SCMA470 Risk Analysis and Credibility

Pairote Satiracoo

2021-09-14

Contents

Chapter 1

Basic Probability Concepts

1.1 Random Variables

Definition 1.1. *Let S be the sample space of an experiment. A real-valued function $X : S \to \mathbb{R}$ is called a **random variable** of the experiment if, for each interval $I \subset \mathbb{R}$, $\{s : X(s) \in I\}$ is an event.*

Random variables are often used for the calculation of the probabilities of events. The real-valued function $P(X \leq t)$ characterizes X, it tells us almost everything about X. This function is called the **cumulative distribution function** of X. The cumulative distribution function describes how the probabilities accumulate.

Definition 1.2. *If X is a random variable, then the function F defined on $\mathbb{R}$ by*

$$F(x) = P(X \leq x)$$

*is called the **cumulative distribution function** or simply **distribution function (c.d.f)** of X.*

Functions that define the probability measure for discrete and continuous random variables are the probability mass function and the probability density function.

Definition 1.3. *Suppose X is a discrete random variable. Then the function*

$$f(x) = P(X = x)$$

*that is defined for each x in the range of X is called the **probability mass function** (p.m.f) of a random variable X.*

Definition 1.4. *Suppose X is a continuous random variable with c.d.f F and there exists a nonnegative, integrable function f, $f : \mathbb{R} \to [0, \infty)$ such that*

$$F(x) = \int_{-\infty}^{x} f(y)\, dy$$

*Then the function f is called the **probability density function** (p.d.f) of a random variable X.*

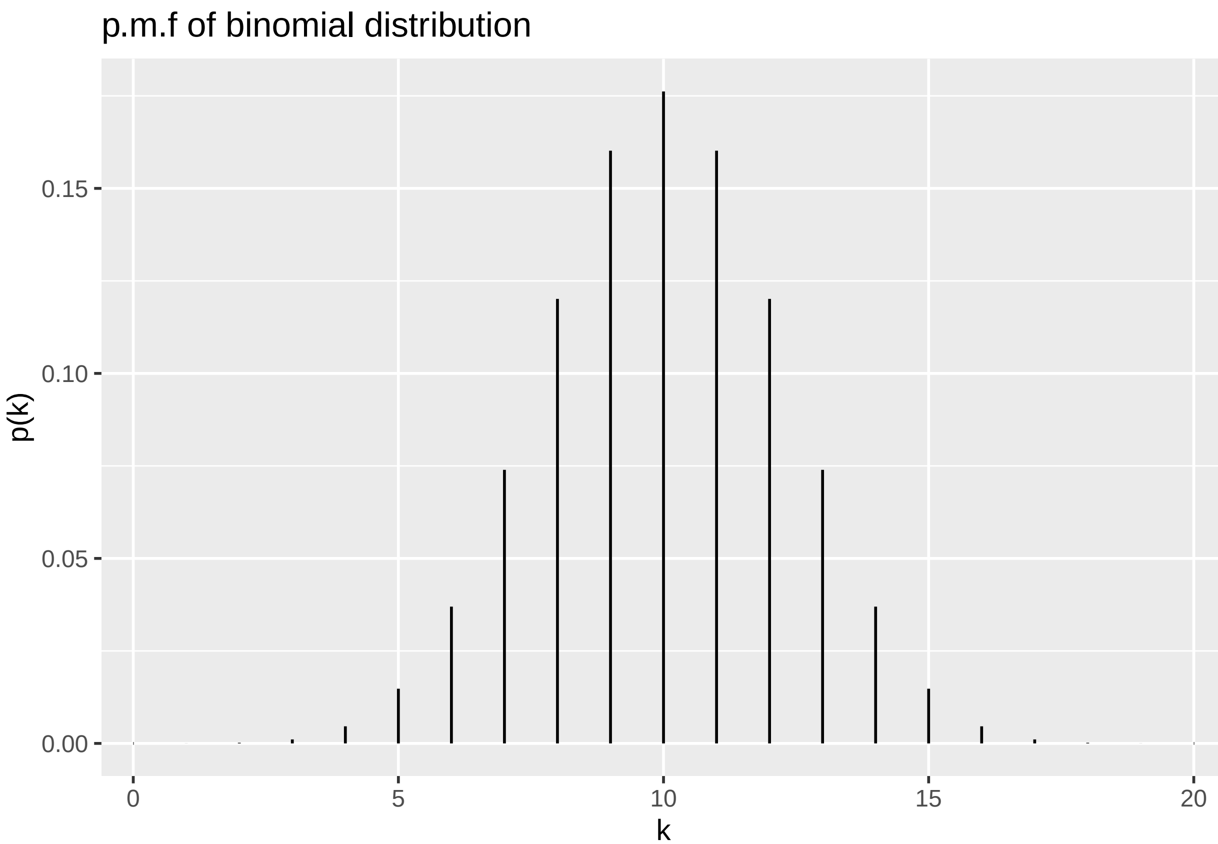
1.1.1 R Functions for Probability Distributions

In R, density, distribution function, for the Poisson distribution with parameter λ is shown as follows:

Distribution	Density function: $P(X = x)$	Distribution function: $P(X \leq x)$	Quantile function (inverse c.d.f.)	random generation
Poisson	`dpois(x, lambda, log = FALSE)`	`ppois(q, lambda, lower.tail = TRUE, log.p = FALSE)`	`qpois(p, lambda, lower.tail = TRUE, log.p = FALSE)`	`rpois(n, lambda)`

For the binomial distribution, these functions are pbinom, qbinom, dbinom, and rbinom. For the normal distribution, these functions are pnorm, qnorm, dnorm, and rnorm. And so forth.

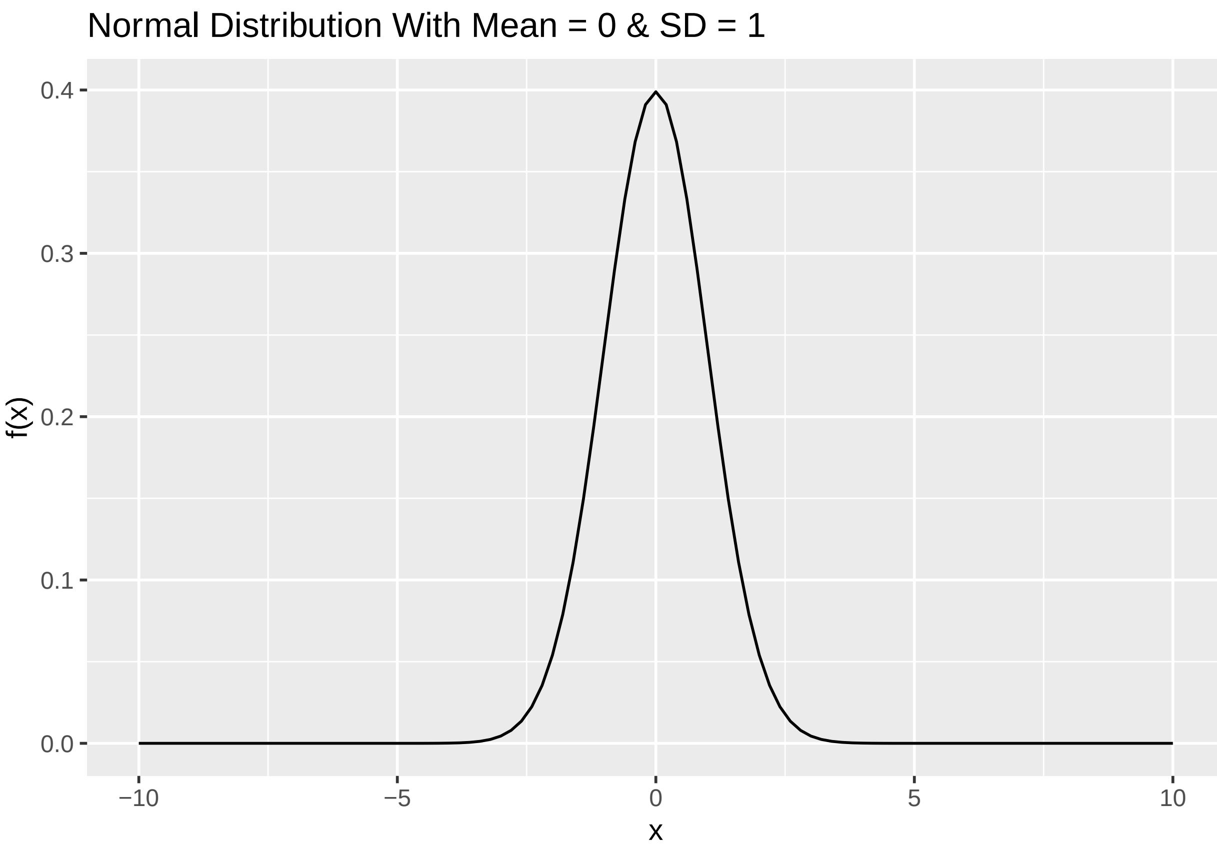
```
library(ggplot2)
x <- 0:20
myData <- data.frame( k = factor(x), pK = dbinom(x, 20, .5))
ggplot(myData,aes(k,ymin=0,ymax=pK)) +
  geom_linerange() + ylab("p(k)") +
  scale_x_discrete(breaks=seq(0,20,5)) +
  ggtitle("p.m.f of binomial distribution")
```



To plot continuous probability distribution in R, we use stat_function to add the density function as its arguement. To specify a different mean or standard deviation, we use the **args** parameter to supply new values.

```
library(ggplot2)
df <- data.frame(x=seq(-10,10,by=0.1))
ggplot(df) +
    stat_function(aes(x),fun=dnorm, args = list(mean = 0, sd = 1))  +
    labs(x = "x", y = "f(x)",
```

```
    title = "Normal Distribution With Mean = 0 & SD = 1")
```



1.2 Expectation

Definition 1.5. *The* ***expected value*** *of a discrete random variable X with the set of possible values A and probability mass function $f(x)$ is defined by*

$$\mathrm{E}(X) = \sum_{x \in A} x f(x)$$

The **expected value** of a random variable X is also called the mean, or the mathematical expectation, or simply the expectation of X. It is also occasionally denoted by $\mathrm{E}[X]$, μ_X, or μ.

Note that if each value x of X is weighted by $f(x) = P(X = x)$, then $\sum_{x \in A} x f(x)$ is nothing but the weighted average of X.

Theorem 1.1. Let X be a discrete random variable with set of possible values A and probability mass function $f(x)$, and let g be a real-valued function. Then $g(X)$ is a random variable with

$$\mathrm{E}[g(X)] = \sum_{x \in A} g(x) f(x)$$

Definition 1.6. *If X is a continuous random variable with probability density function f , the* ***expected value*** *of X is defined by*

$$\mathrm{E}(X) = \int_{-\infty}^{\infty} x f(x)\, dx$$

Theorem 1.2. • *Let X be a continuous random variable with probability density function $f(x)$; then for any function $h : \mathbb{R} \to \mathbb{R}$,*

$$\mathrm{E}[h(X)] = \int_{-\infty}^{\infty} h(x)\, f(x)\, dx$$

*

Theorem 1.3. Let X be a random variable. Let $h_1, h_2, ..., h_n$ be real-valued functions, and $a_1, a_2, \dots, a_n$ be real numbers. Then

$$\mathrm{E}[a_1 h_1(X) + a_2 h_2(X) + \cdots + a_n h_n(X)] = a_1 \mathrm{E}[h_1(X)] + a_2 \mathrm{E}[h_2(X)] + ... + a_n \mathrm{E}[h_n(X)]$$

Moreover, if a and b are constants, then

$$\mathrm{E}(aX + b) = a\mathrm{E}(x) + b$$

1.3 Variances of Random Variables

Definition 1.7. *Let X be a discrete random variable with a set of possible values A, probability mass function $f(x)$, and $\mathrm{E}(X) = \mu$. then $\mathrm{Var}(X)$ and σ_X, called the **variance** and **standard deviation** of X, respectively, are defined by*

$$\mathrm{Var}(X) = \mathrm{E}[(X-\mu)^2] = \sum_{x \in A} (x-\mu)^2 f(x),$$

$$\sigma_X = \sqrt{\mathrm{E}[(X-\mu)^2]}$$

Definition 1.8. *If X is a continuous random variable with $\mathrm{E}(X) = \mu$, then $\mathrm{Var}(X)$ and σ_X, called the **variance** and **standard deviation** of X, respectively, are defined by*

$$\mathrm{Var}(X) = \mathrm{E}[(X-\mu)^2] = \int_{-\infty}^{\infty} (x-\mu)^2\, f(x)\, dx,$$

$$\sigma_X = \sqrt{\mathrm{E}[(X-\mu)^2]}$$

We have the following important relations

$$\mathrm{Var}(x) = \mathrm{E}(X^2) - (\mathrm{E}(x))^2,$$

$$\mathrm{Var}(aX + b) = a^2\ Var(X), \quad \sigma_{aX+b} = |a| \sigma_X$$

where a and b are constants.

1.4 Moments and Moment Generating Function

Definition 1.9. *For $r > 0$, the rth moment of X (the rth moment about the origin) is $\mathrm{E}[X^r]$, when it is defined. The rth central moment of a random variable X (the rth moment about the mean) is $\mathrm{E}[(X - \mathrm{E}[X])^r]$.*

Definition 1.10. *The skewness of X is defined to be the third central moment,*

$$\mathrm{E}[(X - \mathrm{E}[X])^3],$$

and the coefficient of skewness to be given by

$$\frac{\mathrm{E}[(X - \mathrm{E}[X])^3]}{(\mathrm{Var}[X])^{3/2}}.$$

Definition 1.11. *The coefficient of kurtosis of X is defined by*

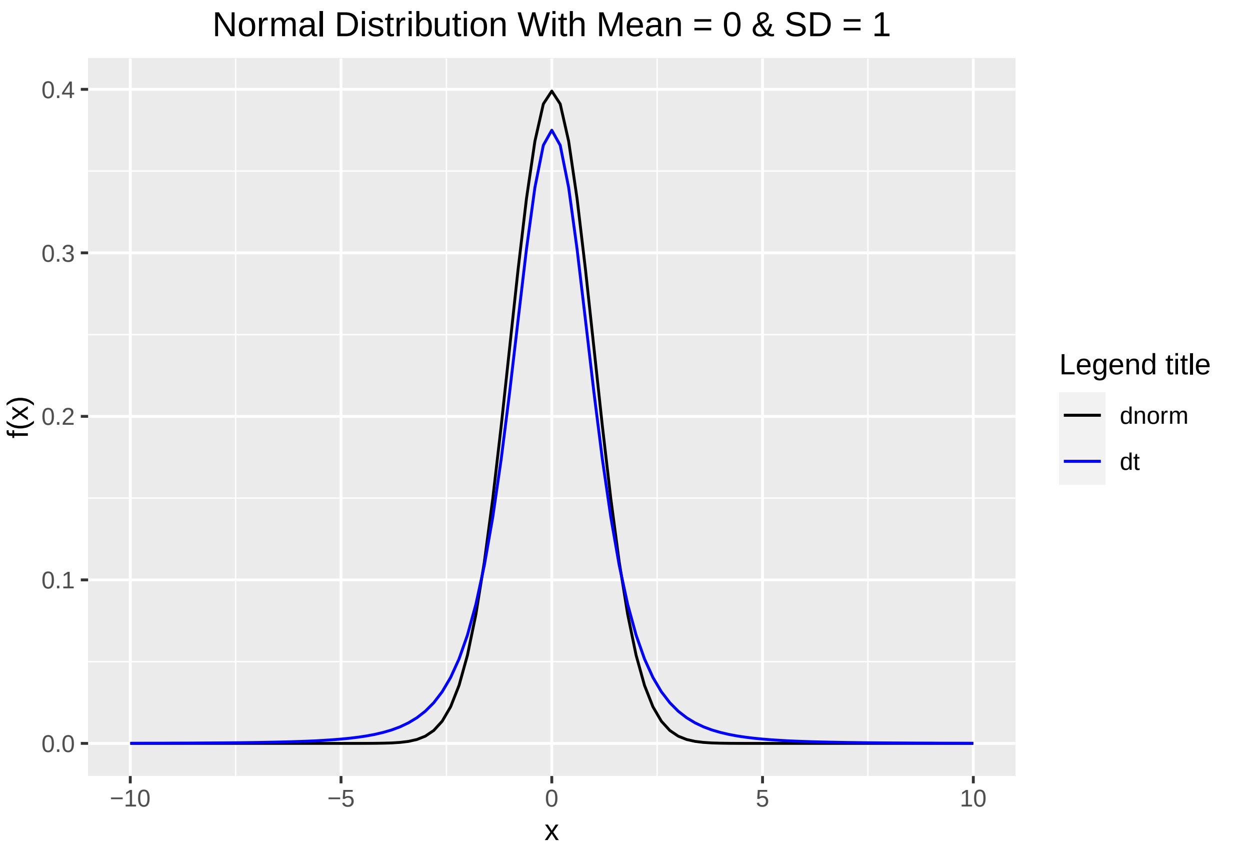
$$\frac{\mathrm{E}[(X - \mathrm{E}[X])^4]}{(\mathrm{Var}[X])^{4/2}}.$$

Note In the formula, subtract from the mean and normalise or divide by the standard deviation center and scale to the standard values. Odd-order moments are increased if there is a long tail to the right and decreased if there is a long tail to the left, while even-order moments are increased if either tail is long. A negative value of the coefficient of skewness that the distribution is skewed to the left, or negatively skewed, meaning that the deviations above the mean tend to be smaller than the deviations below the mean, and vice versa. If the coefficent of skewness is close to zero, this could mean symmetry,

Note The fourth moment measures the fatness in the tails, which is always positive. The kurtosis of the standard normal distribution is 3. Using the standard normal distribution as a benchmark, the excess kurtosis of a random variable is defined as the kurtosis minus 3. A higher kurtosis corresponds to a larger extremity of deviations (or outliers), which is called excess kurtosis.

The following diagram compares the shape between the normal distribution and Student's t-distribution. Note that to use the legend with the `stat_function` in ggplot2, we use `scale_colour_manual` along with `colour =` inside the `aes()` as shown below and give names for specific density plots.

```
library(ggplot2)
df <- data.frame(x=seq(-10,10,by=0.1))
ggplot(df) +
    stat_function(aes(x, colour = "dnorm"),fun = dnorm, args = list(mean = 0, sd = 1))  +
    stat_function(aes(x, colour ="dt"),fun = dt, args = list(df = 4)) +
     scale_colour_manual("Legend title", values = c("black", "blue")) +
    labs(x = "x", y = "f(x)",
       title = "Normal Distribution With Mean = 0 & SD = 1") +
  theme(plot.title = element_text(hjust = 0.5))
```



Next we will simulate 10000 samples from a normal distribution with mean 0, and standard deviation 1, then compute and interpret for the skewness and kurtosis, and plot the histogram. Here we also use the function `set.seed()` to set the seed of R's random number generator, this is useful for creating simulations or random objects that can be reproduced.

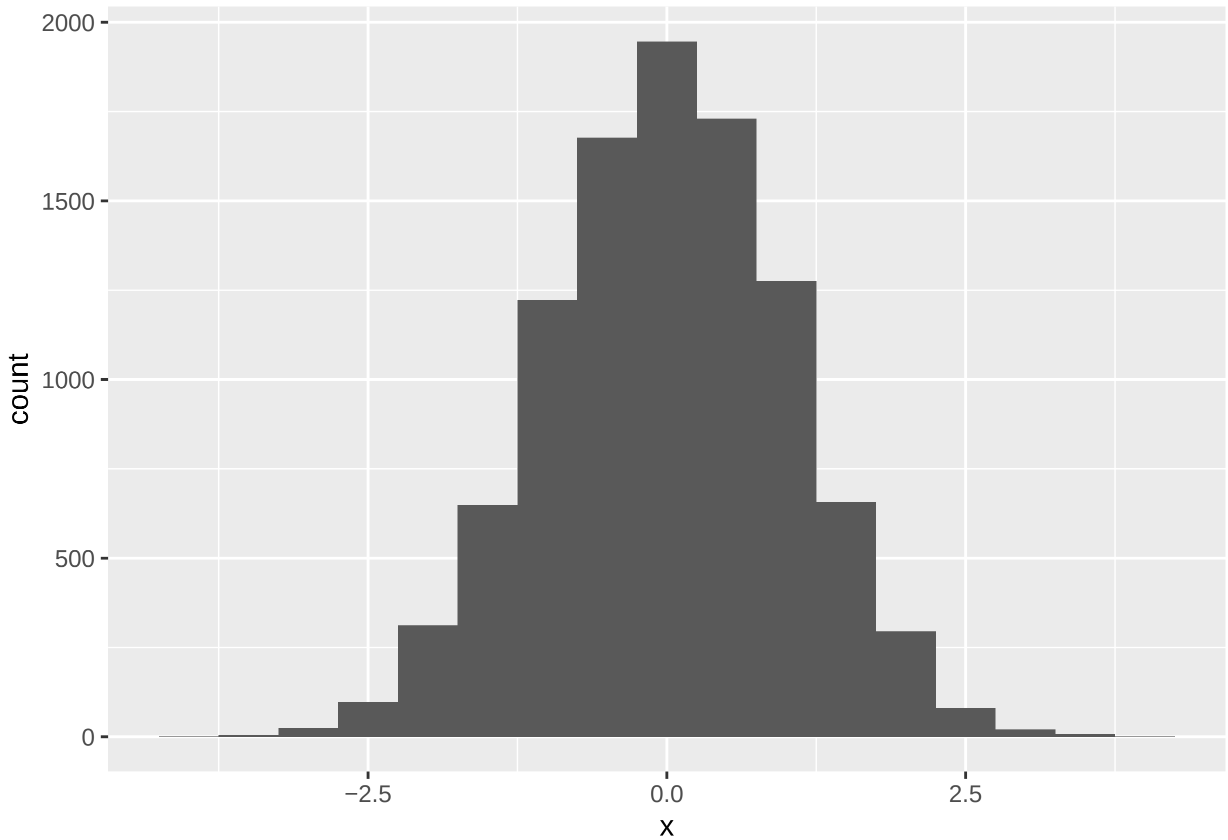
```
set.seed(15) # Set the seed of R's random number generator

#Simulation
n.sample <- rnorm(n = 10000, mean = 0, sd = 1)

#Skewness and Kurtosis
library(moments)
skewness(n.sample)
```

```
## [1] -0.03585812
```

```
kurtosis(n.sample)
```

```
## [1] 2.963189
```

```
ggplot(data.frame(x = n.sample),aes(x)) +
  geom_histogram(binwidth = 0.5)
```



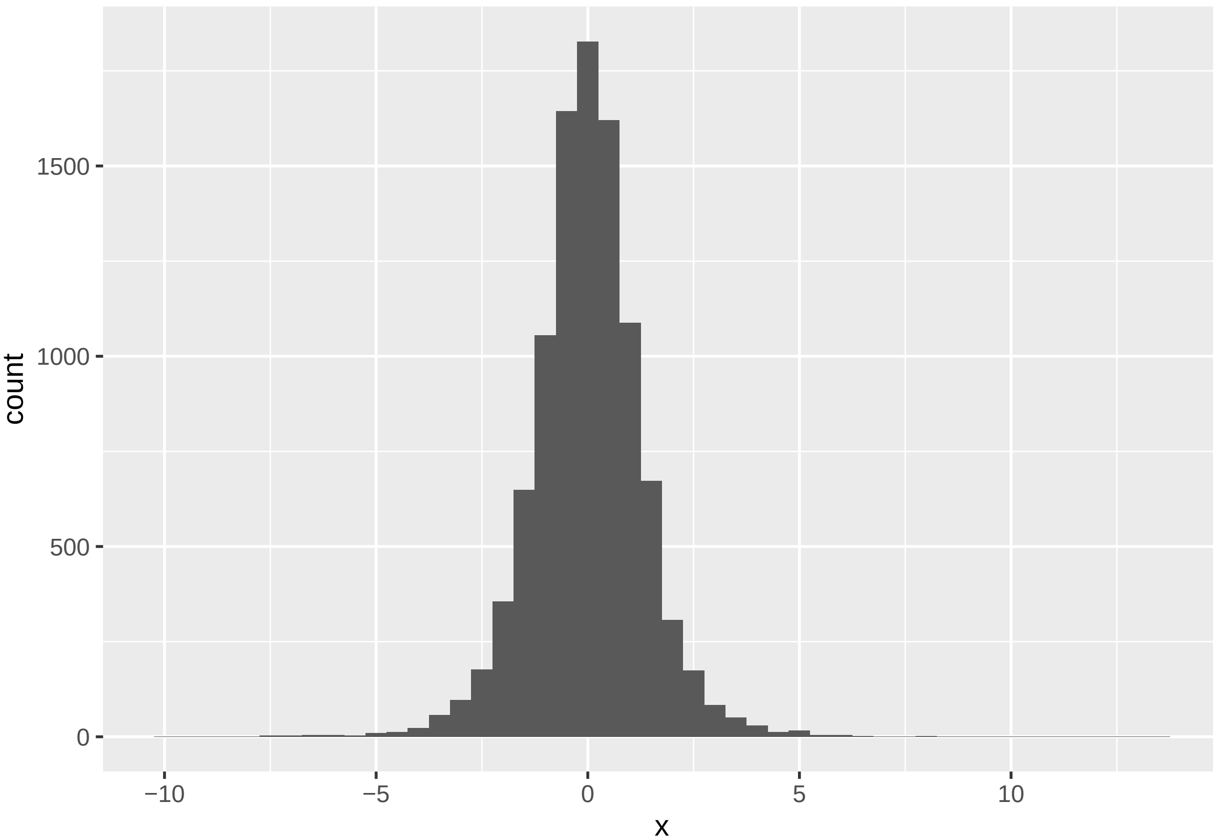

```r
set.seed(15)

#Simulation
t.sample <- rt(n = 10000, df = 5)

#Skewness and Kurtosis
library(moments)
skewness(t.sample)
```

```
## [1] 0.06196269
```

```r
kurtosis(t.sample)
```

```
## [1] 7.646659
```

```r
ggplot(data.frame(x = t.sample),aes(x)) + geom_histogram(binwidth = 0.5)
```



Example Let us count the number of samples greater than 5 from the samples of the normal and Student's t distributions. Comment on your results

eyJsYW5ndWFnZSI6InIiLCJzYW1wbGUiOiIjIFdyaXRlIHlvdXIgY29kZSBoZXJlXG5zZXQuc2VlZCgxNSlcbm4uc2FtcGxlII

Definition 1.12. *The moment generating function (mgf) of a random variable X is defined to be*

$$M_X(t) = E[e^{tX}],$$

if the expectation exists.

Note The moment generating function of X may not defined (may not be finite) for all t in $\mathbb{R}$.

If $M_X(t)$ is finite for $|t| < h$ for some $h > 0$, then, for any $k = 1, 2, \dots$, the function $M_X(t)$ is k-times differentiable at $t = 0$, with

$$M_X^{(k)}(0) = \mathrm{E}[X^k],$$

with $\mathrm{E}[|X|^k]$ finite. We can obtain the moments by succesive differentiation of $M_X(t)$ and letting $t = 0$.

Example 1.1. Derive the formula for the mgf of the standard normal distribution. Hint: its mgf is $e^{\frac{1}{2}t^2}$.

1.5 Probability generating function

Definition 1.13. *For a counting variable N (a variable which assumes some or all of the values $0, 1, 2, \dots$, but no others), The probability generating function of N is*

$$G_N(t) = E[t^N],$$

for those t in $\mathbb{R}$ for which the series converges absolutely.

Let $p_k = P(N = k)$. Then

$$G_N(t) = E[t^N] = \sum_{k=0}^{\infty} t^k p_k.$$

It can be shown that if $E[N] < \infty$ then

$$\mathrm{E}[N] = G'_N(1),$$

and if $E[N^2] < \infty$ then

$$\mathrm{Var}[N] = G''_N(1) + G'_N(1) - (G'_N(1))^2.$$

Moreover, when both pgf and mgf of N are defined, we have

$$G_N(t) = M_N(\log(t)) \quad \text{and } M_N(t) = G_N(e^t).$$

1.6 Multivariate Distributions

When $X_1, X_2, \ldots, X_n$ be random variables defined on the same sample space, a multivariate probability density function or probability mass function
$f(x_1, x_2, \ldots x_n)$ can be defined. The following definitions can be extended to more than two random variables and the case of discrete random variables.

Definition 1.14. *Two random variables X and Y, defined on the same sample space, have a continuous joint distribution if there exists a nonnegative function of two variables, $f(x, y)$ on $\mathbb{R} \times \mathbb{R}$, such that for any region R in the xy-plane that can be formed from rectangles by a countable number of set operations,*

$$P((X, Y) \in R) = \iint_R f(x, y)\, dx\, dy$$

The function $f(x, y)$ is called the **joint probability density function** of X and Y.

Let X and Y have joint probability density function $f(x, y)$. Let f_Y be the probability density function of Y. To find f_Y in terms of f, note that, on the one hand, for any subset B of R,

$$P(Y \in B) = \int_B f_Y(y)\, dy,$$

and on the other hand, we also have

$$P(Y \in B) = P(X \in (-\infty, \infty), Y \in B) = \int_B \left(\int_{-\infty}^{\infty} f(x, y)\, dx \right) dy.$$

We have

$$f_Y(y) = \int_{-\infty}^{\infty} f(x, y)\, dx \tag{1.1}$$

and

$$f_X(x) = \int_{-\infty}^{\infty} f(x, y)\, dy \tag{1.2}$$

Definition 1.15. *Let X and Y have joint probability density function $f(x, y)$; then the functions f_X and f_Y in* (1.1) and (1.2) *are called, respectively, the* ***marginal probability density functions*** *of X and Y.*

Let X and Y be two random variables (discrete, continuous, or mixed). The **joint probability distribution function**, or **joint cumulative probability distribution function**, or simply the joint distribution of X and Y, is defined by

$$F(t,u) = P(X \leq t, Y \leq u)$$

for all $t, u \in (-\infty, \infty)$.

The marginal probability distribution function of X, F_X, can be found from F as follows:

$$F_X(t) = \lim_{n\to\infty} F(t,u) = F(t,\infty)$$

and

$$F_Y(u) = \lim_{n\to\infty} F(t,u) = F(\infty,u)$$

The following relationship between $f(x,y)$ and $F(t,u)$ is as follows:

$$F(t,u) = \int_{-\infty}^{u}\int_{-\infty}^{t} f(x,y)\,dx\,dy.$$

We also have

$$\mathrm{E}(X) = \int_{-\infty}^{\infty} x f_X(x)\,dx, \quad \mathrm{E}(Y) = \int_{-\infty}^{\infty} y f_Y(y)\,dy$$

Theorem 1.4. Let $f(x,y)$ be the joint probability density function of random variables X and Y. If h is a function of two variables from $\mathbb{R}^2$ to $\mathbb{R}$, then $h(X,Y)$ is a random variable with the expected value given by

$$\mathrm{E}[h(X,Y)] = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} h(x,y)\,f(x,y)\,dx\,dy$$

provided that the integral is absolutely convergent.

As a consequence of the above theorem, for random variables X and Y,

$$\mathrm{E}(X+Y) = \mathrm{E}(X) + \mathrm{E}(Y)$$

1.7 Independent random variables

Definition 1.16. *Two random variables X and Y are called independent if, for arbitrary subsets A and B of real numbers, the events $\{X \in A\}$ and $\{Y \in B\}$ are **independent**, that is, if*

$$P(X \in A, Y \in B) = P(X \in A)P(Y \in B).$$

Theorem 1.5. Let X and Y be two random variables defined on the same sample space. If F is the joint probability distribution function of X and Y, then X and Y are independent if and only if for all real numbers t and u,

$$F(t,u) = F_X(t)F_Y(u).$$

Theorem 1.6. Let X and Y be jointly continuous random variables with joint probability density function $f(x,y)$. Then X and Y are independent if and only if

$$f(x,y) = f_X(x)f_Y(y).$$

Theorem 1.7. Let X and Y be independent random variables and $g : \mathbb{R} \to \mathbb{R}$ and $h : \mathbb{R} \to \mathbb{R}$ be real-valued functions; then $g(X)$ and $h(Y)$ are also independent random variables.

As a consequence of the above theorem, we obtain

Theorem 1.8. Let X and Y be independent random variables. Then for all real-value functions $g : \mathbb{R} \to \mathbb{R}$ and $h : \mathbb{R} \to \mathbb{R}$,

$$\mathrm{E}[g(X)h(Y)] = \mathrm{E}[g(X)]\mathrm{E}[h(Y)]$$

1.8 Conditional Distributions

Let X and Y be two continuous random variables with the joint probability density function $f(x, y)$. Note that the case of discrete random variables can be considered in the same way. When no information is given about the value of Y, the marginal probability density function of X, $f_X(x)$ is used to calculate the probabilities of events concerning X. However, when the value of Y is known, to find such probabilities, $f_{X|Y}(x|y)$, the conditional probability density function of X given that $Y = y$ is used and is defined as follows:

$$f_{X|Y}(x|y) = \frac{f(x, y)}{f_Y(y)}$$

provided that $f_Y(y) > 0$. Note also that the conditional probability density function of X given that $Y = y$ is itseof a probability density function, i.e.

$$\int_{-\infty}^{\infty} f_{X|Y}(x|y)\, dx = 1.$$

Note that the conditional probability distribution function of X given that $Y = y$, the conditional expectation of X given that $Y = y$ can be as follows:

$$F_{Y|X}(x|y) = P(X \leq x|Y = y) = \int_{-\infty}^{x} f_{X|Y}(t|y)\, dt$$

and

$$\mathrm{E}(X|Y = y) = \int_{-\infty}^{\infty} x f_{X|Y}(x|y)\, dx,$$

where $f_Y(y) > 0$.

Note that if X and Y are independent, then $f_{X|Y}$ coincides with f_X because

$$f_{X|Y}(x|y) = \frac{f(x, y)}{f_Y(y)} = \frac{f_X(x)f_Y(y)}{f_Y(y)} = f_X(x).$$

1.9 Covariance

The notion of the variance of a random variable X, $\mathrm{Var}(X) = \mathrm{E}[(X - \mathrm{E}(X))^2]$ measures the average magnitude of the fluctuations of the random variable X from its expectation, $\mathrm{E}(X)$. This quantity measures the dispersion, or spread, of the distribution of X about its expectation. Now suppose that X and Y are two jointly distributed random variables. Covariance is a measure of how much two random variables vary together.

Let us calculuate $\mathrm{Var}(aX+bY)$ the joint spread, or dispersion, of X and Y along the $(ax+by)$-direction for arbitrary real numbers a and b:

$$\mathrm{Var}(aX+bY) = a^2\mathrm{Var}(X) + b^2\mathrm{Var}(Y) + 2ab\mathrm{E}[(X-\mathrm{E}(X))(Y-\mathrm{E}(Y))].$$

However, $\mathrm{Var}(X)$ and $\mathrm{Var}(Y)$ determine the dispersions of X and Y independently; therefore, $\mathrm{E}[(X-\mathrm{E}(X))(Y-\mathrm{E}(Y))]$ is the quantity that gives information about the joint spread, or dispersion, X and Y .

Definition 1.17. *Let X and Y be jointly distributed random variables; then the **covariance** of X and Y is defined by*

$$\mathrm{Cov}(X,Y) = \mathrm{E}[(X-\mathrm{E}(X))(Y-\mathrm{E}(Y))].$$

Note that for random variables X, Y and Z, and $ab > 0$, then the joint dispersion of X and Y along the $(ax+by)$-direction is greater than the joint dispersion of X and Z along the $(ax+bz)$-direction if and only if $\mathrm{Cov}(X,Y) > \mathrm{Cov}(X,Z)$.

Note that

$$\mathrm{Cov}(X,X) = \mathrm{Var}(X).$$

Moreover,

$$\mathrm{Cov}(X,Y) = \mathrm{E}(XY) - \mathrm{E}(X)\mathrm{E}(Y).$$

Properties of covariance are as follows: for arbitrary real numbers a, b, c, d and random variables X and Y,

$$\mathrm{Var}(aX+bY) = a^2\mathrm{Var}(X) + b^2\mathrm{Var}(Y) + 2ab\mathrm{Cov}(X,Y).$$

$$\mathrm{Cov}(aX+b, cY+d) = acCov(X,Y)$$

For random variables $X_1, X_2, ..., X_n$ and $Y_1, Y_2, ..., Y_m$,

$$\mathrm{Cov}(\sum_{i=1}^{n} a_i X_i, \sum_{j=1}^{m} b_j Y_j) = \sum_{i=1}^{n}\sum_{j=1}^{m} a_i\, b_j\, \mathrm{Cov}(X_i, Y_j).$$

If $\mathrm{Cov}(X,Y) > 0$, we say that X and Y are positively correlated. If $\mathrm{Cov}(X,Y) < 0$, we say that they are negatively correlated. If $\mathrm{Cov}(X,Y) = 0$, we say that X and Y are uncorrelated.

If X and Y are independent, then

$$\mathrm{Cov}(X,Y) = 0.$$

However, the converse of this is not true; that is, two dependent random variables might be uncorrelated.

1.10 Correlation

A large covariance can mean a strong relationship between variables. However, we cannot compare variances over data sets with different scales. A weak covariance in one data set may be a strong one in a different data set with different scales. The problem can be fixed by dividing the covariance by the standard deviation to get the correlation coefficient.

Definition 1.18. *Let X and Y be two random variables with $0 < \sigma_X^2, \sigma_Y^2 < \infty$. The covariance between the standardized X and the standardized Y is called the correlation coefficient between X and Y and is denoted $\rho = \rho(X,Y)$,*

$$\rho(X,Y) = \frac{\mathrm{Cov}(X,Y)}{\sigma_X \sigma_Y}.$$

Note that

- $\rho(X,Y) > 0$ if and only if X and Y are positively correlated;
- $\rho(X,Y) < 0$ if and only if X and Y are negatively correlated; and
- $\rho(X,Y) = 0$ if and only if X and Y are uncorrelated.
- $\rho(X,Y)$ roughly measures the amount and the sign of linear relationship between X and Y.

In the case of perfect linear relationship, we have $\rho(X,Y) = \pm 1$. A correlation of 0, i.e. $\rho(X,Y) = 0$ does not mean zero relationship between two variables; rather, it means zero linear relationship.

Some importants properties of correlation are

$$-1 \leq \rho(X,Y) \leq 1$$

$$\rho(aX+b, cY+d) = \text{sign}(ac)\rho(X,Y)$$

1.11 Model Fitting

The contents in this section are taken from Gray and Pitts.

To fit a parametric model, we have to calculate estimates of the unknown parameters of the probability distribution. Various criteria are available, including the method of moments, the method of maximum likelihood, etc.

1.12 The method of moments

The method of moments leads to parameter estimates by simply matching the moments of the model, $\mathrm{E}[X], \mathrm{E}[X^2], \mathrm{E}[X^3], \dots$, in turn to the required number of corresponding sample moments calculated from the data $x_1, x_2, \dots, x_n$, where n is the number of observations available. The sample moments are simply

$$\frac{1}{n}\sum_{i=1}^{n} x_i, \quad \frac{1}{n}\sum_{i=1}^{n} x_i^2, \quad \frac{1}{n}\sum_{i=1}^{n} x_i^3, \dots .$$

It is often more convenient to match the mean and central moments, in particular matching $\mathrm{E}[X]$ to the sample mean $\bar{x}$ and $\mathrm{Var}[X]$ to the sample variance

$$s^2 = \frac{1}{n-1}\sum_{i=1}^{n}(x_i - \bar{x})^2.$$

An estimate produced using the method of moments is called an MME, and the MME of a parameter θ, say, is usually denoted $\tilde{\theta}$.

1.13 The method of maximum likelihood

The method of maximum likelihood is the most widely used method for parameter estimation. The estimates it produces are those values of the parameters which give the maximum value attainable by the likelihood function, denoted L, which is the joint probability mass or density function for the data we have (under the chosen parametric distribution), regarded as a function of the unknown parameters.

In practice, it is often easier to maximise the loglikelihood function, which is the logarithm of the likelihood function, rather than the likelihood itself. An estimate produced using the method of maximum likelihood is called an MLE, and the MLE of a parameter θ, say, is denoted $\hat{\theta}$. MLEs have many desirable theoretical properties, especially in the case of large samples.

In some simple cases we can derive MLE(s) analytically as explicit functions of summaries of the data. Thus, suppose our data consist of a random sample $x_1, x_2, \dots, x_n$, from a parametric distribution whose parameter(s) we want to estimate. Some straightforward cases include the following:

- the MLE of λ for a $Poi(\lambda)$ distribution is the sample mean, that is $\hat{\lambda} = \bar{x}$
- the MLE of λ for an $Exp(\lambda)$ distribution is the reciprocal of the sample mean, that is $\hat{\lambda} = 1/\bar{x}$

1.14 Goodness of fit tests

We can assess how well the fitted distributions reflect the distribution of the data in various ways. We should, of course, examine and compare the tables of frequencies and, if appropriate, plot and compare empirical distribution functions. More formally, we can perform certain statistical tests. Here we will use the Pearson chi-square goodness-of-fit criterion.

1.15 the Pearson chi-square goodness-of-fit criterion

We construct the test statistic

$$\chi^2 = \frac{\sum (O-E)^2}{E},$$

where O is the observed frequency in a cell in the frequency table and E is the fitted or expected frequency (the frequency expected in that cell under the fitted model), and where we sum over all usable cells.

The null hypothesis is that the sample comes from a specified distribution.

The value of the test statistic is then evaluated in one of two ways.

1. We convert it to a P-value, which is a measure of the strength of the evidence against the hypothesis that the data do follow the fitted distribution. **If the P-value is small enough, we conclude that the data do not follow the fitted distribution – we say "the fitted distribution does not provide a good fit to the data" (and quote the P-value in support of this conclusion)**.
2. We compare it with values in published tables of the distribution function of the appropriate χ^2 distribution, and if the value of the statistic is high enough to be in a tail of specified size of this reference distribution, we conclude that the fitted distribution does not provide a good fit to the data.

1.16 Kolmogorov-Smirnov (K-S) test.

The K-S test statistic is the maximum difference between the values of the ecdf of the sample and the cdf of the fully specified fitted distribution.

The course does not emphasis on the Goodness of Fit Test. Please refer to the reference text for more details.

Chapter 2

Loss distributions

2.1 Introduction

The aim of the course is to provide a fundamental basis which applies mainly in general insurance. General insurance companies' products are short-term policies that can be purchased for a short period of time. Examples of insurance products are

- motor insurance;
- home insurance;
- health insurance; and
- travel insurance.

In case of an occurrence of an insured event, two important components of financial losses which are of importance for management of an insurance company are

- the number of claims; and
- the amounts of those claims.

Mathematical and statistical techniques used to model these sources of uncertainty will be discussed. This will enable insurance companies to

- calculate premium rates to charge policy holders; and
- decide how much reserve should be set aside for the future payment of incurred claims.

In the chapter, statistical distributions and their properties which are suitable for modelling claim sizes are reviewed. These distribution are also known as loss distributions. In practice, the shape of loss distributions are positive skew with a long right tail. The main features of loss distributions include:

- having a few small claims;
- rising to a peak;
- tailing off gradually with a few very large claims.

2.2 Exponential Distribution

A random variable X has an exponential distribution with a parameter $\lambda > 0$, denoted by $X \sim \mathrm{Exp}(\lambda)$ if its probability density function is given by

$$f_X(x) = \lambda e^{-\lambda x}, \quad x > 0.$$

Example 2.1. *Let $X \sim Exp(\lambda)$ and $0 < a < b$.*

1. *Find the distribution $F_X(x)$.*
2. *Express $P(a < X < B)$ in terms of $f_X(x)$ and $F_X(x)$.*
3. *Show that the moment generating function of X is*

$$M_X(t) = \left(1 - \frac{t}{\lambda}\right)^{-1}, \quad t < \lambda.$$

4. *Derive the r-th moment about the origin* $\mathrm{E}[X^r]$.
5. *Derive the coefficient of skewness for X.*
6. *Simulate a random sample of size $n = 200$ from $X \sim Exp(0.5)$ using the command `sample = rexp(n, rate = lambda)` where n and λ are the chosen parameter values.*
7. *Plot a histogram of the random sample using the command `hist(sample)` (use help for available options for `hist` function in R).*

Solution: The code for questions 6 and 7 is given below. The histogram can be generated from the code below.

```
# set.seed is used so that random number generated from different simulations are the same.
# The number 5353 can be set arbitrarily.
set.seed(5353)

nsample <- 200
data_exp <- rexp(nsample, rate = 0.5)

dataset <- data_exp
hist(dataset, breaks=100,probability = TRUE, xlab = "claim sizes"
     , ylab = "density", main = paste("Histogram of claim sizes" ))

hist(dataset, breaks=100, xlab = "claim sizes"
     , ylab = "count", main = paste("Histogram of claim sizes" ))
```

Copy and paste the code above and run it.

eyJsYW5ndWFnZSI6InIiLCJzYW1wbGUiOiJzZXQuc2VlZCg1MzUzKVxuXG5uc2FtcGxlIDwtIDIwMFxuZGF0YV9leHAgP

eyJsYW5ndWFnZSI6InIiLCJzYW1wbGUiOiJzZXQuc2VlZCg1MzUzKVxuXG5uc2FtcGxlIDwtIDIwMFxuZGF0YV9leHAgP

Notes

1. The exponential distribution can used to model the inter-arrival time of an event.
2. The exponential distribution has an important property called **lack of memory**: if $X \sim \mathrm{Exp}(\lambda)$, then the random variable $X - w$ conditional on $X > w$ has the same distribution as X, i.e.

$$X \sim \mathrm{Exp}(\lambda) \Rightarrow X - w | X > w \sim \mathrm{Exp}(\lambda).$$

We can use R to plot the probability density functions (pdf) of exponential distributions with various parameters λ, which are shown in Figure 2.1. Here we use `scale_colour_manual` to override defaults with scales package (see cheat sheet for details).

```
library(ggplot2)
ggplot(data.frame(x=c(0,10)), aes(x=x)) +
  labs(y="Probability density", x = "x") +
  ggtitle("Exponential distributions") +
  theme(plot.title = element_text(hjust = 0.5)) +
  stat_function(fun=dexp,geom ="line", args = (mean=0.5), aes(colour = "0.5")) +
  stat_function(fun=dexp,geom ="line", args = (mean=1), aes(colour = "1")) +
  stat_function(fun=dexp,geom ="line", args = (mean=1.5), aes(colour = "1.5")) +
  stat_function(fun=dexp,geom ="line", args = (mean=2), aes(colour = "2")) +
  scale_colour_manual(expression(paste(lambda, " = ")), values = c("red", "blue", "green", "orange"))
```

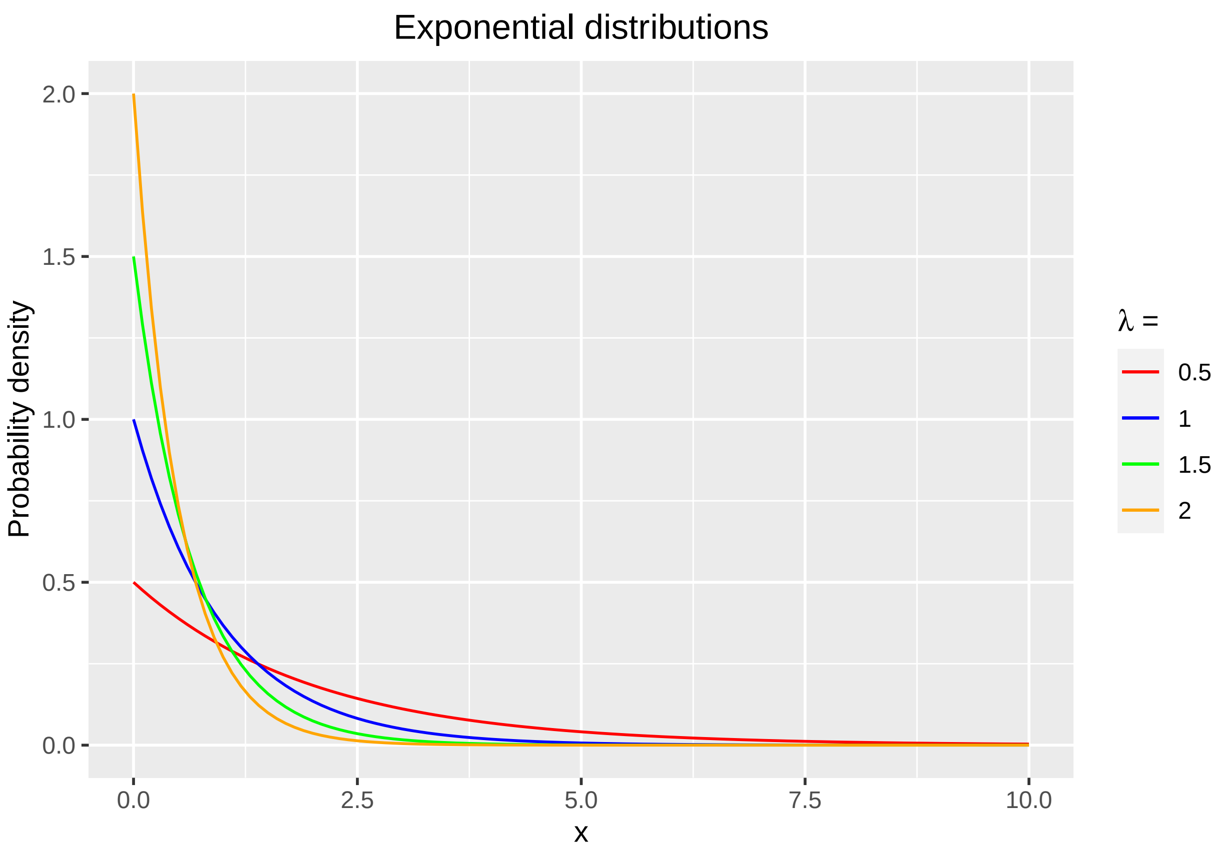


Figure 2.1: The probability density functions (pdf) of exponential distributions with various parameters lambda.

2.3 Gamma distribution

A random variable X has a gamma distribution with parameters $\alpha > 0$ and $\lambda > 0$, denoted by $X \sim \mathcal{G}(\alpha, \lambda)$ or $X \sim \text{gamma}(\alpha, \lambda)$ if its probability density function is given by

$$f_X(x) = \frac{\lambda^\alpha}{\Gamma(\alpha)} x^{\alpha-1} e^{-\lambda x}, \quad x > 0.$$

The symbol Γ denotes the gamma function, which is defined as

$$\Gamma(\alpha) = \int_0^\infty x^{\alpha-1} e^{-x}\, dx, \quad \text{for } \alpha > 0.$$

It follows that $\Gamma(\alpha+1) = \alpha\Gamma(\alpha)$ and that for a positive integer n, $\Gamma(n) = (n-1)!$.

The properties of the gamma distribution are summarised.

- The mean and variance of X are

$$\mathrm{E}[X] = \frac{\alpha}{\lambda} \text{ and } \mathrm{Var}[X] = \frac{\alpha}{\lambda^2}$$

- The r-th moment about the origin is

$$\mathrm{E}[X^r] = \frac{1}{\lambda^r}\frac{\Gamma(\alpha+r)}{\Gamma(\alpha)}, \quad r > 0.$$

- The moment generating function (mgf) of X is

$$M_X(t) = \left(1 - \frac{t}{\lambda}\right)^{-\alpha}, \quad t < \lambda.$$

- The coefficient of skewness is

$$\frac{2}{\sqrt{\alpha}}.$$

Notes 1. The exponential function is a special case of the gamma distribution, i.e. $\mathrm{Exp}(\lambda) = \mathcal{G}(1, \lambda)$

2. If α is a positive integer, the sum of α independent, identically distributed as $\mathrm{Exp}(\lambda)$, is $\mathcal{G}(\alpha, \lambda)$.

3. If $X_1, X_2, \ldots, X_n$ are independent, identically distributed, each with a $\mathcal{G}(\alpha, \lambda)$ distribution, then

$$\sum_{i=1}^n X_i \sim \mathcal{G}(n\alpha, \lambda).$$

4. The exponential and gamma distributions are not fat-tailed, and **may not provide a good fit** to claim amounts.

Example 2.2. *Using the moment generating function of a gamma distribution, show that the sum of independent gamma random variables with the same scale parameter λ, $X \sim \mathcal{G}(\alpha_1, \lambda)$ and $Y \sim \mathcal{G}(\alpha_2, \lambda)$, is $S = X + Y \sim \mathcal{G}(\alpha_1 + \alpha_2, \lambda)$.*

Solution: Because X and Y are independent,

$$\begin{aligned} M_S(t) &= M_{X+Y}(t) = M_X(t) \cdot M_Y(t) \\ &= (1 - \frac{t}{\lambda})^{-\alpha_1} \cdot (1 - \frac{t}{\lambda})^{-\alpha_2} \\ &= (1 - \frac{t}{\lambda})^{-(\alpha_1+\alpha_2)}. \end{aligned}$$

Hence $S = X + Y \sim \mathcal{G}(\alpha_1 + \alpha_2, \lambda)$.

The probability density functions (pdf) of gamma distributions with various shape parameters α and rate parameter $\lambda = 1$ are shown in Figure 2.2.

```
ggplot(data.frame(x=c(0,20)), aes(x=x)) +
  labs(y="Probability density", x = "x") +
  ggtitle("Gamma distribution")  +
  theme(plot.title = element_text(hjust = 0.5)) +
  stat_function(fun=dgamma, args=list(shape=2, rate=1), aes(colour = "2")) +
  stat_function(fun=dgamma, args=list(shape=6, rate=1) , aes(colour = "6")) +
  scale_colour_manual(expression(paste(lambda, " = 1 and ", alpha ," = ")), values = c("red", "blue"))
```

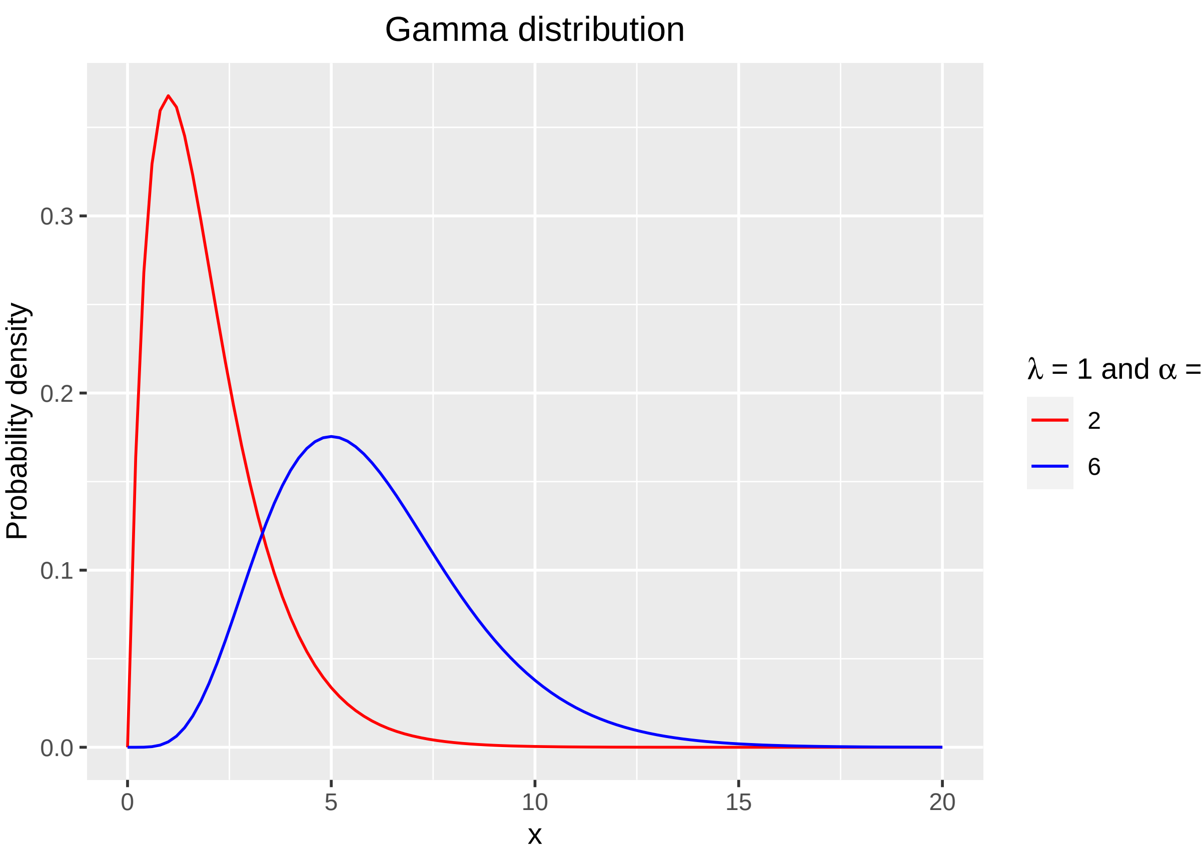


Figure 2.2: The probability density functions (pdf) of gamma distributions with various shape alpha and rate parameter lambda = 1.

2.4 Lognormal distribution

A random variable X has a lognormal distribution with parameters μ and σ^2, denoted by $X \sim \mathcal{LN}(\mu, \sigma^2)$ if its probability density function is given by

$$f_X(x) = \frac{1}{\sigma x\sqrt{2\pi}} \exp\left(-\frac{1}{2}\left(\frac{\log(x)-\mu}{\sigma}\right)^2\right), \quad x > 0.$$

The following relation holds:

$$X \sim \mathcal{LN}(\mu, \sigma^2) \Leftrightarrow Y = \log X \sim \mathcal{N}(\mu, \sigma^2).$$

The properties of the lognormal distribution are summarised.

- The mean and variance of X are

$$\mathrm{E}[X] = \exp\left(\mu + \frac{1}{2}\sigma^2\right) \text{ and } \mathrm{Var}[X] = \exp\left(2\mu + \sigma^2\right)\left(\exp(\sigma^2) - 1\right).$$

- The r-th moment about the origin is

$$\mathrm{E}[X^r] = \exp\left(r\mu + \frac{1}{2}r^2\sigma^2\right).$$

- The moment generating function (mgf) of X is not finite for any positive value of t.
- The coefficient of skewness is

$$\left(\exp(\sigma^2) + 2\right)\left(\exp(\sigma^2) - 1\right)^{1/2}.$$

The probability density functions (pdf) of gamma distributions with various shape parameters α and rate parameter $\lambda = 1$ is shown in Figure 2.3.

```
ggplot(data.frame(x=c(0,10)), aes(x=x)) +
  labs(y="Probability density", x = "x") +
  ggtitle("lognormal distribution")  +
  theme(plot.title = element_text(hjust = 0.5)) +
  stat_function(fun=dlnorm, args = list(meanlog = 0, sdlog = 0.25), aes(colour = "0.25")) +
  stat_function(fun=dlnorm, args = list(meanlog = 0, sdlog = 1), aes(colour = "1")) +
  scale_colour_manual(expression(paste(mu, " = 0 and ", sigma, "= ")), values = c("red", "blue"))
```

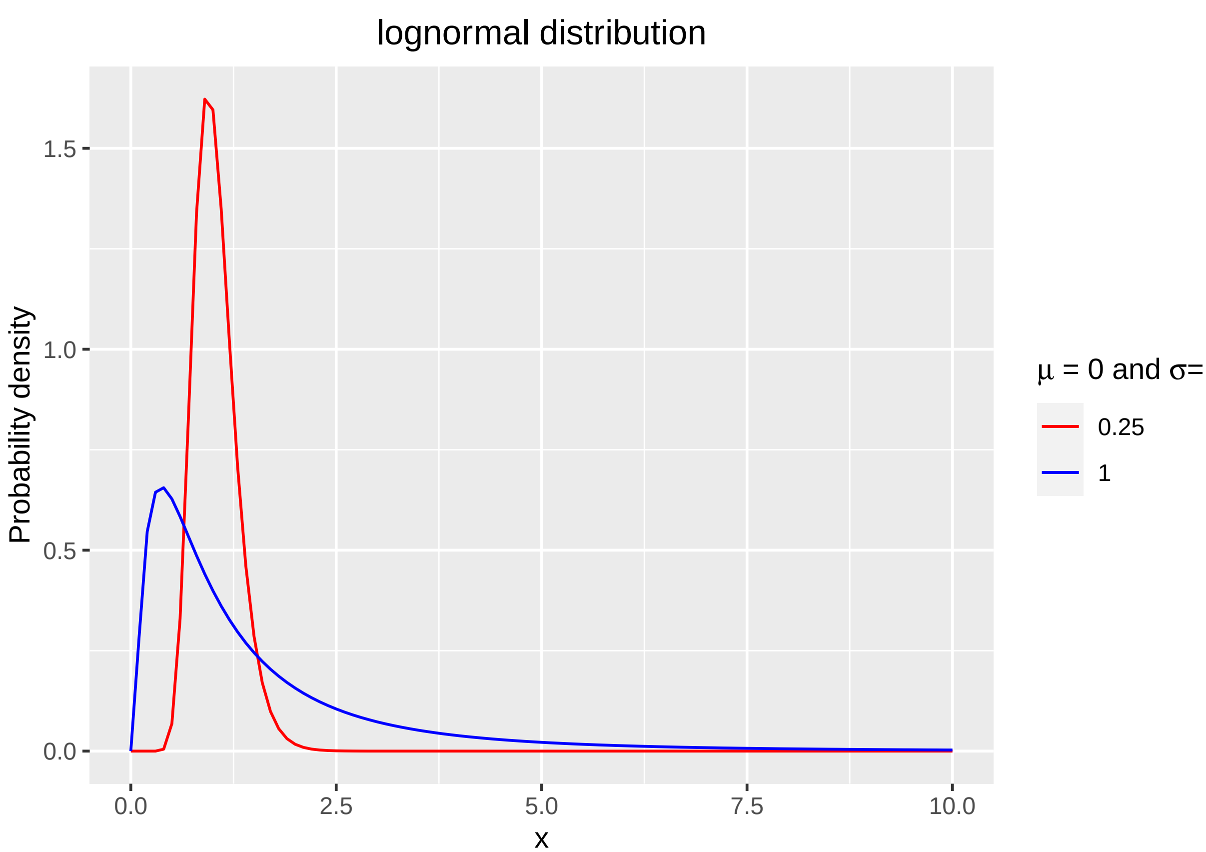


Figure 2.3: The probability density functions (pdf) of lognormal distributions with mu = 0 and sigma = 0.25 or 1.

2.5 Pareto distribution

A random variable X has a Pareto distribution with parameters $\alpha > 0$ and $\lambda > 0$, denoted by $X \sim \text{Pa}(\alpha, \lambda)$ if its probability density function is given by

$$f_X(x) = \frac{\alpha\lambda^\alpha}{(\lambda + x)^{\alpha+1}}, \quad x > 0.$$

The distribution function is given by

$$F_X(x) = 1 - \left(\frac{\lambda}{\lambda + \alpha}\right)^\alpha, \quad x > 0.$$

The properties of the Pareto distribution are summarized.

- The mean and variance of X are

$$\mathrm{E}[X] = \frac{\lambda}{\alpha - 1}, \alpha > 1 \text{ and } \mathrm{Var}[X] = \frac{\alpha\lambda^2}{(\alpha-1)^2(\alpha-2)}, \alpha > 2.$$

- The r-th moment about the origin is

$$\mathrm{E}[X^r] = \frac{\Gamma(\alpha - r)\Gamma(1 + r)}{\Gamma(\alpha)}\lambda^r, \quad 0 < r < \alpha.$$

- The moment generating function (mgf) of X is not finite for any positive value of t.
- The coefficient of skewness is

$$\frac{2(\alpha+1)}{\alpha-3}\sqrt{\frac{\alpha-2}{\alpha}}, \quad \alpha > 3.$$

Note 1. The following conditional tail property for a Pareto distribution is useful for reinsurance calculation. Let $X \sim \text{Pa}(\alpha, \lambda)$. Then the random variable $X - w$ conditional on $X > w$ has a Pareto distribution with parameters α and $\lambda + w$, i.e.

$$X \sim \text{Pa}(\alpha, \lambda) \Rightarrow X - w | X > w \sim \text{Pa}(\alpha, \lambda + w).$$

2. The lognormal and Pareto distributions, in practice, provide a better fit to claim amounts than exponential and gamma distributions.
3. Other loss distribution are useful in practice including **Burr, Weibull and loggamma distributions**.

```
library(actuar)
ggplot(data.frame(x=c(0,60)), aes(x=x)) +
  labs(y="Probability density", x = "x") +
  ggtitle("Pareto distribution")  +
  theme(plot.title = element_text(hjust = 0.5)) +
  stat_function(fun=dpareto, args=list(shape=3, scale=20), aes(colour = "alpha = 3, lambda = 20")) +
  stat_function(fun=dpareto, args=list(shape=6, scale=50), aes(colour = "alpha = 6, lambda = 50")) +
  scale_colour_manual("Parameters", values = c("red", "blue"), labels = c(expression(paste(alpha, " = 3
```

Example 2.3. *Consider a data set consisting of 200 claim amounts in one year from a general insurance portfolio.*

1. *Calculate the sample mean and sample standard deviation.*
2. *Use the method of moments to fit these data with both exponential and gamma distributions.*

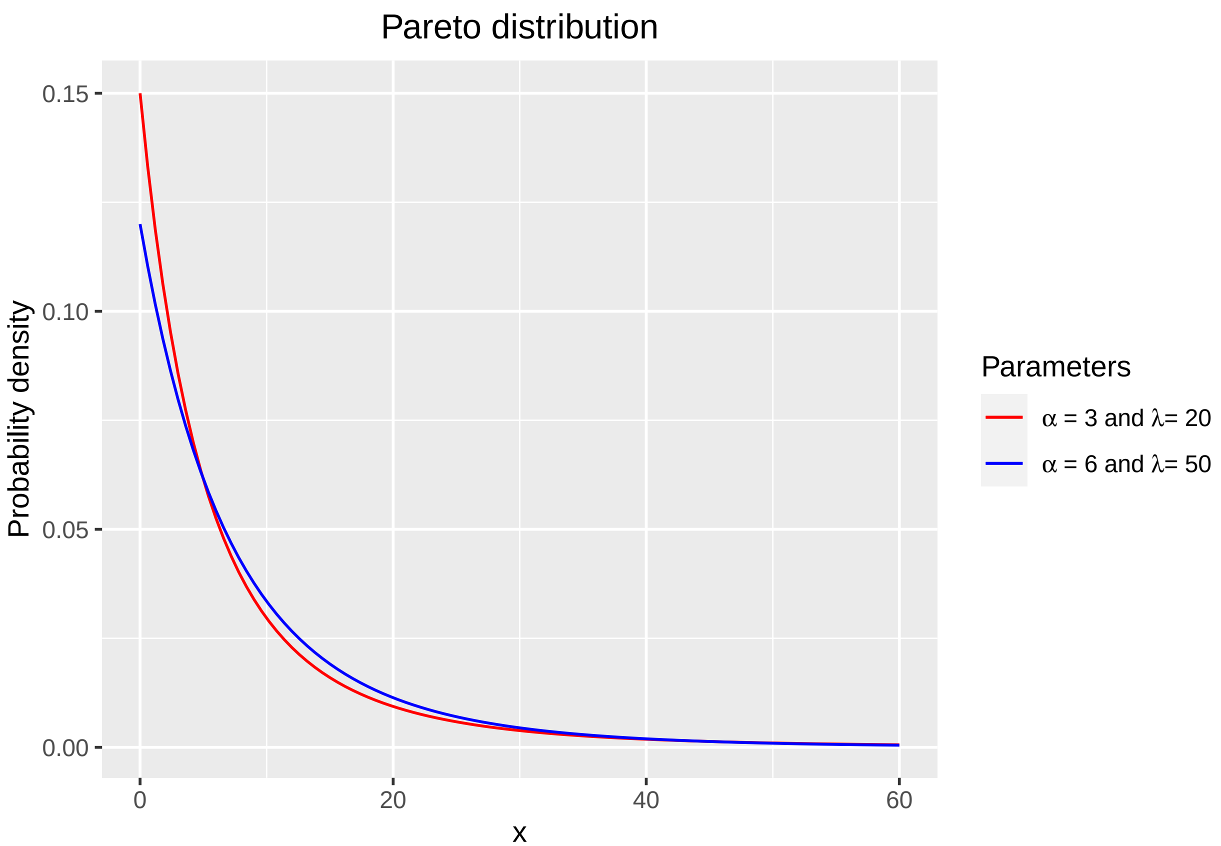


Figure 2.4: The probability density functions (pdf) of Pareto distributions with various shape alpha and rate parameter lambda = 1.

3. *Calculate the boundaries for groups or bins so that the expected number of claims in each bin is 20 under the fitted exponential distribution.*
4. *Count the values of the observed claim amounts in each bin.*
5. *With these bin boundaries, find the expected number of claims when the data are fitted with the gamma, lognormal and Pareto distributions.*
6. *Plot a histogram for the data set along with fitted exponential distribution and fitted gamma distribution. In addition, plot another histogram for the data set along with fitted lognormal and fitted Pareto distribution.*
7. *Comment on the goodness of fit of the fitted distributions.*

Solution: 1. Given that $\sum_{i=1}^{n} x_i = 206046.4$ and $\sum_{i=1}^{n} x_i^2 = 1,472,400,135$, we have

$$\bar{x} = \frac{\sum_{i=1}^{n} x_i}{n} = \frac{206046.4}{200} = 1030.232.$$

The sample variance and standard deviation are

$$s^2 = \frac{1}{n-1}\left(\sum_{i=1}^{n} x_i^2 - \frac{(\sum_{i=1}^{n} x_i)^2}{n}\right) = 6332284,$$

and

$$s = 2516.403.$$

2. We calculate estimates of unknown parameters of both exponential and gamma distributions by the method of moments. We simply match the mean and central moments, i.e. matching $\mathrm{E}[X]$ to the sample mean $\bar{x}$ and $\mathrm{Var}[X]$ to the sample variance.

   The MME (moment matching estimation) of the required distributions are as follows:

   - the MME of λ for an $\mathrm{Exp}(\lambda)$ distribution is the reciprocal of the sample mean,

     $$\tilde{\lambda} = \frac{1}{\bar{x}} = 0.000971.$$

   - the MMEs of α and λ for a $\mathcal{G}(\alpha, \lambda)$ distribution are

     $$\tilde{\alpha} = \left(\frac{\bar{x}}{s}\right)^2 = 0.167614,$$
     $$\tilde{\lambda} = \frac{\tilde{\alpha}}{\bar{x}} = 0.000163.$$

   - the MMEs of μ and σ for a $\mathcal{LN}(\mu, \sigma^2)$ distribution are

     $$\tilde{\sigma} = \sqrt{\ln\left(\frac{s^2}{\bar{x}^2} + 1\right)} = 1.393218,$$
     $$\tilde{\mu} = \ln(\bar{x}) - \frac{\tilde{\sigma}^2}{2} = 5.967012.$$

   - the MMEs of α and λ for a $\mathrm{Pa}(\alpha, \lambda)$ distribution are

     $$\tilde{\alpha} = 2\left(\frac{s^2}{\bar{x}^2}\right)\frac{1}{(\frac{s^2}{\bar{x}^2} - 1)} = 2.402731,$$
     $$\tilde{\lambda} = \bar{x}(\tilde{\alpha} - 1) = 1445.138.$$

3. The upper boundaries for the 10 groups or bins so that the expected number of claims in each bin is 20 under the fitted exponential distribution are determined by

$$\Pr(X \leq \text{upbd}_j) = \frac{j}{10}, \quad j = 1, 2, 3, \dots, 9.$$

With $\tilde{\lambda}$ from the MME for an $\text{Exp}(\lambda)$ from the previous,

$$\Pr(X \leq x) = 1 - \exp(-\tilde{\lambda} x).$$

We obtain

$$\text{upbd}_j = -\frac{1}{\tilde{\lambda}} \ln\left(1 - \frac{j}{10}\right).$$

The results are given in Table 2.1.

4. The following table shows frequency distributions for observed and fitted claims sizes for exponential, gamma, and also lognormal and Pareto fits.

Table 2.1: Frequency distributions for observed and fitted claims sizes.

Range	Observation	Exp	Gamma	Lognormal	Pareto
(0,109]	60	20	109.4	36	31.9
(109,230]	31	20	14.3	34.4	27.8
(230,367]	25	20	9.7	26	24.2
(367,526]	17	20	7.8	20.5	21.2
(526,714]	14	20	6.8	16.6	18.6
(714,944]	13	20	6.3	13.9	16.4
(944,1240]	6	20	6.2	11.9	14.6
(1240,1658]	7	20	6.5	10.8	13.2
(1658,2372]	10	20	7.7	10.4	12.5
(2372,∞)	17	20	25.4	19.5	19.4

5. Let X be the claim size.

- The expected number of claims for the fitted exponential distribution in the range $(a, b]$ is

$$200 \cdot \Pr(a < X \leq b) = 200(e^{-\tilde{\lambda} a} - e^{-\tilde{\lambda} b}).$$

In our case, the expected frequencies under the fitted exponential distribution are given in the third column of Table 2.1.

- (Excel) The expected number of claims for the fitted gamma distribution in the range $(a, b]$ is

$$200 \cdot \left(\text{GAMMADIST}\left(b, \tilde{\alpha}, \frac{1}{\tilde{\lambda}}, \text{TRUE}\right) - \text{GAMMADIST}\left(a, \tilde{\alpha}, \frac{1}{\tilde{\lambda}}, \text{TRUE}\right)\right).$$

The expected frequencies under the fitted gamma distribution are given in the fourth column of Table 2.1.

- (Excel) For the fitted lognormal, the expected number of claims in the range $(a, b]$ can be obtained from

$$200 \cdot \left(\text{NORMDIST}\left(\frac{LN(b) - \tilde{\mu}}{\tilde{\sigma}}\right) - \text{NORMDIST}\left(\frac{LN(a) - \tilde{\mu}}{\tilde{\sigma}}\right)\right).$$

- For the fitted Pareto distribution, the expected number of claims in the range $(a, b]$ can be obtained

from
$$200\left[\left(\frac{\tilde\lambda}{\tilde\lambda+a}\right)^{\tilde\alpha}-\left(\frac{\tilde\lambda}{\tilde\lambda+b}\right)^{\tilde\alpha}\right].$$

6. The histograms for the data set with fitted distributions are shown in Figures 2.5 and 2.6.

7. Comments:

   1. The high positive skewness of the sample reflects the fact that SD is large when compared to the mean. Consequently, the exponential distribution may not fit the data well.

   2. Five claims (2.5%) are greater than 10,000, which is one of the main features of the loss distribution.

   3. The fit is poor for the exponential distribution, as we see that the model under-fits the data for small claims up to 367 and over-fits for large claims between 944 to 2372. The gamma fit is again poor. We see that the model over-fits for small claims between 0-109 and under-fits for claims 230 and 944.

   4. Which one of the lognormal and Pareto distributions provides a better fit to the observed claim data?

```
library(stats)
library(MASS)
library(ggplot2)

xbar <- mean(dat$claims)
s <- sd(dat$claims)

# MME of alpha and lambda for Gamma distribution
alpha_tilde <- (xbar/s)^2
lambda_tilde <- alpha_tilde/xbar

ggplot(dat) + geom_histogram(aes(x = claims, y = ..density..), bins = 90 , fill = "grey", color = "black
  stat_function(fun=dexp, geom ="line", args = (rate = 1/mean(dat$claims)), aes(colour = "Exponential"))
  stat_function(fun=dgamma, geom ="line", args = list(shape = alpha_tilde ,rate = lambda_tilde), aes(col
```

```
library(actuar)
# MME of mu and sigma for lognormal distribution

sigma_tilda <- sqrt(log(  var(dat$claims)/mean(dat$claims)^2 +1  ))  # gives \tilde\sigma
mu_tilda <- log(mean(dat$claims)) - sigma_tilda^2/2      # gives \tilde\mu

# MME of alpha and lambda for Pareto distribution
alpha_tilda <- 2*var(dat$claims)/mean(dat$claims)^2 * 1/(var(dat$claims)/mean(dat$claims)^2 - 1) #/tild
lambda_tilda <- mean(dat$claims)*(alpha_tilda -1)



ggplot(dat) + geom_histogram(aes(x = claims, y = ..density..), bins = 90 , fill = "grey", color = "black
  stat_function(fun=dlnorm, geom ="line", args = list(meanlog = mu_tilda, sdlog = sigma_tilda), aes(colo
  stat_function(fun=dpareto, geom ="line", args = list(shape = alpha_tilda, scale = lambda_tilda), aes(c
  scale_color_discrete(name="Fitted Distributions")
```

Let us plot the histogram of claim sizes with fitted exponential and gamma distributions in this interaction area. Note that the data set is stored in the variable `dat`.

eyJsYW5ndWFnZSI6InIiLCJwcmVfZXhlcmNpc2VfY29kZSI6ImxpYnJhcnkoc3RhdHMpXG5saWJyYXJ5KE1BU1MpXG5sa

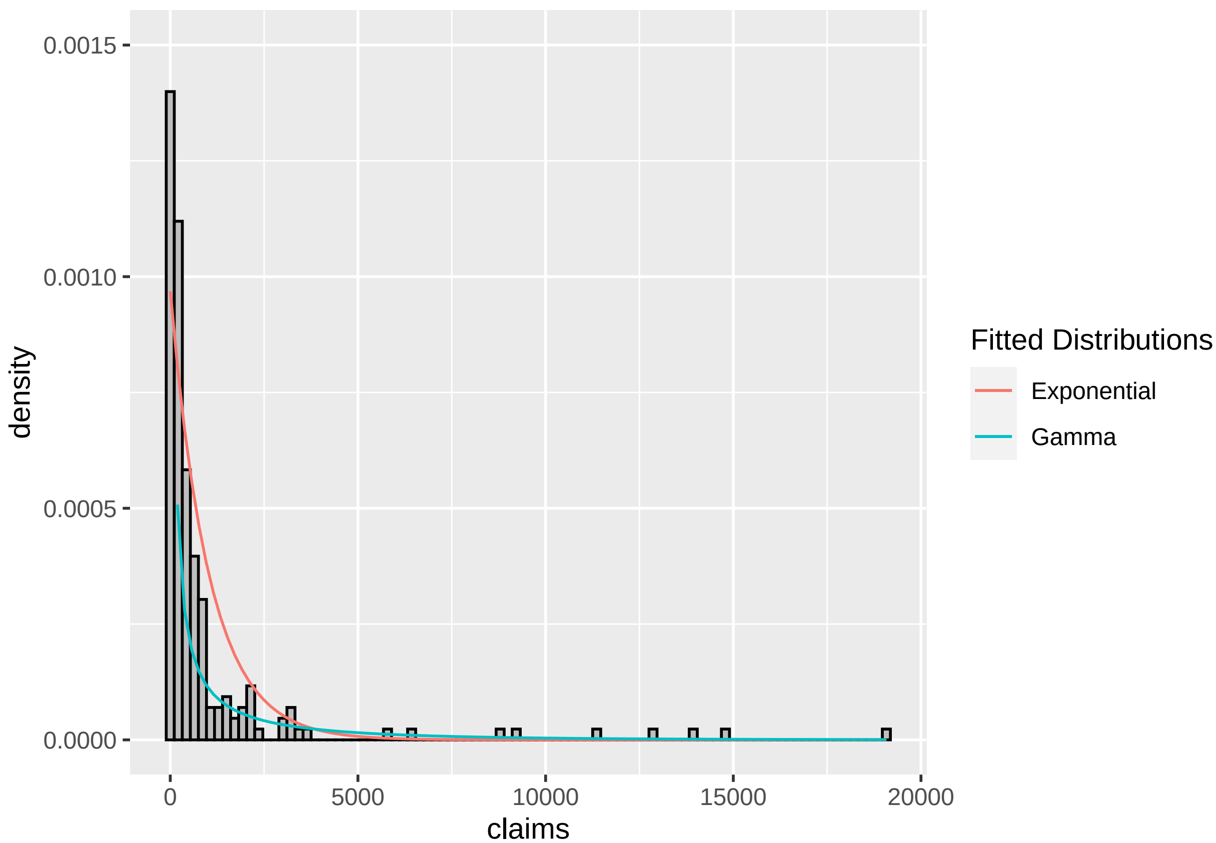


Figure 2.5: Histogram of claim sizes with fitted exponential and gamma distributions.

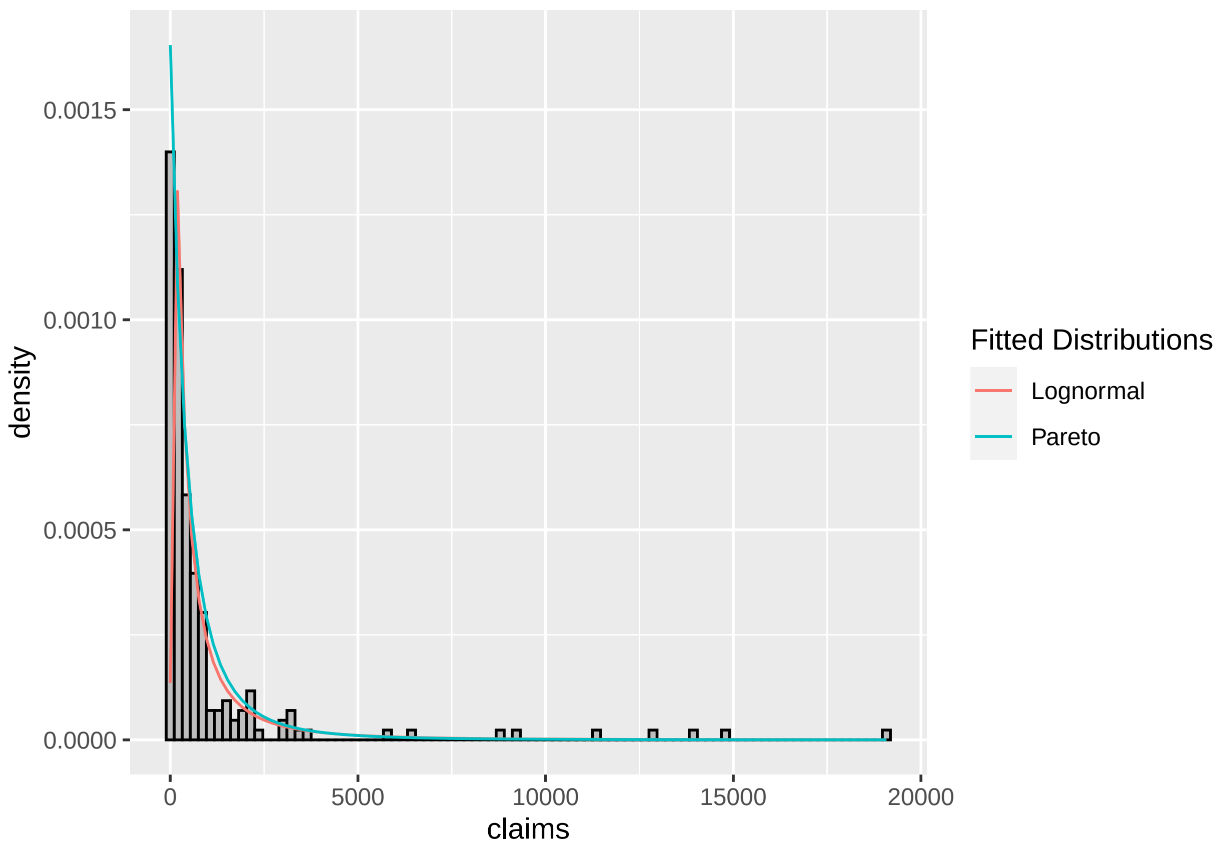


Figure 2.6: Histogram of claim sizes with fitted lognormal and pareto distributions.

The following code can be used to obtain the expected number of claims for the fitted exponential distribution and perform goodness-of-fit test.

eyJsYW5ndWFnZSI6InIiLCJwcmVfZXhlcmNpc2VfY29kZSI6ImxpYnJhcnkoc3RhdHMpXG5saWJyYXJ5KE1BU1MpXG5sa

Chapter 3

Deductibles and reinsurance

3.1 Introduction

In this chapter, we will introduce the concept of risk-sharing. We will consider two types of risk-sharing including deductibles and reinsurance. The purpose of risk sharing is to spread the risk among the parties involved. For example,

1. A policyholder purchases automobile insurance with a deductible. The policyholder is responsible for some of the risk , and transfer the larger portion of the risk to the insurer. The policyholder will submit a claim when the loss exceeds the deductible.
2. A direct insurer can pass on some of the risks to another insurance company known as a reinsurer by purchasing insurance from the reinsurer. It will protect the insurer from paying large claims.

The main goals of the chapter include the derivation of the distribution and corresponding moments of the claim amounts paid by the policyholder, direct insurer and the reinsurer in the presence of risk-sharing arrangements. In addition, the effects of risk-sharing arrangements will reduce the mean and variability of the amount paid by the direct insurer, and also the probability that the insurer will be involved on very large claims.

3.2 Deductibles

The insurer can modify the policy so that the policyholder is responsible for some of the risk by including a deductible (also known as policy excess).

Given a financial loss of X and a deductible of d,

- the insured agrees to bear the first amount of d of any loss X, and only submits a claim when X exceeds d.
- the insurer will pay the remaining of $X - d$ if the loss X exceeds d.

For example, suppose a policy has a deductible of 1000, and you incur a loss of 3000 in a car accident. You pay the deductible of 1000 and the car insurance company pays the remaining of 2000.

Let X be the claim amount, V and Y the amounts of the claim paid by the policyholder, the (direct) insurer, respectively, i.e.

$$X = V + Y.$$

So the amount paid by the policyholder and the insurer are given by

$$V = \begin{cases} X & \text{if } X \leq d \\ d & \text{if } X > d, \end{cases}$$

$$Y = \begin{cases} 0 & \text{if } X \leq d \\ X - d & \text{if } X > d. \end{cases}$$

The amounts V and Y can also be expressed as

$$V = \min(X, d), \quad Y = \max(0, X - d).$$

The relationship between the policyholder and insurer is similar to that between the insurer and reinsurer. Therefore, the detailed analysis of a policy with a deductible is analogous to reinsurance, which will be discussed in the following section.

3.3 Reinsurance

Reinsurance is insurance purchased by an insurance company in order to protect itself from large claims. There are two main types of reinsurance arrangement:

1. excess of loss reinsurance; and
2. proportional reinsurance.

3.4 Excess of loss reinsurance

Under excess of loss reinsurance arrangement, the direct insurer sets a certain limit called a retention level $M > 0$. For a claim X,

- the insurance company pays any claim in full if $X \leq M$; and
- the reinsurer (or reinsurance company) pays the remaining amount of $X - M$ if $X > M$.

The position of the reinsurer under excess of loss reinsurance is the same as that of the insurer for a policy with a deductible.

Let X be the claim amount, V, Y and Z the amounts of the claim paid by the policyholder, (direct) insurer and reinsurer, respectively, i.e.

$$X = V + Y + Z.$$

In what follows, without stated otherwise, we consider the case in which there is no deductible in place, i.e. $V = 0$ and

$$X = Y + Z.$$

So the amount paid by the direct insurer and the reinsurer are given by

$$Y = \begin{cases} X & \text{if } X \leq M \\ M & \text{if } X > M, \end{cases}$$

$$Z = \begin{cases} 0 & \text{if } X \leq M \\ X - M & \text{if } X > M. \end{cases}$$

The amounts Y and Z can also be expressed as

$$Y = \min(X, M), \quad Z = \max(0, X - M).$$

Example 3.1. Suppose a policy has a deductible of 1000 and the insurer arrange excess of loss reinsurance with retention level of 10000. A sample of loss amounts in one year consists of the following values, in unit of Thai baht:

$$3000, 800, 25000, 5000, 20000.$$

Calculate the total amount paid by:*

1. *the policyholder;*
2. *the insurer; and*
3. *the reinsurer.*

Solution:

The total amounts paid by

- the policyholders :
$$1000 + 800 + 1000 + 1000 + 1000 = 4800.$$
- The insurer :
$$2000 + 0 + 10000 + 4000 + 10000 = 26000.$$
- The reinsurer :
$$0 + 0 + 14000 + 0 + 9000 = 23000.$$

3.5 Mixed distributions

In the subsequent sections, we will derive the probability distribution of the random variables Y and Z, which are the insurer's and reinsurer's payouts on claims. Their distributions are neither purely continuous, nor purely discrete. First we start with some important properties of such random variables.

A random variable U which is partly discrete and partly continuous is said to be a mixed distribution. The distribution function of U, denoted by $F_U(x)$ is continuous and differentiable except for some values of x in a countable set S. For a mixed distribution U, there exists a function $f_U(x)$ such that

$$F_U(x) = \Pr(U \le x) = \int_{-\infty}^{x} f_U(x)dx + \sum_{x_i \in S, x_i \le x} \Pr(U = x_i).$$

The expected value of $g(U)$ for some function g is given by

$$\mathrm{E}[g(U)] = \int_{-\infty}^{\infty} g(x) f_U(x)\, dx + \sum_{x_i \in S} g(x_i) \Pr(U = x_i). \tag{3.1}$$

It is the sum of the integral over the intervals at which $f_U(x)$ is continuous and the summation over the points in S.

The function $f_U(x)$ is not the probability density function of U because $\int_{-\infty}^{\infty} f_U(x)dx \neq 1$. In particular, it is the derivative of $F_U(x)$ at the points where $F_U(x)$ is continuous and differentiable.

Recall that X denotes the claim amount and Y and Z be the amounts of the claim paid by the insurer and reinsurer. The distribution function and the density function of the claim amount X are denoted by F_X and $f_X(x)$, where we assume that X is continuous. In the following examples, we will derive the distribution, mean and variance of the random variables Y and Z. Furthermore, both random variables Y and Z are examples of mixed distributions.

Example 3.2. *Let F_Y denote the distribution function of $Y = \min(X, M)$. It follows that*

$$F_Y(x) = \begin{cases} F_X(x) & \text{if } x < M \\ 1 & \text{if } x \geq M \end{cases}.$$

Hence, the distribution function of Y is said to be a mixed distribution.

Solution: From $Y = \min(X, M)$, if $y < M$, then

$$F_Y(y) = \Pr(Y \leq y) = \Pr(X \leq y) = F_X(y).$$

If $y \geq M$, then

$$F_Y(y) = \Pr(Y \leq y) = 1,$$

which follows because $\min(X, M) \leq M$.

Hence, Y is mixed with a density function $f_X(x)$, for $0 \leq x < M$ and a mass of probability at M, with $Pr(Y = M) = 1 - F_X(M)$. The last equality follows from

$$\begin{aligned} \Pr(Y = M) &= \Pr(X > M) \\ &= 1 - \Pr(X \leq M) = 1 - F_X(M). \end{aligned}$$

Example 3.3. *Show that*

$$\mathrm{E}[Y] = \mathrm{E}[\min(X, M)] = \mathrm{E}[X] - \int_0^\infty y f_X(y + M)\, dy.$$

$\mathrm{E}[Y]$ is the expected payout by the insurer.

$$\begin{aligned} \mathrm{E}[Y] &= \mathrm{E}[\min(X, M)] \\ &= \int_0^\infty \min(X, M) \cdot f_X(x)\, dx \\ &= \int_0^M x \cdot f_X(x)\, dx + \int_M^\infty M \cdot f_X(x)\, dx \\ &= \int_0^M x \cdot f_X(x)\, dx + \int_M^\infty x \cdot f_X(x)\, dx + \int_M^\infty (M - x) \cdot f_X(x)\, dx \\ &= \mathrm{E}[X] + \int_M^\infty (M - x) \cdot f_X(x)\, dx \\ &= \mathrm{E}[X] + \int_0^\infty (-y) \cdot f_X(y + M)\, dy \\ &= \mathrm{E}[X] - \int_0^\infty y \cdot f_X(y + M)\, dy \end{aligned}$$

Note Under excess of loss reinsurance arrangement, the mean amount paid by the insurer is reduced by the amount equal to $\int_0^\infty y f_X(y + M)\, dy$.

Example 3.4. *Let X be an exponential distribution with parameter λ and $Y = \min(X, M)$. Then*

$$F_Y(x) = \begin{cases} 1 - e^{-\lambda x} & \text{if } x < M \\ 1 & \text{if } x \geq M \end{cases}.$$

A plot of the distribution function F_Y is given in Figure 1. Hence, Y is a mixed distribution with a density function $f_Y(x) = f_X(x)$ for $0 < x < M$ and a probability mass at M is $\Pr(Y = M) = 1 - F_X(M)$.

Using (3.1), the expected value of Y, $\mathrm{E}[Y]$ is given by

$$\mathrm{E}[Y] = \int_0^M x f_X(x)\, dx + M(1 - F_X(M)).$$

Example 3.5. *Let F_Z denote the distribution function of $Z = \max(0, X - M)$. It follows that*

$$F_Z(x) = \begin{cases} F_X(M) & \textit{if } x = 0 \\ F_X(x+M) & \textit{if } x > 0 \end{cases}.$$

Hence, the distribution function of Z is a mixed distribution with a mass of probability at 0.

Solution: The random variable Z is the **reinsurer's payout** which also include **zero claims**. Later we will consider only **reinsurance claims**, which involve the reinsurer, i.e. claims such that $X > M$.

The distribution of Z can be derived as follows:

- For $x = 0$,
$$F_Z(0) = \Pr(Z = 0) = \Pr(X \le M) = F_X(M).$$
- For $x > 0$,
$$\begin{aligned} F_Z(x) &= \Pr(Z \le x) = \Pr(\max(0, X - M) \le x) \\ &= \Pr(X - M \le x) = \Pr(X \le x + M) = F_X(x + M). \end{aligned}$$

Example 3.6. *Let X be an exponential distribution with parameter λ and $Z = \max(0, X - M)$. Derive and plot the probability distribution F_Z for $\lambda = 1$ and $M = 2$.*

Example 3.7. *Show that*

$$\mathrm{E}[Z] = \mathrm{E}[\max(0, X - M)] = \int_M^\infty (x - M) f_X(x)\, dx = \int_0^\infty y f_X(y + M)\, dy.$$

Comment on the result.

Solution: The expected payout on the claim by the reinsurer, $\mathrm{E}[Z]$, and can also be found directly as follows:

$$\begin{aligned} \mathrm{E}[Z] &= \mathrm{E}[\max(0, X - M)] \\ &= \int_0^M 0 \cdot f_X(x)\, dx + \int_M^\infty (X - M) \cdot f_X(x)\, dx \\ &= 0 + \int_0^\infty y \cdot f_X(y + M)\, dy. \end{aligned}$$

It follows from the previous results that

$$\mathrm{E}[X] = \mathrm{E}[Y + Z] = \mathrm{E}[Y] + \mathrm{E}[Z].$$

Example 3.8. *Let the claim amount X have exponential distribution with mean $\mu = 1/\lambda$.*

1. *Find the proportion of claims which involve the reinsurer.*
2. *Find the insurer's expected payout on a claim.*
3. *Find the reinsurer's expected payout on a claim.*

Solution: 1. The proportion of claims which involve the reinsurer is

$$\Pr(X > M) = 1 - F_X(M) = e^{-\lambda M} = e^{-M/\mu}.$$

2. The insurer's expected payout on a claim can be calculated by

$$\begin{aligned}
\mathrm{E}[Y] &= \mathrm{E}[X] - \int_0^\infty y \cdot \lambda e^{-\lambda(y+M)}\, dy \\
&= \mathrm{E}[X] - e^{-\lambda M} \int_0^\infty y \cdot \lambda e^{-\lambda \cdot y}\, dy \\
&= \mathrm{E}[X] - e^{-\lambda M}\mathrm{E}[X] \\
&= (1 - e^{-\lambda M})\mathrm{E}[X].
\end{aligned}$$

3. It follows from the above result that the reinsurer's expected payout on a claim is $e^{-\lambda M}\mathrm{E}[X]$.

Example 3.9. *An insurer covers an individual loss X with excess of loss reinsurance with retention level M. Let $f_X(x)$ and $F_X(x)$ denote the pdf and cdf of X, respectively.*

1. *Show that the variance of the amount paid by the insurer on a single claim satisfies:*

$$\mathrm{Var}[\min(X, M)] = \int_0^M x^2 f_X(x)\, dx + M^2(1 - F_X(M)) - (\mathrm{E}[\min(X, M)])^2.$$

2. *Show that the variance of the amount paid by the reinsurer on a single claim satisfies:*

$$\mathrm{Var}[\max(0, X - M)] = \int_M^\infty (x - M)^2 f_X(x)\, dx - (\mathrm{E}[\max(0, X - M)])^2.$$

3.6 The distribution of reinsurance claims

In practice, the reinsurer involves only claims which exceed the retention limit, i.e. $X > M$. Information of claims which are less or equal to M may not be available to the reinsurer. The claim amount Z paid by the reinsurer can be modified accordingly to take into account of non-zero claim sizes.

Recall from Example 3.1, there are only three claims whose amounts exceed the retention level of . Such claims, consisting of , 9000 and 23000 which involves the reinsurer are known as **reinsurance claims**.

Let $W = Z|Z > 0$ be a random variable representing the amount of a non-zero payment by the reinsurer on a reinsurance claim. The distribution and density of W can be calculated as follows: for $x > 0$,

$$\begin{aligned}
\Pr[W \le x] &= \Pr[Z \le x | Z > 0] \\
&= \Pr[X - M \le x | X > M] \\
&= \frac{\Pr[M < X \le x + M]}{\Pr[X > M]} \\
&= \frac{F_X(x + M) - F_X(M)}{1 - F_X(M)}.
\end{aligned}$$

Differentiating with respect to x, we obtain the density function of W as

$$f_W(x) = \frac{f_X(x + M)}{1 - F_X(M)}.$$

Hence, the mean and variance can be directly obtained from the density function of W.

3.7 Proportional reinsurance

Under excess of loss reinsurance arrangement, the direct insurer pays a fixed proportion α, called the proportion of the risk retained by the insurer, and the reinsurer pays the remainder of the claim.

Let X be the claim amount, Y and Z the amounts of the claim paid by the policyholder, (direct) insurer and reinsurer, respectively, i.e.

$$X = Y + Z.$$

So the amount paid by the direct insurer and the reinsurer are given by

$$Y = \alpha X, \quad Z = (1 - \alpha)X.$$

Both of the random variables are scaled by the factor of α and $1 - \alpha$, respectively.

Example 3.10. *Derive the distribution function and density function of Y.*

Solution: Let X has a distribution function F with density function f. The distribution function of Y is given by

$$\Pr(Y \leq x) = F(x/a).$$

Hence, the density function is

$$f_Y(x) = \frac{1}{a} f(x/a).$$

You can get more examples from Tutorials.

Chapter 4

Collective Risk Model

Mathematical models of the total amount of claims from a portfolio of policies over a short period of time will be presented in this chapter. The models are referred to as short term risk models. Two main sources of uncertainty including the claim numbers and claim sizes will be taken into consideration. We will begin with the model for aggregate (total) claims or collective risk models.

We define the following random variables:

- S denotes total amount of claims from a portfolio of policies in a fixed time interval, for e.g. one year,
- N represents the number of claims, and
- X_i denotes the amount of the ith claim.

Then the total claims S is given by

$$S = X_1 + ... + X_N.$$

The following assumptions are made for deriving the collective risk model:

1. $\{X_i\}_{i=1}^{\infty}$ are independent and identically distributed with distribution function F_X.
2. N is independent of $\{X_i\}_{i=1}^{\infty}$.

The distribution of the total claim S is said to be a compound distribution. The properties of the compound distribution will be given in the Section 2.

Note The distribution of S can be derived by using convolution technique. In general, the closed form expressions for the compound distribution do not exist so we will mainly concern with the moments of S. For more details about convolution, see Gray and Pitts (2012).

4.1 Conditional expectation and variance formulas

Some useful properties of conditional expectation and conditional variance are given. The conditional expectaion formula is

$$E[E[X|Y]] = E[X].$$

The conditional variance of X given Y is defined to be

$$\begin{aligned} Var[X|Y] &= Var[Z] \text{ where } Z = X|Y \\ &= E[(Z - E[Z])^2] = E[Z^2] - (E[Z])^2 \\ &= E[(X - E[X|Y])^2|Y] \\ &= E[X^2|Y] - (E[X|Y])^2. \end{aligned}$$

The conditional variance formula is

$$Var[X] = E[Var[X|Y]] + Var[E[X|Y]]. \tag{4.1}$$

Example 4.1. Show that

$$Var[X] = E[Var[X|Y]] + Var[E[X|Y]].$$

Solution:

Consider the terms on the right-hand side of (4.1). We have

$$\begin{aligned} E[Var[X|Y]] &= E\left[E[X^2|Y] - (E[X|Y])^2\right] \\ &= E[X^2] - E\left[(E[X|Y])^2\right], \end{aligned}$$

and

$$\begin{aligned} Var[E[X|Y]] &= Var[Z] \text{ where } Z = E[X|Y] \\ &= E[(E[X|Y])^2] - (E[E[X|Y]])^2 \\ &= E[(E[X|Y])^2] - (E[X])^2 \end{aligned}$$

Adding both terms gives the required result.

Example 4.2. In three coloured boxes - Red, Green and Blue, each box has two bags. The bags of Red box contain and respectively, those of Green box contain and , and those of Blue contain and . A box is chosen at random in such a way that Pr(Red) = Pr(Green) = Pr(Blue) = 1/3. A fair coin is tossed to determined which bag to be chosen from the chosen box. Let X be the value of the contents of the chosen bag.

1. Find the distribution of X.
2. Find $E[X]$ and $Var[X]$.
3. Use the conditional expectation and conditional variance formulas to verify your results.

Solution: 1. The distribution of X can be obtained by using the law of total probability: for example

$$\begin{aligned} P(X = 1) &= P(X = 1, R) + P(X = 1, G) + P(X = 1, B) \\ &= P(X = 1|R) \cdot P(R) + P(X = 1|G) \cdot P(G) + P(X = 1|B) \cdot P(B) \\ &= \frac{1}{2} \cdot \frac{1}{3} + \frac{1}{2} \cdot \frac{1}{3} + \frac{1}{2} \cdot \frac{1}{3} = \frac{1}{2}. \end{aligned}$$

Similarly, we have

$$P(X = 1) = \frac{1}{2}, \quad P(X = 2) = P(X = 5) = P(X = 10) = \frac{1}{6}.$$

2. It follows that

$$E[X] = \frac{10}{3}, \quad Var[X] = \frac{98}{9}.$$

3. We first calculate

$$\begin{aligned} E[X|R] &= \frac{1}{2} \cdot (1 + 2) = \frac{3}{2} \\ E[X|G] &= \frac{1}{2} \cdot (1 + 5) = 3 \\ E[X|B] &= \frac{1}{2} \cdot (1 + 10) = \frac{11}{2}. \end{aligned}$$

We have

$$\begin{aligned} E[X] &= E[X|R] \cdot P(R) + E[X|G] \cdot P(G) + E[X|B] \cdot P(B) \\ &= \frac{1}{3} \cdot (\frac{3}{2} + 3 + \frac{11}{2}) = \frac{10}{3}. \end{aligned}$$

4.2 The moments of a compound distribution S

The moments and moment generating function of S can be easily derived from the conditional expectation formula.

4.2.1 The mean of S

Let m_k be the kth moment of X_1, i.e. $E[X_1^k] = m_k$. Conditional on $N = n$, we have

$$E[S|N = n] = E[\sum_{i=1}^{n} X_i] = \sum_{i=1}^{n} E[X_i] = nE[X_i] = n \cdot m_1.$$

Hence, $E[S|N] = Nm_1$ and

$$E[S] = E[E[S|N]] = E[Nm_1] = E[N]m_1 = E[N] \cdot E[X_1].$$

It is no surprise that the mean of the total claims is the product of the means of the number of claims and the mean of claim sizes.

4.2.2 The variance of S

Using the fact that $\{X_i\}_{i=1}^{\infty}$ are independent, we have

$$Var[S|N = n] = Var[\sum_{i=1}^{n} X_i] = \sum_{i=1}^{n} Var[X_i] = nVar[X_i] = n(m_2 - m_1^2),$$

and $Var[S|N] = N(m_2 - m_1^2)$. It follows that

$$\begin{aligned} Var[S] &= E[Var[S|N]] + Var[E[S|N]] \\ &= E[N(m_2 - m_1^2)] + Var[Nm_1] \\ &= E[N](m_2 - m_1^2) + Var[N]m_1^2. \end{aligned}$$

Example 4.3. Show that $M_S(t) = M_N(\log(M_X(t)))$.

Solution:

First, consider the following conditional expectation:

$$\begin{aligned} E\left[e^{tS}|N = n\right] &= E\left[e^{t(X_1+X_2+\cdots X_n)}\right] \\ &= E\left[e^{tX_1}\right] \cdot E\left[e^{tX_2}\right] \cdots E\left[e^{tX_n}\right], \text{ since } X_1, X_2 \ldots, X_n \text{ are independent} \\ &= (M_X(t))^n. \end{aligned}$$

Hence $E\left[e^{tS}|N\right] = (M_X(t))^N$.

From the definition of the moment generating function,

$$\begin{aligned} M_S(t) &= E[e^{tS}] \\ &= E\left[E[e^{tS}|N]\right] \\ &= E\left[(M_X(t))^N\right] \\ &= E\left[Exp(N \cdot \log(M_X(t))\right] \\ &= M_N(\log(M_X(t)))(\text{ since } M_X(t) = E[e^{tX}]). \end{aligned}$$

4.3 Special compound distributions

4.3.1 Compound Poisson distributions

Let N be a Poisson distribution with the parameter λ, i.e. $N \sim Poisson(\lambda)$ and $\{X_i\}_{i=1}^{\infty}$ are independent and identically distributed with distribution function F_X. Then $S = X_1 + ... + X_N$ is said to have a compound Poisson distribution and denote by $\mathcal{CP}(\lambda, F_X)$.

Note The same terminology can be defined similarly for other distributions, for e.g. if N has a negative binomial distribution, then S is said to have a compound negative binomial distribution.

Example 4.4. Let $S \sim \mathcal{CP}(\lambda, F_X)$. Show that

1. $E[S] = \lambda m_1$,
2. $Var[S] = \lambda m_2$,
3. $M_S(t) = Exp(\lambda(M_X(t) - 1))$.
4. The third central moment $E[(S - E[S])^3] = \lambda m_3$, and hence

$$Sk[S] = \frac{\lambda m_3}{(\lambda m_2)^{3/2}},$$

where m_k be the kth moment of X_1

Solution: 1. $E[S] = E[N] \cdot E[X] = \lambda m_1$,

2. $Var[S] = E[N](m_2 - m_1^2) + Var[N]m_1^2 = \lambda(m_2 - m_1^2) + \lambda m_1^2 = \lambda m_2$,
3. From

$$\begin{aligned} M_S(t) &= M_N(\log(M_X(t))) \\ &= Exp\left(\lambda\left(e^{\log(M_X(t))} - 1\right)\right), \text{ since } M_N(t) = Exp(\lambda(e^t - 1)) \\ &= Exp(\lambda(M_X(t) - 1)). \end{aligned}$$

4. The third central moment $E[(S - E[S])^3] = \lambda m_3$, and hence

$$Sk[S] = \frac{\lambda m_3}{(\lambda m_2)^{3/2}}.$$

In particular, we have

$$\begin{aligned} E[(N - E[N])^3] &= E\left[N^3 - 3N^2 \cdot E[N] + 3N \cdot (E[N])^2 - (E[N])^3\right] \\ &= E[N^3] - 3E[N^2] \cdot E[N] + 2(E[N])^3 \\ &= M_N'''(0) - 3M_N''(0) \cdot M_N'(0) + 2(M_N'(0))^3 \end{aligned}$$

For $N \sim Poisson(\lambda)$, $M_N(t) = Exp(\lambda(e^t - 1))$. By differentiating $M_N(t)$ and evaluating at $t = 0$, we can show that

$$M'(0) = \lambda, \quad M''(0) = \lambda(1 + \lambda), \quad M'''(0) = \lambda(1 + 3\lambda + \lambda^2).$$

Hence, $E[(N - E[N])^3] = \lambda$.

Similarly,

$$E[(S - E[S])^3] = E[S^3] - 3E[S^2] \cdot E[S] + 2(E[S])^3$$

In addition, $M_S(t) = Exp(\lambda(M_X(t) - 1))$. By differentiating $M_S(t)$ we can show that

$$M_S'''(t) = \lambda M_X'''(t)M_S(t) + 2\lambda M_X''(t)M_S'(t) + \lambda M_X'(t)M_S''(t).$$

Evaluating $M_S'''(t)$ at $t = 0$ results in

$$M_S'''(0) = E[S^3] = \lambda m_3 + 3E[S] \cdot E[S^2] - 2(E[S])^3,$$

which gives

$$\begin{aligned} E[(S - E[S])^3] &= E[S^3] - 3E[S^2] \cdot E[S] + 2(E[S])^3 \\ &= \lambda m_3. \end{aligned}$$

Example 4.5. Let S be the aggregate annual claims for a risk where $S \sim \mathcal{CP}(10, F_X)$ and the individual claim amounts have a $Pa(4, 1)$ distribution. Calculate $E[S], Var[S]$ and $Sk[S]$.

Solution: Since $X \sim Pa(4, 1)$ with $\alpha = 4$ and $\lambda = 1$, we have

$$\begin{aligned} E[X^r] &= \frac{\Gamma(\alpha - r) \cdot \Gamma(1 + r) \cdot \lambda^r}{\Gamma(\alpha)} \\ E[X] &= \frac{\lambda}{\alpha - 1} = \frac{1}{4 - 1} = \frac{1}{3} \\ E[X^2] &= \frac{\Gamma(2) \cdot \Gamma(3) \cdot \lambda^2}{\Gamma(4)} = \frac{1}{3} \\ E[X^3] &= \frac{\Gamma(1) \cdot \Gamma(4) \cdot \lambda^3}{\Gamma(4)} = 1. \end{aligned}$$

We have

$$\begin{aligned} E[S] &= \lambda E[X] = \frac{10}{3} \\ Var[S] &= \lambda E[X^2] = \frac{10}{3} \\ Sk[S] &= \frac{\lambda E[X^3]}{(\lambda E[X^2])^{3/2}} = \frac{10}{(10/3)^{3/2}} = 1.6432. \end{aligned}$$

An important property of independent, but not necessarily identically distributed, compound Poisson random variables is that the sum of a fixed number of them is also a compound Poisson random variable.

Example 4.6. Let $S_1, \dots, S_n$ be independent compound Poisson random variables, with parameters λ_i and F_i. Then $S = \sum_{i=1}^n S_i$ has a compound distribution with parameter

$$\lambda = \sum_{i=1}^{n} \lambda_i,$$

and

$$F = \frac{1}{\lambda} \sum_{i=1}^{n} \lambda_i F_i.$$

Solution: Exercise.

Note The compound Poisson distribution is the most often used in practice. It possesses the additivity of independent compound Poisson distributions (as shown in Example 4.6, and the expressions of the first three moments are very simple.

4.3.2 Compound negative binomial distributions

A useful discrete random variable that can be used for modelling the distributions of claim numbers is a negative binomial distribution. A random variable N has a negative distribution with parameters k and p, denoted by $N \sim NB(k, p)$ if its probability mass function is given by

$$f_N(n) = \Pr(N = n) = \frac{\Gamma(k + n)}{\Gamma(n + 1)\Gamma(k)} p^k (1 - p)^n \quad n = 0, 1, 2, \dots.$$

It can be interpreted as the probability of getting n failures before the kth success occurs in a sequence of independent Bernoulli trials with probability of success p.

Example 4.7. Let $N \sim NB(k,p)$. Show that the mean, variance and moment generating function of the compound negative binomial distribution, denoted by $\mathcal{CNB}(k,p,F_X)$, are as follows:

1. $E[S] = \frac{kq}{p} m_1$,
2. $Var[S] = \frac{kq}{p^2}(pm_2 + qm_1^2)$,
3. $M_S(t) = \left(\frac{p}{1-qM_X(t)}\right)^k$,

where m_k be the kth moment of X_1 and $q = 1-p$.

Solution: The results follows from the properties of the negative binomial distribution $N \sim NB(k,p)$:

$$E[N] = \frac{kq}{p}, \quad Var[N] = \frac{kq}{p^2},$$

and the moments of a compound distribution S derived in Section 2.

Note The compound negative binomial distribution is an appropriate to model the heterogeneity of the numbers of claims occurring for different risks. In particular, suppose that for each policy, the number of claims in a year has a Poisson distribution $N|\lambda \sim Poisson(\lambda)$, and that the variation in λ across the portfolio can be modelled using a Gamma distribution $G(\alpha,\lambda)$. Then the number of claims in the year for a policy chosen at random from the portfolio has a negative binomial distribution.

4.3.3 Compound binomial distributions

A compound binomial distribution can be used to model a portfolio of policies, each of which can give rise to at most one claim.

Example 4.8. Consider a portfolio of n independent and identical policies where there is at most one claim on each policy in a year (for e.g. life insurance). Let p be the probability that a claim occurs. Explain that the aggregate sum S in this portfolio has a compound binomial distribution, denoted by $\mathcal{CB}(n,p,F_X)$. Derive the mean, variance and moment generating function of S.

Solution: Since n policies (lives) are independent with the probability p that a claim occurs, the number N of claims on the portfolio in one year has a binomial distribution i.e. $N \sim \text{bi}(n,p)$. If the sizes of the claims are i.i.d. random variables, independent of N, then the total amount S claimed on this policy in one year has a compound binomial distribution.

The mean, variance and the moment generating function of S are as follows:

$$\begin{aligned} E[S] &= npm_1, \\ Var[S] &= npm_2 - np^2m_1^2, \\ M_S(t) &= (q + pM_X(t))^n, \end{aligned}$$

where m_k be the kth moment of X_1 and $q = 1-p$.

4.4 The effect of reinsurance

The effect of reinsurance arrangements on an aggregate claims distribution will be presented. Let S denotes the total aggregate claims from a risk in a given time, S_I and S_R denote the insurance and reinsurance aggregate claims, respectively. It follows that

$$S = S_I + S_R.$$

4.4.1 Proportional reinsurance

Recall that under proportional reinsurance arrangement, a fixed proportion α is paid by the direct insurer and the remainder of the claim is paid by the reinsurer. It follows that

$$S_I = \sum_{i=1}^{N} \alpha X_i = \alpha S$$

and

$$S_R = \sum_{i=1}^{N} (1-\alpha) X_i = (1-\alpha) S,$$

where X_i is the amount of the ith claim.

Notes

1. Both direct insurer and the reinsurer are involved in paying each claim.
2. Both have unlimited liability unless a cap on the claim amount is arranged.

Example 4.9. Aggregate claims from a risk in a given time have a compound Poisson distribution with Poisson parameter $\lambda = 10$ and an individual claim amount distribution that is a Pareto distribution, $Pa(4, 1)$. The insurer has effected proportional reinsurance with proportion retained $\alpha = 0.8$.

1. Find the distribution of S_I and S_R and their means and variances.
2. Compare the variances $Var[S_I] + Var[S_R]$ and $Var[S]$. Comment on the results obtained.

Solution: 1. We have

$$S_I = \sum_{i=1}^{N} (\alpha X_i) = \alpha \sum_{i=1}^{N} X_i = \alpha \cdot S,$$
$$S_R = \sum_{i=1}^{N} ((1-\alpha) X_i) = (1-\alpha) \sum_{i=1}^{N} X_i = (1-\alpha) \cdot S,$$

since both insurer and reinsurer are involved in paying each claim, i.e. $Y_i = \alpha X_i$ and $Z_i = (1-\alpha)X_i$. It follows that $S_I \sim \mathcal{CP}(10, F_Y)$ and $S_R \sim \mathcal{CP}(10, F_Z)$.

2. As we can show that if $X \sim Pa(\beta, \lambda)$, then $W = kX \sim Pa(\beta, k\lambda)$. For $X \sim Pa(4, 1)$ with $\beta = 4$ and $\lambda = 1$, we have $Y_i \sim Pa(\beta, \alpha \cdot \lambda) = Pa(4, 0.8)$ and

$$\begin{aligned}
E[S_I] &= 10 \cdot E[Y_i] = 10 \cdot \frac{\alpha \cdot \lambda}{\beta - 1} \\
&= \frac{8}{3}, \\
Var[S_I] &= 10 \cdot E[Y_i^2] = 10 \cdot \frac{\Gamma(\beta-2) \cdot \Gamma(1+2) \cdot (\alpha \cdot \lambda)^2}{\Gamma(\beta)} \\
&= 10 \cdot \frac{\Gamma(2) \cdot \Gamma(3) \cdot (\alpha \cdot \lambda)^2}{\Gamma(4)} = 10 \cdot \frac{2!}{3!} \cdot (0.8)^2 \\
&= \frac{32}{15}
\end{aligned}$$

Alternatively, we can calculate by using the properties of the expectation and variance as follows:

$$\begin{aligned}
E[S_I] &= E[\alpha S] = \alpha \cdot E[S] = \alpha \cdot \lambda \cdot E[X] = 10 \cdot 0.8 \cdot \frac{1}{3} = \frac{8}{3}, \\
Var[S_I] &= Var[\alpha S] = \alpha^2 \cdot Var[S] \\
&= \alpha^2 \cdot \lambda \cdot E[X^2] = \frac{32}{15}.
\end{aligned}$$

Similarly,

$$E[S_R] = E[(1-\alpha)S] = (1-\alpha)\cdot E[S] = \frac{2}{3},$$
$$Var[S_R] = Var[(1-\alpha)S] = (1-\alpha)^2\cdot Var[S] = \frac{2}{15}.$$

Note that $E[S_I] + E[S_R] = E[S]$, while $Var[S_I] + Var[S_R] = \frac{34}{15} < Var[S] = \frac{10}{3}$.

4.4.2 Excess of loss reinsurance

Recall that under excess of loss reinsurance arrangement, the direct insurer has effected excess of loss reinsurance with retention level $M > 0$. For a claim X,

- the insurance company pays any claim in full if $X \leq M$; and
- the reinsurer (or reinsurance company) pays the remaining amount of $X - M$ if $X > M$.

It follows that

$$S_I = \sum_{i=1}^{N} Y_1 + Y_2 + ... + Y_N = \sum_{i=1}^{N} \min(X_i, M) \tag{4.2}$$

and

$$S_R = \sum_{i=1}^{N} Z_1 + Z_2 + ... + Z_N = \sum_{i=1}^{N} \max(0, X_i - M), \tag{4.3}$$

where X_i is the amount of the ith claim. When $N = 0$, we set $S_I = 0$ and $S_R = 0$.

Note S_R can equal 0 even if $N > 0$. This occurs when all claims do not exceed M and hence the insurer pays the full amounts of claims.

As discussed in the previous section, the reinsurer is involved only claims which exceed the retention limit (a claim such that $X > M$). Such claims are called **reinsurance claims**. Taking in account of counting only non-zero claims, we can rewrite S_R as follows. Let N_R be the number of insurance (non-zero) claims for the reinsurer and W_i be the amount of the ith non-zero payment by the reinsurer. The aggregate claim amount paid by the reinsurer can be written as

$$S_R = \sum_{i=1}^{N_R} W_i.$$

:::{.example} By using the probability generating function, show that if $N \sim Poisson(\lambda)$, then the distribution $N_R \sim Poisson(\lambda \pi_M)$ where $\pi_M = \Pr(X_j > M)$. ::: **Solution:** Define the indicator random variable $\{I_j\}_{j=1}^{\infty}$, where

$$I_j = \begin{cases} 1 & \text{if } X_j > M \\ 0 & \text{if } X_j \leq M. \end{cases}$$

Therefore,

$$N_R = \sum_{j=1}^{N} I_j.$$

The variable N_R has a compound distribution with its probability generating function

$$P_{N_R}(r) = P_N[P_I(r)],$$

where P_I is the probability generating function of the indicator random variable. It can be shown that

$$P_I(r) = 1 - \pi_M + \pi_M r,$$

where $\pi_M = \Pr(I_j = 1) = \Pr(X_j > M) = 1 - F(M)$.

Note In the above example, one can derive the distribution of N_R by using the moment generating function:

$$M_{N_R}(t) = M_N(\log M_I(t)),$$

where M_N and M_I are the moment generating functions of N and I. Note also that

$$M_I(t) = 1 - \pi_M + \pi_M Exp(t).$$

4.4.3 Compound Poisson distributions under excess of loss reinsurance

Assume that aggregate claim amount $S \sim \mathcal{CP}(\lambda, F_X)$ has a compound Poisson distribution. Under excess of loss reinsurance with retention level M, it follows from (4.2) and (4.3) that

1. $S_I \sim \mathcal{CP}(\lambda, F_Y)$,

   where $f_Y(x) = f_X(x)$ for $0 < x < M$ and $\Pr(Y = M) = 1 - F_X(M)$.
2. $S_R \sim \mathcal{CP}(\lambda, F_Z)$,

   where $F_Z(0) = F_X(M)$ and $f_Z(x) = f_X(x + M), x > 0$.
3. Excluding zero claims, $S_R \sim \mathcal{CP}(\lambda\,(1 - F_X(M)), F_W)$,

   where $f_W(x) = \dfrac{f_X(x+M)}{1 - F_X(M)}, x > 0$.

Example 4.10. Suppose that S has a compound Poisson distribution with Poisson parameters $\lambda = 10$ and the claim sizes have the following distribution

x	1	2	5	10
$\Pr(X = x)$	0.4	0.3	0.2	0.1

The insurer enters into an excess of loss reinsurance contract with retention level $M = 4$.

1. Show that $S_I \sim \mathcal{CP}(\lambda, F_Y)$.
2. Show that $S_R \sim \mathcal{CP}(\lambda, F_Z)$.
3. By excluding zero claims, show that the S_R can also be expressed as $S_R \sim \mathcal{CP}(\lambda\, p, F_W)$ where $p = \Pr(X > M)$.
4. Find the mean and variance of the aggregate claim amount for both insurer and reinsurer.

Example 4.11. Suppose that S has a compound Poisson distribution with Poisson parameters $\lambda = 40$ and the claim sizes have a Pareto distribution $Pa(3, 4)$. The insurer has an excess of loss reinsurance contract in place with retention level $M = 2$. Find the mean and variance of the aggregate claim amount for both insurer and reinsurer.

Chapter 5

Tutorials

5.1 Tutorial 1

1. Using the method of moments, calculate the parameter values for the gamma, lognormal and Pareto distributions for which
$$\mathrm{E}[X] = 500 \quad \text{and} \quad \mathrm{Var}[X] = 100^2.$$

2. Show that if $X \sim \mathrm{Exp}(\lambda)$, then the random variable $X - w$ conditional on $X > w$ has the same distribution as X, i.e.
$$X \sim \mathrm{Exp}(\lambda) \Rightarrow X - w | X > w \sim \mathrm{Exp}(\lambda).$$

3. Derive an expression for the variance of the $\mathrm{Pa}(\alpha, \lambda)$ distribution. (Hint: using the pdf)

4. Show that the MLE (the maximum likelihood estimation) of λ for an $\mathrm{Exp}(\lambda)$ distribution is the reciprocal of the sample mean, i.e. $\hat{\lambda} = 1/\bar{x}$.

5. Claims last year on a portfolio of policies of a risk had a lognormal distribution with parameter $\mu = 5$ and $\sigma^2 = 0.4$. It is estimated that all claims will increase by 15% next year. Find the probability that a claim next year will exceed 1000.

5.2 Tutorial 2

1. Claims occur on a general insurance portfolio independently and at random. Each claim is classified as being of "Type A" or "Type B". Type A claim amounts are distributed $\mathrm{Pa}(3, 400)$ and Type B claim amounts are distributed $\mathrm{Pa}(4, 1000)$. It is known that 90% of all claims are of Type A.

   Let X denote a claim chosen at random from the portfolio.

   1. Calculate $\Pr(X > 1000)$.
   2. Calculate $\mathrm{E}[X]$ and $\mathrm{Var}[X]$.
   3. Let Y have a Pareto distribution with the same mean and variance as X. Calculate $\Pr(Y > 1000)$.
   4. Comment on the difference in the answers found in 1.1 and 1.2.

2. An insurer covers an individual loss X with excess of loss reinsurance with retention level M. Let Y and Z be random variables representing the amounts paid by the insurer and reinsurer, respectively, i.e. $X = Y + Z$. Show that $\mathrm{Cov}[Y, Z] \geq 0$ and deduce that
$$\mathrm{Var}[X] \geq \mathrm{Var}[Y] + \mathrm{Var}[Z].$$

Comment on the results obtained.

3. Claim amounts from a general insurance portfolio are lognormally distributed with mean 200 and variance 2916. Excess of loss reinsurance with retenton level 250 is arranged. Calculate the probability that the reinsurer is involved in a claim.

4. Show that if $X \sim \mathrm{Pa}(\alpha, \lambda)$, then the random variable $X - d$ conditional on $X > d$ has a pareto distribution with parameters α and $\lambda + d$, i.e.

$$X \sim \mathrm{Pa}(\alpha, \lambda) \Rightarrow X - d|X > d \sim \mathrm{Pa}(\alpha, \lambda + d).$$

5. Consider a portfolio of motor insurance policies. In the event of an accident, the cost of the repairs to a car has a Pareto distribution with parameters α and λ. A deductible of 100 is applied to all claims and a claim is always made if the cost of the repairs exceeds this amount. A sample of 100 claims has mean 200 and standard deviation 250.

    1. Using the method of moments, estimate α and λ.

    2. Estimate the proportion of accidents that do not result in a claim being made.

    3. The insurance company arranges excess of loss reinsurance with another insurance company to reduce the mean amount it pays on a claim to 160. Calculate the retention limit needed to achieve this.

Chapter 6

Interactive Lecture

Some *significant* applications are demonstrated in this chapter.

6.1 DataCamp Light

By default, `tutorial` will convert all R chunks.

eyJsYW5ndWFnZSI6InIiLCJzYW1wbGUiOiJhIDwtIDJcbmIgPC0gM1xuYSArIGIifQ==

eyJsYW5ndWFnZSI6InIiLCJwcmVfZXhlcmNpc2VfY29kZSI6ImxpYnJhcnkoc3RhdHMpXG5saWJyYXJ5KE1BU1MpXG5sa