Queen Mary
University of London

MTH5126 Statistics for Insurance

Dr Lei Fang

lei.fang@qmul.ac.uk

Week 1

Great credit and thanks to D. Boland, G. Ng and F. Parsa, previous lecturers, for their excellent work in producing the original version of these notes.

Loss distributions

Loss distributions

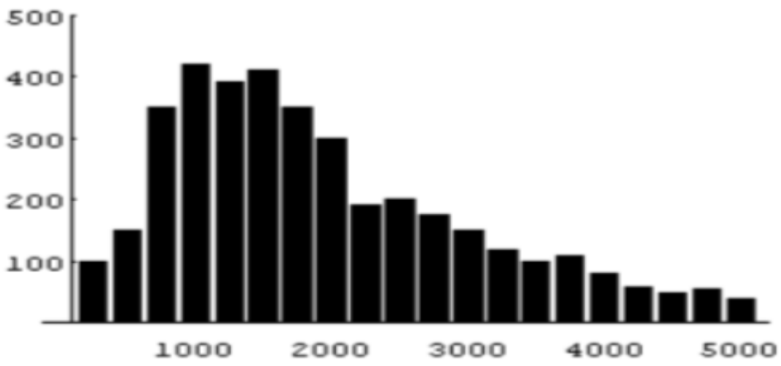
Loss distributions are statistical distributions that are used to model claim amounts.

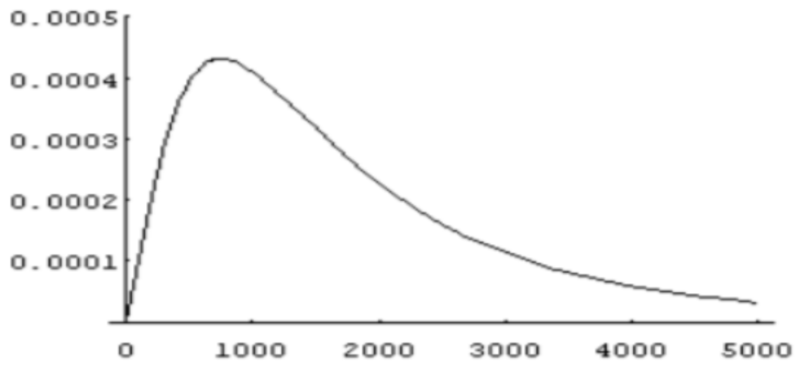
The key assumption in all the models studied here is that:

the occurrence of a claim and the amount of a claim can be studied separately.

Thus, a claim occurs according to some simple model for events occurring in time, then the amount of the claim is chosen from a distribution describing the claim amount.

Loss distributions





1. The frequency of claim amounts when plotted against size might look like the above graph.
2. Statistical distributions can be used to approximate this graph.

1. We might decide to use a loss distribution like the above as an approximation to the claims arising in the graph in the left.
2. The object is to describe the variation in claim amounts by finding a loss distribution that adequately describes the claims that actually occur.

Loss distributions

- In practice, the exact claims distribution will never be known.
- 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 now be estimated using the claim amount records by an appropriate method such as *maximum likelihood*.
- Complications will arise if large claims have been limited (by *reinsurance*) or some small claims have not been lodged (*policy excess*).
- A lot of studies have been performed on the kind of distribution that can be used to describe the variation in claim amounts.
- The general conclusion is that claims distributions tend to be **positively skewed** and **long-tailed**.

- *Can you think of a statistical distribution that is positively skewed?*

Loss distributions

- Moment Generating Functions for the Exponential, Gamma and Normal distributions are reasonably straightforward to derive.
- While these three distributions can be used to model losses, in practice there are a wide variety of other distributions that may also be used:

  1. The Lognormal distribution
  2. The Pareto distribution
  3. The Burr distribution
  4. The Weibull distribution

Moment generating functions (Revision)

The **n^{th} moment** of a random variable X is defined to be **$E[X^n]$**.
The **n^{th} central moment** of X is defined to be **$E[(X - E(X))^n]$**.

Definition of a MGF: The moment generating function (MGF) of a random variable X is a function $M_X(t)$ defined as:

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

We say that the MGF of X exists if there is a positive constant *a* such that $M_X(t)$ is finite for all $t \in [-a, a]$.

Why is the MGF useful?

1. The MGF gives us all the moments of X, hence the name!
2. The MGF (if it exists) uniquely determines the distribution. If 2 random variables have the same MGF, then they must have the same distribution. So if you find the MGF of a random variable, you will have determined its distribution.

Moment generating functions (Revision)

The **skewness** and **coefficient of skewness** of a random variable X are respectively:

$$\text{skew}(X) = E[(X-\mu)^3]$$

where μ is the mean of the distribution and

$$\text{coeff of skew}(X) = \frac{\text{skew}(X)}{[\text{var}(X)]^{3/2}}\ .$$

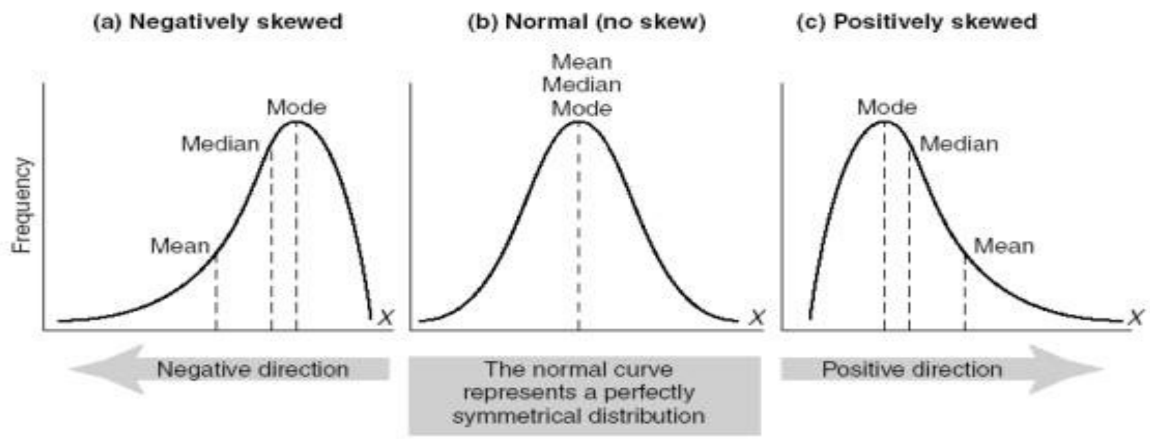


FIGURE 15.6 Examples of normal and skewed distributions

Queen Mary
University of London

MGF

Example 1: MGF of the Uniform(0, 1) random variable

If Y is a Uniform(0, 1) random variable, find its MGF.

Answer:

$$M_Y(t) = E[e^{tY}] = \int_0^1 e^{ty}\, dy = \frac{e^t - 1}{t}$$

Note that we always have

$$M_Y(0) = E[e^{0.Y}] = 1$$

So $M_Y(t)$ is well-defined for all $t \in \mathbb{R}$.

MGF

Finding moments from the MGF

Remember the Taylor series for e^x: for all $x \in \mathbb{R}$, we have:

$$e^x = 1 + x + \frac{x^2}{2!} + \frac{x^3}{3!} + \ldots = \sum_{k=0}^{\infty} \frac{x^k}{k!}$$

Now we can write:

$$e^{tX} = \sum_{k=0}^{\infty} \frac{(tX)^k}{k!} = \sum_{k=0}^{\infty} \frac{X^k t^k}{k!}$$

And so we have:

$$M_x(t) = E[e^{tX}] = \sum_{k=0}^{\infty} E[X^k] \frac{t^k}{k!}$$

We conclude that the kth moment of X is the coefficient of $\frac{t^k}{k!}$ in the Taylor series of $M_X(t)$.

Queen Mary University of London

MGF

Example 2: Finding the kth moment of the Uniform(0,1) r.v. from its MGF

If $Y \sim Uniform(0,1)$, find $E[Y^k]$ using $M_Y(t)$.

Answer:

Having $M_Y(t)$ from Example 1, we have:

$$M_Y(t) = \frac{e^t - 1}{t} = \frac{1}{t}\left(\sum_{k=0}^{\infty} \frac{t^k}{k!} - 1\right) = \frac{1}{t}\left(\sum_{k=1}^{\infty} \frac{t^k}{k!}\right) = \sum_{k=1}^{\infty} \frac{t^{k-1}}{k!} = \sum_{k=0}^{\infty} \frac{1}{k+1} \frac{t^k}{k!}$$

So the coefficient of $\frac{t^k}{k!}$ in the Taylor series for $M_Y(t)$ is $\frac{1}{k+1}$ and:

$$E[Y^k] = \frac{1}{k+1}$$

MGF
Generating moments with calculus

From calculus we remember that the coefficient of $\frac{t^k}{k!}$ in the Taylor series of $M_X(t)$ is obtained by taking the kth derivative of $M_X(t)$ and evaluating it at t = 0, that is $E[X^k] = \frac{d^k}{dt^k} M_x(t)|_{t=0}$

Deriving moments

We can obtain all moments of X^k from its MGF:

$$M_X(t) = \sum_{k=0}^{\infty} E[X^k] \frac{t^k}{k!}$$

$$E[X^k] = \frac{d^k}{dt^k} M_X(t)|_{t=0}$$

The Exponential Distribution

A random variable X has an Exponential distribution with parameter $\lambda > 0$ if it has CDF:

$$F(x) = 1 - e^{-\lambda x}, \qquad x > 0$$

where CDF is the cumulative distribution function and is defined by $P(X \leq x)$.

- We write $X \sim Exp(\lambda)$
- For the exponential distribution the PDF is: $f(x) = \lambda e^{-\lambda x}, x > 0$
- We know that the mean is $1/\lambda$ and the variance is $1/\lambda^2$ but how are these derived?

The Exponential distribution

MGF

Question: Find the MGF of X, $M_X(t)$ and all of its moments $E[X^k]$.

Answer:

$$
\begin{aligned}
M_X(t) &= E[e^{tX}] \\
&= \int_0^\infty \lambda e^{-\lambda x} e^{tx}\, dx \\
&= \left[\frac{-\lambda}{\lambda - t} e^{-(\lambda - t)x}\right]_0^\infty, t < \lambda \\
&= \frac{\lambda}{\lambda - t}, t < \lambda
\end{aligned}
$$

The Exponential distribution

MGF

Answer:

This is often written in the form $\frac{1}{1-\frac{t}{\lambda}}$ for $t < \lambda$ because then it is easy to see:

$$= \sum_{k=0}^{\infty} \left(\frac{t}{\lambda}\right)^k$$

for $|\frac{t}{\lambda}| < 1$

$$= \sum_{k=0}^{\infty} \frac{k!}{\lambda^k} \frac{t^k}{k!}$$

$$\therefore E[X^k] = \frac{k!}{\lambda^k}$$

for $k = 0, 1, 2, \ldots$

Exercise: Use the MGF for the Exponential to derive the mean and variance.

The Gamma distribution

The random variable X has a gamma distribution with parameters $\alpha > 0$ and $\lambda > 0$ if it has PDF:

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

In that case we write $X \sim Ga(\alpha, \lambda)$ or $(X \sim Gamma(\alpha, \lambda))$

The mean and variance of X are:

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

and

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

The Gamma distribution

MGF

Question: Show that the MGF of the gamma distribution is $M_X(t) = (1 - \frac{t}{\lambda})^{-\alpha}$

Answer:

$$M_X(t) = E[e^{tX}] = \int_0^\infty e^{tx} \frac{1}{\Gamma(\alpha)} \lambda^\alpha x^{\alpha-1} e^{-\lambda x} dx$$

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

$$= \left(\frac{\lambda}{\lambda - t}\right)^\alpha \int_0^\infty \frac{1}{\Gamma(\alpha)} (\lambda - t)^\alpha x^{\alpha-1} e^{-(\lambda-t)x} dx$$

This makes the whole integral the PDF of a $Gamma(\alpha, \lambda - t)$ distribution, which is always 1. So we have:

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

Queen Mary University of London

The Gamma distribution

Estimating Gamma probabilities using Chi-square distribution

- Because there is no closed form for the CDF of a Gamma distribution, it is not easy to find Gamma probabilities directly.
- However, a rough estimate can be obtained by using the Chi-square distribution.
- We state the following result without proof. This result can be used in all instances for finding probabilities from the Gamma distribution as direct integration is not usually possible.

Relationship between gamma and chi-squared distributions

If $X \sim Gamma(\alpha, \lambda)$ and 2α is an integer, then:

$$2\lambda X \sim \chi^2_{2\alpha}$$

If Y follows a chi-square distribution with degree of freedom 2α, then

$$M_Y(t) = (1 - 2t)^{-\alpha} \quad \text{for } t < \frac{1}{2}$$

The Gamma distribution

Estimating gamma probabilities using chi-square distribution

Question:

(i) Show that if X has a $Gamma(10, 4)$ distribution, then the random variable $Y = 8X$ has a χ^2_{20} distribution.

(ii) Hence find approximately the probability that X is greater than 4.375.

Answer (i):

The MGF of a $Gamma(10, 4)$ distribution is $(1 - \frac{t}{4})^{-10}$.

So the MGF of $Y = 8X$ is:

$M_Y(t) = E(e^{tY}) = E(e^{8tX})$

$= M_X(8t) = (1 - 2t)^{-10}$

This is the MGF of a χ^2_{20} distribution.

So by using the uniqueness property of MGFs, Y has this distribution.

The Gamma distribution

Estimating gamma probabilities using chi-square distribution

Answer (ii): Find $P(X > 4.375)$

Expressing then the required probability in terms of Y:

$$P(X > 4.375) = P(Y > 35) = P(\chi^2_{20} > 35)$$

From the statistical tables, this is

1 - 0.9799 = 0.0201

Check using R. Find P(X > 4.375) by typing in the following and clicking ‘Run’.

```
pgamma(4.375,10,4,lower.tail=FALSE)
```

What is the R output that you get?

Answer: `0.02010428`

Queen Mary University of London

The Normal distribution

MGF

- Since loss distributions tend to be positively skewed, the Normal distribution is of limited use for modelling loss distributions because of its symmetry.
- Deriving the formula for the MGF of a $N(0, 1)$ random variable is left as a worksheet exercise.
- Here, we state without proof the MGF of a $N(\mu, \sigma^2)$ random variable.

$$M_X(t) = \exp\left[\mu t + \frac{1}{2}\sigma^2 t^2\right]$$

The Lognormal distribution

The definition of the lognormal distribution is quite straightforward.

X has a lognormal distribution if $logX$ has a normal distribution.

When $logX \sim N(\mu, \sigma^2), X \sim LN(\mu, \sigma^2)$

- The mean and variance are:

$$mean = exp(\mu + \frac{1}{2}\sigma^2)$$

$$variance = exp(2\mu + \sigma^2)[exp(\sigma^2) - 1]$$

- Lognormal probabilities can be evaluated by expressing them as standard normal probabilities and looking up the values in the *Tables.*

The Pareto distribution

A random variable X has the Pareto distribution with parameters $\alpha > 0$ and $\lambda > 0$ if it has CDF:

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

In that case we write $X \sim Pa(\alpha, \lambda)$

It is easily checked by differentiating $F(x)$ with respect to x that the Pareto distribution has PDF:

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

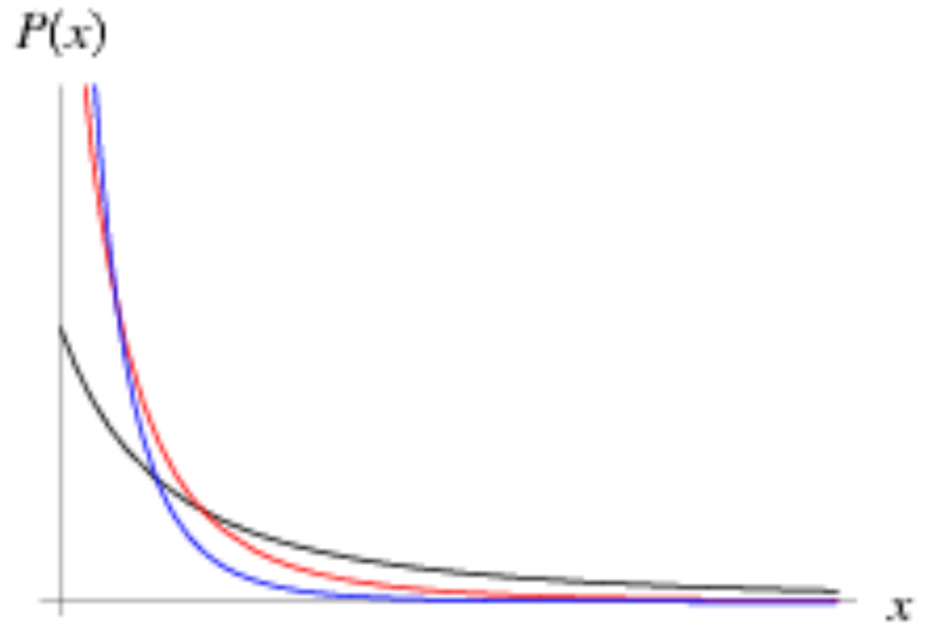
The Pareto distribution

The Pareto distribution is both positively skewed and long-tailed.
The mean is:

$$\frac{\lambda}{\alpha - 1}$$

and the variance is:

$$\frac{\lambda^2 \alpha}{(\alpha - 1)^2(\alpha - 2)}$$



P(x) in the graph here refers to probability density function.

Queen Mary University of London

The Burr distribution (transformed Pareto)

The CDF of the Pareto distribution $Pa(\alpha, \lambda)$ is:

$F(x) = 1 - \frac{\lambda^\alpha}{(\lambda+x)^\alpha}, x > 0$

We can introduce a futher parameter $\gamma > 0$ by setting:

$F(x) = 1 - \frac{\lambda^\alpha}{(\lambda+x^\gamma)^\alpha}, x > 0$

- This is the CDF of the Burr distribution (also called the transformed Pareto).
- The addition of the γ parameter gives extra flexibility when we are fitting to data.

The Burr distribution (transformed Pareto)

Example: Deriving the median

(i) Find a general formula for the median of the Burr distribution.
(ii) Find the median of the Burr distribution that has parameters $\lambda = 800$, $\alpha = 4.5$ and $\gamma = 0.75$

Answer (i):

(i) We want to solve the equation $F(x) = \frac{1}{2}$. This means:

$$\left(\frac{\lambda}{\lambda + x^{\gamma}}\right)^{\alpha} = \frac{1}{2}$$

or:

$$x^{\gamma} = \lambda\left(2^{\frac{1}{\alpha}} - 1\right)$$

$\Longrightarrow$

$$x = \left[\lambda\left(2^{\frac{1}{\alpha}} - 1\right)\right]^{\frac{1}{\gamma}}$$

The Burr distribution (transformed Pareto)

Example: Deriving the median

(i) Find a general formula for the median of the Burr distribution.
(ii) Find the median of the Burr distribution that has parameters $\lambda = 800, \alpha = 4.5$ and $\gamma = 0.75$

Answer (ii):

(ii) Substituting in $\lambda = 800, \alpha = 4.5, \gamma = 0.75$ we get:

$$x = \left[800\left(2^{\frac{1}{4.5}} - 1\right)\right]^{\frac{1}{0.75}}$$

$$= 680.414.$$

- As an exercise you should calculate the mean of this distribution to 1,782.7.
- There is a large difference to the median.
- What does this mean for the skewness of this distribution?

Queen Mary
University of London

The Weibull distribution

- The Pareto distribution is a distribution with an upper tail that tends to 0 as a power of x.
- This gives a distribution with a much heavier tail than the exponential.
- The expressions for the upper tails of the exponential and Pareto distributions are:

Exponential:

$$P(X > x) = e^{-\lambda x}$$

Pareto:

$$P(X > x) = \left(\frac{\lambda}{\lambda + x}\right)^{\alpha}$$

- So if we want to choose a model with a thick tail so as not to underestimate the probability of a large claim, we might well choose the Pareto distribution to model our claims (assuming that it is a suitable distribution in other respects).
- However, these are not the only types of tail.

The Weibull distribution

There is another possibility:

Set:

$$P(X > x) = e^{-\lambda x^{\gamma}}, \gamma > 0$$

There are now 2 cases:

1. If $\gamma < 1$ then a distribution with a tail intermediate in weight between the exponential and Pareto will be obtained.

2. If $\gamma > 1$ the upper tail will be lighter than the exponential.

(Note that $\gamma = 1$ is just the exponential distribution itself).

This distribution is called the Weibull distribution - it is quite a flexible distribution and can be used as a model for losses in insurance, usually with $\gamma < 1$.

The Weibull distribution

A random variable X has a Weibull distribution with parameters $c > 0$ and $\gamma > 0$ if it has CDF:

$F(x) = 1 - e^{-cx^{\gamma}}, x > 0$

In this case we write $X \sim W(c, \gamma)$

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

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

The mean is $c^{-\frac{1}{\gamma}} \Gamma(1 + \gamma^{-1})$

Queen Mary
University of London

Estimation

- The more important statistical distributions from the general insurance world have been introduced in the preceding couple of lectures.
- Having met these distributions we note that each distribution involves one or more parameters which determine the distribution's location and spread.
- For distributions like the Weibull and Burr we had an additional parameter to determine the shape.
- The parameters of a distribution are seldom known a priori and they need to be estimated from claims data before the distribution itself can be applied to a particular problem.
- The function of the observations which we choose, to estimate a parameter, is known as an **estimator** and the numerical value obtained from it, using a particular set of data is called an **estimate**.
- Often, several different functions will suggest themselves as estimators (for example the sample mean and the sample median can both be used to estimate the mean of a Normal population).
- We need some criteria to decide which to use.

Estimation

Estimator criteria

- The estimator should be *unbiased*, so that its expectation is equal to the true value of the parameter. Thus, on average, the estimate is equal to the underlying parameter. The estimator should not provide estimates which are, on average, too high or too low.
- The estimator should be *consistent*. By this we mean that, for any small quantity ε, the probability that the absolute value of the deviation from the true parameter value is less than ε and tends to 1 as $n \to \infty$. Thus, for an estimate based on a large number of observations, there is a very small probability that its value will differ seriously from the true value of the parameter.
- The estimator should be *efficient*. That is, the variance of the estimator should be minimal.

Estimation - Point estimates

Statisticians have devised different standard methods for producing point estimates of parameters, including:

1) The method of moments
2) Maximum likelihood
3) Method of percentiles

- When sample sizes are large, they all tend to provide more or less the same answers, even in more complicated cases.
- In other circumstances, however, markedly different results can emerge, and the three criteria on the previous slide are often used by statisticians in deciding which estimator to use.
- The maximum likelihood method usually provides estimators which are quite satisfactory as far as the above criteria are concerned.
- However, the method frequently produces equations which are rather awkward to solve and, for this reason, the simpler method of moments is often preferred, despite the fact that the resulting estimators may be sub-optimal in respect of the above three criteria.
- Of the three methods listed, the method of moments is perhaps the most readily understood.

Method of moments

✓ To obtain a method of moments estimator for a parameter of the distribution we are trying to fit, We equate the corresponding sample and population moments of the distribution.

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:
$m_j = E(X^j|\theta)$, which is a function of the unknown parameter, θ, being estimated.
n = the sample size
x_i = the i^{th} value in the sample

The estimate for the parameter θ, 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.

Method of moments

Example 1: Estimating one parameter

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:

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

- The LHS is the 1st non-central moment for the population
- The RHS is the 1st non-central moment for the sample
- *i.e.* we would equate the first non-central moments for the population and the sample.

Method of moments

Example 2: Estimating two parameters

If we are trying to find estimates for two parameters (e.g. if we are fitting a Gamma distribution and need to obtain estimates for both parameters), we would solve the simultaneous equations:

$$E(X) = \frac{1}{n}\sum_{i=1}^{n} x_i \quad \text{and} \quad E(X^2) = \frac{1}{n}\sum_{i=1}^{n} x_i^2$$

- *i.e.* we would equate the first two non-central moments for the population and the sample.

In fact, in the two-parameter case, estimates of parameters are often obtained by equating population and sample means and variances. If we use the n-denominator sample variance:

$$s_n^2 = \frac{1}{n}\sum_{i=1}^{n}(x_i - \bar{x})^2 = \frac{1}{n}\left[\sum_{i=1}^{n} x_i^2 - n\bar{x}^2\right]$$

- this would give the same estimates as would be obtained by equating the first two non-central moments.

Method of moments

Example 3: Estimating the two parameters of a Lognormal distribution

Question: The table shows the claim sizes of a sample of 100 claims on an insurance company. Assuming that the Log-normal distribution is a suitable model:

a) Obtain estimates of its parameters μ and σ

b) Estimate the probability that a particular claim exceeds £4,000.

Claim size (£)	Number of claims
0-400	2
400-800	24
800-1200	32
1200-1600	21
1600-2000	10
2000-2400	6
2400-2800	3
2800-3200	1
3200-3600	1
over 3600	0
Total	100

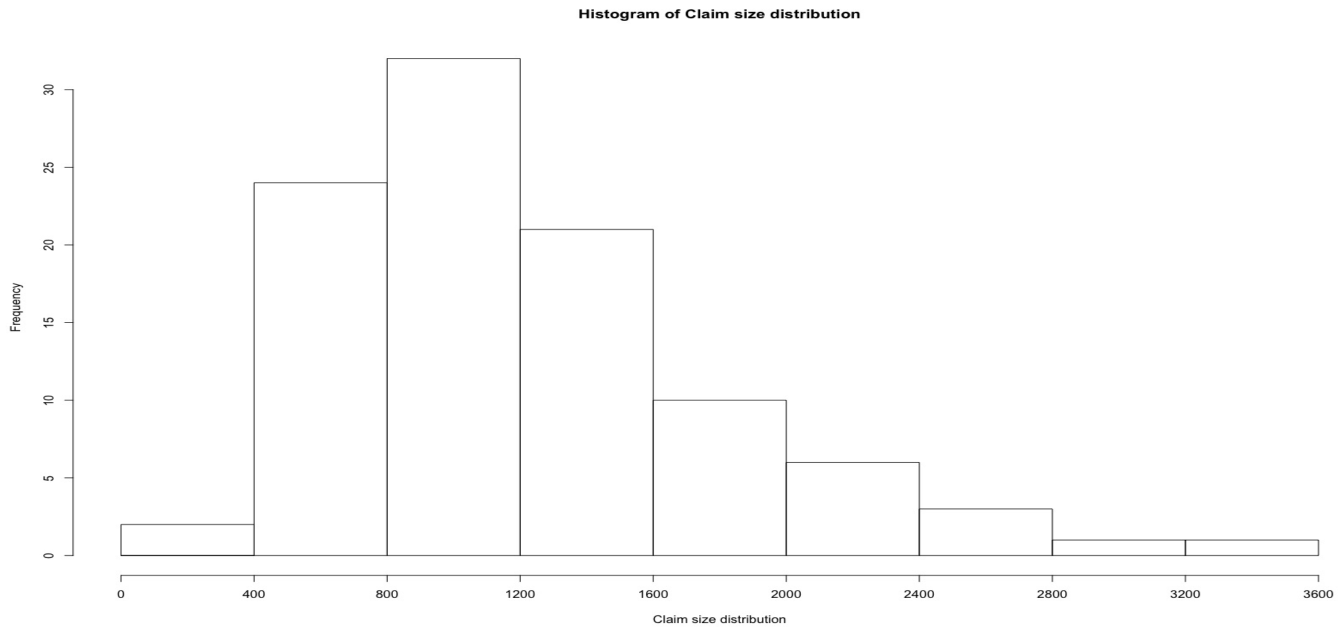
Queen Mary University of London

Method of moments

Example 3: Estimating the two parameters of a Lognormal distribution

Answer:

The skewness of the observed claim size distribution is evident when the histogram representing the data is drawn:



Method of moments

Example 3: Estimating the two parameters of a Lognormal distribution

Answer:

a) **Sample mean**

Assuming that the number of claims in the right-hand column can be said to refer to claims with sizes equal to the mid-point of the respective claim size interval (which may not be a very accurate assumption with a skew distribution like this one), we obtain the mean claim size of the observed distribution as follows:

$$\text{Sample mean} = £\left(200 \times \frac{2}{100} + 600 \times \frac{24}{100} + \ldots + 3400 \times \frac{1}{100}\right)$$

$$= £1216$$

Queen Mary University of London

Method of moments

Example 3: Estimating the two parameters of a Lognormal distribution

Answer:

a) **Sample variance**

The variance of the observed claim size distribution can be calculated as follows:

$$s_n^2 = \frac{1}{n}\sum_{i=1}^{n}(x_i - \overline{x})^2 = \frac{1}{n}\left[\sum_{i=1}^{n}x_i^2 - n\overline{x}^2\right]$$

$$= \sum_x x^2 P(X = x) - (\text{mean})^2$$

$$= \left(200^2 \times \frac{2}{100} + 600^2 \times \frac{24}{100} + \ldots + 3400^2 \times \frac{1}{100}\right) - 1216^2$$

$$= 362944$$

Method of moments

Example 3: Estimating the two parameters of a Lognormal distribution

Answer:

a) Population mean and variance

The mean and variance of a *Lognormal* (μ, σ^2) distribution are:

$$mean = exp(\mu + \frac{1}{2}\sigma^2)$$

$$variance = exp(2\mu + \sigma^2)[exp(\sigma^2) - 1]$$

Method of moments

Example 3: Estimating the two parameters of a Lognormal distribution

Answer:

a) Now we can equate our sample moments to our population moments.

$$exp(\mu + \frac{1}{2}\sigma^2) = 1216$$

$$exp(2\mu + \sigma^2)[exp(\sigma^2) - 1]) = 362944$$

Squaring the first of these equations and dividing the second equation by this square we obtain:

$$exp(\sigma^2) - 1 = 0.2455$$

from which,

$$\sigma = 0.469$$

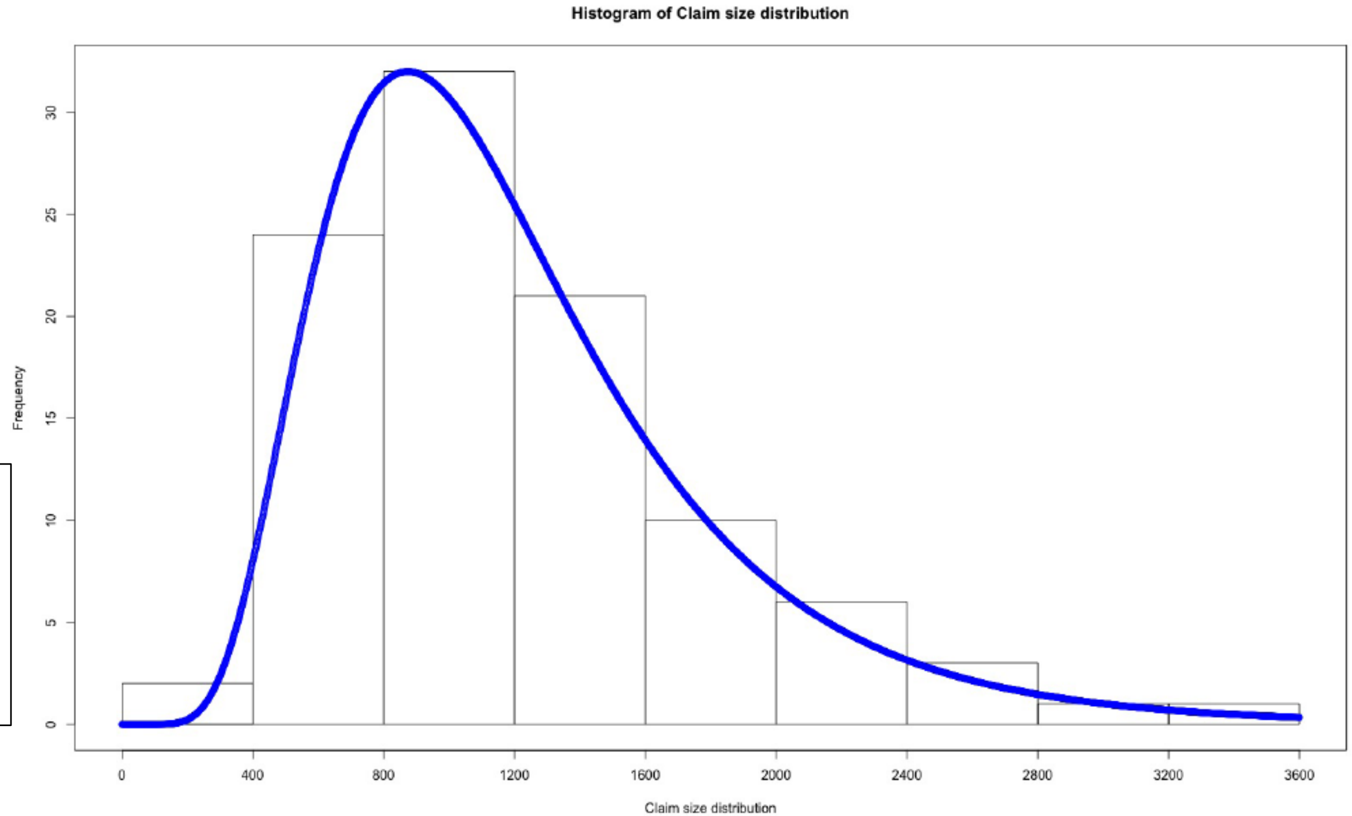
$$\mu = 6.993$$

Method of moments

Example 3: Estimating the two parameters of a Lognormal distribution

Answer: a)

The blue curve shows the fitted distribution with
$\mu = 6.993$,
and
$\sigma = 0.469$



Method of moments

Example 3: Estimating the two parameters of a Lognormal distribution

Answer: b)

Estimate the probability that a particular claim exceeds £4,000.

From part ***a:***

$$X \sim Lognormal(6.993,\ 0.469^2)$$

so

$$\ln X \sim Normal(6.993,\ 0.469^2)$$

we can write

$$\frac{\ln X - 6.993}{0.469} \sim Normal(0,\ 1).$$

Therefore

$$P(X > 4{,}000) = P\left(\frac{\ln X - 6.993}{0.469} > \frac{\ln 4{,}000 - 6.993}{0.469}\right)$$

$$= P(Z > 2.774)$$

$$= 0.0028$$

where Z is the standard normal r.v.

Queen Mary University of London

Maximum likelihood estimation

The likelihood function of a random variable, X, is the probability (or PDF) of a specific observation given a certain value of the parameter, θ.

The maximum likelihood estimate (MLE) is the one that yields the highest probability (or PDF), i.e. that maximises the likelihood function.

For our sample the likelihood function $L(\theta)$ can be expressed as:

$$L(\theta) = \prod_{i=1}^{n} P(X = x_i|\theta)$$

for a discrete random variable, X, or

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

for a continuous random variable, X.

Maximum likelihood estimation

- To determine the Maximum Likelihood Estimate (MLE), the likelihood function needs to be maximised.
- Often it is practical to consider the Log-likelihood function:

$$l(\theta) = log(L(\theta)) = \sum_{i=1}^{n} logP(X = x_i|\theta)$$

for a discrete random variable, X, or

$$l(\theta) = log(L(\theta)) = \sum_{i=1}^{n} logf(x_i|\theta)$$

for a continuous random variable, X.

Maximum likelihood estimation

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

$$\frac{d}{d\theta} l(\hat{\theta}) = 0$$

Question: What if there is more than one parameter that you are trying to estimate?

Answer: Where there is more than one parameter, the MLEs for each parameter can be determined by taking partial derivatives of the log-likelihood function and setting each to zero.

Maximum likelihood estimation

4 steps to finding a MLE

Step 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

$$f(x_1|\theta)\, f(x_2|\theta) \ldots f(x_n|\theta),$$

where $f(x|\theta)$ is the PDF of $X|\theta$, a continuous random variable or

$$P(X = x_1|\theta)P(X = x_2|\theta), \ldots, P(X = x_n|\theta)$$

in the case where $X|\theta$ is a discrete random variable.

Queen Mary University of London

Maximum likelihood estimation

4 steps to finding a MLE

Step 2 Take natural logs.
This should simplify the resulting algebra.

Step 3 Maximise 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 equal to zero.

Step 4 Solve the resulting equations to find the MLEs of your parameters.

You should check that the values you have found do indeed maximise the function by taking second derivatives - forgetting to do this is a common student mistake.

Maximum likelihood estimation

The Exponential distribution

Example:

An insurance company uses an Exponential distribution to model the cost of repairing insured vehicles that are involved in accidents. Let λ denote the parameter of the Exponential distribution. Find the maximum likelihood estimate of λ, given that the average cost of repairing a sample of 1,000 vehicles was £2,200.

Answer:

Let $X_1, X_2, \ldots, X_n$ denote the individual repair costs (where n = 1000). The likelihood of obtaining these values for the costs, if they come from an Exponential distribution with parameter λ is:

$$L = \prod_{i=1}^{n} \lambda e^{-\lambda x_i} = \lambda^n e^{-\lambda \sum x_i} = \lambda^n e^{-\lambda n\bar{x}}$$

Queen Mary University of London

Maximum likelihood estimation

The Exponential distribution

Answer (continued):

To find the MLE, we need to find the value of λ that maximises the likelihood, or, alternatively, the value that maximises the log-likelihood.

$$log(L) = nlog(\lambda) - \lambda n\bar{x}.$$

Differentiating to look for stationary points:

$$\frac{\partial}{\partial\lambda} log(L) = \frac{n}{\lambda} - n\bar{x}$$

Setting this to zero gives :

$$\hat{\lambda} = \frac{1}{\bar{x}}$$

ie

$$\hat{\lambda} = \frac{1}{2200}$$

Queen Mary
University of London

Maximum likelihood estimation

The Exponential distribution

Answer (continued):

The second derivative is

$$\frac{\partial^2}{\partial \lambda^2} log(L) = -\frac{n}{\lambda^2} < 0$$

Which shows that this is a *maximum*.

Maximum likelihood estimation

The Gamma distribution

The moments have a simple form and so the method of moments is very easy to apply.

The MLEs for a gamma distribution cannot be obtained in closed form (i.e. in terms of elementary functions) but the moment estimators can be used as initial estimators in the search of MLEs.

Maximum likelihood estimation

The Normal distribution

Both the method of moments and MLE are straightforward to apply in this case.
Both give the obvious answers:

$$\hat{\mu} = \bar{y}$$

$$\hat{\sigma}^2 = \frac{1}{n}\sum_{i=1}^{n}(y_i - \bar{y})^2$$

Maximum likelihood estimation

The Lognormal distribution

Estimation for the lognormal distribution is straightforward since μ and σ^2 may be estimated using the log-transformed data.

Let $x_1, x_2, \ldots, x_n$ be the observed values and let $y_i = log x_i$.

The MLE's of μ and σ^2 are $\bar{y}$ and s_y^2, where the subscript y signifies a sample variance (n-denominator) computed on the y values.

$$\hat{\sigma}^2 = \frac{1}{n} \sum_{i=1}^{n} (y_i - \bar{y})^2$$

Maximum likelihood estimation

The Pareto distribution

The method of moments is very easy to apply in the case of the Pareto distribution but the estimators obtained in this way will tend to have rather large standard errors.

This is mainly because S^2, the sample variance, has a very large variance.

However, the method does provide initial estimators 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.

Example:

A random sample of claims with $n = 20$ from a distribution believed to be Pareto with parameters α and λ gives values such that:

$\sum x = 1,508$ and $\sum x^2 = 257,212$.

Estimate α and λ using the method of moments.

Maximum likelihood estimation

The Pareto distribution

Answer:

We first work out the population moments. The first and second moments of the Pareto distribution are:

$$E(X) = \frac{\lambda}{\alpha - 1}$$

Rearranging the variance formula to find $E(X^2)$, we have:

$$E(X^2) = \text{var}(X) + [E(X)^2] = \frac{2\lambda^2}{(\alpha-1)(\alpha-2)}$$

Queen Mary
University of London

Maximum likelihood estimation

The Pareto distribution

Answer (continued):

We next equate the population moments with the sample moments.

$$\frac{\lambda}{\alpha - 1} = \frac{1508}{20}$$

and

$$\frac{2\lambda^2}{(\alpha - 1)(\alpha - 2)} = \frac{257,212}{20}$$

Substituting the first equation into the second, we get:

$$\frac{2 \times 75.4^2(\alpha - 1)}{\alpha - 2} = 12,860.6$$

Solving this equation for α we get $\alpha = 9.630$ and hence $\lambda = 650.676$

Maximum likelihood estimation

The Weibull and Burr distribution

Neither the method of moments nor maximum likelihood is straightforward to apply if both of the parameters *c* and γ are unknown.

If we at least know γ then MLE can be used.

However, when we don't, we can use a method called the ***method of percentiles***.

Method of percentiles

- **This involves equating sample percentiles to the distribution function.**
- **For example we can equate the 25th and 75th sample percentiles to the population quartiles.**

This corresponds to the way in which sample moments are equated to population moments in the method of moments.

The method of percentiles

- In the method of moments the first two moments are used if there are two unknown parameters.
- In a similar fashion when using the method of percentiles, the median would be used if there was just one parameter to estimate.
- With two parameters the choice is less clear but lower and upper quartiles is a widely-accepted choice.

Example:

Estimate c and γ 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.

Answer:

The two equations for c and γ are:

$$F(401) = 1 - e^{(-c \times 401^{\gamma})} = 0.25$$

$$F(2,836.75) = 1 - e^{(-c \times 2,836.75^{\gamma})} = 0.75$$

The method of percentiles

Answer (continued):

These two equations can be rewritten as:

$$-c \times 401^{\gamma} = log(0.75)$$

and

$$-c \times 2,836.75^{\gamma} = log(0.25)$$

On division we find that $\tilde{\gamma} = 0.8038$ and hence $\tilde{c} = 0.002326$, where $\sim$ denotes the percentile estimate.

Note that $\tilde{\gamma}$ is less than 1, indicating a fatter tail than the exponential distribution gives.

Goodness-of-fit

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

Example:

A breakdown of the repair costs of vehicles from our earlier example revealed the following numbers in different bands:

£0 - £1000 :	**200**
£1000 - £2000:	**300**
£2000 - £3000:	**250**
£3000 - £4000:	**150**
£4000 - £5000:	**100**
£5000+:	**0**

Use this information to test whether the exponential distribution provides a good model for the individual repair costs.

Goodness-of-fit

Answer:

- We are testing:

H_0: The costs come from an exponential distribution.

H_1: The costs do not come from an exponential distribution.

- In order to apply the chi-square test we need to determine the “expected” numbers.

i.e. the numbers in each band if the costs are from an Exponential distribution.

Goodness-of-fit

Answer (continued):

- From earlier example, we estimated the value of lambda as 1/2200.

  Using our estimate of the value of $\lambda = \frac{1}{2,200}$, the probability that an individual repair cost will fall in the £2,000 - £3,000 band is:

$$\int_{2,000}^{3,000} \lambda e^{-\lambda x} dx = \left[-e^{-\lambda x}\right]_{2,000}^{3,000} = e^{-2,000\lambda} - e^{-3,000\lambda} = 0.1472$$

and the expected number for this band is: $1,000 \times 0.1472 = 147.2$

The expected numbers for all the bands can be calculated in a similar way, giving: 365.3, 231.8, 147.2, 93.4, 59.3, 103.0

Goodness-of-fit

Answer (continued):

- The value of the chi-square statistic can then be calculated:

$$\chi^2 = \frac{(200 - 365.3)^2}{365.3} + \frac{(300 - 231.8)^2}{231.8} + \ldots + \frac{(0 - 103.0)^2}{103.0} = 331.89$$

- We have 6 bands but we equated the totals and estimated one parameter so that means 4 degrees of freedom.
- Our observed value of the chi-square statistic far exceeds 14.86, the upper 99.5% point of a chi-square with 4 d.f.
- So we can reject H_0 and conclude that these repair costs do not conform to an exponential distribution.

Goodness-of-fit

- Given a set of data, we fit a distribution and estimate the parameters of the distribution using an appropriate method method of moments, MLE or method of percentiles.
- We can check the fit in R by plotting a histogram of the data and superimposing the density function of the fitted distribution.
- Better yet, we can plot an empirical density function from the data using the function `density`, and add the true density function of the fitted distribution.
- A better way is to use the `qqplot` function to compare the sample data to simulated values from the fitted model distribution. A straight diagonal line indicates perfect fit:
- `qqplot(<simulated values from fitted distribution>, <values from data>)`
- `abline(0,1)`

- Note that `abline(a,b)` adds a line y=bx +a onto an existing plot.