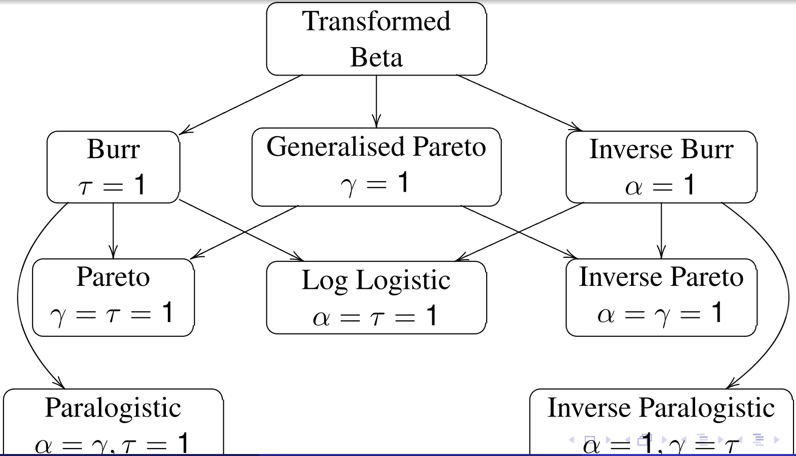
ACSC/STAT 4703, Actuarial Models II
Further Probability with Applications to Actuarial Science
FALL 2023

Toby Kenney

In Class Examples

Transformed Beta Distribution

$$f_X(x) = \left(\frac{\Gamma(\alpha + \tau)}{\Gamma(\alpha)\Gamma(\tau)} \right) \frac{\gamma \left(\frac{x}{\theta}\right)^{\gamma\tau}}{x \left(1 + \left(\frac{x}{\theta}\right)^{\gamma}\right)^{\alpha+\tau}}$$



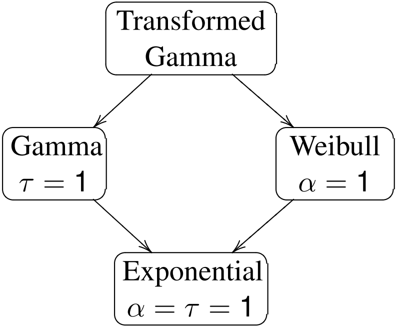
The Transformed Gamma and Inverse Transformed Gamma Families (Revision)

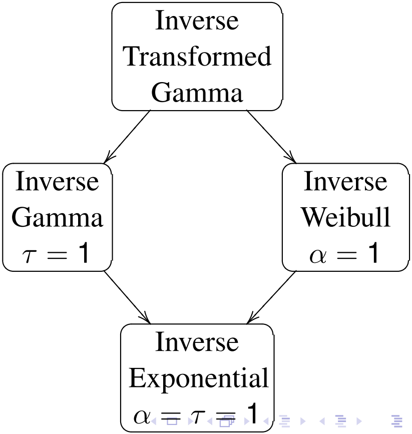
Transformed Gamma

$$f_X(x) = \frac{\tau\left(\frac{x}{\theta}\right)^{\alpha\tau} e^{-\left(\frac{x}{\theta}\right)^{\tau}}}{x\Gamma(\alpha)}$$

Inverse Transformed Gamma

$$f_X(x) = \frac{\tau\left(\frac{\theta}{x}\right)^{\alpha\tau} e^{-\left(\frac{\theta}{x}\right)^{\tau}}}{x\Gamma(\alpha)}$$





Methods to Create New Distributions

Transformation

- Adding a constant
- Multiplication by a constant
- Raising to a power
- Exponentiation

Combining Distributions

- Convolution
- Mixing
- Splicing

5.2 Creating New Distributions— Transformations

Scale and Location Transformations

- Many distributions include scale and location parameters.
- Location parameters inappropriate for non-negative distributions.
- Scale can represent change of unit or inflation.
- Sometimes need to standardise variables for asymptotic results.

$F^*(x) = F\left(\frac{x-\mu}{\sigma}\right) \quad f^*(x) = \frac{1}{\sigma} f\left(\frac{x-\mu}{\sigma}\right)$

Raising to a Power

- Can make values positive — e.g. Chi-square test.
- Can reduce skewness — e.g. Box-Cox transformation.

$$F_{X^{2n}}(x) = F_X\left(x^{\frac{1}{2n}}\right) - F_X\left(-x^{\frac{1}{2n}}\right) \qquad f_{X^{2n}}(x) = \frac{x^{\frac{2n-1}{2n}}}{2n}\left(f_X\left(x^{\frac{1}{2n}}\right) + f_X\left(-x^{\frac{1}{2n}}\right)\right)$$

$$F_{X^\alpha}(x) = F_X\left(x^{\frac{1}{\alpha}}\right) \qquad f_{X^\alpha}(x) = \frac{x^{1-\frac{1}{\alpha}}}{\alpha} f_X\left(x^{\frac{1}{\alpha}}\right)$$

Exponentiation

- Converts sums into products.

$F^*(x) = F(\log(x)) \quad f^*(x) = \frac{1}{x} f(\log(x))$

Question 1

Let X follow a beta distribution. Calculate the density function of a standardised version of X.

Question 2

Calculate the pdf of the square of a standardised gamma function.

Question 3

Every day, the value of a particular investment increases by $X\%$ where X has mean 0.04 and variance 5. What is the distribution of the value after 1 year?

5.2 Creating New Distributions— Combining

Convolution

- Often deal with sums of independent random variables.
- Sometimes from the same distribution — e.g. normal, gamma.

$$f_{X+Y}(x) = \int f_X(y) f_Y(x-y)\, dy$$

Mixing

- Marginal distribution with unobserved covariate.

$$f(x) = \int f_\theta(x)\pi(\theta)\, d\theta$$

Splicing

- Mixture with disjoint supports.
- Used to allow different distributions for tail and main part of data.

$$f(x) = \begin{cases} f_1(x) & x < C \\ f_2(x) & x \geqslant C \end{cases}$$

Question 4

An insurer sells two policies. Aggregate losses from each policy are assumed to follow a Pareto distribution with $\alpha = 4$ and $\theta = \$10,000$. What is the probability that aggregate losses from both policies exceed \$50,000?

Question 5

The aggregate losses on an auto insurance policy follow a Pareto distribution with $\alpha = 2.5$, and θ varying between policyholders. For a randomly chosen policyholder, Θ follows a gamma distribution with $\alpha = 4$ and $\theta = 800$.

(a) What is the probability that the aggregate losses on a random policy exceed $10,000?

(b) What is the expected aggregate loss for a random individual?

(c) What are the VaR and TVaR for the aggregate loss of a random policy at the 0.99 level?

Question 6

An actuary is modelling aggregate claims. For aggregate claims less than \$10,000 a normal distribution with mean \$4,000 and standard deviation \$3,000 can be used, because of the central limit theorem. For larger aggregate claims, the actuary decides that aggregate claims larger than \$10,000 should be modelled as following a Pareto distribution with $\alpha = 3$. The probability that aggregate claims exceed \$10,000 is estimated to be 0.08, and the parameter θ for the Pareto distribution is chosen so that the density function of the resulting distribution is continuous. What is the probability under this model that aggregate claims exceed \$25,000?

Deductibles (Revision)

If a policy has a deductible d, then the amount paid for a loss X is $(X - d)_+$.

Dealing with Deductibles (Revision)

- Deductibles reduce claim frequency.
- For severity distribution, we sometimes consider per loss, and sometimes per claim.
- Deductibles always reduce per loss severity, but might increase or decrease per claim severity.

Loss Elimination Ratio (Revision)

The Loss Elimination Ratio is the ratio

$$\frac{\text{Losses paid by policyholder due to deductible}}{\text{Total losses}} = 1 - \frac{\text{Losses paid by insurer}}{\text{Total losses}}$$

8.4 Policy Limits

Policy Limits (Revision)

If a policy has a limit u, then the amount paid for a loss X is $X \wedge u$.

Dealing with Limits (Revision)

- Deductibles are like limits from the point of view of the policyholder — with a deductible d, the policyholder pays $X \wedge d$ and the insurer pays $(X - d)_+$ With a limit u, the insurer pays $X \wedge u$ and the policyholder pays $(X - u)_+$.
- Limits decrease the effect of inflation.

Increased limits factors

- Relative increase in premium caused by increasing policy limit.
- Several difficulties in estimating ILF:
  - Loss development factors increase with policy limit.
  - Trend factors tend to increase with policy limit.
  - Risk to insurer increases faster than expected claim.
  - Some expenses are fixed; some vary with premium; some vary with policy size.
- Historical policy limits affect the data in several ways:
  - Insurance company records generally censor data at policy limits.
  - Limit can impact settlement amounts — e.g. lawyers might aim for policy limit.
  - Adverse selection
- These data issues can be mitigated by only considering policies with limits at least as high as the limit under consideration, provided there is sufficient data.

Question 7

An insurance company has the following data on its policies:

Policy limit	Losses Limited to			
	50,000	100,000	500,000	1,000,000
50,000	10,000			
100,000	34,000	41,000		
500,000	23,000	26,000	31,000	
1,000,000	11,000	12,300	13,400	17,000

Use this data to calculate the ILFs.

IRLRPCI 5.2 Increased limits factors

Loss Development

- Loss development factors tend to be larger for larger claims.
- They should be estimated from datasets with a single limit.

Trend Factors

- Lower policy limits reduce the effects of inflation.
- Different trend factors should be calculated for each policy limit.
- For higher policy limits, the larger variance and smaller data set can mean estimates are not credible, so data from other policy limits may need to be used.

Risk

- Higher policy limits increase risk more than premium.
- Typically risk load should be increased to compensate for this increased risk.

Question 8

For a certain line of insurance, the loss amount per claim follows an exponential distribution with mean $a\theta$, where a is the exposure. The policy has a limit l, which is currently set at 5θ per unit of exposure. Losses increase by an inflation rate of 10%. Calculate the percentage increase in expected total payments per claim.

Question 9

An insurance company models the number of claims on its policies as following a Poisson distribution with parameter $\lambda = 100$. Losses follow a Pareto distribution with $\alpha = 3$ and $\theta = 10,000$. The policies have a policy limit per claim of \$50,000. The insurer models aggregate losses as following a normal distribution, and sets its total premiums at the 95th percentile of the agregate loss distribution.
(a) Calculate the current risk loading as a percentage of the gross rate.
(b) Calculate risk loading as a percentage of the gross rate if the company increases the policy limit to \$100,000 per claim.

Question 10

An insurance company charges a risk charge equal to the square of the average loss amount, divided by 50,000. It has the following data on a set of claims from policies with limit $1,000,000.

Interval	No. of claims	Total claimed
$(0, 10,000]$	2,300	6,850,000
$(10,000, 100,000]$	900	13,600,000
$(100,000, 500,000]$	140	19,400,000
$(500,000, 1,000,000]$	25	18,600,000

Calculate the ILF from $100,000 to $500,000, and to $1,000,000.

IRLRPCI 5.2 Increased limits factors

Expenses

- Expenses tend to be subdivided into fixed costs and costs that vary.
- Some expenses are proportional to premium, other variable expenses will increase non-linearly with premium, e.g. adjustment expenses.

Loss Distributions

- Parametric loss distributions make calculating ILFs easier.
- To fit parametric distributions case reserves should be used for open claims, because time to settlement is not independent of loss size.
- Case reserves from very recent claims can be subjective, so it is often a good idea to ignore data from most recent years.

Question 11

An insurer finds that the pure premium ILF from \$1,000 to \$1,000,000 is 4.62. What is the Loss elimination ratio of a \$1,000 deductible for a policy with limit \$1,000,000?

Question 12

An insurer sells policies with limits \$1,000,000 and \$2,000,000. The trend factor for losses limited to \$1,000,000 is 1.052. The trend factor for losses limited to \$2,000,000 is 1.044. The insurer's loading for policies with limit \$1,000,000 is 25%. For policies with limit \$2,000,000, the insurer buys reinsurance from a reinsurer. The ILF from \$1,000,000 to \$2,000,000 decreases from 1.36 in 2021 to 1.35 in 2022. What is the reinsurer's loading on this reinsurance.

Problem

Given a sample

$$x_1, \ldots, x_N$$

How do we model the largest losses, from very few samples?

Block Maxima

1. Divide the sample into $\frac{N}{n}$ blocks of size n. (By time or at random).
2. Let $m_{n,i}$ be the maximum of block i of size n. We therefore have the sample of block maxima:

$$m_{n,1}, \ldots, m_{n,\frac{N}{m}}$$

Points over Threshold

1. Choose a threshold T — usually a high empirical quantile.
2. Restrict to the sample points

$$\{x_i | x_i > T\}$$

SN2 5.3 Distribution of Block Maxima

Question 13

(a) Simulate a large sample from a standard normal distribution. Divide this sample into blocks of size n, for varying values of n, and calculate the block maxima.

(b) Fit scale and location functions to the distributions of block maxima.

(c) Use the fitted function to rescale the distributions of the block maxima, and compare the rescaled distributions.

Question 14

Repeat Question 13 for an exponential distribution.

Generalised Extreme Value Distribution

Theorem (Fisher-Tippet-Gnedenko Theorem)

If M_n is the maximum of a sample of n i.i.d. random variables with distribution function F, and there are functions c_n and d_n of n such that the distributions of $\frac{M_n-d_n}{c_n}$ converge in distribution to a non-degenerate distribution, then for a certain choice of c_n and d_n, that distribution has CDF of the form

$$H_\xi(x) = \begin{cases} e^{-(1+\xi x)^{-\frac{1}{\xi}}} & \text{if } \xi \neq 0 \\ e^{-e^{-x}} & \text{if } \xi = 0 \end{cases}$$

Extreme Value Distributions

$\xi > 0$	Fréchet distribution	$F(x) = \exp\left(-\left(\frac{x-\mu}{\theta}\right)^{-\alpha}\right)$
$\xi = 0$	Gumbel distribution	$F(x) = \exp\left(-\exp\left(-\frac{x-\mu}{\theta}\right)\right)$
$\xi < 0$	Weibull EV distribution	$F(x) = \exp\left(-\left(\frac{\mu-x}{\theta}\right)^{\tau}\right)$

Question 15

when X_i follow a log-normal distribution with parameters μ and σ^2, the values of d_n are given as the solutions to $\log(d_n) d_n^{\frac{\log(d_n)}{2}} = n$, and $c_n = \frac{1}{\log(d_n)}$. Find the corresponding value of ξ.

Generalised Extreme Value Distribution

Theorem

Let X have survival function $S(x)$. Its distribution function F is in the Maximum Domain of Attraction of H_ξ if and only if $\lim_{n\to\infty} nS(c_n x + d_n) = -\log H_\xi(x)$.

Generalised Extreme Value Distribution

Question 16

(a) What are the appropriate values of c_n and d_n for a Weibull distribution, and what is the corresponding value of ξ?

(b) For a Weibull distribution with $\tau = 0.6$ and $\theta = 10$, for $n = 100$ and $n = 1000$, what is the probability that $M_n > 100$ using the GEV approximation? What is the exact probability?

Generalised Extreme Value Distribution

Question 17

A reinsurance company estimates that the annual aggregate losses in millions on a certain portfolio is in the MDA of a GEV distribution with $\xi = 2$ and the normalising constants are $c_n = (2+n)^{\frac{1}{3}}$ and $d_n = (4+n)^{\frac{2}{5}}$. The reinsurer is selling stop-loss reinsurance on this portfolio with attachment point $20,000,000. What is the probability that it will need to pay a claim in the next 100 years?

SN2 5.3.3 GEV distributions

Fréchet Distribution ($\xi > 0$)

$$F(X) = \exp\left(-\left(\frac{x-\mu}{\theta}\right)^{-\frac{1}{\xi}}\right)$$

- Bounded below by $\mu - \frac{\theta}{\xi}$.
- Fat-tailed.
- $\alpha = \frac{1}{\xi}$ called the tail-index of the distribution.
- Larger ξ (or smaller α) correspond to fatter tails.
- For any distribution in the MDA of the Fréchet distribution, only moments that are $< \alpha$ are finite. Thus any distribution with all moments is not in the MDA of the Fréchet distribution.
- Pareto, t and Burr distributions in Fréchet MDA.
- Distribution with survival function $S(x)$ in Fréchet MDA if and only $S(x) = x^{-\frac{1}{\xi}}L(x)$, where for any $t > 0$, $\lim_{x\to\infty}\frac{L(tx)}{L(x)} = 1$.

SN2 5.3.3 GEV distributions

Gumbel Distribution ($\xi = 0$)

$$F(X) = \exp\left(-\exp\left(-\left(\frac{x-\mu}{\theta}\right)\right)\right)$$

- Unbounded.
- Fat-tailed.
- For any distribution in the MDA of the Gumbel distribution, all finite moments exist.
- Gumbel MDA contains a range of distributions including light-tailed such as normal and exponential, and heavier tailed such as gamma and log-normal. Some of these distributions are bounded below.

SN2 5.3.3 GEV distributions

Weibull EV Distribution ($\xi < 0$)

$$F(X) = \exp\left(-\exp\left(-\left(\frac{x-\mu}{\theta}\right)\right)\right)$$

- Bounded above by $x < \mu - \frac{1}{\xi}$.
- If Y follows a Weibull EV distribution, then $1 + \xi Y$ follows a Weibull distribution.
- Beta and Uniform distributions in Weibull EV MDA.

Estimating ξ

- c_n and d_n unknown & hard to estimate, but without normalisation, M_n follows a scaled translated GEV distribution.
- Estimate parameters from block maxima by maximum likelihood.
- Estimation of ξ should be consistent for different choices of n.
- For small n, GEV asymptotics may not apply. For large n, the resulting sample size may be too small.
- GEV density

$$h_{\xi,\theta,\mu}(x) = \frac{1}{\theta}\left(1+\xi\left(\frac{x-\mu}{\theta}\right)\right)^{-\left(1+\frac{1}{\xi}\right)} \exp\left(-\left(1+\xi\left(\frac{x-\mu}{\theta}\right)\right)^{-\frac{1}{\xi}}\right)$$

- Log-likelihood $l(\xi,\mu,\theta)$:

$$-k\log(\theta) - \left(1+\frac{1}{\xi}\right)\sum_{i=1}^{k}\log\left(1+\xi\left(\frac{m_j-\mu}{\theta}\right)\right) - \sum_{i=1}^{k}\left(1+\xi\left(\frac{m_j-\mu}{\theta}\right)\right)^{-\frac{1}{\xi}}$$

SN2 5.3.4 Estimating the GEV parameter

Question 18

(a) Simulate a sample of 1,000,000 normal random variables. Use the fit.GEV function from the R package QRM to estimate the parameters for a range of different block sizes n.

(b) Repeat this 100 times for each block size to find the distribution of the estimated parameter values.

Introduction

- Model distribution of excess losses $Y_d = X - d | X > d$.
- Surival function given by $S_d(y) = \frac{S_X(y+d)}{S_X(d)}$.
- PDF (or PMF) given by $f_d(y) = \frac{f_X(y+d)}{S_X(d)}$.
- Mean excess loss $e(d) = \mathbb{E}(X - d | X > d) = \frac{\mathbb{E}(X) - \mathbb{E}(X \wedge d)}{S_X(d)}$

Survival Function

$$S(x) = \begin{cases} \left(1 + \xi \frac{x}{\beta}\right)^{-\frac{1}{\xi}} & \xi \neq 0 \\ e^{-\frac{x}{\beta}} & \xi = 0 \end{cases}$$

Where $\beta > 0$, $0 \leqslant x$ and $x \leqslant -\frac{\beta}{\xi}$ for $\xi < 0$.

Notes

- For $\xi > 0$, this is a Pareto distribution with $\alpha = \frac{1}{\xi}$ and $\theta = \frac{\beta}{\xi}$.
- For $\xi = 0$, this is an exponential distribution.
- For $\xi < 0$, this is a scaled beta distribution.
- For $\xi > 0$, kth moment exists only for $k < \frac{1}{\xi}$.
- $G_{\xi,1}(x) = 1 + \log H_\xi(x)$ where H_ξ is the CDF of the GEV.

Question 19

Show that the mean excess loss function of a GPD distribution is a straight line whenever it is defined.

Theorem (Pickands-Balkema-De Haan Theorem)

Let F denote the cdf of a random variable X with upper bound $x_{sup} \leqslant \infty$. We have $F \in MDA(H_\xi)$ if and only if there is some function $\beta_d \geqslant 0$ such that:

$$\lim_{d \to x_{sup}} \sup_{d \leqslant x \leqslant x_{sup}} \left| F_d(x) - G_{\xi,\beta_d}(x) \right| = 0$$

Notes

- This means we can approximate the excess loss distribution by a GPD, for sufficiently large losses.
- For GPD, MEL is a linear function, so one approach to decide whether GPD approximation is appropriate is to estimate MEL function, and decide when it becomes linear.

Question 20

(a) Simulate 100,000 values from an inverse gamma distribution with $\alpha = 4$. Calculate the empirical MEL as a function of threshold.

(b) Plot the empirical MEL on a graph, and see where the linear approximation becomes reasonable.

(c) Use this to estimate the probability that a random loss exceeds the threshold by at least 1, and compare this to the true value.

Question 21

Are there any distributions with linear MEL function, except for the GPD?

SN2 5.4.2 Generalised Pareto Distribution

Question 22

An insurance company estimates that the 95th percentile of a loss distribution is $4,200 and that above this point, the GPD approximation applies. The company estimates that the GPD parameters are $\xi = 0.4$ and $\beta = 300$ for this *d*. Estimate the VaR and TVaR for this loss distribution at the 0.99 level.

Question 23

Show that if F with positive support is in the MDA of a Fréchet distribution, with parameter ξ, then the mean excess loss of $\log(X)$ converges to ξ.

SN2 5.4.4 The Hill Estimator

Hill Estimator

- Let $x_{(1)} \leqslant x_{(2)} \leqslant \ldots \leqslant x_{(n)}$ be the order statistics of a sample.
- The Hill estimator is

$$\hat{\alpha}_j^H = \left(\sum_{k=j+1}^{n} \frac{\log(x_{(k)})}{n-j+1} - \log(x_{(j)}) \right)^{-1} \qquad \hat{S}^H(x) = \frac{j}{n} \left(\frac{x}{x_{(n-j)}} \right)^{-\hat{\alpha}_j^H}$$

Notes

- Get different estimates for different values of j.
- For small j, GPD approximation may be poor.
- For large j sample size may be too small.
- Plot values of $\hat{\alpha}_j^H$ for a range of j.

Question 24

Simulate 100,000 data points from an inverse Weibull distribution with $\tau = 8.7$ and $\theta = 200$. Plot the Hill estimator against *j*. Compare the Hill estimator with the MLE estimator of α based on different cut-offs. (You can use the `fit.GPD` function from the `R` package `QRM` for this.)

Question 25

Calculate the probability function of a mixed Poisson distribution with mixing distribution a Gamma distribution with shape α and scale θ.

7.1 Compound Frequency Distributions

Probability Generating Functions (Revision)

- For a random variable X, the p.g.f. is given by $P(z) = \mathbb{E}(z^X)$.
- For independant X and Y, $P_{X+Y}(z) = P_X(z)P_Y(z)$.
- Related to m.g.f. by $P_X(z) = M_X(\log(z))$.
-

Distribution	$P_X(z)$
Binomial	$(1 - p(1 - z))^n$
Poisson	$e^{-\lambda(1-z)}$
Negative Binomial	$(1 + \beta(1 - z))^{-r}$

Compound Distributions

- Primary distribution N has p.g.f. $P(z)$. Secondary distribution X has p.g.f. $Q(z)$.
- Compound distribution has p.g.f. $P(Q(z))$.
- This is the distribution of $X_1 + \ldots + X_N$, where X_i are i.i.d. and independent of N.

Question 26

Consider a compound distribution where the primary distribution is a member of the $(a, b, 0)$ distribution. Find a recurrence relation between the probabilities of the compound distribution.

Question 27

Calculate the probabilities of each of the values 0, 1, and 2 of a compound Poisson-Poisson distribution with parameters λ_1 and λ_2.

Question 28

Show that the binomial-geometric and negative binomial-geometric with r a positive integer, give the same distribution.

Question 29

Show that a compound Poisson-logarithmic distribution gives the same distribution as the negative binomial distribution.

Question 30

Show that a sum of independant compound Poisson random variables is another compound Poisson random variable.

Question 31

(a) Calculate the skewness of a compound Poisson distribution in terms of the first three moments of the secondary distribution.
(b) Use this to calculate the skewness of the Poisson-ETNB distribution.

Characteristic Functions and Infinite Divisibility

Definition

For a random variable X, the *characteristic function* ϕ_X is given by

$$\phi_X(z) = \mathbb{E}(e^{izx}) = \mathbb{E}(\cos(zX) + i\sin(zX))$$

This is similar to the moment generating function, but it exists for all distributions.

Definition

A distribution with characteristic function $\phi(z)$ is *infinitely divisible* if for any positive integer n, there is another distribution with characteristic function $\phi_n(z)$ such that $(\phi_n(z))^n = \phi(z)$.

This is equivalent to the same statement for the probability generating function or the moment generating function if they exist.

Question 32

Which of the following distributions are infinitely divisible?

a) gamma
b) inverse gamma
c) inverse Gaussian
d) binomial

9.3 The Compound Model for Aggregate Claims

Revision

- The number of losses N is a discrete random variable.
- Each loss amount X_i is assumed i.i.d. and independent of N.
- The aggregate loss is $S = X_1 + \cdots + X_N$.
- To get the aggregate loss from first principles, we can use

$$f_S(x) = \sum_{n=0}^{\infty} P(N = n) f_{S|N}(x|N = n)$$

- In practice, computation prohibits this approach.
- In a very small number of cases, the distribution can be simplified to a finite mixture.
- When the primary distribution is from the $(a, b, 1)$ class and the secondary distribution is arithmetic, there is a recurrence formula (see Question 26) for the compound distribution.

Question 33

An individual loss distribution is normal with mean 100 and standard deviation 35. The total number of losses N has the following distribution:

n	$P(N=n)$
0	0.4
1	0.3
2	0.2
3	0.1

What is the probability that the aggregate losses exceed 130?

Question 34

Calculate the probability density function of the aggregate loss distribution if claim frequency follows a negative binomial distribution with $r = 2$ and severity follows an exponential distribution.

Question 35

An insurance company models the number of claims it receives as a negative binomial distribution with parameters $r = 15$ and $\beta = 2.4$. The severity of each claim follows an exponential distribution with mean \$3,000. What is the net-premium for stop-loss insurance with an attachment point of \$204,000?

Question 36

Suppose that the total number of claims follows a negative binomial distribution with $r = 2$ and $\beta = 3$. Suppose that the severity of each claim (in thousands of dollars) follows a zero-truncated ETNB distribution with $r = -0.6$ and $\beta = 7$. What is the probability that the aggregate loss is at most 3? Calculate this from first principles.

The Recursive Method

Theorem

Suppose the severity distribution is a discrete distribution with probability function $f_X(x)$ for $x = 0, 1, \ldots, m$ (m could be infinite) and the frequency distribution is a member of the $(a, b, 1)$ class with probabilities $p_k, k = 0, 1, 2, \ldots$ satisfying $p_k = \left(a + \frac{b}{k}\right) p_{k-1}$ for all $k \geqslant 2$.
Then the aggregate loss distribution is given by

$$f_S(x) = \frac{(p_1 - (a+b)p_0)f_X(x) + \sum_{y=1}^{x \wedge m} \left(a + \frac{by}{x}\right) f_X(y) f_S(x-y)}{1 - af_X(0)}$$

Question 37

Let the number of claims follow a Poisson distibution with $\lambda = 2.4$ and the severity of each claim follow a negative binomial distribution with $r = 10$ and $\beta = 2.3$. What is the probability that the aggregate loss is at most 3?

Question 38

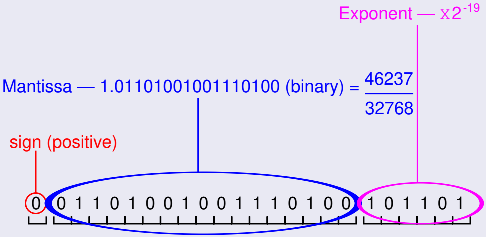
An insurance company offers car insurance. The number of losses a driver experiences in a year follows a negative binomial random variable with $r = 0.2$ and $\beta = 0.6$. The size of each loss (in hundreds of dollars) is modelled as following a zero-truncated ETNB distribution with $r = -0.6$ and $\beta = 3$. The policy has a deductible of $1,000 per loss. What is the probability that the company has to pay out at least $400 in a single year to a driver under such a policy?

Question 39

The number of claims an insurance company receives is modelled as a compound Poisson distribution with parameter $\lambda = 6$ for the primary distribution and $\lambda = 0.1$ for the secondary distribution. Claim severity (in thousands of dollars) is modelled as following a zero-truncated logarithmic distribution with parameter $\beta = 4$. What is the probability that the total amount claimed is more than \$3,000.

6.4.3 Computation Issues

How Computers Think of Numbers



Problems That can Arise

- There is a smallest representable positive number. This is very small, but for an aggregate of a large number of losses, the probability of zero can be smaller than this value, leading to underflow.
- Numbers are rounded to the limited accuracy. If we subtract a number from a very close number, most of the accuracy may be lost.

6.4.3 Computation Issues— Dealing with Underflow

Starting Above 0

- Approximate $f_S(x) = 0$ for $x < k$.
- As we don't know $f_S(k)$, start with $f_S(k) = 1$, and rescale later.
- Use the recurrence to compute $f_s(x)$ for $k < x < u$.
- Rescale so that $\sum_{x=k}^{u} f_s(x) = 1$.
- Common practice: let $k, u = \mu \pm 6\sigma$ so $P(X \in [k, u]) \approx 1$.

Convolution

- Primary distribution is divisible (∞ly if Poisson or n.b.).
- This means we can subdivide $N = N_1 + \cdots + N_k$.

$$
\begin{aligned}
S &= X_1 + \cdots + X_N \\
&= \underbrace{X_1 + \cdots + X_{N_1}} + \underbrace{X_{N_1+1} + \cdots + X_{N_1+N_2}} + \cdots + \underbrace{X_{N-N_k+1} + \cdots + X_N} \\
&= S_1 + \cdots + S_k
\end{aligned}
$$

- Each S_i compound so computed using the recurrence.
- Compute S from S_i by repeated convolution. (Easiest if $k = 2^m$).

Question 40

The number of claims an insurance company receives is modelled as a Poisson distribution with parameter $\lambda = 96$. The size of each claim is modelled as a zero-truncated negative binomial distribution with $r = 4$ and $\beta = 2.2$. Calculate the approximated distribution of the aggregate claims:

(a) By starting the recursion at a value of k six standard deviations below the mean.

(b) By solving for a rescaled Poisson distribution with $\lambda = 12$ and convolving the solution up to 96.

Answer to Question 40

R-Code for (a)

```
ans<-1
ans<-as.vector(ans)
for(n in 2:2000){
  temp<-0
  for(i in 1:(n-1)){%
    temp<-temp+16*i*(i+1)*(i+2)*(i+3)/(n+240)*0.6875^i*
        0.3125^4*ans[n-i]/(1-0.3125^4)
  }
  ans<-c(ans,temp)
}
```

R-Code for (b)

```
ConvolveSelf<-function(n){
  convolution<-vector("numeric",2*length(n))
  for(i in 1:(length(n))){
    convolution[i]<-sum(n[1:i]*n[i:1])
  }
  for(i in 1:(length(n))){
    convolution[2*length(n)+1-i]<-sum(n[length(n)+1-(1:i)
       ]*n[length(n)+1-(i:1)])
  }
  return(convolution)
}

d24<-ConvolveSelf(ans2)
d48<-ConvolveSelf(d24)
d96<-ConvolveSelf(d48)
plot(dist1,d96[241:2240])
```

Question 41

If the primary distribution is binomial with $n = 7$ and $p = 0.8$, and the secondary distribution has probability mass function

$$f_X(x) = \begin{cases} 0.21 & \text{if } x = 0 \\ 0.41 & \text{if } x = 1 \\ 0 & \text{if } x = 2 \\ 0.38 & \text{if } x = 3 \end{cases}$$

use the recurrence relation to compute the aggregate loss distribution.

Constructing Arithmetic Distributions

Method of Rounding

$$f_{X^a}(x) = P(x - 0.5 \leqslant X < x + 0.5)$$

Method of Local Moment Matching

- Divide $\mathbb{N}$ into intervals of length k: $[0, k], [k, 2k], \ldots$.
- For each interval $I = [nk, (n+1)k]$, calculate $P(X \in I)$ and $\mathbb{E}(X^i | X \in I)$ for $i = 1, \ldots, k$.
- Construct values $q_{nk}, q_{nk+1}, \ldots q_{(n+1)k}$ such that $\frac{\sum_{m=nk}^{(n+1)k} q_m m^i}{\sum_{m=nk}^{(n+1)k} q_m} = \mathbb{E}(X^i | X \in I)$ for $i = 1, \ldots, k$.
- For $m \neq nk$, there is a unique interval with a value of q_m. Let $f_X(m)$ be this value. If $m = nk$, then we have one value of q_{nk} from the interval $[(n-1)k, nk]$, and one from the interval $[nk, (n+1)k]$. Let $f_x(nk)$ be the sum of these values.

Question 42

Let X follow an exponential distribution with mean θ. Approximate this with an arithmetic distribution ($h = 1$) using:
(a) The method of rounding.
(b) The method of local moment matching, matching 2 moments on each interval.

Why is Model Selection Important?

- Using a wrong model will lead to wrong conclusions.

Advantages of Graphical Approaches

- Looking at graphs tells us not only whether the model fit is good, but also where the model fit is good or bad.
- Many tests only detect particular deviations from the model, and miss other deviations.
- Your eyes have fewer bugs than your `R` code.

Advantages of Testing or Score-based Approaches

- It is hard to judge how much deviation from the expected distribution should occur by chance.
- Formal tests or scores are harder to manipulate, and easier to defend to regulators.

16.3 Graphical Comparison of Density and Distribution Functions

Question 43

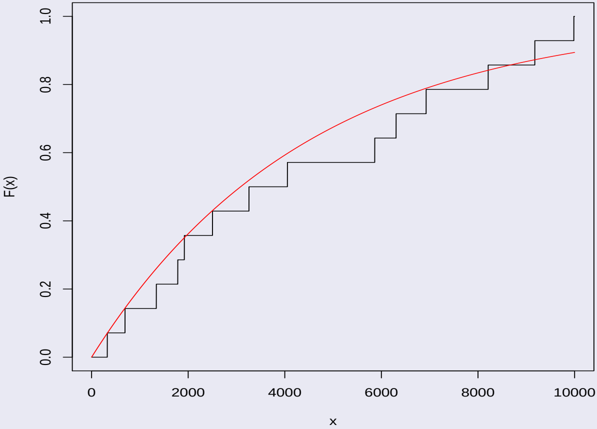
An insurance company is modeling claim severity. It collects the following data points:

325	692	1340	1784	1920	2503	3238	4054	5862
6304	6926	8210	9176	9984				

By graphically comparing distribution functions, assess the appropriateness of a Pareto distribution for modeling this data. The MLE estimates for the parameters of the Pareto distribution are $\alpha = 934.25$, $\theta = 4156615$

16.3 Graphical Comparison of Density and Distribution Functions

Answer to Question 43



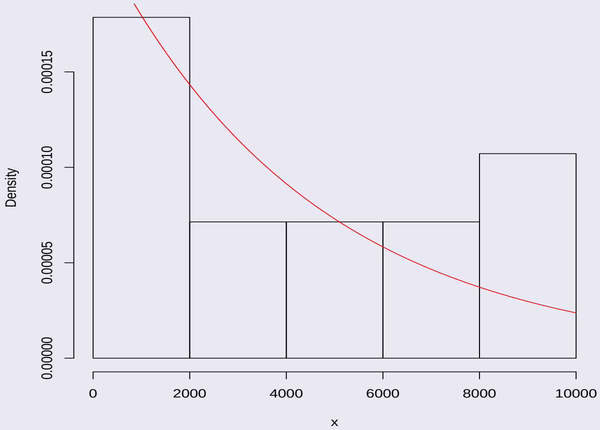
Question 44

For the data from Question 43:

325	692	1340	1784	1920	2503	3238	4054	5862
6304	6926	8210	9176	9984				

Graphically compare density functions to assess the appropriateness of a Pareto distribution for modeling this data.

Answer to Question 44



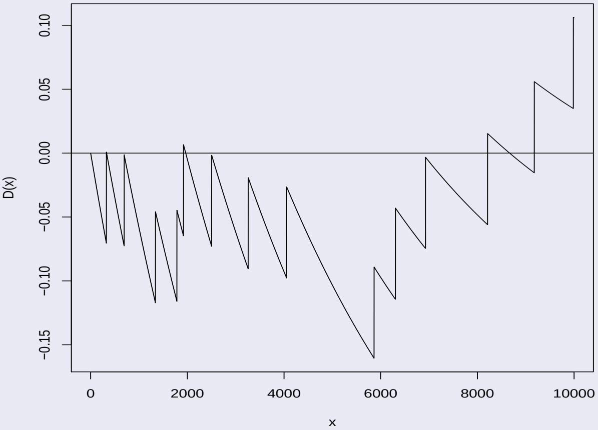
Question 45

For the data from Question 43:

325 692 1340 1784 1920 2503 3238 4054 5862
6304 6926 8210 9176 9984

By Graphing the differnce $D(x) = F^*(x) - F_n(x)$, assess the appropriateness of a Pareto distribution for modeling this data.

Answer to Question 45



Question 46

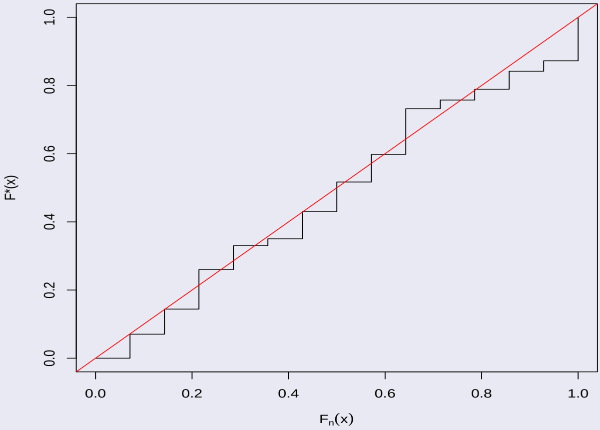
For the data from Question 43:

325	692	1340	1784	1920	2503	3238	4054	5862
6304	6926	8210	9176	9984				

Use a *p*-*p* plot to assess the appropriateness of a Pareto distribution for modeling this data.

16.3 Graphical Comparison of Density and Distribution Functions

Answer to Question 46

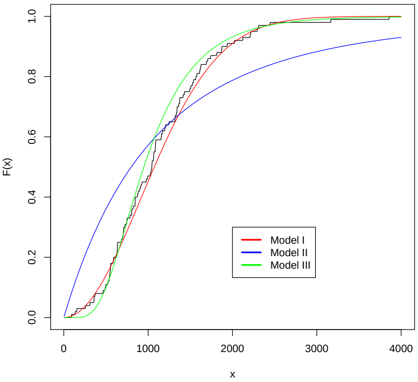


16.3 Graphical Comparison of Density and Distribution Functions

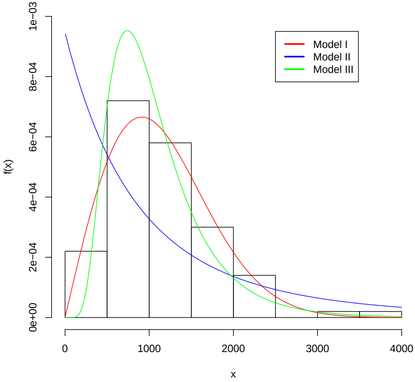
Question 47

An insurance company is modelling a data set. It is considering 3 models, each with 1 parameter to be estimated. On the following slides are various diagnostic plots of the fit of each model.
Determine which model they should use for the data in the following situations. Justify your answers.
(a) Which model should they choose if accurately estimating (right-hand) tail probabilities is most important?
(b) The company is considering imposing a deductible, and therefore wants to model the distribution very accurately on small values of x.

Models



F(x)
0.0
0.2
0.4
0.6
0.8
1.0
0
1000
2000
3000
4000
x
Model I
Model II
Model III



f(x)
0e+00
2e−04
4e−04
6e−04
8e−04
1e−03
0
1000
2000
3000
4000
x
Model I
Model II
Model III

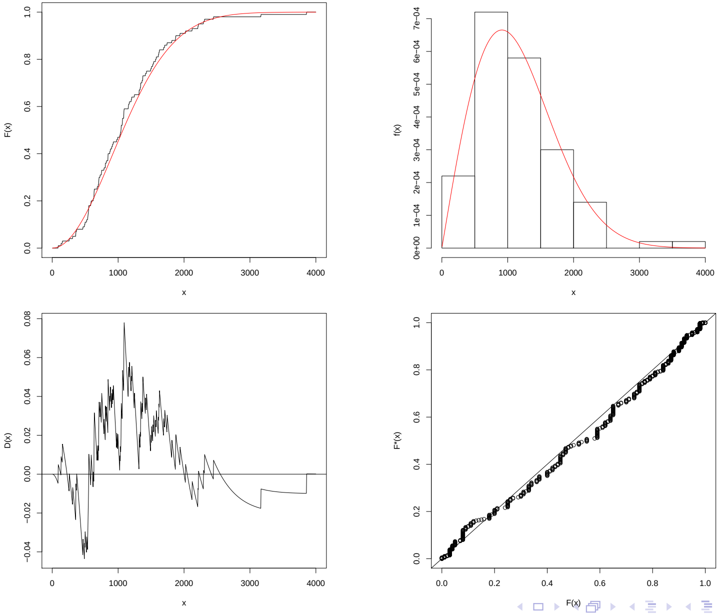
16.3 Graphical Comparison of Density and Distribution Functions

Question 48

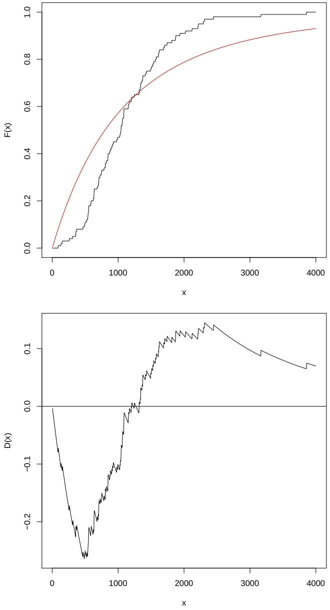
For each of the models on the following three slides, determine which of the statements below best describes the fit between the model and the data:

i. The model distribution assigns too much probability to high values and too little probability to low values.
ii. The model distribution assigns too much probability to low values and too little probability to high values.
iii. The model distribution assigns too much probability to tail values and too little probability to central values.
iv. The model distribution assigns too much probability to central values and too little probability to tail values.

Model I

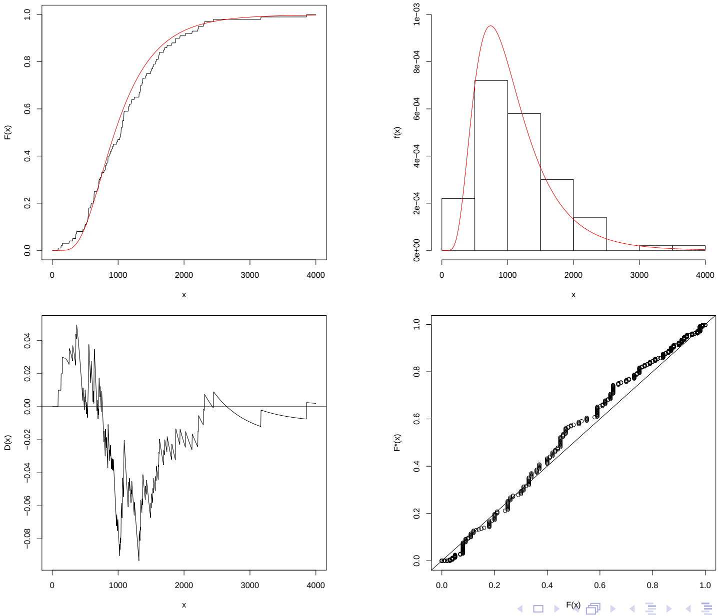


Model II



F(x)
0.0
0.2
0.4
0.6
0.8
1.0
0
1000
2000
3000
4000
x
f(x)
0e+00
1e−04
2e−04
3e−04
4e−04
5e−04
6e−04
7e−04
0
1000
2000
3000
4000
x
D(x)
−0.2
−0.1
0.0
0.1
0
1000
2000
3000
4000
x
F*(x)
0.0
0.2
0.4
0.6
0.8
0.0
0.2
0.4
0.6
0.8
1.0
F(x)

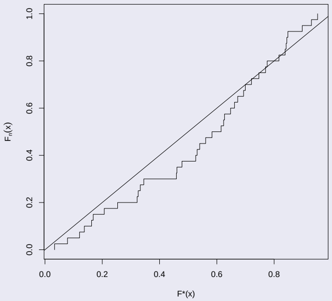
Model III



F(x)
0.0
0.2
0.4
0.6
0.8
1.0
0
1000
2000
3000
4000
x
f(x)
0e+00
2e−04
4e−04
6e−04
8e−04
1e−03
0
1000
2000
3000
4000
x
D(x)
−0.08
−0.06
−0.04
−0.02
0.00
0.02
0.04
0
1000
2000
3000
4000
x
F*(x)
0.0
0.2
0.4
0.6
0.8
1.0
0.0
0.2
0.4
0.6
0.8
1.0
F(x)

Question 49

An insurance company wants to know whether an exponential distribution is a good fit for a sample of 40 claim severities. It estimates $\theta = 5.609949$, and draws the following p-p plot:

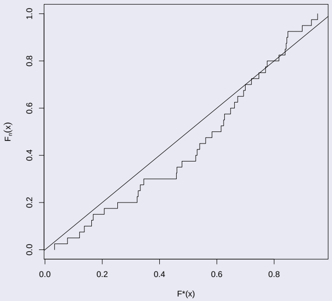


How many of the samples they collected were more than 10?

16.3 Graphical Comparison of Density and Distribution Functions

Question 50

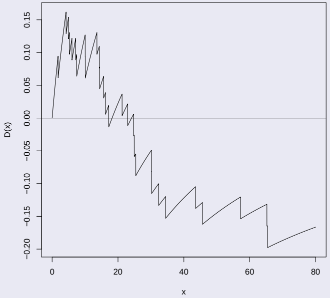
An insurance company wants to know whether an exponential distribution is a good fit for a sample of 40 claim severities. It estimates $\theta = 5.609949$, and draws the following p-p plot:



How many of the samples they collected were less than 3?

Question 51

An insurance company wants to know whether a Pareto distribution with $\theta = 15$ is a good fit for a sample of 30 claim severities. It estimates $\alpha = 0.8725098$ and draws the following plot of $D(x)$:



How many of the samples they collected were less than 10?

Hypothesis Tests

We test the following hypotheses:

H_0: The data came from a population with the given model.

H_1: The data did not come from a population with the given model.

16.4 Hypothesis Tests

Kolmogorov-Smirnov test

$$D = \max_{t \leqslant x \leqslant u} |F_n(x) - F(x)|$$

Anderson-Darling test

$$A^2 = n \int_t^u \frac{(F_n(x) - F(x))^2}{F(x)(1 - F(x))} f(x)\, dx$$

Chi-square Goodness-of-fit test

- Divide the range into separate regions, $t = c_0 < c_1 < \cdots < c_n = u$.
- Let O_i be the number of samples in the interval $[c_{i-1}, c_i)$.
- Let E_i be the expected number of sample in the interval $[c_{i-1}, c_i)$.

$$X^2 = \sum_{i=1}^{n} \frac{(O_i - E_i)^2}{E_i}$$

Question 52

For the data from Question 43:

325	692	1340	1784	1920	2503	3238	4054	5862
6304	6926	8210	9176	9984				

Test the goodness of fit of the model using:
(a) The Kolmogorov-Smirnov test.
(b) The Anderson-Darling test.

Answer to Question 52

(b)

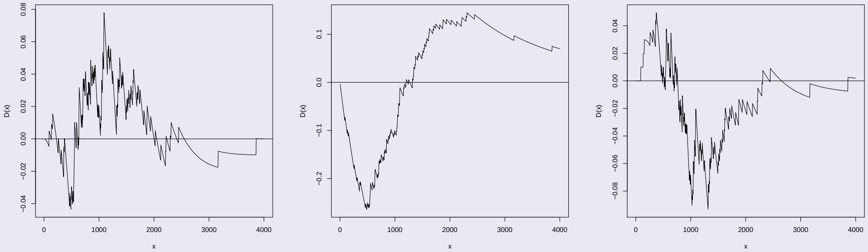
$$A^2 = -nF^*(u) + \sum_{j=0}^{k}(1 - F_n(y_j))^2(\log(1 - F^*(y_j)) - \log(1 - F^*(y_{j+1}))$$

$$+ n\sum_{j=1}^{k} F_n(y_j)^2(\log(F^*(y_{j+1})) - \log(F^*(y_{j+1})))$$

x	$F_n(x)$	$F^*(x)$	term	x	$F_n(x)$	$F^*(x)$	term
325	0.0714	0.0704	0.0748	4054	0.5714	0.5978	0.1407
692	0.1429	0.1440	0.1190	5862	0.6429	0.7320	0.0267
1340	0.2143	0.2600	0.0726	6304	0.7143	0.7573	0.0323
1784	0.2857	0.3303	0.0204	6926	0.7857	0.7889	0.0532
1920	0.3571	0.3504	0.0803	8210	0.8571	0.8417	0.0309
2503	0.4286	0.4302	0.0876	9176	0.9286	0.8726	0.0215
3238	0.5000	0.5169	0.0822	9984	1.0000	0.8937	0.1124

Question 53

Recall Question 47, where a company was deciding between three models. The $D(x)$ plots are below:



If the company uses the Kolmogorov-Smirnov statistic to decide the best model, which will it choose?

Question 54

An insurance company records the following claim data:

Claim Amount	Frequency
0–5,000	742
5,000–10,000	1304
10,000–15,000	1022
15,000–20,000	830
20,000–25,000	211
More than 25,000	143

Use a Chi-square test to determine whether Claim size follows an exponential distribution. The best mean for the exponential distribution is $\theta = 9543.586$.

Likelihood Ratio test

The Likelihood ratio test compares two nested models — $\mathcal{M}_0$ and $\mathcal{M}_1$.

Hypotheses

H_0: The simpler model describes the data as well as the more complicated model.

H_1: The more complicated model describes the data better than the simpler model.

We compute the parameters from both models by maximum likelihood. The test statistic is.

$$2(l_{\mathcal{M}_1}(x;\theta_1) - l_{\mathcal{M}_0}(x;\theta_0))$$

Under H_0, for large n, this follows a Chi-square distribution with degrees of freedom equal to the difference in number of parameters.

Question 55

An insurance company observes the following sample of claim data:

382	596	920	1241	1358	1822	2010	2417	2773
3002	3631	4120	4692	5123				

Use a likelihood ratio test to determine whether an exponential or a Weibull distribution fits this data better.
The maximum likelihood estimates for the Weibull distribution are $\tau = 1.695356$ and $\theta = 2729.417$.

Study note: Information Criteria

Basic Idea

- For natural measures of fit (log-likelihood, KS test statistic, AD test statistic, etc.) more complicated models produce better fit.
- This is (at least partly) because they are fitting noise in the data.
- We can compensate for this by adding a penalty term to penalise model complexity.

Two Common Approaches

- Akaike Information Criterion (AIC): $l(\theta; x) - p$
- Schwarz Bayesian Criterion(SBC)/Bayesian Information Criterion (BIC): $l(\theta; x) - \frac{p}{2}\log(n)$

where p is the number of estimated parameters, and n is the sample size.

Question 56

Recall Question 55, where we had a sample

382	596	920	1241	1358	1822	2010	2417	2773
3002	3631	4120	4692	5123				

for which the Weibull distribution has a log-likelihood of -120.7921. Use AIC and BIC to determine whether an inverse exponential distribution is a better fit for the data.

16.5 Selecting a Model

Comments on Model Selection

- Try to pick a model with as few parameters as possible. (Parsimony)
- Choice of model depends on the aspects that are important. Even if a formal test is used, the choice of which test depends on the aspects that are important.
- Aim is generalisability. The model should apply to future data. (Models which fit the given data well, but not new data are said to overfit.)
- Trying large numbers of models will lead to one which fits well just by chance.
- Experience is a valuable factor in deciding on a model.
- Sometimes knowledge of the underlying process may lead to a particular model (e.g. binomial).

Credibility Theory (Revision)

Problem

- Policyholders are all different, so average rates from large populations will be wrong for individual policyholders.
- Not enough data to estimate a reliable rate for individual policyholder or group.

Limited Fluctuation Credibility

- Determine how much experience is needed to reliably estimate a premium for an individual policyholder or group.
- For groups with less experience, take a weighted average $Z\overline{X} + (1 - Z)\mu$, where $Z = \sqrt{\frac{n}{n_0}}$ is the group's credibility.

Problems with Limited Fluctuation Credibility

- No theoretical justification.
- Parameters r and p chosen arbitrarily.

18 Greatest Accuracy Credibility

Assumptions

- Each policyholder has a risk parameter Θ, which is a random variable, but is assumed constant for that particular policyholder.
- Individual values of Θ can never be observed.
- The distribution of this risk parameter Θ has density (or mass) function $\pi(\theta)$, which is known. (We will denote the distribution function $\Pi(\theta)$.)
- For a given value $\Theta = \theta$, the conditional density (or mass) of the loss distribution $f_{X|\Theta}(x|\theta)$ is known.

18.2 Conditional Distributions and Expectation

Conditional Distributions (revision)

$$f_{X|\Theta}(x|\theta) = \frac{f_{X,\Theta}(x,\theta)}{\int f_{X,\Theta}(y,\theta)\,dy}$$

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

Conditional Expectation (revision)

$$\mathbb{E}(X) = \mathbb{E}(\mathbb{E}(X|\Theta))$$

$$\mathrm{Var}(X) = \mathbb{E}(\mathrm{Var}(X|\Theta)) + \mathrm{Var}(\mathbb{E}(X|\Theta))$$

Question 57

An insurance company models drivers as falling into two categories: frequent and infrequent. 75% of drivers fall into the frequent category. The number of claims made per year by a driver follows a Poisson distribution with parameter 0.4 for frequent drivers and 0.1 for infrequent drivers.

(a) Calculate the expectation and variance of the number of claims in a year for a randomly chosen driver.

(b) Calculate the expectation and variance of the number of claims in a year for a randomly chosen driver who made no claims in the previous year.

Question 58

The aggregate health claims (in a year) of an individual follows an inverse gamma distribution with $\alpha = 3$ and θ varying between individuals. The distribution of θ is a Gamma distribution with parameters $\alpha = 3$ and $\theta = 100$.

(a) Calculate the expected total health claims for a random individual.

(b) If an individual's aggregate claims in two consecutive years are \$112 and \$240, calculate the expected aggregate claims in the third year.

Question 59

The number of claims made by an individual in a year follows a Poisson distribution with parameter Λ. Λ varies between individuals, and follows a Gamma distribution with $\alpha = 0.5$ and $\theta = 2$.
(a) Calculate the expected number of claims for a new policyholder.
(b) Calculate the expected number of claims for a policyholder who has made m claims in the previous n years.

Question 60

The number of claims made by an individual in a year follows a Poisson distribution with parameter Λ. Λ varies between individuals, and follows a Pareto distribution with $\alpha = 4$ and $\theta = 3$. [This has mean 1 and variance 2, like the Gamma distribution from Question 59.] Calculate the expected number of claims for a policyholder who has made m claims in the previous n years.

Answer to Question 60

	Pareto Prior					Gamma Prior			
	1	2	3	4		1	2	3	4
0	0.433	0.294	0.224	0.182	0	0.333	0.200	0.143	0.111
1	0.926	0.607	0.458	0.369	1	1.000	0.600	0.429	0.333
2	1.479	0.940	0.700	0.561	2	1.667	1.000	0.714	0.556
3	2.087	1.289	0.951	0.758	3	2.333	1.400	1.000	0.778
4	2.749	1.654	1.208	0.958	4	3.000	1.800	1.286	1.000
5	3.457	2.034	1.472	1.163	5	3.667	2.200	1.571	1.222
6	4.207	2.426	1.742	1.370	6	4.333	2.600	1.857	1.444
7	4.992	2.829	2.018	1.581	7	5.000	3.000	2.143	1.667
8	5.807	3.242	2.298	1.795	8	5.667	3.400	2.429	1.889
9	6.648	3.664	2.583	2.011	9	6.333	3.800	2.714	2.111

Problems with Bayesian Approach

- Difficult to Compute.
- Sensitive to exact model specification.
- Difficult to perform model selection for the unobserved risk parameter Θ.

Approach

- Credibility premium is a linear combination of book premium and personal history.

$$\alpha_0 + \sum_{i=1}^{n} \alpha_i X_i$$

- Coefficients are chosen to minimise Mean Squared Error (MSE)

$$\mathbb{E}\left(\mu(\Theta) - \left(\alpha_0 + \sum_{i=1}^{n} \alpha_i X_i\right)\right)^2$$

Question 61

Show that the solution which minimises the MSE satisfies:

$$\mathbb{E}(X_{n+1}) = \alpha_0 + \sum_{i=1}^{n} \alpha_i \mathbb{E}(X_i)$$

$$\mathrm{Cov}(X_i, X_{n+1}) = \sum_{j=1}^{n} \alpha_j \,\mathrm{Cov}(X_i, X_j)$$

Question 62

Suppose the X_i all have the same mean, the variance of X_i is σ^2, and the covariance $\mathrm{Cov}(X_i, X_j) = \rho$. Calculate the credibility estimate for X_{n+1}.

Question 63

Suppose we have observations $X_1, \ldots, X_n$ and $Y_1, \ldots, Y_m$, which are the aggregate annual claims for each of two cars driven by an individual. We assume:

$$\begin{aligned}
\mathbb{E}(X_i) &= \mu \\
\mathbb{E}(Y_i) &= \nu \\
\mathrm{Var}(X_i) &= \sigma^2 \\
\mathrm{Var}(Y_i) &= \tau^2 \\
\mathrm{Cov}(X_i, X_j) &= \rho \quad \text{for } i \neq j \\
\mathrm{Cov}(Y_i, Y_j) &= \zeta \quad \text{for } i \neq j \\
\mathrm{Cov}(X_i, Y_j) &= \xi
\end{aligned}$$

Calculate the credibility estimate for $X_{n+1} + Y_{m+1}$.

Assumptions

- $X_1, \ldots, X_n$ are i.i.d. conditional on Θ.

We then define:

$\mu(\theta)=\mathbb{E}(X|\Theta=\theta)$ $\mu=\mathbb{E}(\mu(\Theta))$

$\nu(\theta)=\mathrm{Var}(X|\Theta=\theta)$ $\nu=\mathbb{E}(\nu(\Theta))$

$a=\mathrm{Var}(\mu(\Theta))$

Solution

$$\mathbb{E}(X_i)=\mu \qquad \mathrm{Var}(X_i)=\nu+a$$

$$\mathrm{Cov}(X_i,X_j)=a$$

Recall from Question 62, that the solution to this is:

$$\hat{\mu}=\frac{\left(\frac{v}{a}\right)}{n+\left(\frac{v}{a}\right)}\mu+\frac{n}{n+\left(\frac{v}{a}\right)}\overline{X}$$

Question 64

An insurance company offers group health insurance to an employer. Over the past 5 years, the insurance company has provided 851 policies to employees. The aggregate claims from these policies are \$121,336. The usual premium for such a policy is \$326. The variance of hypothetical means is 23,804, and the expected process variance is 84,036. Calculate the credibility premium for employees of this employer.

Question 65

An insurance company offers car insurance. One policyholder has been insured for 10 years, and during that time, the policyholder's aggregate claims have been $3,224. The book premium for this policyholder is $990. The expected process variance is 732403 and the variance of hypothetical means is 28822. Calculate the credibility premium for this driver next year.

18.6 The Bühlmann-Straub Model

Assumptions

- Each observation X_i (expressed as loss per exposure) has a (known) exposure m_i. The conditional variance of X_i is $\frac{v(\theta)}{m_i}$.

$$\mathrm{Cov}(X_i, X_j) = a$$
$$\mathrm{Var}(X_i) = \frac{v}{m_i} + a$$

Solution

$$\alpha_0 = \frac{\left(\frac{v}{a}\right)}{m + \frac{v}{a}}\mu \qquad \alpha_i = \frac{m_i}{m + \frac{v}{a}}$$
$$\hat{\mu} = \frac{\left(\frac{v}{a}\right)}{m + \frac{v}{a}}\mu + \frac{m}{m + \frac{v}{a}}\overline{X}$$

where $\overline{X}$ is the weighted mean $\sum_{i=1}^{n} \frac{m_i}{m} X_i$.

Question 66

For a group life insurance policy, the number of lives insured and the total aggregate claims for each of the past 5 years are shown in the following table:

Year	1	2	3	4	5
Lives insured	123	286	302	234	297
Agg. claims	0	\$300,000	\$200,000	\$200,000	\$300,000

The book rate for this policy premium is \$1,243 per life insured. The variance of hypothetical means is 120,384 and the expected process variance is 81,243,100. Calculate the credibility premium per life insured for the next year of the policy.

Question 67

A policyholder holds a landlord's insurance on a rental property. This policy is in effect while the property is rented out. The company has the following experience from this policy:

Year	1	2	3	4	5	6
Months rented	3	11	8	12	6	9
Agg. claims	0	\$10,000	0	0	\$4,000	0

The standard premium is \$600 per year for this policy. The variance of hypothetical means is 832076, and the expected process variance is 34280533 (both for annual claims). Calculate the credibility premium for the following year using the Bühlmann-Straub model.

Question 68

Show that if the Bayes premium is a linear function of X_i, and the expectation and variance of X are defined, then the Bayes premium is equal to the credibility premium.

Question 69

Show that if the model distribution is from the linear exponential family, and the prior is the conjugate prior, with $\frac{\pi(\theta_1)}{r'(\theta_1)} = \frac{\pi(\theta_0)}{r'(\theta_0)}$, where θ_0 and θ_1 are the upper and lower bounds for θ, then the Bayes premium is a linear function in X.

19 Empirical Bayes Parameter Estimation

Approach

- Estimate the distribution of Θ from the data.
- Use this estimate to calculate the credibility estimate of μ.

Two possibilities

Either: We do not have a good model for the conditional or prior distribution. We only need the variances, so we estimate them non-parametrically.

or: We have a parametric model, such as a Poisson distribution, which allows us to estimate the variance more efficiently (assuming the model is accurate).

Question 70

An insurance company has the following aggregate claims data on a new type of insurance policy:

No.	Year 1	Year 2	Year 3	Year 4	Year 5	Mean	Variance
1	336	0	528	0	0	172.80	60595.2
2	180	234	0	2,642	302	671.60	1225822.8
3	0	0	528	361	0	177.80	62760.2
4	443	729	1,165	0	840	635.40	192962.3
5	0	0	0	0	0	0.00	0.0
6	196	482	254	303	0	247.00	30505.0
7	927	0	884	741	604	633.60	140653.7
8	0	601	105	130	327	232.60	56385.3

(a) Estimate the expected process variance and the variance of hypothetical means.
(b) Calculate the credibility premiums for each policyholder next year.

Theorem

Let $X_1, \ldots, X_n$ have means $M_1, \ldots, M_n$ respsectively. Let the M_i have mean μ, and let $X_i|M_1$ have variance $\frac{\sigma^2}{m_i}$ where all m_i are known. Let $m = \sum_{i=1}^{n} m_i$.
We can obtain the following unbiassed estimators for μ and σ^2:

$$\hat{\mu} = \frac{\sum_{i=1}^{n} m_i X_i}{m}$$

$$\hat{\sigma^2} = \frac{\sum_{i=1}^{n} m_i (X_i - \hat{\mu})^2}{n-1}$$

Question 71

Let $X_1, \ldots, X_n$ all have mean μ, and let X_i have variance $\frac{\sigma_i^2}{m_i}$, where all m_i are known, and let $\sigma = \mathbb{E}(\sigma_i^2)$. Let M_i have variance τ^2. Let $m = \sum_{i=1}^n m_i$. Let $m = \sum_{i=1}^n m_i$.
Show that the following is an estimator for the Variance of Hypothetical Means:

$$\hat{\mu} = \frac{\sum_{i=1}^n m_i X_i}{m}$$

$$\mathrm{V\hat{H}M} = \frac{\sum_{i=1}^n m_i (X_i - \hat{\mu})^2 - (n-1)\hat{\sigma}^2}{m - \frac{\sum m_i{}^2}{m}}$$

Question 72

An insurance company offers a group-life policy to 3 companies. These are the companies' exposures and aggregate claims (in millions) for the past 4 years:

Co.		Year 1	Year 2	Year 3	Year 4	Total
1	Exp	769	928	880	1,046	3,623
	Claims	1.3	1.5	0.8	1.7	5.3
2	Exp	1,430	1,207	949	1,322	4,908
	Claims	1.0	0.9	0.6	1.5	4.0
3	Exp	942	1,485	2,031	1,704	6,162
	Claims	1.1	1.4	1.9	2.0	6.4

Calculate the credibility premiums per life for each company in the fifth year.

Question 73

In a particular year, an insurance company observes the following claim frequencies:

No. of Claims	Frequency
0	3951
1	1406
2	740
3	97
4	13
5	3

Assuming the number of claims an individual makes follows a Poisson distribution, calculate the credibility estimate for number of claims for an individual who has made 6 claims in the past 3 years.

Question 74

Assume annual claims from one policyholder follow a Poisson distribution with mean Λ. The last 4 years of claims data are:

Claims	0	1	2	3	4	5	6	7	8	9
1 year	3951	1406	740	97	13	3	0	0	0	0
2 years	3628	2807	1023	461	104	13	4	0	1	0
3 years	2967	4032	2214	890	734	215	131	22	0	2
4 years	1460	2828	2204	985	747	358	194	43	8	0

Calculate the credibility estimate of Λ for an individual who made 2 claims in the last 3 years of coverage.

Question 75

Claim frequency in a year for an individual follows a Poisson with parameter Λt where Λ is the individual's risk factor and t is the individual's exposure in that year. An insurance company collects the following data:

Policy-holder	Year 1		Year 2		Year 3		Year 4	
	Exp	claims	Exp	claims	Exp	claims	Exp	claims
1	45	12	10	6	45	14	14	2
2	27	0	12	0	74	0	27	0
3	10	9	293	149	14	6	13	5
4	10	0	14	3	17	2	6	2

In year 5, policyholder 3 has 64 units of exposure. Calculate the credibility estimate for claim frequency for policyholder 3.

Question 76

An insurer is reviewing aggregate claims data from last year. It assumes that average aggregate claims for an individual follows an exponential distribution with parameter varying between individuals. The insurer has data from 1000 policyholders and finds that the average aggregate claim is $689 and the standard deviation is $832. What is the credibility premium for an individual who claimed $462 last year?

SN1.1 Introduction (Revision)

Reasons for Delays

- Delays in Reporting
- Claims Processing Delays
- Legal Procedings

Approaches (revision)

- Case-based estimation — Use adjustment estimates for each claim. Usually only used for very large claims.
- Expected Loss Ratio — Estimate losses from earned premiums.
- Aggregate Run-off Triangle Methods — Project future from past. For example Chain-Ladder and Bornhuetter-Fergusson methods.

Approaches

- Credibility methods — Weighted average of two estimates.
- Frequency-Severity — Separately estimate frequency & severity.
- Parametric methods — Based on a parametric model. Usually give same estimates, but better inference.

Run-off Triangles (Revision)

Accident Year	Development Year 0	1	2	3	4	5
0	801	962	887	728	560	77
1	879	1043	968	802	606	
2	957	1155	1057	852		
3	1033	1238	1144			
4	1119	1340				
5	1207					

- Entries give payments made in each development year, for each accident year.
- Antidiagonals correspond to payments made in a single calendar year.
- Assume at least AY0 is closed. Tail-factor methods exist if not.
- Often work with cumulative payments by summing rows of triangle.

Notation

- $X_{i,j}$ — incremental payments made in development years 0–j for claims in accident year i.
- $C_{i,j}$ — cumulative payments made in development years 0–j for claims in accident year i. That is $C_{i,j} = \sum_{k=0}^{j} X_{i,k}$.

Method

- Let $f_{i.j} = \frac{C_{i,j+1}}{C_{i,j}}$.
- Estimate an average $\hat{f}_j$ for each j, either a direct average of $f_{i,j}$, or a weighted average $\hat{f}_j = \frac{\sum_{i=0}^{I-j-1} C_{i,j+1}}{\sum_{i=0}^{I-j-1} C_{i,j}}$.
- Use these $\hat{f}_j$ to estimate all unknown $C_{i,j}$ from $C_{i,j-1}$.

Question 77

For the run-off table:

Accident Year	Development Year 0	1	2	3	4	5
0	2006	2098	2321	1795	1387	431
1	2104	2204	2418	1893	1474	
2	2196	2321	2533	1959		
3	2314	2426	2659			
4	2425	2563				
5	2503					

Estimate the future losses. (For convenience, this table is in the file "`RunOff1.txt`".)

Question 78

Suppose that claim inflation over the previous 5 years was given by the following indices

Year	0–1	1–2	2–3	3–4	4–5
Inflation (%)	2	4	7	5	1

Where Year 0 represents Accident Year 0. Recalculate the expected future claims from Question 77, adjusted for this inflation.

$\hat{C_{i,J}}$ as an Expected Value

Assumptions

- There is a development factor f_j such that $\mathbb{E}(C_{i,j+1}|\mathcal{D}_{i+j}) = f_j C_{i,j}$
- $C_{i,j}$ and $C_{l,k}$ are independent when $i \neq l$ for all j and k.

Theorem

Under the above assumptions:

(a) $\hat{f}_j$ *is an unbiased estimator for* f_j.

(b) $\mathbb{E}(\hat{f}_0 \hat{f}_1 \cdots \hat{f}_j) = f_0 f_1 \ldots f_j$ *for all* j.

(c) $\hat{C_{i,J}}$ *is an unbiased estimator.*

SN1 3.2 Testing Chain-Ladder Assumptions

Correlated Development Factors

- Can use Pearson Correlation coefficients between years.
- Significance based on normality and same variance.
- Alternatively, can use Spearman rank correlation coefficient.
- Weaker test but more robust.
- Does not test correlation of the original variables.

Calendar Year Effects

- Can arise from changes to settlement policy or claim inflation.
- Simple test: rank the estimates $\hat{f}_{i,j}$ for each j, and count values above or below the median on each antidiagonal.

Multiple Testing

- Statistical tests reject true hypotheses α% of the time.
- If we conduct many statistical tests (e.g. one test for each year) the expected number of rejections is too high.
- There are methods to correct for this.

Question 79

Test the assumptions of the Chain-ladder method in Question 77.

Question 80

Plot the estimated cumulative development factors for the adjusted run-off table in Question 78.

SN1 3.3 Bornhuetter-Fergusson Method (Revision)

Bornhuetter-Ferguson method

1. Calculate the expected ultimate claim payments (using expected ultimate loss ratio times earned premiums)
2. Calculate loss development factors using chain-ladder method
3. Work backwards from expected ultimate payments using loss development factors to get expected loss development.

Question 81

Recall Question 77, where the mean loss development factors were

Year	1/0	2/1	3/2	4/3	5/4
$\hat{f}_j$	2.051335	1.562058	1.279541	1.169903	1.044863

Suppose the expected loss ratio is 0.81, and the earned premiums are

Accident Year	0	1	2	3	4	5
Earned Prem.	11980	12105	12610	13240	14370	14600

Use the Bornhuetter Fergusson method to calculate the loss reserves needed for each accident year.

SN1 3.5 Bühlmann-Straub Credibility Reserves

Assumptions

- Losses from AY i follow distribution, with unknown θ_i drawn i.i.d.
- Given $\theta_i, \theta_{i'}$, $X_{i,j}$ and $X_{i',j'}$ are independent.
- If $\theta_i = \theta_{i'}$ then $X_{i,j}$ and $X_{i',j}$ are identically distributed.
- $\mathbb{E}(X_{i,j}|\theta_i) = \gamma_j \mu(\theta_i)$ and $\mathrm{Var}(X_{i,j}|\theta_i) = \gamma_j \nu(\theta_i)$.

Method

- Estimate EPV $v = \mathbb{E}(\nu(\Theta))$ and VHM $a = \mathrm{Var}(\mu(\Theta))$.

$$\hat{v} = \frac{1}{I}\sum_{i=0}^{I-1}\frac{1}{I-i}\sum_{j=0}^{I-i}\hat{\gamma}_j\left(\frac{X_{ij}}{\hat{\gamma}_j} - \hat{C_{i,J}}\right)^2 \qquad \hat{a} = \frac{\sum_{i=0}^{I}\hat{\beta}_{I-i}\left(\hat{C}_{i,J} - \overline{C}\right)^2 - I\hat{v}}{\sum_{i=0}^{I}\hat{\beta}_{I-i} - \frac{1}{\sum_{i=0}^{I}\hat{\beta}_{I-i}}\sum_{i=0}^{I}\hat{\beta}_{I-i}^2}$$

- Estimate credibility $Z_i = \frac{\hat{\beta}_{I-i}}{\hat{\beta}_{I-i} + \frac{\hat{v}}{\hat{a}}}$ for each year.
- Estimate $\mu = \mathbb{E}(\mu(\Theta))$, by $\hat{\mu} = \frac{\sum_{i=0}^{I} Z_i \hat{C}_{i,J}}{\sum_{i=0}^{I} Z_i}$.
- Estimate $\hat{C}_{i,J}^{\mathrm{BS}} = Z_i \hat{C}_{i,J} + (1 - Z_i)\hat{\mu}$.
- Estimate $\hat{C}_{i,J}^{\mathrm{BS2}} = C_{i,I-i} + (1 - \hat{\beta}_j)\hat{C}_{i,J}^{\mathrm{BS}}$.

SN1 3.5 Bühlmann-Straub Credibility Reserves

Comments

- $\hat{C}^{\mathrm{BS2}}_{i,j} = Z_i^* \hat{C}_{i,j} + (1 - Z_i^*)\hat{\mu}$ for $Z_i^* = 1 - (1 - \hat{\beta}_{i,I-i})(1 - Z_i)$.
- Big assumption that μ is the same for all accident years. Can change $X_{i,j}$ to per-premium losses to make this assumption more reasonable.

Question 82

For the run-off table from Question 77, use the Bühlman-Straub method to estimate the total reserve payments needed.

SN1 3.5 The Poisson Model

Assumptions

- $X_{i,j}$ are independent for all i and j.
- $X_{i,j} \sim Po(\mu_i \gamma_j)$ for some $\mu_i, \gamma_j > 0$ such that $\sum_{j=0}^{J} \gamma_j = 1$.

Results

- These are the same assumptions as BF. Thus, if μ_i is given *a priori*, we can use the same estimates $\hat{\beta}_i$ and $\hat{\gamma}_i$.
- $C_{i,j}$ and $C_{i',j'}$ independent whenever $i \neq i'$.
- $\mathbb{E}(C_{i,j+1}|\mathcal{D}_j) = C_{i,j} + (\beta_{j+1} - \beta_j)\mu_i$
- $\mathbb{E}(C_{i,J}|\mathcal{D}) = C_{i,j} + (1 - \beta_j)\mu_i$

Question 83

Show that the MLE estimate from the data under the Poisson model gives the chain-ladder estimate for average loss reserves, and calculate the variance of outstanding claims under the Poisson model.

Assumptions

1. For $i \neq i'$, and any j, j', $C_{i,j}$ and $C_{i',j'}$ are independent.
2. For Accident Year i, $(C_{i,j})_{j=0,\dots,J}$ is a Markov chain (meaning for $k < j$, $C_{i,j+1}$ and $C_{i,k}$ are conditionally independent given $C_{i,j}$).
3. $\mathbb{E}(C_{i,j+1}|C_{i,j}) = f_j C_{i,j}$ for some factor f_j.
4. $\mathrm{Var}(C_{i,j+1}|C_{i,j}) = \sigma_j^2 C_{i,j}$ for some σ_j^2.

Comments

- (1) and (3) are the assumptions for the chain-ladder method.
- For $j \leqslant I - 2$, unbiased estimator $\hat{\sigma}_j^2 = \dfrac{1}{I-1-j} \displaystyle\sum_{i=0}^{I-1-j} C_{ij} \left(f_{ij} - \hat{f}_j\right)^2$.
- When $J = I$, we cannot estimate σ_{J-1}^2. Mack suggests using $\hat{\sigma}_{J-1}^2 = \min\left(\sigma_{J-2}^2, \sigma_{J-3}^2, \frac{\sigma_{J-2}^4}{\sigma_{J-3}^2}\right)$.

SN1 4 Mack's Model— Estimating the Variance

Theorem

The process variance for $C_{i,J}|C_{i,I-i}$ is approximated by

$$\mathrm{Var}(C_{i,J}|C_{i,I-i}) \approx \hat{C}_{i,J}^2 \sum_{j=I-i}^{J-1} \frac{\hat{\sigma}_j^2}{\hat{f}_j^2 \hat{C}_{i,j}}$$

Theorem

Under Mack's model, we have

$$\mathbb{E}\left(\left(\hat{C}_{i,J} - \mathbb{E}(C_{i,j}|D_I)\right)^2\right) \approx \hat{C}_{i,J}^2 \sum_{j=I-i}^{J} \frac{\hat{\sigma}_j^2}{\hat{f}_j^2 S_j}$$

and

$$\mathbb{E}\left(\left(\hat{C}_{i,J} - \mathbb{E}(C_{i,j}|D_I)\right)\left(\hat{C}_{i',J} - \mathbb{E}(C_{i',j}|D_I)\right)\right) \approx \hat{C}_{i,J}\hat{C}_{i',J} \sum_{j=I-(i\wedge i')}^{J} \frac{\hat{\sigma}_j^2}{\hat{f}_j^2 S_j}$$

where $S_j = \sum_{i=0}^{I-1-j} C_{i,j}$.

Question 84

(a) Estimate $\hat{\sigma}_j^2$ for all j for the run-off table from Question 77.

(b) Using these, estimate the variance of the outstanding claims.

(c) Estimate the mean squared estimation error for each $\hat{C}_{i,J}$.

(d) Estimate the mean product of estimation errors for each pair $\hat{C}_{i,J}$ and $\hat{C}_{i',J}$, and use this to estimate the total MSE of the outstanding losses. .

SN1. 5 The Overdispersed Poisson Model

Poisson Model as a GLM

- We can rewrite the Poisson model as $\log(\mathbb{E}(X_{i,j})) = c + \alpha_i + \beta_j$.
- This is a Generalised Linear Model with $\log$ link.
- The parameters $\alpha_i = \log(\mu_i)$ and $\beta_i = \log(\gamma_i)$ are estimated by maximum likelihood.

Quasilikelihood and Overdispersion

- Poisson log-likelihood is $x\log(\lambda) - \lambda$ (ignore the $x!$ constant.)
- Derivative of log-likelihood is $\sum_{i,j}\left(\frac{x_{ij}}{\lambda_{ij}} - 1\right)\frac{\partial\lambda_{ij}}{\partial\theta}$
- This assumes variance is equal to x. If we replace the log-likelihood function by the quasilikelihood function whose derivative is $\frac{\partial l}{\partial\theta} = \sum_{i,j}\left(\phi\frac{x_{ij}}{\lambda_{ij}} - 1\right)\frac{\partial\lambda_{ij}}{\partial\theta}$
- This is not the likelihood of an actual discrete distribution, but can be used to approximate a large number of distributions.

Question 85

Fit an overdispersed Poisson model to the data from the run-off table in Question 77.

SN1.6 Frequency-Severity Models

Frequency

- Separately analysing frequency and severity can show trends or outliers which are not apparent in the aggregate claims data.
- Chain-ladder assumptions are reasonable for frequency.
- For frequency, we often need to analyse both reported and settled claims. Reported claims usually have fast development.
- Separate estimates of reported and settled claims inconsistent.
- Better (but still inconsistent) approach: Estimate γ_j^S — proportion of claims settled in development year j. Estimate claims settled as $\hat{\gamma}_j^s \hat{C}_{i,J}^R$, where $\hat{C}_{i,J}^R$ is estimated total reported claims.

Severity

- Common approach for severity is to calculate cumulative average severity. The chain-ladder assumptions here are dubious.
- Longer settlement times usually correlated with larger claims. Thus, better to calculate incremental average claim cost.

Question 86

An actuary is reviewing the following loss development triangles in the files:

```
ClaimsReportedRunOff.txt
ClaimsSettledRunOff.txt
AggregateSettledPaymentsRunOff.txt
```

(a) Estimate the outstanding claim settlements using the chain-ladder method on reported claims, and using the proportion of estimated reported claims.

(b) Estimate the aggregate reserves using the average cumulative losses and the average incremental losses per claim.

3.9 Rate Changes

Overall Rate Change (Revision)

- Loss cost method: $\text{New average gross rate} = \frac{\text{New Average Loss Cost}}{1-\text{Expense Ratio}}$
- Loss ratio method: $\text{Rate Change} = \frac{\text{Expected Effective Loss Ratio}}{\text{Permissible Loss Ratio}} - 1$

Risk Classification Differential Changes

- Rate manual consists of rate for base cell, and for each variable, a vector of differentials — multiplicative factors.

Question 87

An insurer has three classes of risk - low, medium and high. Its experience from the previous year is shown in the table below.

Risk Class	Current differential	Earned premiums	Loss payments
Low	0.74	1,300	1,100
Medium	1	4,300	3,900
High	1.46	1,600	1,400

Calculate the new differentials for the coming year.

Question 88

An insurer has base rate \$46.30 and expense ratio is 20%. Its experience from the previous year is shown in the table below.

		Earned Premiums		Loss Payments	
		Male	Female	Male	Female
Differential		1	0.88	1	0.88
Low	0.74	900	1,100	1,050	850
Medium	1	4,700	4,400	4,100	3,900
High	1.46	1,900	1,400	1,200	1,100

(a) Calculate the new differentials.

(b) If the base premium is adjusted by the loss ratio for this year, calculate the loss ratio with the new differentials.

(c) What base premium would give the desired loss ratio?

(d) What would the new premiums be if the original differentials had been 1.66 for female, 0.34 for low-risk and 1.89 for high-risk?

Question 89

The categories and differentials for three factors in 2022 were:

Age		Sex		Health Status	
Young	1	Female	1	Healthy	1
Old	1.74	Male	1.18	Unhealthy	1.49

Earned premiums in accident year 2022 were:

	Female			Male		
	Healthy	Unhealthy	Total	Healthy	Unhealthy	Total
Young	3,600	1,800	5,400	3,200	1,700	4,900
Old	7,300	6,900	14,200	5,300	5,800	11,100
Total	10,900	8,700	19,600	8,500	7,500	16,000

After reviewing the data from 2022, the new differentials are

Old	1.63	Male	1.14	Unhealthy	1.57

The total losses were $29,000. Calculate the percentage change in the base premium which achieves a loss ratio of 0.8 with the new differentials.