Chapter 3

*Loss distributions*

0 Introduction

- Insurance companies need to investigate claims experience and apply mathematical techniques for many purposes.
  - rating
  - reserving
  - reinsurance arrangements
  - solvency
In this chapter we will look at loss distributions, which are a mathematical method of modelling individual claims.

- introduce some new statistical distributions
- see how these can be ”fitted” to observed claims data
- test for goodness of fit
- use the fitted loss distributions to estimate probabilities
• The total amount of claims in a particular time period is a quantity of fundamental importance to the proper management of an insurance company.

• The **key assumption** in all the models studied here is that the occurrence of a claim and the amount of a claim can be studied separately.
• See the figures on page 2. The above figure is the frequency of individual claims when plotted against size. The statistical distributions in this chapter are used to approximate this distribution, which is called a loss distribution.
• First we will look at the distribution of the size of an *individual* claim. Later we will consider the *total* claim amount over a period of time (usually a year), which is called the **aggregate claim amount**.
• A range of statistical techniques can be used to describe the distribution of random variables. The object is to describe the variation in claim amounts by finding a loss distribution that adequately describes the claims that actually occur. This can be done at two levels:

– 1\textsuperscript{st} LEVEL Assume that the claims arise as realizations from a known distribution. Knowledge of the claim amount process would be complete, and interest would then center on the consequences for insurance.

* For example, claims above a certain level might trigger some reinsurance arrangements or claims below a certain level might never be lodged if a policy excess was in force.
– 2\textsuperscript{nd} LEVEL A standard method of proceeding is to assume that the claims distribution is a member of a certain family. The parameters of the family must be estimated using the claim amount records by an appropriate method. Complications will arise if large claims have been limited (reinsurance) or some small claims have not been lodged (policy excess).

• Claims distributions tend to be \textbf{positively skewed} and \textbf{long tailed}. (See the figures on page 2.)
1 MGFs and basic loss distributions

• In this section we will review some of the properties of the statistical distributions we will need for modelling claims distributions.

• Moment generating functions (MGFs) of
  – exponential distribution
  – gamma distribution
  – normal distribution
1.1 The exponential distribution

- A random variable $X$ has the exponential distribution with parameter $\lambda$ if the distribution function
  \[ F(x) = 1 - e^{-\lambda x}, \quad \lambda > 0 \text{ and } x > 0 \]
  and is written $X \sim Exp(\lambda)$.

- The mean, variance, and the MGF of $X$ are, respectively,
  \[
  \begin{align*}
  E(X) &= \frac{1}{\lambda} \\
  \text{Var}(X) &= \frac{1}{\lambda^2} \\
  M(t) &= \frac{\lambda}{\lambda - t}, \quad t < \lambda.
  \end{align*}
  \]

- The restriction $t < \lambda$ is real because $E(e^{tX})$ does not exist for $t \geq \lambda$. (The integral diverges.)
1.2 The gamma distribution

- The random variable $X$ has a gamma distribution with parameters $\alpha$ and $\lambda$ if the density function

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

and is written $X \sim Ga(\alpha, \lambda)$. (OR $Gamma(\alpha, \lambda)$, $\Gamma(\alpha, \lambda)$.)

- $Ga(1, \lambda) \equiv Exp(\lambda)$

- The mean, variance, and the MGF of $X$ are, respectively,

$$E(X) = \frac{\alpha}{\lambda}$$

$$Var(X) = \frac{\alpha}{\lambda^2}$$

$$M(t) = \left(\frac{\lambda}{\lambda - t}\right)^\alpha, \quad t < \lambda.$$
• MGF can be used to derive the formulae for the mean, the variance and higher moments of the distribution.

• Because there is no closed form (i.e. no simple formula) for the cumulative distribution function of the gamma distribution, it is not easy to find gamma probabilities directly. However, a rough estimate of gamma tail probabilities can be obtained by using the table of the $\chi^2$ distribution. If $X \sim Ga(\alpha, \lambda)$, then $2\lambda X \sim \chi^2_{2\alpha}$. 
1.3 The normal distribution

- The normal distribution is perhaps less useful here in modelling loss distributions because of its symmetry. (loss distributions tend to be asymmetrical).

- The MGF of the normal distribution is

\[ M(t) = \exp(\mu t + \frac{1}{2} \sigma^2 t^2) \]
2 Other loss distributions

• We will discuss in this section
  – The Pareto and generalized Pareto distributions
  – The lognormal distribution
  – The Weibull distribution
  – The Burr distribution

• None of the distributions in this section have an MGF that is easy to derive or use.
2.1 The Pareto and generalized Pareto distributions

- A random variable $X$ has the Pareto distribution with parameters $\alpha$ and $\lambda$ if
  \[
  F(x) = 1 - \left( \frac{\lambda}{\lambda + x} \right)^\alpha
  \]
  and is written $X \sim Pa(\alpha, \lambda)$.

- Its PDF is
  \[
  f(x) = \frac{\alpha \lambda^\alpha}{(\lambda + x)^{\alpha+1}} \quad x > 0, \; \alpha > 0, \; \lambda > 0.
  \]

- Question 3.5 on page 7.
  Verify the formula given in the tables for the mean of the Pareto distribution. For what values of the parameters is this formula valid?

- Example on page 8.
• The **generalized Pareto distribution** has PDF:

\[
f(x) = \frac{\Gamma(\alpha + k)\lambda^\alpha x^{k-1}}{\Gamma(\alpha)\Gamma(k)(\lambda + x)^{\alpha+k}}
\]

\[x > 0, \alpha > 0, \lambda > 0, k > 0.\]

• The two-parameter version is a special case of the generalized Pareto distribution, with the third parameter \(k = 1\).
2.2 The lognormal distribution

- $X$ has a lognormal distribution if $\log X$ has a normal distribution.
- When $\log X \sim N(\mu, \sigma^2)$, $X \sim LN(\mu, \sigma^2)$.
- The mean and variance of the lognormal distribution can be proved using the MGF of the normal distribution or by direct integration.
- Example on pages 9-10.
2.3 The Weibull distribution

- A random variable $X$ has a Weibull distribution with parameters $c$ and $\gamma$ if the distribution function

$$F(x) = 1 - \exp(-cx^\gamma)$$

and it is written $X \sim W(c, \gamma)$.

- The PDF of the $W(c, \gamma)$ distribution is

$$f(x) = c\gamma x^{\gamma-1} \exp(-cx^\gamma), \quad x > 0, \ c > 0, \ \gamma > 0.$$ 

- The expressions for the upper tails of the exponential, the Pareto distributions and the Weibull distribution are:

  - exponential, $P(X > x) = \exp(-\lambda x)$
  - Pareto, $P(X > x) = (\lambda/(\lambda + x))^\alpha$
  - Weibull, $P(X > x) = \exp(-\lambda x^\gamma)$
• If $\gamma < 1$, a distribution with a tail intermediate in weight between the exponential and the Pareto will be obtained. While if $\gamma > 1$, the upper tail will be lighter than the exponential ($\gamma = 1$ is the exponential distribution).

• The tail distribution defines the Weibull distribution, a very flexible distribution, which can be used as a model for losses in insurance, usually with $\gamma < 1$.

• See the figure on page 12.
2.4 The Burr distribution

- A random variable $X$ has a Burr distribution with parameters $\alpha$, $\lambda$ and $\gamma$ if the distribution function

$$F(x) = 1 - \frac{\lambda^\alpha}{(\lambda + x^\gamma)^\alpha}$$

and it is written $X \sim Burr(\alpha, \lambda, \gamma)$.

- cf. $Pa(\alpha, \lambda)$

- Example on page 14.
3 Estimation

• The methods of maximum likelihood, moments and percentiles can be used to fit distributions to sets of data. The fit of the distribution can be tested formally by using a $\chi^2$ test.
3.1 The method of moments

- For a distribution with \( r \) parameters, the moments are as follows:

\[
m_j = \frac{1}{n} \sum_{i=1}^{n} x_i^j, j = 1, 2, \ldots, r
\]

where \( m_j = \text{E}(X^j|\theta) \), a function of the unknown parameter, \( \theta \), being estimated,

- The estimate for the parameter, \( \theta \), can be determined by solving the equation above.
  Where there is more than one parameter, they can be determined by solving the simultaneous equations for each \( m_j \).

- To obtain a method of moments estimator for a parameter, we equate the corresponding sample and population noncentral moments.
• If we were trying to estimate the value of a single parameter, and we had a sample of \( n \) claims whose sizes were \( x_1, x_2, \ldots, x_n \), we would solve the equation:

\[
E(X) = \frac{1}{n} \sum_{i=1}^{n} x_i
\]

\( i.e. \) we would equate the first noncentral moments for the population and the sample.

• If we were trying to find estimates for two parameters (for example if we were fitting a gamma distribution and needed to find estimates for both parameters), we would solve the simultaneous equations:

\[
E(X) = \frac{1}{n} \sum_{i=1}^{n} x_i
\]

\[
E(X^2) = \frac{1}{n} \sum_{i=1}^{n} x_i^2
\]
• In fact in the two parameter case, estimates are usually obtained by equating sample and population means and variances. If we define the sample variance to have a denominator \( n \), this will give the same estimates as would be obtained by equating the first two noncentral moments.

• More generally, we use as many equations of the form

\[
E(X^k) = \frac{1}{n} \sum_{i=1}^{n} x_i^k
\]

as are needed to determine estimates of the relevant parameters.
3.2 Maximum Likelihood Estimation

• The likelihood function of a variable, $x$, is the probability of observing what was observed given a hypothetical value of the parameter, $\theta$. The maximum likelihood estimate (MLE) is the one that yields the highest probability, i.e. that which maximizes the likelihood function.
- The likelihood function \( L(\theta) \) can be expressed as:

\[
L(\theta) = \prod_{i=1}^{n} f(x_i|\theta)
\]

To determine the MLE the above expression needs to be maximized.

Often it is practical to consider the log-likelihood function:

\[
I(\theta) = \log L(\theta) = \sum_{i=1}^{n} \log f(x_i|\theta)
\]

If \( I(\theta) \) can be differentiated with respect to \( \theta \), the MLE, expressed as \( \hat{\theta} \), satisfies the expression:

\[
\frac{d}{d\theta} I(\hat{\theta}) = 0
\]
• Where there is more than one parameter, the MLE for each parameter can be determined by taking partial derivatives of the log-likelihood function and setting each to zero.
• The steps involved in finding a maximum likelihood estimate (MLE) are as follows:

1. Write down the likelihood function for the available data.
   If the likelihood is based on a set of known values $x_1, x_2, \ldots, x_n$, then the likelihood function will take the form $f(x_1)f(x_2)\ldots f(x_n)$, where $f(x)$ is the PDF (or PF if it’s discrete) of the distribution that is to be fitted.

2. Take logs. This will usually simplify the algebra.

3. Maximize the log-likelihood function. This usually involves differentiating the log-likelihood function with respect to each of the unknown parameters, and setting the resulting expression(s) equal to zero.

4. Solve the resulting equation(s) to find the MLEs of the parameters.
3.3 The exponential and gamma distributions

- It is possible to use the method of maximum likelihood (ML) or the method of moments to estimate the parameter of the exponential distribution.

- Example on pages 18-19.
• The MLEs for the gamma distribution cannot be obtained in closed form (i.e. in terms of elementary functions) but the moment estimators can be used as initial estimates in the search for the maximum likelihood estimates.

• It is more convenient to obtain maximum likelihood estimates for the gamma distribution using a different parameterisation.
  Set $\mu = \alpha/\lambda$ and estimate the parameters $\alpha$ and $\mu$. Then recover the MLE of $\lambda$ by setting $\hat{\lambda} = \hat{\alpha}/\hat{\mu}$. 
3.4 The normal distribution

- The method of moments and maximum likelihood estimation are both straightforward to apply in this case.

- Both give the obvious answers:

  \[ \hat{\mu} = \bar{x} \]

  and

  \[ \hat{\sigma}^2 = \frac{1}{n} \sum_{i=1}^{n} (x_i - \bar{x})^2 \]

- Note that the estimate for the population variance is \( \frac{n - 1}{n} \times \) the sample variance.
3.5 The Pareto and generalized Pareto distributions

- The method of moments is very easy to apply in the case of the Pareto distribution, but the estimates obtained in this way will tend to have rather large standard errors, mainly because $S^2$, the sample variance, has a very large variance. However, the method does provide initial estimates for more efficient methods of estimation that may not be so simple to apply, like maximum likelihood, where numerical methods may need to be used.

- **Question 3.12 on page 20.**

  A random sample of claims with $n = 20$ from a distribution believed to be Pareto with parameters $\alpha$ and $\lambda$ gives values such that $\sum x = 1,508$ and $\sum x^2 = 257,212$. Estimate $\alpha$ and $\lambda$ using the method of moments.
• For the generalized Pareto, things are not quite so easy.
As for estimation, the CDF does not exist in closed form, so the method of percentiles is not available. ML can be used, but again suitable computer software is required. The method of moments can provide initial estimates for any iterative scheme.
3.6 The lognormal distribution

- Estimation for the lognormal distribution is straightforward since $\mu$ and $\sigma^2$ may be estimated using the log transformed data.

- Alternatively the method of moments can be used.

- Example on page 21.
3.7 The Weibull and Burr distributions

- For the Weibull distribution, neither the method of moments nor maximum likelihood is elementary to apply if both $c$ and $\gamma$ are unknown (although if a computer is available, as would be the case in practice, the equations are simple enough). In the case where $\gamma$ has the known value $\gamma^*$, maximum likelihood is easy enough.

c.f. The Weibull distribution and the exponential distribution.
• The distribution function of the $W(c, \gamma)$ distribution is an elementary function, and a simple method of estimation of both $c$ and $\gamma$ can be based on this fact.
  The method involves equating selected sample percentiles to the distribution function; for example, equate the quartiles, the 25th and 75th percentiles, to the population quartiles.
  This corresponds to the way in which sample moments are equated to population moments in the method of moments.

• This method will be referred to as the method of percentiles.
• In the method of moments the first two moments are used if there are two unknown parameters, and this seems intuitively reasonable.
In a similar fashion, the median would be used if there were one parameter to estimate.
With two parameters, the best procedure is less clear, but the lower and upper quartiles seem a sensible choice.

• **Example on pages 22-23.**

  Estimate $c$ and $\gamma$ in the Weibull distribution using the method of percentiles, where the first sample quartile is 401 and the third sample quartile is 2,836,75.

• **Example on pages 23-24.**
• For the Burr distribution, since the CDF exists in closed form, it may be possible to fit the Burr distribution to data by using the method of percentiles.
ML will certainly require the use of computer software that allows non-linear optimization.

• **Question 3.14 on pages 24.**
4 Goodness-of-fit-tests

• One method of testing whether a given loss distribution provides a good model for the observed claim amounts is to apply a $\chi^2$ goodness-of-fit test.

• Example on pages 25-26.
5 Mixture distributions

- The exponential distribution is one of the simplest models for insurance losses. Suppose that each individual in a large insurance portfolio incurs losses according to an exponential distribution. Practical knowledge of almost any insurance portfolio reveals that the means of these various distributions will differ among the policyholders. Thus the description of the losses in the portfolio is that each loss follows its own exponential distribution, i.e. the exponential distributions have means that differ from individual to individual.
• So rather than having identically distributed claim amounts $X \sim \text{Exp}(\lambda)$, what we have are claim amounts whose distributions are conditional on the value of $\lambda$, i.e. $X|\lambda \sim \text{Exp}(\lambda)$.

• A description of the variation among the individual means must now be found. One way to do this is to assume that the exponential means themselves follow a distribution.
• In the exponential case, it is convenient to make the following assumption.
Let \( \lambda_i = 1/\theta_i \) be the reciprocal of the mean loss for the \( i \)-th policyholder. Assume that the variation among the \( \lambda_i \) can be described by a known gamma distribution \( Ga(\alpha, \delta) \), i.e. assume that
\[
\lambda \sim Ga(\alpha, \delta)
\]
where
\[
f(\lambda) = \frac{\delta^\alpha}{\Gamma(\alpha)} \lambda^{\alpha-1} \exp(-\delta \lambda), \quad \lambda > 0.
\]

• Notice that the purpose here is to describe the aggregate losses over the whole portfolio.
• In this problem of describing the losses over the whole portfolio, the \( Ga(\alpha, \delta) \) distribution is used to average the exponential distributions.

• The \( Ga(\alpha, \delta) \) distribution is referred to as the **mixing distribution** and the resulting loss distribution as a **mixture distribution**.

• To find the overall distribution of claim amounts, we need to work out the marginal distribution of \( X \).
The marginal distribution of $X$ is

$$f_X(x) = \int f_{X,\lambda}(x, \lambda) d\lambda$$

$$= \int f_\lambda(\lambda) f_{X|\lambda}(x|\lambda) d\lambda$$

$$= \int_0^\infty \frac{\delta^\alpha}{\Gamma(\alpha)} \lambda^{\alpha-1} \exp(-\delta \lambda) \times \lambda \exp(-\lambda x) d\lambda$$

$$= \frac{\delta^\alpha}{\Gamma(\alpha)} \int_0^\infty \lambda^{\alpha} \exp[-(x + \delta)\lambda)]d\lambda$$

$$= \frac{\delta^\alpha}{\Gamma(\alpha)} \times \frac{\Gamma(\alpha + 1)}{(x + \delta)^{\alpha+1}} (aGa(\alpha + 1, x + \delta) \text{integral})$$

$$= \frac{\alpha \delta^\alpha}{(x + \delta)^{\alpha+1}}$$

which can be recognized as the Pareto distribution, $Pa(\alpha, \delta)$. 
• This result gives a very nice interpretation of the Pareto distribution: \( Pa(\alpha, \delta) \) arises when exponentially distributed losses are averaged using a \( Ga(\alpha, \delta) \) mixing distribution.

• **Example on page 29.**
• If we regard the generalized Pareto as the result of mixing together $\text{Gamma}(k, \lambda)$ distributions whose $\lambda$ parameters come from a $\text{Gamma}(\alpha, \delta)$ distribution, then we can find the mean of the generalized Pareto by calculating:

$$E(X) = \int_0^\infty \frac{k}{\lambda} \times \frac{1}{\Gamma(\alpha)} \delta^\alpha \lambda^{\alpha-1} e^{-\delta \lambda} d\lambda$$

where $k/\lambda$ is the mean of the $\text{Gamma}(k, \lambda)$ distribution.
• **Question 3.15 on page 30.**

The annual number of claims from an individual policy in a portfolio has a $Poisson(\theta)$ distribution. The variability in $\theta$ among policies is modelled by assuming that over the portfolio, individual values of $\theta$ have a $\Gamma(\alpha, \delta)$ distribution. Derive the mixture distribution for the annual number of claims from each policy in the portfolio.

• **Question 3.16 on page 30.**

Claim numbers from individual policies in a portfolio have a $Bin(n, p)$ distribution. The parameter $p$ varies over the portfolio with a $Beta(\alpha, \beta)$ distribution. Find the mixture distribution.