Modeling Loss Severity

A short course authored by the Actuarial Community

19 Jan 2021

Important Severity Distributions

Severity Distributions

- Severity - Amount, or size, of an individual loss or claim for an insured event
- We will consider continuous probability distributions in these overheads to model severity

Important Severity Distributions

Three important loss severity distributions:

- Gamma
  - Fits medium tail lines like physical damage auto and homeowners well
  - Member of "exponential family of distributions"
- Pareto
- GB2 - Generalized Beta of the Second Kind

Gamma Distribution

- Two positive parameters, α and θ
- Probability density function (pdf) is 0 for $x \leq 0$ and for $x > 0$

$$f(x) = \frac{(\frac{x}{\theta})^{\alpha} e^{-x/\theta}}{x\Gamma(\alpha)} = \frac{1}{\theta^{\alpha}\Gamma(\alpha)} x^{\alpha-1} e^{-x/\theta}$$

- $\Gamma(\cdot)$ is the gamma function, defined as

$$\Gamma(\alpha) = \int_0^{\infty} x^{\alpha-1} e^{-x} \, dx$$

- For a positive integer n, $\Gamma(n) = (n-1)!$
- For more general arguments, one needs to rely on numerical integration to evaluate $\Gamma(\cdot)$. Two main exceptions are:
  - For any $\alpha > 0$, $\Gamma(\alpha + 1) = \alpha\Gamma(\alpha)$
  - $\Gamma(0.5) = \sqrt{\pi}$
- If $\alpha = 1$, gamma distribution reduces to the exponential distribution

Gamma Moments

- Define the kth *raw moment* to be

$$E(X^k) = \int_0^\infty x^k f(x) dx$$

- Using a change of variable, $t = x/\theta$, we have

$$\begin{aligned} E(X^k) &= \frac{1}{\theta^\alpha \Gamma(\alpha)} \int_0^\infty x^{\alpha+k-1} \exp(-x/\theta)\ dx \\ &= \frac{\theta^{\alpha+k}}{\theta^\alpha \Gamma(\alpha)} \int_0^\infty t^{\alpha+k-1} \exp(-t) dt \\ &= \frac{\theta^k}{\Gamma(\alpha)} \Gamma(\alpha + k). \end{aligned}$$

- With $k = 1$, we have $E(X) = \theta \frac{\Gamma(\alpha+1)}{\Gamma(\alpha)} = \alpha\theta$
- $E(X^2) = \theta^2 \alpha(\alpha + 1)$, $\mathrm{Var}(X) = \theta^2 \alpha$, If k is an integer: $E(X^k) = \theta^k (\alpha + k - 1) \cdots \alpha$

Important Severity Distributions

Three important loss severity distributions:

- Gamma
- Pareto
  - Fits longer tail lines like injury liability in auto and workers' compensation well
  - Simple to work with analytically (hence can provide intuition as we develop theory and explain theory to others)
- GB2 - Generalized Beta of the Second Kind

Two-Parameter Pareto Distribution

- Two positive parameters α and θ. The pdf, for $x > 0$, is

$$\mathrm{f}(x) = \frac{\alpha\theta^{\alpha}}{(x+\theta)^{\alpha+1}}$$

and the moments are

$$E(X^k) = \frac{\theta^k k!}{(\alpha-1)\cdots(\alpha-k)}$$

- Unlike gamma, there is a simple expression for the cumulative distribution function (cdf)

$$\mathrm{F}(x) = \int_0^x \mathrm{f}(y)dy = 1 - \left(\frac{\theta}{x+\theta}\right)^{\alpha}$$

- It is easy to compute quantiles

Important Severity Distributions

Three important loss severity distributions:

- Gamma
- Pareto
- GB2 - Generalized Beta of the Second Kind
  - A four parameter distribution family, complex
  - Yet, many severity distributions can be expressed as a special case of this distribution (good for programming)
  - Some applications have been fit well by GB2 where others do not seem to work

GB2 - Generalized Beta of the Second Kind

- Distribution with **four positive parameters**: α, τ, γ, θ
- Pdf, for $x > 0$, is

$$f(x) = \frac{\Gamma(\alpha+\tau)}{\Gamma(\alpha)\Gamma(\tau)} \frac{\gamma\,(x/\theta)^{\gamma\tau}}{x\,[1+(x/\theta)^{\gamma}]^{\alpha+\tau}}$$

with moments

$$E(X^k) = \theta^k \frac{\Gamma(\tau+\frac{k}{\gamma})\Gamma(\alpha-\frac{k}{\gamma})}{\Gamma(\alpha)\Gamma(\tau)}.$$

GB2 Special Cases

GB2 captures many other distributions, either as special cases or as limiting results:

- Special Case: Pareto Distribution. Use GB2 distribution with $\gamma = \tau = 1$
- Limiting Case: Generalized Gamma Distribution

Replace θ by $\theta\tau^{1/\gamma}$

One can show that

$$\lim_{\tau \to \infty} f_{GB2}(x; \theta\tau^{1/\gamma}, \alpha, \tau, \gamma) = f(x),$$

a generalized gamma pdf

REVIEW

In this section, you learned how to define and apply three fundamental severity distributions:

- gamma,
- Pareto, and
- generalized beta distribution of the second kind.

Methods of Creating New Distributions: Random Variable Transformations

Creating Severity Distributions Using Transformations

- In this section, consider distributions that are created by transforming the random variable of a distribution:
  - Multiplication by a constant ($Y = cX$)
  - Raising to a power ($Y = X^{\tau}$)
  - Exponentiation ($Y = e^{X}$)
- In next section, consider ways of combining distributions to form a distribution of interest:
  - Mixing
  - Splicing

Multiplication by a Constant

- Multiplying a random variable by a positive constant
  - Think of X as this year's losses and assume that we have an 8% inflation rate. We can model next year's losses as $Y = 1.08X$
  - Can readily go from dollars to thousands of dollars ($c = 1/1000$) or from dollars to Euros
- More generally, let $Y = cX$ and use

$$
\begin{aligned}
F_Y(y) &= \Pr(Y \leq y) = \Pr\left(X \leq \frac{y}{c}\right) = F_X\left(\frac{y}{c}\right) \\
f_Y(y) &= \frac{1}{c}\, f_X\left(\frac{y}{c}\right)
\end{aligned}
$$

Scale Distributions

- In a scale distribution, the transformed variable $Y = cX$ has a distribution from the same family as the random variable X
- Many loss distributions are scale distributions
- Typically, one uses θ as the scale parameter

If X comes from a distribution with parameter θ, then $Y = cX$ has the same distribution with scale parameter $\theta^* = c\theta$

- Gamma distribution is an example of a scale distribution

Raising to a Power

- Consider $Y = X^{\tau}$. We examine three cases:

$$\begin{array}{rl} \tau > 0 & \textbf{transformed} \\ \tau = -1 & \textbf{inverse} \\ \tau < 0 & \textbf{inverse transformed} \end{array}$$

- *Special Case: Exponential Distribution*. Suppose that X has an exponential distribution with parameter θ^* and consider $Y = 1/X$
- Cdf of Y is

$$F_Y(y) = \Pr(Y \leq y) = \Pr(\frac{1}{X} \leq y) = \Pr(X \geq \frac{1}{y}) = \exp\left(-\frac{1}{y\theta^*}\right).$$

- Define a new parameter $\theta = \frac{1}{\theta^*}$. With this notation,

$$F_Y(y) = \Pr(Y \leq y) = \exp\left(-\frac{\theta}{y}\right).$$

- This is an inverse exponential distribution with parameter θ

Exponential to get a Weibull

Example: Transforming an Exponential to get a Weibull.

- Start with $X \sim$ exponential distribution with parameter 1. Define transformed random variable with positive parameters τ and θ:

$$Y = \theta X^{1/\tau}$$

- This has distribution

$$\begin{aligned} F_Y(y) &= \Pr(Y \leq y) \\ &= \Pr(X^{1/\tau} \leq \frac{y}{\theta}) = \Pr(X \leq \left(\frac{y}{\theta}\right)^{\tau}) \\ &= 1 - \exp\left(-\left(\frac{y}{\theta}\right)^{\tau}\right), \end{aligned}$$

known as a Weibull distribution

Exponentiation

- Another type of transformation involves exponentiating a random variable so that $Y = e^X$
- Develop the distribution of the new random variable through the cdf

$$F_Y(y) = \Pr(Y \leq y) = \Pr(e^X \leq y) = \Pr(X \leq \ln y) = F_X(\ln y)$$

and the pdf

$$f_Y(y) = \frac{1}{y} f_X(\ln y).$$

- If $X \sim$ normal, then $Y = e^X \sim$ a lognormal distribution

Methods of Creating New Distributions: Combining Distributions

Discrete Mixture Severity Distributions

- *Definition.* Let $X_1, \ldots, X_k$ be random variables and define

$$Y = \begin{cases} X_1 & \textit{with probability } \alpha_1 \\ \vdots & \vdots \\ X_k & \textit{with probability } \alpha_k \end{cases}$$

Here, $\alpha_j > 0$ and $\alpha_1 + \cdots + \alpha_k = 1$. Then, Y is a k-point mixture random variable

Cdf is

$$F_Y(y) = \alpha_1 F_{X_1}(y) + \cdots + \alpha_k F_{X_k}(y)$$

with mean

$$E(Y) = \alpha_1 E(X_1) + \cdots + \alpha_k E(X_k).$$

Discrete Mixture Severity Distributions

- *Example from Exam M Spring 05 #34*

The distribution of a loss, X, is a two-point mixture:

- With probability 0.8, X has a two-parameter Pareto distribution with $\alpha = 2$ and $\theta = 100$.
- With probability 0.2, X has a two-parameter Pareto distribution with $\alpha = 4$ and $\theta = 3000$.
- Calculate $Pr(X \leq 200)$.
- Calculate $E(X)$.

Continuous Mixtures for Severity

- Infinite number of subgroups within a population

Each subgroup has $\mathrm{F}(\cdot|\theta)$ (e.g., exponential) but with a parameter θ that accounts for population differences

- Assume the random variable Θ has pdf $f_\Theta(\theta)$
- Cdf:

$$\begin{aligned}\mathrm{F}_X(x) = \Pr(X \le x) &= E_\Theta[\Pr(X \le x|\Theta)] \\ &= \int \Pr(X \le x|\theta) f_\Theta(\theta) d\theta \\ &= \int \mathrm{F}_{X|\Theta}(x|\theta) f_\Theta(\theta) d\theta\end{aligned}$$

- Pdf:

$$f_X(x) = \int f_{X|\Theta}(x|\theta) f_\Theta(\theta) d\theta$$

Special Case: Gamma Mixtures of Exponentials

- Suppose $X|\Theta \sim \text{exponential}(\frac{1}{\Theta})$:

$$f_{X|\Theta}(x|\theta) = \theta e^{-\theta x}$$

- Suppose $\Theta \sim \text{gamma}(\alpha, \beta)$

$$f_{\Theta}(\theta) = \frac{1}{\Gamma(\alpha)\beta^{\alpha}} \theta^{\alpha-1} e^{-\theta/\beta}$$

- Pdf of X is

$$\begin{aligned} f_X(x) &= \int f_{X|\Theta}(x|\theta) f_{\Theta}(\theta) d\theta \\ &= \frac{1}{\Gamma(\alpha)\beta^{\alpha}} \int_0^{\infty} \theta^{\alpha} e^{-\theta(x+1/\beta)} d\theta = \frac{\alpha\beta}{(1+x\beta)^{\alpha+1}} \end{aligned}$$

- This is a Pareto distribution with parameters α and $\theta = 1/\beta$

Mixture Expectations

- Law of iterated expectation:

$$E(X) = E_{\Theta}[\mathrm{E}(X|\Theta)]$$

- This is easily extended to kth moment:

$$E(X^k) = E_{\Theta}[E(X^k|\Theta)]$$

Law of total variance:

$$Var(X) = E_{\Theta}[Var(X|\Theta)] + Var_{\Theta}[E(X|\Theta)]$$

Splicing

- Join (splice) together different probability density functions to form a pdf over support of a random variable

$$f_X(x) = \begin{cases} \alpha_1 f_1(x) & c_0 < x < c_1 \\ \alpha_2 f_2(x) & c_1 < x < c_2 \\ \vdots & \vdots \\ \alpha_k f_k(x) & c_{k-1} < x < c_k \end{cases}$$

$\alpha_1 + \alpha_2 \cdots + \alpha_k = 1$

Each f_j is a pdf, so that $\int_{c_{j-1}}^{c_j} f_j(x)dx = 1$

c_j's are typically known

REVIEW

In this section, you learned how to:

- Understand connections among the distributions
- Give insights into when a distribution is preferred when compared to alternatives
- Provide foundations for creating new distributions

Coverage Modifications: Deductibles and Limits

Risk Retention Framework

- Now consider the following framework:
  - Policyholder or insured suffers a loss of amount X
  - Under an insurance contract, the insurer is obligated to cover a portion of X, denoted as Y
  - Y represents the insurer's claim payment
- We introduce standard mechanisms that insurers use to reduce, or mitigate, their risk, including deductibles and policy limits
- Further, we examine how the distribution of the insurer's obligations depends on these mechanisms

Risk Retention Function

- Known risk retention function $g(\cdot)$ maps the amount insured to the amount retained by the insurer, that is, $Y = g(X)$
- **Special Case 1. Deductible (d)**

$$g(x) = (x-d)_+ = \begin{cases} 0 & x \leq d \\ x-d & x > d. \end{cases}$$

Notation "$(\cdot)_+$" means "Take the positive part of"

- **Special Case 2. Limit (u)**

$$g(x) = x \wedge u = \begin{cases} x & x \leq u \\ u & x > u. \end{cases}$$

Notation "$\wedge$" means "take the minimum of"

- **Special Case 3. Coinsurance**. Define $Y = g(X) = cX$. Typically, $0 < c < 1$, and so represents the proportion of claims retained by the insurer

Information Set for Deductibles

- Specify what type of information is available to the insurer
- **Special Case 4. Policyholder Deductible**. Define:

$$g_P(x) = \begin{cases} \text{undefined/not observed} & x \leq d \\ x - d & x > d \end{cases}$$

- Insurance only pays amounts in excess of the deductible d. If a loss is less than the deductible, the insurer does not observe the loss.
  - Random variable $Y^P = g_P(X)$ is the claim that an insurer observes
  - "P" subscript indicates that the retained loss is on a per payment basis
  - For case where a claim of zero is observed for losses $X \leq d$, terminology per loss is used.
  - Notation is $Y^L = (X - d)_+$

Distributions of Retained Risks - Deductible

- Consider two types of ordinary deductible:
- Cost (amount of payment) per loss event

$$Y^L = (X - d)_+ = \begin{cases} 0 & X \le d \\ X - d & X > d \end{cases}$$

- Cost (amount of payment) per payment event

$$Y^P = \begin{cases} \textit{undefined} & X \le d \\ X - d & X > d \end{cases}$$

Example. Exponential Distribution. Suppose that the loss X has cumulative distribution function $F(x) = 1 - \exp(-x/1000)$. Compute the cdf and pdf for Y^L and Y^P with $d = 250$

Coverage Modifications: Expectations of Retained Risks

Limited Expected Value

- Use a generic "u" for the upper limit. Expected value of limited loss variable $(X \wedge u)$ is

$$E(X \wedge u) = \int_0^u (1 - F(x))\, dx = \int_0^u S(x) dx.$$

Pareto Policy Limit. Recall

$$1 - F(x) = S(x) = \Pr(X > x) = \left(\frac{\theta}{x+\theta}\right)^{\alpha}$$

with mean $\mathrm{E}\,(X) = \frac{\theta}{\alpha-1}$. Thus, the limited expected value is

$$\begin{aligned}
E(X \wedge u) &= \theta^{\alpha} \int_0^u (x+\theta)^{-\alpha} dx = \theta^{\alpha} \left. \frac{(x+\theta)^{-\alpha+1}}{-\alpha+1} \right|_0^u \\
&= \theta^{\alpha} \left(\frac{\theta^{-\alpha+1} - (u+\theta)^{-\alpha+1}}{\alpha - 1} \right) \\
&= \frac{\theta}{\alpha - 1} \left\{ 1 - \left(\frac{\theta}{u+\theta} \right)^{\alpha-1} \right\}.
\end{aligned}$$

Pareto Deductible

- Claim amount on a per loss basis is $Y^L = (X-d)_+$ for a deductible d
- To calculate $E(X-d)_+$, use $X \wedge d + (X-d)_+ = X$
- For the Pareto distribution, recall $E(X) = \frac{\theta}{\alpha-1}$ and

$$E(X \wedge d) = \frac{\theta}{\alpha-1}\left\{1 - \left(\frac{\theta}{d+\theta}\right)^{\alpha-1}\right\}.$$

Thus,

$$\begin{aligned} E(X-d)_+ &= E(X) - E(X \wedge d) = \frac{\theta}{\alpha-1} - \frac{\theta}{\alpha-1}\left\{1 - \left(\frac{\theta}{d+\theta}\right)^{\alpha-1}\right\} \\ &= \frac{\theta}{\alpha-1}\left\{\left(\frac{\theta}{d+\theta}\right)^{\alpha-1}\right\}. \end{aligned}$$

Mean Excess Loss

- For the per payment random variable associated with the policyholder deductible case,

$$g_P(x) = \begin{cases} \text{undefined/not observed} & x \leq d \\ x - d & x > d \end{cases}$$

we can calculate the expectation as

$$e_X(d) = e(d) = \mathrm{E}\left(X - d | X > d\right)$$

- $e_X(d)$ is the mean excess loss that We can write this as

$$\begin{aligned} e(d) &= E(X - d|X > d) \\ &= \frac{\int_d^\infty (x-d)f(x)dx}{S(d)} \\ &= \frac{E(X-d)_+}{S(d)} \\ &= \frac{\int_d^\infty S(x)dx}{S(d)} \end{aligned}$$

Example. Exam M Fall 2005, Exercise 26

For an insurance:

- Losses have the density function

$$f_X(x) = \begin{cases} 0.02x & 0 < x < 10 \\ 0 & \text{elsewhere} \end{cases}$$

- Insurance has an ordinary deductible of 4 per loss
- Y^P is the claim payment per payment random variable

Calculate $E[Y^P]$

Summary of Limited Loss Variables

Random Variable	**Expectation**
Excess loss random variable $Y = X - d$ if $X > d$ left truncated	$e_X(d) = E(Y) = \mathrm{E}(X - d \mid X > d)$ mean excess loss function mean residual life function complete expectation of life $e_X^k(d) = E\left[(X-d)^k \mid X > d\right]$
$(X-d)_+ = \begin{cases} 0 & X \le d \\ X - d & X > d \end{cases}$ left-censored and shifted variable	$E\left(X-d\right)_+ = e(d)S(d)$ $E\left(X-d\right)_+^k = e^k(d)S(d)$
$min(X,d) = X \wedge d = \begin{cases} X & X \le d \\ d & X > d \end{cases}$ limited loss variable	$E\left(X \wedge d\right) -$ limited expected value right censored

Note that $(X-d)_+ + (X \wedge d) = X$. Thus, $E\left(X-d\right)_+ + E\left(X \wedge d\right) = E(X)$

For nonnegative, continuous random variables,

$$E\left(X \wedge d\right) = \int_0^d S\left(x\right) dx \quad \text{and} \quad E(X-d)_+ = \int_d^\infty S\left(x\right) dx$$

LER and More Risk Retention

Loss Elimination Ratio (LER)

- Consider an ordinary deductible, cost (amount of payment) per loss event
- Loss elimination ratio at deductible d is

$$
\begin{aligned}
LER &= \frac{\mathrm{E}\left(X \wedge d\right)}{E(X)} \\
&= \frac{\text{limited exp value}}{\text{exp value}}
\end{aligned}
$$

What fraction of the losses have been eliminated by introducing the deductible?

Example. Losses have a lognormal distribution with $\mu = 6$ and $\sigma = 2$. There is a deductible of 2,000

Determine the loss elimination ratio

Risk Retention Function II

- Combining three special cases of coverage modifications (deductible, limit, coinsurance) results in

$$g(x) = \begin{cases} 0 & x \leq d \\ c(x-d) & d < x \leq u \\ c(u-d) & x > u. \end{cases}$$

- Think about these as parameters in a contract between a policyholder and an insurer and so represent modifications of the underlying contract
- Note: $g(X) = c\left((X \wedge u) - (X \wedge d)\right)$

Estimating Severity Distributions

Maximum Likelihood Estimation

- Let $\mathrm{f}(\cdot;\boldsymbol{\theta})$ be pmf if X is discrete or pdf if continuous
- Define the likelihood function,

$$L(\boldsymbol{\theta}) = L(\mathbf{x};\boldsymbol{\theta}) = \prod_{i=1}^{n} \mathrm{f}(x_i;\boldsymbol{\theta}),$$

- Define the log-likelihood function,

$$l(\boldsymbol{\theta}) = l(\mathbf{x};\boldsymbol{\theta}) = \ln L(\boldsymbol{\theta}) = \sum_{i=1}^{n} \ln \mathrm{f}(x_i;\boldsymbol{\theta}),$$

- Value of $\boldsymbol{\theta}$, say $\widehat{\theta}_{\mathbf{MLE}}$, that maximizes $L(\boldsymbol{\theta})$, or equivalently $l(\boldsymbol{\theta})$, is the maximum likelihood estimate (mle) of θ

Example: Single-Parameter Pareto

- Suppose $X_1, \dots, X_n$ represent a random sample from a single-parameter Pareto distribution with cdf:

$$\mathrm{F}(x) = 1 - \left(\frac{500}{x}\right)^{\alpha}, \quad x > 500$$

- There is a single parameter of $\boldsymbol{\theta} = \alpha$
- Corresponding pdf is $\mathrm{f}(x) = 500^{\alpha}\alpha x^{-\alpha-1}$
- Log-likelihood function is

$$l(\boldsymbol{\alpha}) = \sum_{i=1}^{n} \ln \mathrm{f}(x_i; \alpha) = n\alpha \ln 500 + n \ln \alpha - (\alpha + 1) \sum_{i=1}^{n} \ln x_i.$$

Asymptotic Normality of MLE

- Consider a distribution (X) with pmf or pdf $f(\cdot; \boldsymbol{\theta})$
- There is only one estimable parameter: $\boldsymbol{\theta} = \theta$
- Theorem: Under mild regularity conditions, as the sample size n approaches infinity, the distribution of the maximum likelihood estimator of θ, $\hat{\theta}$, converges to a normal distribution with mean θ and variance equal to the inverse of the Fisher Information, $I(\theta)$, where:

$$I(\theta) = \; -\mathrm{E}_X\left[\frac{\partial^2}{\partial\theta^2} l(\theta)\right]$$

If all observations (X) are identically distributed:

$$I(\theta) = -n\,\mathrm{E}_X\left[\frac{\partial^2}{\partial\theta^2}\ln(f(X;\theta))\right]$$

Delta Method

- Consider a distribution (X) with pmf or pdf $f(\cdot;\boldsymbol{\theta})$
- There is only one estimable parameter: $\boldsymbol{\theta} = \theta$
- From the previous slide, as $n \to \infty$:

$$\hat{\theta} \sim N\left(\mu = \theta, \sigma^2 = [I(\theta)]^{-1}\right)$$

- Delta Method: Consider a function of θ, $g(\theta)$

$g(\hat{\theta})$ is the maximum likelihood estimator of $g(\theta)$

As $n \to \infty$:

$$g(\hat{\theta}) \sim N\left(\mu = g(\theta), \sigma^2 = \left(\frac{\partial g}{\partial \theta}\right)^2 [I(\theta)]^{-1}\right)$$

REVIEW

In this section, you learned how to:

- Define a likelihood for a sample of observations from a continuous distribution
- Define the maximum likelihood estimator for a random sample of observations from a continuous distribution
- Estimate parametric distributions based on grouped, censored, and truncated data