Lecture 1

INSTITUTE OF ACTUARIAL & QUANTITATIVE STUDIES

Class: TY BSc

Subject : Statistical & Risk Modelling - 3

Subject Code:

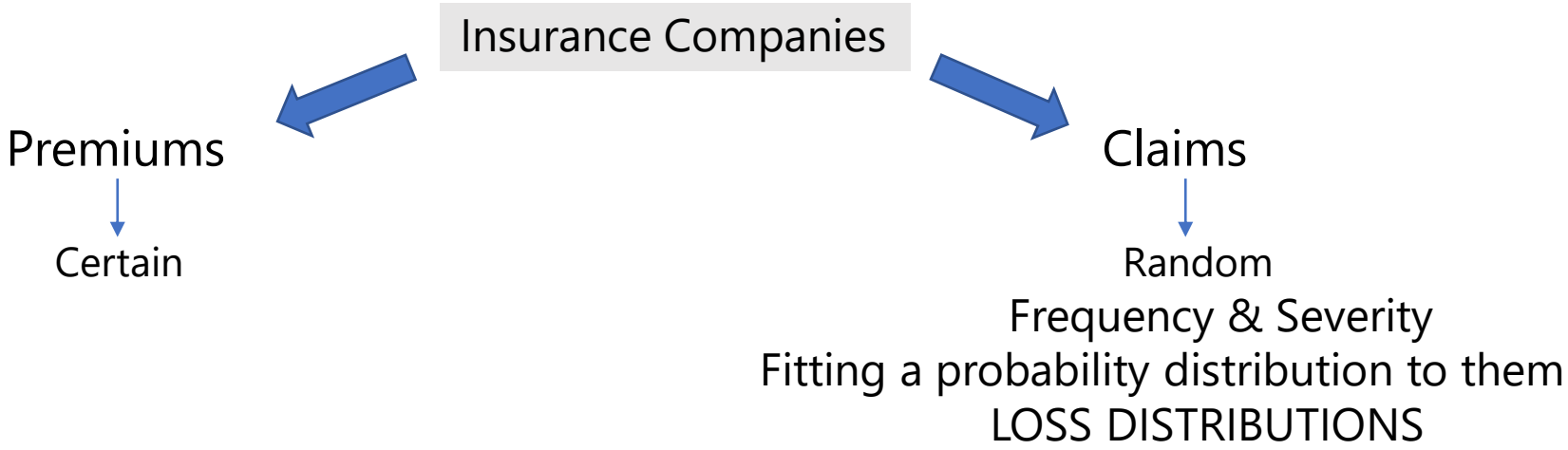
Chapter: Unit 1 Chapter 1

Chapter Name: Loss distributions in insurance risk management

Agenda

1. Introduction
2. Generating functions
   a. PGF
   b. MGF
   c. CGF
3. Loss distributions – Exponential, Gamma, Normal, Log-Normal, Pareto, Three-parameter Pareto distribution, Burr, Weibull
4. Basic Distributional Quantities
   a. Method of moments
   b. Maximum Likelihood Estimator
   c. Method of Percentiles
5. Goodness-of-fit

1.0 Introduction



HELPS IN:

i. Premium calculation
ii. Reserve calculation
iii. Solvency
iv. Reinsurance arrangement

2.0 Generating Functions

1) Probability Generating Functions

The probability generating function of a discrete random variable is a power series representation (the generating function) of the probability mass function of the random variable.

Pgf is used for **discrete distributions**.

$$P_X(t) = E[t^X] = \sum_{x=0}^{\infty} p(x)t^X$$

2.0 Generating Functions

2) Moment Generating Functions

A moment generating function (MGF) can be used to generate moments (about the origin) of the distribution of a random variable (discrete or continuous). Although the moments of most distributions can be determined directly by evaluation using the necessary integrals or summation, utilising moment generating functions sometimes provides considerable simplifications.

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

2.0 Generating Functions

3) Cumulant generating functions

A cumulant generating function (CGF) takes the moment of a sequence of numbers that describes the distribution in a useful, compact way. The first cumulant is the mean, the second the variance, and the third cumulant is the skewness or third central moment.

$$C_X(t) = \ln M_X(t)$$

3.1 Exponential Distribution

A random variable X has the exponential distribution with parameter $\lambda > 0$ if it has CDF

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

In that case, we write $X \sim Exp(\lambda)$.

E(X) =

var(X) =

MGF =

Practical Application:

Normally used to determine the inter-event times.

Example- time until next claim, time between 2 claims.

3.2 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 $X \sim Ga(\alpha, \lambda)$. The mean and variance of X are

MGF =

CGF =

$E(X)$ =

Var(X) =

Skew(X) =

Practical Application:

Normally used to model size of Insurance claims.

3.3 Normal distribution

Normal distribution is a type of continuous probability distribution for a real-valued random variable

$f(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2}$

E(X) =

var(X) =

MGF =

Practical Application:

Used for approximation, fitting distribution to symmetrical data.

3.4 Lognormal distribution

The definition of the lognormal distribution is very simple: X has a lognormal distribution if $\log(X)$ has a normal distribution. When $\log(X) \sim N(\mu, \sigma^2)$, $X \sim LogN(\mu, \sigma^2)$

E(X) =

var(X) =

MGF =

Practical Application:

Black-Scholes Model for option pricing, to model income distribution of people.

Normal distribution → LOG(X) → **Lognormal distribution**

3.5 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

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

Practical Application:

Few large, many small scenarios (like distribution of wealth, human population in cities and villages.)

To derive mean and variance we will have to understand Three-Parameter Pareto Distribution.

3.6 Three-parameter Pareto distribution

The pdf of three-parameter Pareto distribution is:

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

Pareto distribution

Parameter k (shape)

Three-parameter Pareto distribution

3.6 Three-parameter Pareto distribution

Pareto Distribution:

E(X) =

Var(X) =

Median =

Three Parameter Pareto:

E(X) =

3.7 Burr distribution

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

This is the CDF of the transformed Pareto or Burr distribution.

Median =

$E(X^k) =$

E(X) =

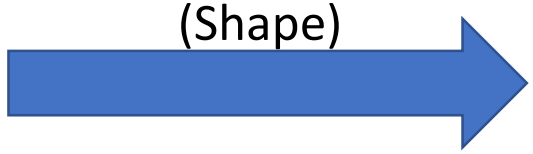
var(X)

Practical Application:

Flexible to fit, can be used instead of Normal distribution if data is skewed.

Pareto distribution

Parameter γ (Shape)



Burr distribution

3.8 Weibull distribution

To control flexibility of exponential distributions i.e. To increase/ decrease frequency of extreme events, we can use the Weibull distribution by adding a shape parameter γ.

Exponential F(x) =1 - exp(−λx)

There is a further possibility. Set

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

E(X) =

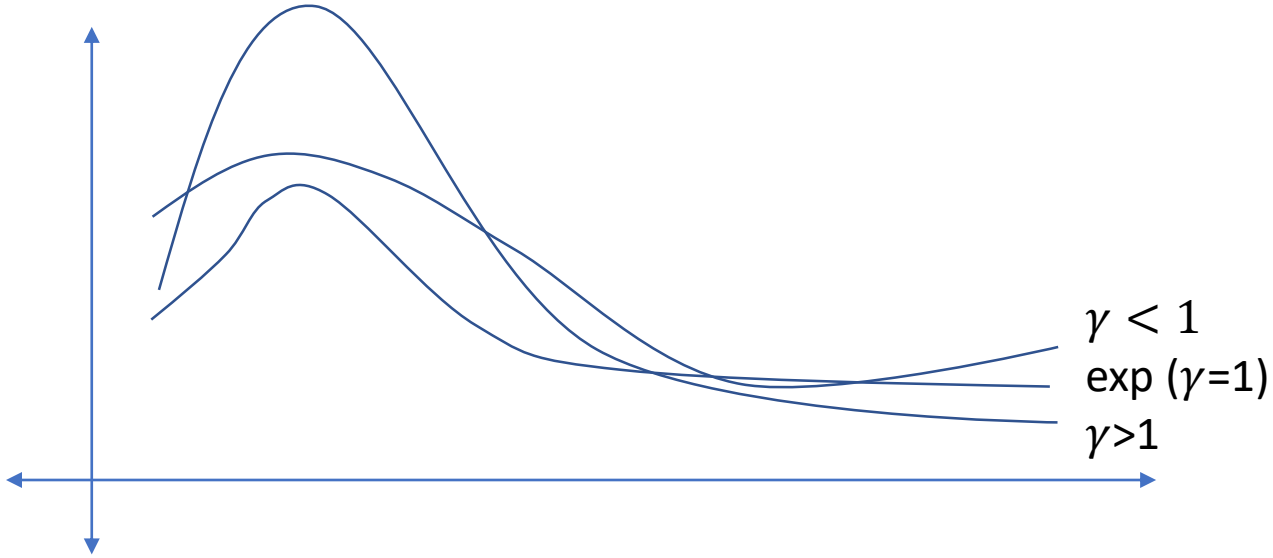
var(X) =

Median =

Practical Application:

Used in Extreme Value Theory, can be used to control flexibility of exponential distribution, inter-event times with non-constant rates.

3.8 Weibull distribution



Exponential distribution → Parameter γ → **Weibull distribution**

4.0 Methods Of Estimation

- Practically, we will not have a ready distribution for claims.
- We need to fit a distribution to the available data.
- For this we will need to estimate the parameters of such distributions.

- Methods of estimation

1. Method of moments
2. Maximum Likelihood Estimator
3. Method of Percentiles

4.1 Method of moments

The method of moments is a way to estimate population parameters, like the population mean or the population standard deviation. We simply use sample data to estimate population parameters.

1st parameter – E(X):

$$\bar{x} = \frac{1}{N} \sum_{i=1}^{N} x_i$$

2nd parameter – Var(X):

$$S^2 = \sum_{i=1}^{n} (x_i - \mu)^2 / N$$

4.1 Method of moments

Advantages

- Method of moments is simple (compared to other methods like the maximum likelihood method) and can be performed by hand.

Disadvantages

- The parameter estimates may be inaccurate. This is more frequent with smaller samples and less common with large samples.
- The method may not result in sufficient statistics. In other words, it may not take into account all of the relevant information in the sample.

4.2 Maximum likelihood estimation

Maximum likelihood estimation (MLE) is a technique used for estimating the parameters of a given distribution, using some observed data.

Using a limited sample of the population, we find particular values of the mean and variance such that the observation is the most likely result to have occurred.

For this we define a likelihood function.
The likelihood function of a random variable, X, will give us the probability (or PDF) using a hypothetical parameter, θ.

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

4.2 Maximum likelihood estimation

To determine the MLE, the likelihood function needs to be maximised. Often it is practical to consider the log-likelihood function

$$l(\theta) = \ln L(\theta) = \sum_{i=1}^{n} \ln P(X_i = x_i|\theta) \text{for discrete random variable X}$$

$$l(\theta) = \ln L(\theta) = \sum_{i=1}^{n} \ln f(x_i|\theta) \text{for continuous random variable X}$$

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

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

STEPS to find MLE :

1. Find L such that L= $\prod_{i=1}^{n} f(x_i)$
2. lnL
3. $\frac{d\ln L}{dx_i} = 0$, $\hat{x_i}$ = _____ (MLE)
4. $\frac{d^2 \ln L}{dx_i^2} < 0$, maximum

4.2 Maximum likelihood estimation

Advantages

Simple to apply and The method is statistically well understood.

Lower variance than other methods (i.e. estimation method least affected by sampling error) and unbiased as the sample size increases.

Able to analyze statistical models with different characters on the same basis. Maximum likelihood provides a consistent approach to parameter estimation problems. This means that maximum likelihood estimates can be developed for a large variety of estimation situations.

Disadvantages

Computationally intensive and so extremely slow (though this is becoming much less of an issue)

Frequently requires strong assumptions about the structure of the data

The estimates that are obtained using this method are often biased. That is, they contain a systematic error of estimation. This is true for small samples. The optimality properties may not apply for small samples.

MLE is inapplicable for the analysis of non-regular populations (Non-regular distributions are models where a parameter value is constrained by a single observed value).

Question

CT6 September 2008 Q11
Losses on a portfolio of insurance policies in 2006 are assumed to have an exponential distribution with parameter λ. In 2007 loss amounts have increased by a factor *k* (so that a loss incurred in 2007 is *k* times an equivalent loss incurred in 2006).

(i) Show that the distribution of loss amounts in 2007 is also exponential and determine the parameter of the distribution. [3]

Over the calendar years 2006 and 2007 the insurer had in place an individual excess of- loss reinsurance arrangement with a retention of *M*. Claims paid by the insurer were:

2006: 4 amounts of *M* and 10 claims under *M* for a total of 13,500.
2007: 6 amounts of *M* and 12 claims under *M* for a total of 17,000.

(ii) Show that the maximum likelihood estimate of λ is: $\hat{\lambda} = \dfrac{22}{13,500 + \dfrac{17,000}{k} + 4M + \dfrac{6M}{k}}$

Question

(iii) The insurer is negotiating a new reinsurance arrangement for 2008. The retention was set at 1600 when the current arrangement was put in place in 2006. Loss inflation between 2006 and 2007 was 10% (i.e. k = 1.1) and further loss inflation of 5% is expected between 2007 and 2008.

(a) Use this information to calculate $\hat{\lambda}$.

(b) The insurer wishes to set the retention M' for 2008 such that the expected (net of re-insurance) payment per claim for 2008 is the same as the expected payment per claim for 2006. Calculate the value of M' , using your estimate of λ from (iii)(a).

4.3 Method of Percentile

Method of Percentile indicate the values below which a certain percentage of the data in a data set is found.
The method involves equating selected sample percentiles to the distribution function; for example, equate the sample quartiles, 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. 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 (although the theoretical basis for this is not so clear). In a similar fashion, when using the method of percentiles, 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.

Advantage

The main advantage of using percentiles is that unusually high values (like whiskers in boxplots) are not included into the averaging calculations. This means that statistics include more relevant data.
In the example of the 95th-percentile, 5% of the highest measured values are discarded for the statistical report.

5.0 Goodness-of-fit test

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

Recall that the formula for the test statistic is $\sum \frac{(O-E)^2}{E}$, $where$

- O is the observed number in a particular category
- E is the corresponding expected number predicted by the assumed probabilities the sum is over all possible categories.

A high value for the total indicates that the overall discrepancy is quite large and would lead us to reject the model.