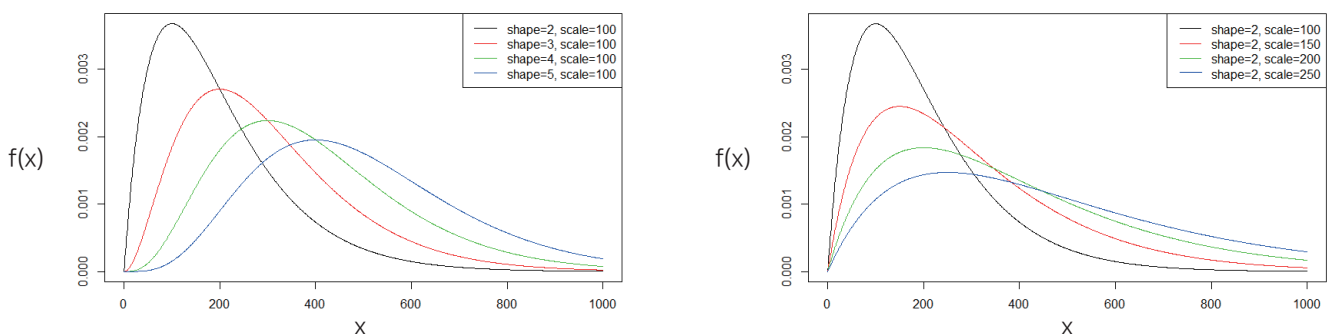
정규분포 그래프를 보면 음의 값에도 확률이 부여되어 있으므로 심도를 모델링하는데 문제가 있어 보인다. 그래서 정규분포를 이용하여 심도를 모델링 하려면 분산에 비해 높은 평균값을 부여하는 제한조건을 두어 음의 값에 부여되는 확률이 0에 가깝게 해야 한다. 대안으로 정규분포를 지수변형 시킨 로그정규분포를 사용하기도 한다. 로그정규분포는  $x$  축의 범위가 항상 양의 값을 가지기 때문에 정규분포에서의 문제가 해소된다. 로그정규분포의 대부분의 확률은 좌측에 집중되기 때문에 오른쪽으로 꼬리가 긴 (Right-skewed) 특성을 지닌다.

### 2. 감마분포(Gamma Distribution)

감마분포는 모수  $\alpha, \theta$  로 구성되며 확률분포와 모수값에 따른 그래프는 다음과 같다.

**Definition.** 감마분포의 확률밀도함수

$$\text{감마분포: } f(x) = \frac{(x/\theta)^\alpha}{x\Gamma(\alpha)} e^{-x/\theta}, \text{ for } x > 0.$$



[그림2] 각 모수 값(Shape, Scale) 변화에 따른 감마분포 형태

확률분포를 살펴보면 확률변수는 항상 0 이상의 값을 가진다. 이는 보험회사 입장에서 심도를 관측하기에 좋은 조건이다(손해액이 음수 일리는 없다). [그림2]의 좌측 그래프는 감마분포의 shape parameter인  $\alpha$ 가 바뀔에 따라 분포가 어떻게 변형되는지 나타내고 [그림2]의 우측 그래프는 scale parameter인  $\theta$  값에 따라 분포가 어떻게 변형되는지 나타내고 있다. 각 모수의 값이 증가 할수록 평균과 분산이 커지는 것을 관측할 수 있다.

감마분포는  $\alpha=1$ 일 때 지수분포가 되며  $\alpha=\frac{n}{2}, \theta=2$ 일 때는 카이제곱분포가 되는 특징이 있다. 보험회사의 입장에서 중요한 감마분포의 특징은 꼬리의 확률이 낮다(light-tailed)는 것이다. 감마분포의 모든 양의  $k$ 에 대하여 적률(Moments)이 존재하여 light-tailed 분포로 분류되고 높은 심도에 낮은 확률을 부여한다. 즉, 리스크가 큰 상품이나 커다란 손실이 일어날 가능성이 극히 적다면 감마분포를 모델 분포로 고려해볼 수도 있다.

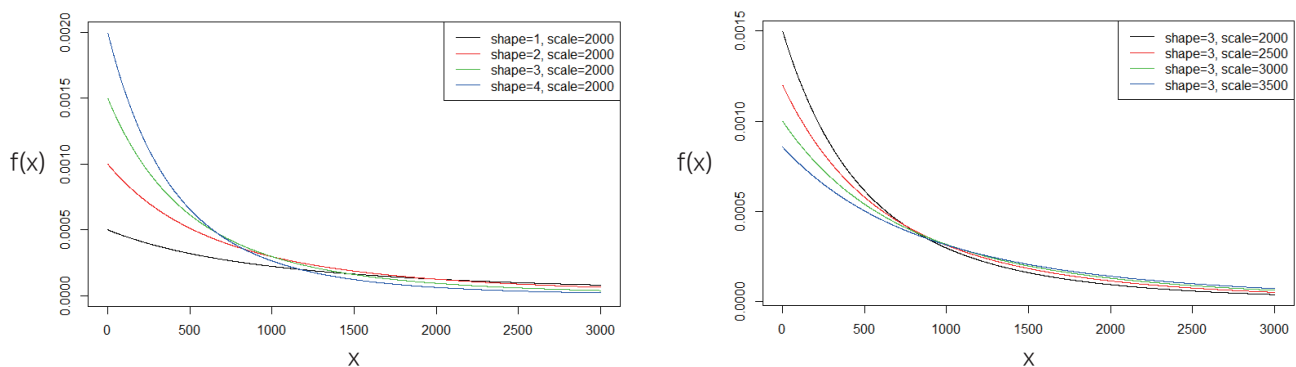
## 8장. 사고 발생 금액은 어떻게 모델링할까?

### 3. 파레토분포(Pareto Distribution)

파레토분포는 모수  $\alpha, \theta$  로 구성되며 확률분포와 모수값에 따른 그래프는 다음과 같다.

**Definition.** 파레토분포의 확률밀도함수

$$\text{파레토분포 : } f(x) = \frac{\alpha\theta^\alpha}{(x+\theta)^{\alpha+1}} \text{ for } x > 0.$$



[그림3] 각 모수 값(Shape, Scale) 변화에 따른 파레토분포 형태

파레토분포는 감마분포와 같이 확률변수가 0 이상의 값을 가진다. 그러나 scale parameter  $\theta$  커질수록 평균값은 작아지고 shape parameter  $\alpha$ 가 작아질수록 평균값이 커진다. 보험회사에서 중요하게 생각하는 파레토분포의 특성은 양의 첨도와 꼬리의 확률이 크다(heavy-tailed)는 점이다.

[그림3]의 그래프들을 보면 확률이 좌측에 밀집해 있고, 심도가 증가 할수록 파레토 분포의 확률밀도함수 값이 감소하는 것을 확인할 수 있다. 큰 심도가 발생하는 구간에서 파레토 분포는 다른 분포들에 비해 확률밀도함수 값이 비교적 높은 편이다. 그래서 심도를 모델링 할 때, 파레토 분포 전체를 이용하기 보다는 큰 손실금액이 발생할 가능성이 높음을 표현해주기 위해 다른 분포와 Mixture된 형태로 사용하곤 한다. 예를 들어, 손실금액이 0~100만원까지는 감마분포로 모델링을 하고 100만원 초과 부분은 파레토 분포로 모델링한다. 그 후 전체 확률의 합이 1이 되도록 각 분포에 가중치를 주어 Mixture된 분포로 심도를 모델링 한다.

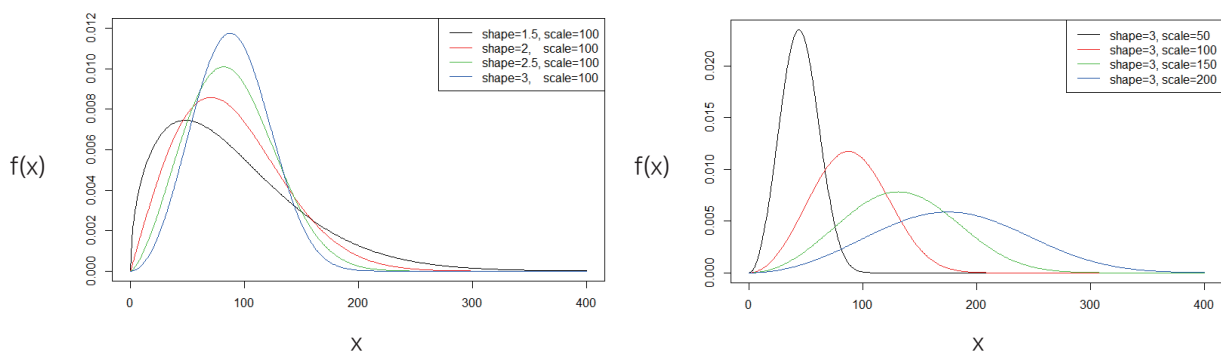
## 8장. 사고 발생 금액은 어떻게 모델링할까?

### 4. 와이블분포(Weibull Distribution)

와이블 분포의 모수는  $\alpha, \theta$  로 이루어지며 확률분포와 모수 값에 따른 그래프는 다음과 같다.

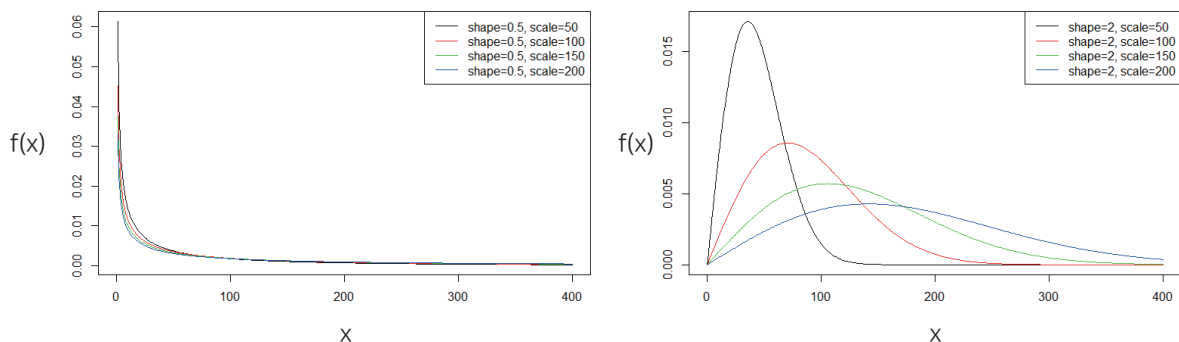
**Definition.** 와이블분포의 확률밀도함수

$$\text{와이블분포 : } f(x) = \frac{\alpha}{\theta} \left( \frac{x}{\theta} \right)^{\alpha-1} e^{-\left( \frac{x}{\theta} \right)^{\alpha}} \text{ for } x > 0.$$



[그림4] 각 모수 값(Shape, Scale) 변화에 따른 와이블분포 형태

와이블 분포 또한 확률변수가 0 이상의 범위에서 존재한다. 그러나 와이블 분포는 파레토, 감마분포와는 다르게 shape parameter  $\alpha$ 의 범위에 따라 light-tailed 또는 heavy-tailed 이 결정된다.  $\alpha$ 가 1 보다 작을 경우 heavy-tailed 분포로 취급되며  $\alpha$ 가 1 보다 클 경우에는 light-tailed 분포가 된다. 즉, 와이블 분포는 감마분포 보다 우측꼬리 부분에 대하여 보수적인 모델링 접근이 가능하다. [그림5]의 좌측 그래프는  $\alpha$ 가 0.5인 heavy-tailed를 나타내고 [그림5]의 우측 그래프는  $\alpha$ 가 2인 light-tailed를 나타낸다.



[그림5] Shape parameter 값에 따른 와이블분포 형태 (light-tailed or heavy-tailed)

★ 쉬어가는 페이지 ★



사진출처 : National Herald

## “금융위기 예측 실패는 정규분포 때문이다?”

세계적 규모의 금융위기가 일어나면 확률로 보아 수만 년에 한 번 있을 만한 일이 생겼다고도 한다. 심지어 1987년 10월 미국에서 있었던 금융상품의 가격 폭락 사태에 대해 전문가들은 그런 규모의 대폭락이 일어날 확률이 10의 160승 분의 1이라고 추정하기도 하였다.

그런데 놀라운 점은 그런 대규모 금융위기가 겨우 지난 몇 십 년 동안만 하더라도 여러 번 일어났다는 사실이다. 1987년 가을, 미국을 비롯한 여러 나라에서 주식가격이 폭락했던 날은 10월 19일, 월요일이었으므로 사람들은 지금도 그날을 ‘블랙 먼데이’라고 부르고 있다.

그런데 그게 마지막이 아니었다. 1998년에는 ‘롱텀 캐피탈 매니지먼트’라는 펀드 회사가 무너졌는데 사람들은 그런 일이 일어날 확률을 나타내려면 1이라는 숫자 뒤에 0이 무려 스물세 개나 붙은 엄청나게 큰 수가 필요하다고들 했다. 그보다 더 큰 위기가 오는 데 수만 년을 기다릴 필요는 없었다. 겨우 10년 뒤인 2008년이 되자 미국의 투자 은행 리먼브라더스가 파산했고 금융위기가 전 세계를 휩쓸었다.

수만 년, 수십만 년에 한 번 일어난다는 대형 사고가 이렇게 자주 일어나다니, 세계 금융시장에 중대한 변화가 생긴 걸까? 아니면 확률 계산에 무슨 심각한 문제가 있었던 걸까? 여러 이유 가운데 확률 계산에 문제가 있다고 생각한 사람들이 집중적으로 비판한 것이 금융상품의 가격이 정규분포를 따를 것이라는 가정이었다.

출처 : 삼성 디스플레이 뉴스룸 조재근 교수

