1 Solvency and pricing

1.1 Introduction

The principal tasks of an actuary in general insurance is solvency and pricing. Solvency is financial control of liabilities under near worst-case scenarios. Target is then the (upper) percentiles q_ϵ of the portfolio risk $\mathcal{X}$, known as the **reserve**, which takes modelling and computation to determine. Examples have been spread through several of the previous chapters, but we shall now discuss general approaches, into which the many details arising in practice can be fed. Evaluation of the reserve takes the entire distribution of $\mathcal{X}$. Monte Carlo is the obvious, *general* tool. A number of problems (but not all) are well handled by simpler Gaussian approximations, sometimes with a correction for skewness added. Computational methods for solvency are outlined in the next two sections.

The second main topic is the pricing of risk, not a purely actuarial subject. There is above all a market side. A company will gladly charge what people are willing to pay! Strategic considerations could influence pricing too, and there are overhead costs to cover. Yet a core is the pure premium $\pi = E(X)$ or $\Pi = E(\mathcal{X})$; i.e. the expected policy or portfolio payout during a certain period of time. Evaluations of those are important not only as basis for pricing, but also as an aid to decision making. Not all risks are worth taking! Pricing or **rating** methods in actuarial science follow two main lines. The first one draws on claim histories of *individuals*. Those with good records are to be considered lower risks (premium reduced), those with bad ones the opposite (premium raised). The traditional approach is through the the theory of **credibility**, a classic presented in Section 10.5. Alternatively, price differentials could be administered to *groups* rather than individuals. What counts now is experience with the group. It could be defined according to age, to what kind of car you own, where your residence are and so on. The natural method is regression. Solvency and pricing under re-insurance schemes are treated at the end.

Numerical examples are used extensively to give a feel for numbers and for how sensitively evaluations depend on assumptions. The ideas of Chapter 7 looming underneath. Risk over longer time horizons is taken up in the next chapter.

1.2 Portfolio liabilities by simple approximation

Introduction

The distribution of the portfolio loss $\mathcal{X}$ becomes normal when the number of policies $J \to \infty$. This is a consequence of the central limit theorem and leads to a straightforward assessments of the reserve that avoid detailed probabilistic modelling (more on that below). The method is useful due to its simplicity, but the underlying conditions are too restrictive for it to be the only one. Normal approximations underestimate risk for small portfolios and in branches with large claim severities. Some of that is rectified by taking the skewness of $\mathcal{X}$ into account, leading to the so-called **NP**-version. The purpose of this section is to review these simple approximation methods, show how they are put to practical use and indicate their accuracy and range of application.

Normal approximations

Let μ be claim intensity and ξ_z and σ_z mean and standard deviation of the individual losses. If they are the same for all policy holders, mean and standard deviation of $\mathcal{X}$ over a period of length T become

$$E(\mathcal{X}) = a_0 J, \qquad \text{and} \qquad \text{sd}(\mathcal{X}) = a_1\sqrt{J};$$

where

$$a_0 = \mu T \xi_z, \qquad \text{and} \qquad a_1 = (\mu T)^{1/2}(\sigma_z^2 + \xi_z^2)^{1/2}; \tag{1.1}$$

see Section 6.3 and Exercise 6.3.1. This leads to to the true percentile q_ϵ being approximated by

$$q_\epsilon^{\text{No}} = a_0 J + a_1 \phi_\epsilon \sqrt{J} \tag{1.2}$$

where ϕ_ϵ is the (upper) ϵ percentile of the the standard normal distribution. Estimates of μ, ξ_z and σ_z are required for the coefficients a_0 and a_1, but the entire claim size distribution is *not* needed. Detailed modelling can at this point be avoided by using the sample mean and the sample standard deviation as estimates $\hat{\xi}_z$ and $\hat{\sigma}_z$. Of course, there are many other ways.

The approximation (1.2) is also valid when μ, ξ_z and σ_z depend on j. The coefficients a_1 and a_1 are then changed to

$$a_0 = \frac{T}{J}\sum_{j=1}^{J}\mu_j \xi_{zj} \qquad \text{and} \qquad a_1 = \sqrt{\frac{T}{J}\sum_{j=1}^{J}\mu_j(\sigma_{zj}^2 + \xi_{zj}^2)}. \tag{1.3}$$

Check that they reduce to (1.1) when all parameters are equal! With all μ_j, ξ_{zj} and σ_{zj} available on file this method gives (when applicable) a quick appraisal of the reserve. Still another version emerges when we regard policy holders (and their parameters) as an independent sample. The most important special case is when $\mu_1, \ldots, \mu_J$ have common mean and standard deviation ξ_μ and σ_μ with ξ_z and σ_z being fixed. The coefficients now become

$$a_0 = \xi_\mu T \xi_z, \qquad \text{and} \qquad a_1 = T^{1/2}\{\xi_\mu(\sigma_z^2 + \xi_z^2) + \sigma_\mu^2 \xi_z^2\}^{1/2}, \tag{1.4}$$

see (**??**) and (**??**) in Section 6.3.

Example: Motor insurance
The Norwegian autmobile portfolio introduced in Chapter 8 is described by the parameters

$$\hat{\xi}_\mu = 5.6\%, \qquad \hat{\sigma}_\mu = 2.0\% \qquad \text{and} \qquad \hat{\xi}_z = 0.30, \qquad \hat{\sigma}_z = 0.35,$$

annual parameters | *unit: 1000 euro*

where those for claim intensity was determined in Section 8.5. Mean and standard deviation for the loss distribution (which exclude personal injuries) were obtained from almost 7000 incidents; see also Section 10.4. That is enough to evaluate the reserve if the normal approximation is applicable. With $J = 10000$ policies (and $T = 1$) the coefficients a_1 and a_2 are obtained from (1.1) and (1.6). After having looked up the Gaussian percentiles this leads to the following assessments (in 1000 euro):

Equal risk calculation			*Unequal claim frequency*	
1860,	1934	and	1860,	1935.
5% *reserve*	1% *reserve*		5% *reserve*	1% *reserve*

Money unit: Million DKK

	Portfolio size: $J = 1000$		Portfolio: $J = 100000$	
	5% *reserve*	1% *reserve*	5% *reserve*	1% *reserve*
Normal	80	100	3860	4060
Normal power	120	160	3900	4120

Table 10.1 Normal and normal power approximations to the reserve under the Danish fire claims.

Note the minor impact of including risk variation among policy holders, the same message as in Section 6.3. Even a quite substantial hetereogeneity (as in the present example) means little for the reserve.

The normal power approximation

Normal approximations are refined by adjusting for skewness in $\mathcal{X}$. In actuarial science this is called the **normal power** (or **NP**) approximation. In reality the *NP* method is the leading term in a series of corrections to the central limit theorem, in statistics and elsewhere known as the Cornish-Fisher expansion; see Feller (1970) for a probabilistic introduction and Hall (1992) for one in statistics. The underlying theory is beyond the scope of this book, but a brief sketch of the structure is indicated in Section 10.7. Only the pure Poisson model is considered below. The extension to the negative binomial and other models is treated in Daykin, Pentikäinen and Pesonen (1994), but as has been argued earlier, the practical impact is limited.

Let γ_z be the skewness coefficient of the claim size distribution. The refined approximation then reads

$$q_\epsilon^{\mathrm{NP}} = q_\epsilon^{\mathrm{No}} + a_2(\phi_\epsilon^2 - 1)/6 \qquad \text{where} \qquad a_2 = \frac{\gamma_z\sigma_z^3 + 3\xi_z\sigma_z^2 + \xi_z^3}{\sigma_z^2 + \xi_z^2}, \tag{1.5}$$

which is justified in Section 10.7. Inserting (1.1) for q_ϵ^{No} yields

$$q_\epsilon^{\mathrm{NP}} = \underbrace{a_0 J + a_1\phi_\epsilon\sqrt{J}}_{\textit{the normal component}} + \underbrace{a_2(\phi_\epsilon^2 - 1)/6}_{\textit{NP correction}} \tag{1.6}$$

which is a series in falling powers of $\sqrt{J}$. The NP correction term is *independent* of portfolio size.

To use the approximation in practice skewness γ_z must be estimated in addition to ξ_z and σ_z (μ as well). There is no new ideas in this. We may fit a parametric family to the historical data or with plenty of data use the sample skewness coefficient introduced in Section 9.2.

Example: Danish fire claims

Consider a portfolio for which

$$\underbrace{\hat{\mu} = 1\%}_{\textit{annual}} \qquad \text{and} \qquad \underbrace{\hat{\xi}_z = 3.385, \quad \hat{\sigma}_z = 8.507,}_{\textit{Unit: Million DKK}} \qquad \hat{\gamma}_z = 18.74.$$

The parameters for claim size are those found for the Danish fire data in Chapter 9 (one million DKK could be around 125 000 euro). With $J = 1000$ and $J = 100000$ policies the assessments

Office type computer with pentium III processor. *Implementation: Fortran 77*

	Portfolio size: $J = 1000$		Portfolio: $J = 100000$	
	Algorithm 3.1	*Algorithm 3.2*	*Algorithm3.1*	Algorithm 3.2.
Emp. dist.	0.005	0.02	0.05	0.03
Gamma	5	17	37	51

*Table 10.2 CPU time (***seconds***) per* **1000** *simulations of portfolio liabiltities.*

of the reserve becomes those in Table 10.1. The NP correction has considerable impact for the small portfolio on the left, raising the 1% the reserve by as much as 60%. That is due to the losses being strongly skewed towards the right (skewness more than 18). When the number of policies is higher, the relative effect is smaller. With 100000 policies the difference between the two methods is of minor importance and their almost common assessment one to be trusted.

But what about the other case? The huge impact of the NP correction on the left in Table 10.1 is ominous and should make us suspicous. Indeed, the more reliable Monte Carlo assessments in the next section match neither. A simple test when using normal reserving could be to calculate the NP term. If it isn't very important, NP reserving probably works.

1.3 Portfolio liabilities by simulation

Introduction

Monte Carlo has several advantages over the methods of the preceeding section. It is more *general* (no restriction on use), more *versatile* (easier to adapt changing circumstances) and better suited long time horizons (Chapter 11). But the method is slow computationally and doesn't it demand the entire claim size distribution whereas the normal approximation could do with only mean and variance? The last point is deceptive. If the portfolio size is so large that the normal distribution provides a reasonable approximation, the claim size distribution doesn't matter outside mean and variance.

What about computational speed? Two simulation algorithms were presented in Section 3.3. Algorithm 3.2 was the more general (risks could be unequal), but it went through the entire portfolio and could therefore appear slower than Algorithm 3.1. An experiment to measure performance is reported in Table 10.2 using a Fortran77 implementation of Algorithm 2.10 (Poisson), Algorithm 4.1 (the empirical distribution function) and Algorithm 9.1a (Gamma). Detailed conditions were

$\mu T = 5\%$ and emprirical distribution or Gamma(α)
10000 historical claims $\alpha = 2$

The results are above all testemony to how fast the empirical distribution function is sampled, Gamma distributions being three or four times slower. The amount of computational work in generating the cost of claims is the same within both algorithms, and when this component dominates, much of their diffences are wiped out. Among the distributions used in this book the Gamma distribution is the most labourious one to sample.

Simulation algorithms

The computation of portfolio liabiltities is one of the most important issues in general insurance,

Distribution		Distribution	
Empirical distribution	Algorithm 4.1	Weibull	Exercise 2.5.1
Pareto mixing	Algorithm 9.2	Fréchet	Exercise 2.5.2
Gamma	Algorithm 9.1a,b	Logistic	Exercise 2.5.3
Log-normal	Algorithm 2.2	Burr	Exercise 2.5.4
Pareto	Algorithm 2.8		

Table 10.3 List of claim size algorithms

and its seems worthwile to collect algorithms spread on several chapters. Consider a situation where claim intensities $\mu_1, \ldots, \mu_J$ are stored on file along with claim size distributions. If Algorithm 2.10 are used for the Poisson sampling, the programming steps can be organized as follows:

Algorithm 10.1 Portfolio liabilities in the general case

0 Input: Poisson parameters $\lambda_j = \mu_j T$ $(j = 1, \ldots, J)$, claim size models, H(z).
1 $\mathcal{X}^* \leftarrow 0$
2 For $j = 1, \ldots, J$ do
3 Draw $U^* \sim$ uniform and $S^* \leftarrow -\log(U^*)$
4 Repeat while $S^* < \lambda_j$
5 Draw claim size Z^* *%Might depend on j*
6 $\mathcal{X}^* \leftarrow \mathcal{X}^* + H(Z^*)$ *% Add loss,*
7 Draw $U^* \sim$ uniform and $S^* \leftarrow S^* - \log(U^*)$ *%Update for Poisson*
8 Return $\mathcal{X}^*$

Poisson sampling has been integrated into the code. The algorithm goes through the entire portfolio and add costs of settling incidents until the citerion on Line 4 is *not* satisfied. There are many different algorithms for Line 5. Table 10.3 lists examples from this book.

Danish fire data: Sensitivity against claim size model

The Danish fire data was examined in Section 9.6 and a number of models were tried. Some seemed to work well, others did not. Table 10.4 shows how the fit or lack of it is passed on to the reserve. Models considered were the empirical distribution function without or with Pareto mixing for the extremes, Pareto, Gamma and log-normal. All were fitted the historical fire claims as described in Chapter 9. The portfolio size were $J = 1000$ with annual claim rate $\hat{\mu} = 1\%$, producing no more than 10 claims per year on average. Ten million simulations were used, making Monte Carlo uncertainty very small indeed.

The situation is then identical to the one on the left in Table 10.1 and testfies to the difficulty of calculating the reserves for small portfolios when the shape of the claim size distribution matters. On its own the empirical distribution function underestimates risk, but mixed with the Pareto distribution it seems to work well and is not overly dependent on where the threshold b is placed. The Gamma distribution on log-scale fitted well in Section 9.5 and lead to similar results. Others that were grossly in error, also leads to strongly deviating reserves. If you compare with the normal power method in Table 10.1 you will discover that it over-shoots at level 5% and

[a] *EDF: The empirical distribuiton* [b] *Thresholds are* 50%, 25%, 10%, 5% [c] *Log-transformed claims*

	EDF[a]	EDF[a] with Pareto above b[b]				Other claim size models		
Reserve		b=10	b=5.6	b=3.0	b=1.8	Pareto	Gamma[c]	Log-normal
5%	72	100	104	105	100	71	94	49
1%	173	200	217	230	225	137	214	61
0.03%	330	590	870	1400	1750	900	1944	84

Table 10.4 Calculated reserves for the Danish fire data. Money unit: Million DKK (about 8 *DKK for one euro).*

under-shoots at 1%.

Reserves at level 0.03% have been added. Luckily those figures are not in demand! The results are a bewildering mess of unstability. What this shows is the extreme difficulty of producing assessments very far out into the tails of a distribution where they become sensitively dependent on modelling details. Although such tiny percentiles are rarely needed with insurance liabilities, they are used by the bureaus rating financial soundness.

1.4 Differentiated pricing I: Using the observable

Introduction

Very young male drivers or owners of fast cars are groups of clients notoriously more risky than others, and it is hardly unfair to charge them more. These are examples of pricing unequally based on experience. The technological development which enhances our possibilities of collecting and storing information with bearing on risk, can only further this practice. Historical records of insurance incidents and their cost are then tied through statistical techniques to circumstances, conditions and people causing them. Among the methods available log-linear regression is arguably the most important one. The purpose of this section is to indicate how Poisson, Gamma and log-normal regression from earlier chapters are put to use.

Explanatory variables are then identified as observations, registrations or measurements $x_1 \dots, x_v$ and linked to claim intensity μ and mean loss per event ξ through

$$\log(\mu) = b_{\mu 0}x_0 + \dots + b_{\mu v}x_v \qquad \text{and} \qquad \log(\xi) = b_{\xi 0}x_0 + \dots + b_{\xi v}x_v,$$

where $b_{\mu 0}, b_{\mu 1} \dots,$ and $b_{\xi 0}, b_{\xi 1}, \dots$ are coefficients. By default $x_0 = 1$, a convention introduced to make formulae neater. The explanatory variables by no means have to be the same for both μ and ξ, but the mathematics becomes simpler to write down if they are, and we can always 'zero' irrelevant ones away; i.e. take $b_{\xi i} = 0$ if (for example) x_i isn't included in the regression for ξ. In motor insurance (example below) the relationship is typically more important for μ than for ξ. Inserting the regression equations for μ and ξ into the pure premium $\pi = \mu\xi$ yields

$$\pi = \exp(\eta) \qquad \text{where} \qquad \eta = (b_{\mu 0} + b_{\xi 0})x_0 + \dots + (b_{\mu v} + b_{\xi v})x_v,$$

where the time of exposure $T = 1$; for general T see Exercise 10.4.1.

Estimates of the pure premium

The regression coefficients are in practice replaced by *estimated* ones $\hat{b}_{\mu i}$ and $\hat{b}_{\xi i}$. The pure premium of a policy with $x_1, \ldots, x_v$ as explanatory variables is then evaluated as

$$\hat{\pi} = \exp(\hat{\eta}) \qquad \text{where} \qquad \hat{\eta} = (\hat{b}_{\mu 0} + \hat{b}_{\xi 0})x_0 + \ldots + (\hat{b}_{\mu v} + \hat{b}_{\xi v})x_v.$$

Commercial software will typically be used to determine the coefficients from historical data (see Section 8.4 and 9.3). Assessments of their error are also provided, but they must be passed on to the estimate of the pure premium itself, and you should know how that is done. Bootstrapping (Section 7.4) *can* be used (as indeed always), but there is also a simpler technique available if there is enough historical data. The estimated regression coefficients are then approximately normal, $\hat{\eta}$ (their sum) becomes nearly normal too and $\hat{\pi}$ has the statistical properties of a log-normal. Take a robust attitude towards the normality assumption. High accuracy in error assessments isn't that important.

There are two sets of estimated coefficients $(\hat{b}_{\mu 0}, \ldots, \hat{b}_{\mu_v})$ and $(\hat{b}_{\xi 0}, \ldots, \hat{b}_{\xi_v})$ coming from two different regression analyses. It is usually unproblematic to assume independence *between* sets, whereas dependence *within* may be very strong indeed. With $\sigma_{\mu ij} = \text{cov}(\hat{b}_{\mu_i}\hat{b}_{\mu_j})$ and $\sigma_{\xi ij} = \text{cov}(\hat{b}_{\xi_i}\hat{b}_{\xi_j})$ mean and variance of $\hat{\eta}$ are approximately

$$E(\hat{\eta}) \doteq \eta \qquad \text{and} \qquad \tau^2 = \text{var}(\hat{\eta}) \doteq \sum_{i=0}^{v}\sum_{j=0}^{v} x_i x_j (\sigma_{\mu ij} + \sigma_{\xi ij}),$$

where the variance formula (??) in Appendix A has been used. This is passed on to $\hat{\pi}$ by the usual formulae for the log-normal; i.e.

$$E(\hat{\pi}) \doteq \pi \exp(\tau^2/2) \qquad \text{and} \qquad \text{sd}(\hat{\pi}) \doteq E(\hat{\pi})\sqrt{\exp(\tau^2) - 1}.$$

Note that $E(\hat{\pi}) > \pi$, and $\hat{\pi}$ is biased upwards. but usually not by very much (see below). Bias and standard deviation is estimated by

$$\underbrace{\hat{\pi}(e^{\hat{\tau}^2/2} - 1)}_{\textit{bias}}, \qquad \underbrace{\hat{\pi}e^{\hat{\tau}^2/2}\sqrt{e^{\hat{\tau}^2} - 1}}_{\textit{standard deviation}} \qquad \text{where} \qquad \hat{\tau}^2 = \sum_{i=0}^{v}\sum_{j=0}^{v} x_i x_j (\hat{\sigma}_{\mu ij} + \hat{\sigma}_{\xi ij}).$$

and were $\hat{\sigma}_{\mu ij}$ and $\hat{\sigma}_{\xi ij}$ are estimates (standard software makes them available). In the formula for $\hat{\tau}^2$ take $\hat{\sigma}_{\xi ij} = 0$ or $\hat{\sigma}_{\xi ij} = 0$ when variable i or j is absent.

Designing regression models

Log-linear regression is a general tool that offers many possibiltities within a framework that adds contributions on logarithmic scale. On the natural scale such specifications are of the form

$$\mu = \underbrace{\mu_0}_{\textit{baseline}} \cdot \underbrace{e^{(b_{\mu 1}+b_{\xi 1})x_1}}_{\textit{variable 1}} \cdots \underbrace{e^{(b_{\mu v}+b_{\xi v})x_v}}_{\textit{variable v}} \qquad \text{where} \qquad \mu_0 = e^{b_{\mu 0}+b_{\xi 0}}.$$

Here μ_0 is claim intensity when $x_1 = \ldots = x_v = 0$, a useful interpretation. The explanatory variables drive intensities up and down independently of each other compared to this baseline. As an example suppose x_1 is binary with 0 for males and 1 for females. Then

$$\underbrace{\mu_m = \mu_0 e^{(b_{\mu 2}+b_{\xi v})x_2} \cdots e^{(b_{\mu v}+b_{\xi v})x_v}}_{\textit{for males}} \qquad \text{and} \qquad \underbrace{\mu_f = \mu_0 e^{b_{\mu 1}+b_{\xi 1}} e^{(b_{\mu 2}+b_{\xi v})x_2} \cdots e^{(b_{\mu v}+b_{\xi v})x_v}}_{\textit{for females}},$$

[a] *Estimated shape of the Gamma distribution:* $\hat{\alpha} = 1.1$

	Intercept	Age	
		≤ 26	> 26
Freq.	-2.43 (.08)	0 (0)	-0.55 (.07)
Size[a]	8.33 (.07)	0 (0)	-0.36 (.06)

	Distance limit on policy (in 1000 km)					
	8	12	16	20	25-30	No limit
Freq.	0 (0)	.17 (.04)	0.28 (.04)	0.50 (.04)	0.62 (.05)	0.82 (.08)
Size[a]	0 (0)	.02 (.04)	0.03 (.04)	0.09 (.04)	0.11 (.05)	0.14 (.08)

	Geographical regions with traffic density from **high** to **low**					
	Region 1	Region 2	Region 3	Region 4	Region 5	Region 6
Freq.	0 (0)	-0.19 (.0.4)	-0.24 (.06)	-0.29 (.04)	-0.39 (.05)	-0.36 (.04)
Size[a]	0 (0)	-0.10 (.0.4)	-0.03 (.05)	-0.07 (.04)	-0.02 (.05)	0.06 (.04)

Table 10.5 Estimated coefficients of claim intensity and claim size for automobile data (standard deviation in parenthesis). Methods: Poisson and Gamma regression.

and $\mu_f/\mu_m = e^{b_{\mu 1}+b_{\xi 1}}$, is fixed and *independent of all other covariates.* The model permits no **interaction** between explanatory variables. The female drivers in Section 8.3 who were more reliable than men when young and less when old is not captured by this.

Modifications are possible. One way is to design suitable crossed categories, a little like age was divided into groups in Section 8.4, see Exercise 10.4.2. The problem with such procedures is that the number of parameters grows rapidly. We shall below examine an example with three variables consisting of 2, 6 and 6 categories. The total number of combinations is then $2 \times 6 \times 6 = 72$ which may not appear much when the historical material is over 200000 policy years. On average each group would have around 2500 policy years behind it, enough for fairly accurate assessments of claim intensities by the elementary estimate (**??**). The problem is that historical data often are very unequally divided among such groups which leads to much random error in some of the estimates. Simplifications through log-linear regression enables us to dampen random error; see also Exercise 10.4.3.

Case study: The Norwegian automobile portfolio
A useful case for illustration is the Norwegian automobile portfolio of Chapter 8. There are around 100000 policies extending two years back with much customer turnover. Almost 7000 claims were registered as basis for claim size modelling. Explanatory variables used are

- age (2 categories that were ≤ 26 and > 26 years)
- driving limit (6 categories)
- geographical region (6 categories).

Driving limit is a proxy for how much people drive. Age is simplified drastically compared to

Age	Distance limit on policy (in 1000 km)					
	80	120	160	200	250-300	No limit
$\leq$ 26 years	365 (6.3)	442 (6.8)	497 (7.5)	656 (8.3)	750 (9.0)	951 (9.8)
> 26 years	148 (2.9)	179 (3.0)	201 (3.7)	265 (3.7)	303 (4.1)	385 (4.3)

Table 10.6 Estimated pure premium (in euro) for Region 1 of the Scandinavian autmobile portfolio (standard deviation in parenthesis)

what you would use in practice. If interaction is neglected, the regression equation for μ becomes

$$\log(\mu) \;=\; b_{\mu 0} \;+\; \underbrace{b_{\mu 1}x_1}_{age} \;+\; \underbrace{\textstyle\sum_{i=2}^{6} b_{\mu 1}(i)x_2(i)}_{distance\ limit} \;+\; \underbrace{\textstyle\sum_{i=2}^{6} b_{\mu 1}(i)x_3(i)}_{region},$$

with a similar relation for ξ. Coding is the same as in Section 8.4. Note that x_1 is 0 or 1 according to the whether the individual is below or above 26. ts of young people appear to be both more frequent and more severe. Regression methods used were Poisson (claim frequency) and Gamma (claim size).

Estimated parameters are shown in Table 10.5. They vary smoothly with the categories. As expected, the more people drive and the heavier the traffic the larger is the risk. Claim frequency fluctuates stronger than claim size (coefficients larger in absolute value). Accidents of young people appear to be both more frequent and more severe. The results in Table 10.5 yield estimates of the pure premia for the 72 groups along with their standard deviation, as explained above. Those for the region with heaviest traffic (Oslo area) is shown in Table 10.6. Estimates are smooth and might be be used as basis for a pricing policy. The log-normal bias varied from 0.2 to 0.5, much smaller than the standard deviation in parenthesis in Table 10.6.

1.5 Differentiated pricing II: The individual record

Introduction

The preceding section differentiated premium between groups, though in a rather crude way: Within a given group, the same to everyone! Is that fair? Perhaps, and yet personal factors we don't observe make real risk vary. Could it be possible to detect risky customers from their own track record? Individual claim history may be *good* (no claims at all) or *bad* (very costly ones in the past). We might reject those that present heavy claims all the time, but perhaps it would be equally good to raise premium instead. But then by how much? That is the kind of question now addressed.

Formally the problem is as follows. Let $X_1, \ldots, X_K$ be the annual claims from a policy holder after K years in the company. Many of those, perhaps even all, would be zero. How should that record, whether good or bad, influence premium charged? If such a problem could be solved, no client would be unwelcome. The risky ones would be answered by sky high rates! Such a rosy picture is only theory, but the problem has nevertheless an elegant solution, known as the theory of **linear credibility**.

Credibility: The approach

Every policy holder is carrier of a list of attributes with impact on risk. Variables like age, sex, or geographical location are relevant, but those are easily observable and therefore better handled through regression methodology. What we have in mind are factors like general ability, power of concentration, recklesness or practical experience. All these things influence the performance of drivers of automobiles and cause risk to vary over the population. Such factors can't be measured or quantified. Their inventory is endless, but luckily details are not needed. It is sufficient to postulate their *existence* and define the distribution of the risk X *conditionally* given a certain underlying ω. Expectation and standard deviation then become

$$\underbrace{\pi(\omega) = E(X|\omega)}_{\textit{conditional pure premium}} \qquad \text{and} \qquad \sigma(\omega) = \text{sd}(X|\omega). \tag{1.7}$$

If $\widehat{\pi(\omega)}$ is an estimate of $\pi(\omega)$, the premium charged might be $(1+\gamma)\widehat{\pi(\omega)}$, where γ is a suitable loading.

To derive the estimate it may appear natural to factorize the individual pure premium in the usual manner as

$$\pi(\omega) = \mu(\omega)E(Z|\omega), \tag{1.8}$$

where both claim frequency $\mu(\omega)$ and expected claim size $E(Z|\omega)$ depend on ω. This approach is not the common one. The traditional theory of credibility starts from the aggregated (typically annual) claim records $X_1, \dots, X_K$ and attacks $\pi(\omega)$ directly without the factorization (1.8), but see Exercise ?.

Credibility: Assumptions and modelling
A long string of conditions is *not* required. We shall treat ω as a *random* variable (or vector). This is plausible and leads to the common factor model of Section 6.2; i.e.

- $X_1 \dots, X_K$ are identically and independently distributed given ω.

Independence given ω follows a main line in property insurance modelling. Note that $X_1, \dots, X_K$ are *dependent* as we observe them, since they are tied to the *same* ω. Indeed, that is the reason past claims $X_1, \dots, X_K$ say anything about a future X at all. Their distribution would not be the same if *learning* is involved (young drivers gaining in experience, for example); see Sundt (1991). Such extensions will not be covered.

We need the so-called **structural parameters** to work from; i.e.

$$\zeta = E\{\pi(\omega)\}, \qquad \upsilon^2 = \text{var}\{\pi(\omega)\}, \qquad \tau^2 = E\{\sigma^2(\omega)\}. \tag{1.9}$$

Here ζ is the *average* pure premium over the entire population policy holders. Both υ and τ represent variation. The former is caused by diversity between individuals; the latter by the actual physical processes underlying the incidents. How the structural parameters are estimated from historical data is shown below.

The two important relationships

$$E(X) = \zeta \qquad \text{and} \qquad \text{var}(X) = \tau^2 + \upsilon^2 \tag{1.10}$$

will be needed later. They are easily derived from the rules of double expectation and double variance in Section 6.3. First note that

$$E(X) = E\{E(X|\omega)\} = E\{\pi(\omega)\} = \zeta,$$

and then

$$\mathrm{var}(X) = E\{\mathrm{var}(X|\omega)\} + \mathrm{var}\{E(X|\omega)\} = E\{\sigma^2(\omega)\} + \mathrm{var}\{\pi(\omega)\} = \tau^2 + \upsilon^2.$$

Estimation by linear credibility
An estimate $\widehat{\pi(\omega)}$ of $\pi(\omega)$ can be derived from the data record $X_1, \ldots, X_K$ and the structural parameters ξ, υ and τ. The standard construction is to minimize mean squared error

$$Q = E\{\widehat{\pi(\omega)} - \pi(\omega)\}^2.$$

This yields (see Section 6.4) the solution

$$\widehat{\pi(\omega)} = E\{\pi(\omega)|X_1, \ldots, X_K\},$$

known as the *general credibility estimate.* Although the most accurate estimate possible, it also requires much more detailed modelling. However, there is a *linear* alternative. Now the estimate is of the form

$$\widehat{\pi_K(\omega)} = b_0 + b_1 X_1 + \ldots + b_K X_K,$$

where $b_0, b_1, \ldots, b_K$ are coefficients determined so that Q is minimized. The advantage is that no more than the three structural parameters are needed.

The symmetry of the problem forces all of $b_1, \ldots, b_K$ to be equal. This must be so since all of $X_1, \ldots, X_K$ carry exactly the same amount of information. But then the best estimates must be of the form

$$\widehat{\pi_K(\omega)} = b_0 + b\bar{X} \qquad \text{where} \qquad \bar{X} = (X_1 + \ldots + X_K)/K. \tag{1.11}$$

The coefficients b_0 and b minimizing Q is derived in Section 10.7, leading to the *linear credibility* estimate

$$\widehat{\pi_K(\omega)} = (1 - w)\zeta + w\bar{X}, \qquad \text{where} \qquad w = \frac{\upsilon^2}{\upsilon^2 + \tau^2/K}. \tag{1.12}$$

Here w is a weight between zero and one, defining a compromise between the *average* pure premium $\zeta = E\{\pi(\omega)\}$ and the data record of the policy holder. There is a close resemblance between this solution and the Bayes estimate of the normal mean in Section 7.6.

The estimate can be understood from many angles. A first observation is that w *increases* with the track record K; i.e. experience counts for more if the customer have been with us for a long time. If the client is new ($K = 0$), then $w = 0$ and $\widehat{\pi_0(\omega)} = \zeta$. With no information available the best that can be done is to use the population average. For other interpretations, see Exercises ? and ?.

Statistical properties
Estimation error in the linear credibility estimate is summarized by

$$E\{\widehat{\pi_K(\omega)} - \pi(\omega)\} = 0 \tag{1.13}$$

and

$$\mathrm{sd}\{\widehat{\pi_K(\omega)} - \pi(\omega)\} = \gamma_K \upsilon, \qquad \text{where} \qquad \gamma_K = (1 + K\upsilon^2/\tau^2)^{-1/2}; \tag{1.14}$$

see Section 10.7 for the proof. The linear credibility estimate is *unbiased*, but as least as important is random error. From (1.14) note that

$$\frac{\mathrm{sd}\{\widehat{\pi_K(\omega)} - \pi(\omega)\}}{\mathrm{sd}\{\widehat{\pi_0(\omega)} - \pi(\omega)\}} = \frac{\gamma_K \upsilon}{\gamma_0 \upsilon} = \gamma_K,$$

and γ_K conveys the benefit of the historical record $X_1, \dots X_K$. What its size might be is indicated next.

Example: When claim intensity varies
Consider a portfolio for which the the claim size distribution is the same for everybody, but where the claim size *intensity* varies randomly between individuals. In motor insurance this captures much of what goes on. Identify μ with ω and introduce

$$E(\mu) = \xi_\mu, \quad \mathrm{sd}(\mu) = \sigma_\mu \qquad \text{and} \qquad E(Z) = \xi_z, \quad \mathrm{sd}(Z) = \sigma_z.$$

For an individual with intensity μ observed over T

$$\pi(\mu) = E(X|\mu) = \mu T \xi_z \qquad \text{and} \qquad \sigma^2(\mu) = \mathrm{var}(X|\mu) = \mu T(\xi_z^2 + \sigma_z^2);$$

see Exercise 6.3.1. The structural parameters (1.9) then become

$$\zeta = \xi_\mu T \xi_z \qquad \upsilon^2 = \sigma_\mu^2 T^2 \xi_z^2 \qquad \tau^2 = \xi_\mu T(\xi_z^2 + \sigma_z^2),$$

and when these expressions are inserted into the coefficient γ_K in (1.14), we obtain

$$\gamma_K = (1 + K\theta_z T \sigma_\mu^2/\xi_\mu)^{-1/2} \qquad \text{where} \qquad \theta_z = \xi_z^2/(\xi_z^2 + \sigma_z^2).$$

Accurate estimation of $\pi(\mu)$ requires γ_K to decline fast as K is raised, and since $\theta_z \leq 1$, much hinges on the ratio σ_μ^2/ξ_μ. Variability in μ has to be quite large for this quantity to be anything but quite small. For the automobile portfolio used earlier $\xi_\mu = 5.6\%$ and $\sigma_\mu = 2\%$ annually (thus $T = 1$), and if $\xi_z = 10000$ and $\sigma_z = 1000$, the standard deviation of the credibility estimate becomes

$$\mathrm{sd}\{\widehat{\pi_K(\omega)} - \pi(\omega)\} = \underset{K=0}{200} \quad \underset{K=10}{193} \quad \underset{K=20}{187}.$$

These are huge errors when

$$\zeta = E\{\pi(\mu)\} = 10000 \times 0.056 = 560.$$

Even 20 years of experience with the same client hasn't reduced uncertainty much. Credibility estimation is an ambitious project of charging premium fair, but it clearly can't be used indiscriminantly.

Finding the structural parameters
Historical data might be of the following form:

1	x_{11}	$\ldots$	x_{1K_1}	$\bar{x_1}$	s_1
.	.	$\ldots$	.	.	.
.	.	$\ldots$	.	.	.
J	x_{J1}	$\ldots$	x_{JK_J}	$\bar{x_J}$	s_J
Policies	*Annual claims*			*mean*	*sd*

There are J policies that have been in the company $K_1, \ldots, K_J$ years. Annual claims from client j are $x_{j1}, \ldots, x_{jK_j}$, i.e. the j'th row of the table, from which mean $\bar{x}_j$ and standard deviation s_j can be calculated. Let $\mathcal{K} = K_1 + \ldots + K_J$. Unbiased, moment estimates of the structural parameters are then

$$\hat{\zeta} = \frac{1}{\mathcal{K}}\sum_{j=1}^{J} K_j \bar{x}_j, \qquad \hat{\tau}^2 = \frac{1}{\mathcal{K}-J}\sum_{j=1}^{J}(K_j-1)s_j^2 \tag{1.15}$$

and

$$\hat{\upsilon}^2 = \frac{\sum_{j=1}^{J}(K_j/\mathcal{K})(\bar{x}_j - \hat{\xi})^2 - \hat{\tau}^2(J-1)/\mathcal{K}}{1 - \sum_{j=1}^{J}(K_j/\mathcal{K})^2}. \tag{1.16}$$

verfication is given in Section 10.7. The expression for $\hat{\upsilon}^2$ does not have to is positive. If it isn't, the pragmatic (and sensible) position is to assume $\upsilon = 0$. Variation in the individual pure premium over the portfolio is then too small to be detected.

1.6 Re-insurance

Introduction
Re-insurance (introduced in Section 3.2) deals with **primary** risks placed with a **cedent** who passes some of it on to **re-insurers** who may in turn go to other re-insurers. At the end there is global network of players dividing risk between them. Re-insurers provide cover to incidents far away both geographically and in terms of intermediaries, but for the original clients at the bottom of the chain all of this is irrelevant. For them re-insurance instruments used higher up are without importance as long as the companies involved are solvent. For cedents these arrangements are ways to spread risk and may also enable small or medium-sized companies to take on heavier responsibilites than its capital base in itself would allow.

Methods change little compared to ordinary insurance. Primary risk rests with the cedents, and their modelling is (of course) the same thing. Cash flows differ, but those are merely modifications through fixed functions $H(z)$ containing the payment clauses and are easily handled by Monte Carlo (Section 3.3). Economic impact may be huge, the methodological not. The purpose of this section is to outline some of the most common contracts and indicate consequences for pricing and solvency.

Types of contracts
Re-insurance can be seen as expenses shared between two or more parties. Contracts may apply to costs of settling single events or to sums of claims affecting the entire portfolio. These losses (denoted Z and $\mathcal{X}$) are then (in obvious mathematical notation) divided between re-insurer and cedent according to

$$\underbrace{Z^{\text{re}} = H(Z), \quad Z^{\text{ce}} = Z - H(Z)}_{\textit{single events}} \qquad \text{and} \qquad \underbrace{\mathcal{X}^{\text{re}} = H(\mathcal{X}), \quad \mathcal{X}^{\text{ce}} = \mathcal{X} - H(\mathcal{X}),}_{\textit{on portfolio level}} \tag{1.17}$$

where $0 \leq H(z) \leq 1$. Here Z^{ce} and $\mathcal{X}^{\text{ce}}$ are the *net* responsibility of the cedent after the re-insurer reimbursements have been subtracted.

One of the most common contracts is the $a \times b$ type considered in Chapter 3. When drawn up in terms of single events re-insurer and cedent responsibilities are

$$Z^{\text{re}} = \begin{matrix} 0, & \text{if } Z < a \\ Z - a, & \text{if } a \leq Z < a + b \\ b - a, & \text{if } Z \geq a + b. \end{matrix} \qquad \text{and} \qquad Z^{\text{ce}} = \begin{matrix} Z, & \text{if } Z < a \\ a, & \text{if } a \leq Z < a + b \\ Z - b, & \text{if } Z \geq a + b, \end{matrix}$$

Note that $Z^{\text{re}} + Z^{\text{ce}}$=Z. The lower bound a is the **retention** limit of the cedent who must cover all claims below that threshold. Responsibility (i.e. Z^{ce}) appears unlimited, but in practice there is usually a maximum insured sum S that makes $Z \leq S$. If a and b are taylored to S, the scheme gives good cedent protection. If the upper bound b (the retention limit of the *re-insurer*) is infinite (rare in practice), the contract is known as **excess of loss**. The $a \times b$ re-insurance is also used with the aggregated claim $\mathcal{X}$ against the portfolio. The expressions for $\mathcal{X}^{\text{re}}$ and $\mathcal{X}^{\text{ce}}$ are similar to those above. If b is infinite, the treaty is now known as **stopp loss**.

Another type of contract is the **proportional** one for which

$$\underbrace{Z^{\text{re}} = cZ, \quad Z^{\text{ce}} = (1 - Z)}_{\textit{single events}} \qquad \text{and} \qquad \underbrace{\mathcal{X}^{\text{re}} = c\mathcal{X}, \quad \mathcal{X}^{\text{ce}} = (1 - c)\mathcal{X}}_{\textit{on portfolio level}} \tag{1.18}$$

Risk is now shared by cedent and re-insurer in a fixed proportion. Suppose there are J separate re-insurance treaties, one for each of J contracts placed with the cedent. Such an arrangement is known as **quota share** if the constant of proportionality c is the same for all policies. In the opposite case we are dealing with **surplus** re-insurance if $c = c_j$ is of the form

$$c_j = \max(0, 1 - \frac{a}{S_j}) \qquad \text{so that} \qquad Z_j^{\text{re}} = \begin{matrix} 0 & \text{if } a \geq S_j \\ (1 - a/S_j)Z_j & \text{if } a < S_j. \end{matrix} \tag{1.19}$$

Here S_j is the maximum insured sum of the j'th primary risk. Note that a (the cedent retention limit) does not depend on j. As S_j increases from a, the re-insurer part grows.

Pricing re-insurance
Examples of pure re-insurance premia are

$$\underbrace{\pi^{\text{re}} = \mu T \xi^{\text{re}} \quad \text{for} \quad \xi^{\text{re}} = E\{H(Z)\}}_{\textit{single event contracts}} \qquad \text{and} \qquad \underbrace{\Pi^{\text{re}} = E\{H(\mathcal{X})\}.}_{\textit{contracts on portfolio level}}$$

Their approximations through Monte Carlo (m runs) are

$$\underset{\textit{single event contracts}}{\pi^{\rm re*} = \frac{\mu T}{m}\sum_{i=1}^m H(Z_i^*)} \qquad \text{and} \qquad \underset{\textit{contracts on portfolio level}}{\Pi^{\rm re*} = \frac{1}{m}\sum_{i=1}^m H(\mathcal{X}_i^*),}$$

and usually this is the simplest way to do it. If you know the ropes, it often takes *less* time to implement Monte Carlo than to work out exact formulas, and of course, the latter may be impossible. On portfolio level simulations $\mathcal{X}^*$ of the total portfolio loss (obtained from Algorithm 3.1 and 3.2) are inserted into the re-insurance contract $H(x)$

For $a \times b$ contracts in terms of *single* events there is a useful formula. If $f(z)$ and $F(z)$ are density and distribution function of Z, then

$$\begin{aligned}\xi^{\rm re} &= \int_a^{a+b}(z-a)f(z)\,dz + \int_{a+b}^{\infty} bf(z)\,dz \\ &= -(z-a)\{1-F(z)\}\Big|_a^{a+b} + \int_a^{a+b}\{1-F(z)\}\,dz + b\{1-F(a+b)\} = \int_a^{a+b}\{1-F(z)\}\,dz\end{aligned}$$

after integration by parts. Writing $F(z) = F_0(z/\beta)$ as in Section 9.2 yields

$$\pi^{\rm re} = \mu T\int_a^{a+b}\{1-F_0(z/\beta)\}dz, \tag{1.20}$$

which is is possible to evaluate under the Pareto distribution; i.e. when $1 - F_0(z) = (1+z)^{-\alpha}$. Then

$$\pi^{\rm re} = \mu T\frac{\beta}{\alpha-1}\left(\frac{1}{(1+a/\beta)^{\alpha-1}} - \frac{1}{(1+(a+b)/\beta)^{\alpha-1}}\right) \qquad \text{for} \qquad \alpha > 0. \tag{1.21}$$

Special treatment is needed for $\alpha = 1$. Numerical example:

$$\mu T = 1\%,\quad a = 50,\quad b = 500\quad \alpha = 2,\quad \beta = 100 \qquad \text{gives} \qquad \pi^{\rm re} = 0.50,$$

which was reproduced to two decimal places with $m = 100000$ simulations. Monte Carlo standard deviation was 0.003 and three decimals would need around $m = 10$ million.

Effect of inflation

Inflation drives claims upwards into the regions where re-insurance treaties apply, and contracts will be mis-priced if the re-insurance premium is not adjusted. The mathematical formulation rests on the rate of inflation I changing the parameter of scale from $\beta = \beta_0$ to $\beta_I = (1+I)\beta_0$, see Section 9.2. As the rest of the model is unchanged, it follows from (1.20) that the corresponding pure premia $\pi_0^{\rm re}$ and $\pi_I^{\rm re}$ are related through

$$\frac{\pi_I^{\rm re}}{\pi_0^{\rm re}} = \frac{\int_a^b\{1-F_0(z/\beta_I)\}dz}{\int_a^b\{1-F_0(z/\beta_0)\}dz},$$

which applies to $a \times b$ contracts in terms of single events. How other types react to inflation is studied among the exercises.

Number of simulations: One million

	Annual claim frequency: 1.05				Annual claim frequency: 5.25			
Upper limit (b)	0	2200	4200	10200	0	2200	4200	10200
Pure premium	0	82	92	100	0	410	460	500
Cedent reserve	2170	590	510	480	6300	3800	1800	1200

Table 10.7 Re-insurance premium and net cedent reserve (1%) under the conditions in the text. Money unit: Million NOK (8 NOK for 1 euro).

For Pareto models with *infinte* b it follows from (1.21) that

$$\frac{\pi_I^{\rm re}}{\pi_0^{\rm re}} = (1+I)\left(\frac{1+a\beta_0^{-1}}{1+a\beta_0^{-1}/(1+I)}\right)^{\alpha-1}.$$

This is not negligable for values of α of some size; try some suitable values if $I = 5\%$, for example. It is also an increasing function of α which means that the *lighter* the tail of the Pareto distribution, the *higher* the impact of inflation. That appears to be a general phenomenon. Another example is

$$\underset{\textit{orginal model}}{Z_0 \sim \text{Gamma}(\alpha)} \qquad \text{and} \qquad \underset{\textit{inflated model}}{Z_I = (1+I)Z_0,}$$

and the pure premia $\pi_0^{\rm re}$ and $\pi_I^{\rm re}$ can be computed by Monte Carlo. When the upper limit b is infinite and $I = 5\%$, the relative change $(\pi_I^{\rm re} - \pi_0^{\rm re})/\pi_0^{\rm re}$ was found to be

9%	23%	76%		17%	46%	169%
$\alpha = 1$	$\alpha = 10$	$\alpha = 100$	and	$\alpha = 1$	$\alpha = 10$	$\alpha = 100$
a median of Z_0				a upper 10% percentile of Z_0		

Note the huge increase in the effect of inflation as α moves from the heavy-tailed $\alpha = 1$ to the light-tailed, almost normal $\alpha = 100$.

Effect on the reserve

Re-insurance may lead to substantials reductions in capital requirements. Money is lost on average, but since the cedent company can get around on less own capital, its value per share could be higher. There is here a decision to be made which must balance extra cost against capital saved. An illustration is given in Table 10.7. Losses were those of the Norwegain pool of natural disasters in Chapter 7 for which a possible distribution was

$$Z \sim \text{Pareto}(\alpha, \beta) \qquad \text{with} \qquad \alpha = 1.71 \text{ and } \beta = 140.$$

Re-insurance was a $a \times b$ arrangement with $a = 200$ and b varied. Maximum cedent responsibility per event is $S = 10200$. Monte Carlo was used for computation.

Table 10.7 shows cedent net reserve against the pure re-insurance premium. With claim frequency 1.05 annually (the actual case) the 1% reserve is down from 2170 to about one fourth in exchange for the premium paid. When claim freqeuency is five-doubled, savings is smaller in percent, but larger in value. How much does the cedent lose by taking out re-insurance? It depends on the deals available in the market. If the premium paid is $(1+\gamma)\pi$ where π is pure premium and γ the loading, the average loss due to re-insurance is

$$\underbrace{(1+\gamma)\pi}_{\textit{premium paid}} \quad - \quad \underbrace{\pi}_{\textit{claims saved}} \quad = \quad \underbrace{\gamma\pi.}_{\textit{net loss}}$$

In practice γ varies enormously. During the decade around the turn of the century the loading went from barely more than zero to 100% and even 200%!

1.7 Mathematical arguments

Section 10.2

The normal power approximation: The NP approximation of Section 10.2 is a special case of the Cornish-Fisher expansion (Hall 1992) which sets up a series of approximations to the percentile q_ϵ of a random sum $\mathcal{X}$. The first two are

$$q_\epsilon \doteq \underbrace{E(\mathcal{X}) + \mathrm{sd}(\mathcal{X})\phi_\epsilon}_{\textit{normal approximation}} + \underbrace{\mathrm{sd}(\mathcal{X})\frac{1}{6}(\phi_\epsilon^2 - 1)\mathrm{skew}(\mathcal{X}).}_{\textit{skewness correction}} \tag{1.22}$$

A fourth term on the right would involve the kurtosis, but that one isn't much in use in property insurance. The approximation become exact as the portfolio size $J \to \infty$. *Relative* error is proportional to $J^{-1/2}$ (skewness omitted) and to J^{-1} (skewness included), which means that skewness adjustments typically enhance accuracy considerably.

Suppose $\mathcal{X}$ is the total portfolio liability based on identical Poisson risks with intensity μ and with ξ_z, σ_z and γ_z as mean, standard deviation and skewness of the claim size distribution. Mean, variance and third order moment of $\mathcal{X}$ are then

$$E(\mathcal{X}) = J\mu T\xi_z, \qquad \mathrm{var}(\mathcal{X}) = J\mu T(\sigma_z^2 + \xi_z^2), \qquad \mu_3(\mathcal{X}) = J\mu T(\gamma_z\sigma_z^3 + 3\sigma_z^2\xi_z + \xi_z^3),$$

where the third order moment is verified below (the other two were derived in Chapter 6, see Exercise 6.3.1). Skewness is $\mu_3(\mathcal{X})/\mathrm{var}(\mathcal{X})^{3/2}$, and some straightforward manipulations yield

$$\mathrm{skew}(\mathcal{X}) = \frac{1}{(J\mu T)^{1/2}}\,\frac{\gamma_z\sigma_z^3 + 3\sigma_z^2\xi_z + \xi_z^3}{(\sigma_z^2 + \xi_z^2)^{3/2}}.$$

The NP approximation (1.6) follows when the formulae for $\mathrm{sd}(\mathcal{X})$ and $\mathrm{skew}(\mathcal{X})$ are inserted into (1.22).

The third order moment of $\mathcal{X}$ Let $\lambda = J\mu T$ be the Poisson parameter for the total number of claims $\mathcal{N}$. The third order moment of $\mu_3(\mathcal{X})$ is then the expectation of

$$\{\mathcal{X} - \lambda\xi_z\}^3 = \{(\mathcal{X} - \mathcal{N}\xi_z) + (\mathcal{N} - \lambda)\xi_z\}^3 = B_1 + 3B_2 + 3B_3 + B_4$$

where

$$B_1 = (\mathcal{X} - \mathcal{N}\xi_z)^3, \qquad B_2 = (\mathcal{X} - \mathcal{N}\xi_z)^2(\mathcal{N} - \lambda)\xi_z,$$
$$B_3 = (\mathcal{X} - \mathcal{N}\xi_z)(\mathcal{N} - \lambda)^2\xi_z^2, \qquad B_4 = (\mathcal{N} - \lambda)\xi_z^3.$$

Expectations of all these terms follow by computing the *conditional* expectation given $\mathcal{N}$ and applying the rule of double expectation. This is simple since $\mathcal{X}$ is a sum of identically and independently distributed random variables. Start with B_1. It follows from a result in Appendix A that the *conditional* third order moment of $\mathcal{X}$ is $\mathcal{N}$ times the third order moment of Z. Hence

$$E(B_1|\mathcal{N}) = \mathcal{N}(EZ_1 - \xi_z)^3 = \mathcal{N}\gamma_z\sigma_z^3 \qquad \text{which yields} \qquad E(B_1) = \lambda\gamma_z\sigma_z^3.$$

Similarly, from the sum of variance formula

$$E(B_2|\mathcal{N}) = \mathcal{N}\sigma_z^2(N - \lambda)\xi_z \qquad \text{and} \qquad E(B_2) = E\{\mathcal{N}(N - \lambda)\}\sigma_z^2\xi_z = \lambda\sigma_z^2\xi_z.$$

It has here been utilized that $E\{\mathcal{N}(N - \lambda)\} = \text{var}(\mathcal{N}) = \lambda$. For the two remaining terms

$$E(B_3|\mathcal{N}) = 0 \qquad \text{so that} \qquad E(B_3) = 0$$

and

$$E(B_4) = E(\mathcal{N} - \lambda)^3\xi_z^3 = \mu_3(\mathcal{N})\xi_z^3 = \lambda\xi_z^3$$

since $\mu_3(\mathcal{N}) = \lambda$; see Section 8.3. Tieing all these expectations together yields

$$E(\mathcal{X} - \lambda\xi_z)^3 = E(B_1) + 3E(B_2) + 3E(B_3) + E(B_4) = \lambda(\gamma_z\sigma_z^3 + 3\sigma_z^2\xi_z + \xi_z^3)$$

which is $\mu_3(\mathcal{X})$.

Section 10.5
The objective is to derive the linear credibility estimate as an optimum one, verify its statistical properties and justify the estimates of the structural parameters.

Statistical properties of $\bar{X}$. Three auxiliary results on the distribution of the average claim are needed:

$$E(\bar{X}) = \zeta, \qquad \text{var}(\bar{X}) = \upsilon^2 + \tau^2/K \qquad \text{cov}\{\bar{X}, \pi(\omega)\} = \upsilon^2 \tag{1.23}$$

The expectation follows from $E(\bar{X}) = E(X_1) = \zeta$. To derive the variance note that

$$E(\bar{X}|\omega) = E(X_1|\omega) = \pi(\omega) \qquad \text{and} \qquad \text{var}(\bar{X}|\omega) = \text{var}(X_1|\omega)/K = \sigma^2(\omega)/K,$$

and the rule of double variance yields

$$\text{var}(\bar{X}) = \text{var}\{E(\bar{X}|\omega)\} + E\{\text{var}(\bar{X}|\omega)\} = \text{var}\{\pi(\omega)\} + E\{\sigma^2(\omega)/K\} = \upsilon^2 + \frac{\tau^2}{K},$$

as asserted. Finally for the covariance

$$E\{(\bar{X} - \eta)(\pi(\omega) - \eta)|\omega\} = E\{\bar{X} - \eta\}\{\pi(\omega) - \eta\} = \{\pi(\omega) - \eta\}^2,$$

and by the rule of double expectation

$$E\{(\bar{X} - \eta)(\pi(\omega) - \eta)\} = E\{\pi(\omega) - \eta\}^2 = \upsilon^2,$$

and the term on the left is $\text{cov}\{\bar{X}, \pi(\omega)\}$.

The credibility formula Let $\widehat{\pi(\omega)}$ be the estimate in (1.11). Then

$$\widehat{\pi(\omega)} - \pi(\omega) = b_0 + b\bar{X} - \pi(\omega) = b_0 - (1 - b\zeta) + b(\bar{X} - \zeta) - (\pi(\omega) - \zeta)$$

after a little reorganization. Hence

$$\{\widehat{\pi(\omega)} - \pi(\omega)\}^2 = \{b_0 - (1 - b\zeta)\}^2 + b^2(\bar{X} - \zeta)^2 + (\pi(\omega) - \zeta)^2$$
$$+2\{b_0 - (1 - b\zeta)\}(\bar{X} - \zeta) + 2\{b_0 - (1 - b\zeta)\}(\pi(\omega) - \zeta) - 2b(\bar{X} - \zeta)(\pi(\omega) - \zeta),$$

and $Q = E\{\widehat{\pi(\omega)} - \pi(\omega)\}^2$ is calculated by taking expectation on both sides. Since $E(\bar{X}) = \zeta$ and $E\pi(\omega) = \zeta$, this yields

$$Q = (b_0 - (1 - b\zeta)^2 + b^2\text{var}(\bar{X}) + \text{var}\{\pi(\omega)\} + 0 + 0 - 2b\text{cov}\{\bar{X}, \pi(\omega)\}$$

and after inserting (1.23) (middle and right) and $\upsilon^2 = \text{var}\{\pi(\omega)\}$ we obtain

$$Q = (b_0 - (1 - b)\zeta)^2 + b^2(\upsilon + \tau^2/K) + \upsilon^2 - 2b\upsilon^2$$

This is minimized by taking

$$b_0 = 1 - b\zeta \qquad \text{and} \qquad b = w = \frac{\upsilon^2}{\upsilon + \tau^2/K},$$

the solution of b_0 being obvious and that for b being found by differentiation afterwards. This yields the credibility estimate $\widehat{\pi_K(\omega)}$ defined in (1.12).

The statistical properties Unbiasedness is a consequence of

$$E\{\widehat{\pi_K(\omega)}\} = E\{(1 - w)\zeta + w\bar{X}\} = (1 - w)\zeta + wE(\bar{X}) = (1 - w)\zeta + w\zeta = \zeta$$

which equals $E\{\pi(\omega)\}$. The variance of the error is calculated by inserting $b_0 = 1 - w\xi$ and $b = w$ in the expression for Q. This yields

$$Q = \left(\frac{\upsilon^2}{\upsilon^2 + \tau^2/K}\right)^2 (\upsilon^2 + \tau^2/K) + \upsilon^2 - 2\frac{\upsilon^2}{\upsilon^2 + \tau^2/K}\upsilon^2 = \frac{\upsilon^2\tau^2/K}{\upsilon^2 + \tau^2/K},$$

so that

$$E\{\widehat{\pi_K(\omega)} - \pi(\omega)\}^2 = Q = \frac{\upsilon^2}{1 + K\upsilon^2/\tau^2}$$

as asserted in (1.14).

The estimates of ξ, τ and υ.
We shall examine the estimates (1.15) and (1.16). The principal part of the argument is to verify unbiasedness. For $\hat{\zeta}$, defined by (1.15)

$$E(\hat{\zeta}) = \sum_{j=1}^{J} \frac{K_j}{\mathcal{K}} E(\bar{x}_j) = \sum_{j=1}^{J} \frac{K_j}{\mathcal{K}} \zeta = \zeta,$$

since $E(\bar{x}_j) = \zeta$ and $K_1 + \ldots + K_J = \mathcal{K}$. For τ we must utilize that s_j^2 is the ordinary empirical variance so that

$$E(s_j^2|\omega) = \sigma^2(\omega),$$

By the rule of double expectation

$$E(s_j^2) = E\{E(s_j^2|\omega)\} = E\{\sigma^2(\omega)\} = \tau^2,$$

and by (1.15) right

$$E(\hat{\tau}^2) = \sum_{j=1}^{J} \frac{K_j - 1}{\mathcal{K} - J} E(s_j^2) = \sum_{j=1}^{J} \frac{K_j - 1}{\mathcal{K} - J} \tau^2 = \tau^2.$$

Finally, to justify the estimate for $\hat{\upsilon}^2$ in (1.16) consider

$$Q_\upsilon = \sum_{j=1}^{J} \frac{K_j}{\mathcal{K}} (\bar{x}_j - \hat{\zeta})^2.$$

Since

$$\hat{\zeta} - \zeta = \sum_{j=1}^{J} \frac{K_j}{\mathcal{K}} (\bar{x}_j - \zeta),$$

it follows that

$$Q_\upsilon = \sum_{j=1}^{J} \frac{K_j}{\mathcal{K}} (\bar{x}_j - \zeta)^2 - (\hat{\zeta} - \zeta)^2.$$

Moreover, we have from (1.22) middle that

$$E(\bar{x}_j - \zeta)^2 = \upsilon^2 + \frac{\tau^2}{K_j}$$

and that

$$E(\hat{\zeta} - \zeta)^2 = \mathrm{var}(\hat{\zeta}) = \sum_{j=1}^{J} \left(\frac{K_j}{\mathcal{K}}\right)^2 \mathrm{var}(x_j) = \sum_{j=1}^{J} \left(\frac{K_j}{\mathcal{K}}\right)^2 (\upsilon^2 + \frac{\tau^2}{K_j}).$$

It follows that

$$E(Q_\upsilon) = \sum_{j=1}^{J} \frac{K_j}{\mathcal{K}} (\upsilon^2 + \frac{\tau^2}{K_j}) - \sum_{j=1}^{J} \left(\frac{K_j}{\mathcal{K}}\right)^2 (\upsilon^2 + \frac{\tau^2}{K_j})$$

or (since $K_1 \ldots + K_J = \mathcal{K}$)

$$E(Q_\upsilon) = \upsilon^2 + \frac{J}{\mathcal{K}} \tau^2 - \upsilon^2 \sum_{j=1}^{J} \left(\frac{K_j}{\mathcal{K}}\right)^2 - \frac{\tau^2}{\mathcal{K}}.$$

Thus

$$E(Q_\upsilon) = 1 - \sum_{j=1}^{J} \left(\frac{K_j}{\mathcal{K}}\right)^2 \upsilon^2 + \frac{J - 1}{M} \tau^2,$$

and the the estimate $\hat{\upsilon}^2$ is determined by solving the equation

$$Q_\upsilon = \sum_{j=1}^{J} \frac{K_j}{\mathcal{K}}(\bar{x}_j - \hat{\eta})^2 = 1 - \sum_{j=1}^{J}\left(\frac{K_j}{\mathcal{K}}\right)^2 \hat{\upsilon}^2 + \frac{J-1}{\mathcal{K}}\hat{\tau}^2.$$

This yields

$$\hat{\upsilon}^2 = \frac{\sum_{j=1}^{J} K_j/\mathcal{K} - \hat{\tau}^2(J-1)/K}{1 - \sum_{j=1}^{J}(K_j/\mathcal{K})^2}.$$

which is (1.16) for $\hat{\upsilon}$. The argument has also shown that $\hat{\upsilon}$ is unbiased.

1.8 Furter reading

1.9 Exertcises

Exercise 1
Consider in the credibility estimate the weight w as defined in (**??**).

a) Show that w is an *increasing* function of υ. Explain why this had to be so.

b) What does the weight become when $\upsilon = 0$ and when $\upsilon \to \infty$? Interprete!

c) Show that w is a *decreasing* function of τ. There is a good good reason for that. What is it?

d) What does the weight become when $\tau = 0$ and when $\tau \to \infty$? Explain once again.

Exercise 2
Consider a policy holder with annual claim frequency

$$\mu = \xi y,$$

where y is gamma distributed with density function

$$g(y) = \frac{\alpha}{\Gamma(\alpha)} y^{\alpha-1} \exp(-y\alpha)$$

as in section 6.6. The client has been in the company for m years. The number of claims is N_k in year k, $k = 1, \ldots, m$. he problem addressed is what we can say about the *future* number of claims N given $N_1, \ldots, N_m$. Assumptions are the model above for μ and $N_1, \ldots, N_m$ being conditionally independent and conditionally Poisson distributed given μ. Thus

$$\Pr(N_k = n) = \frac{\mu^n}{n!}\exp(-\mu).$$

Note that this is the same type of conditions as in section 8.4.

a) Use Bayes' formula (??) to show that the conditional density function of y given $n_1, \ldots, n_m$ is of the form

$$\text{const} \; \times \; y^{m\bar{n}+\alpha-1} \exp\{-y(\alpha + m\xi)\}$$

where

$$\bar{n} = \frac{1}{m}(n_1 + \ldots + n_m)$$

is the average number of claims per year in the past.

Introduce

$$x = y\frac{\alpha + m\xi}{\alpha + m\bar{n}}.$$

b) Prove that the density function of x becomes

$$g(y) = \frac{\beta}{\Gamma(\beta)} x^{\beta-1} \exp(-x\beta)$$

where

$$\beta = \alpha + m\bar{n}.$$

[Hint: Use exercise 2.?]

We have now established that

$$\mu = \left(\xi\frac{\alpha + m\xi}{\alpha + m\bar{n}}\right) x,$$

where x given $n_1, \ldots, n_m$ follows the gamma distribution above.

c) Use b) and a result in section 6.6 to conclude that N given $n_1 \ldots n_m$ is neagtively binomial distributed.

d) From (**??**) and (**??**) conclude that

$$E(N|n_1, \ldots, n_m) = \xi\frac{\alpha + m\xi}{\alpha + m\bar{n}}$$

and

$$\text{var}(N|n_1, \ldots n_m) = (1 + \gamma)\xi\frac{\alpha + m\xi}{\alpha + m\bar{n}},$$

$$\gamma = \frac{\xi}{\alpha + m\xi}$$

Exercise 3

This is a continuation of the preceding exercise. If we follow the notation of section 8.4, the claim frequency of the policy holder is $\mu(\omega)$ and its estimate from the past record is

$$\widehat{\mu(\omega)} = \xi\frac{\alpha + m\xi}{\alpha + m\bar{n}}.$$

a) When is the estimate above and below the portfolio mean ξ? Explain!

b) Show that

$$E\{\widehat{\mu(\omega)} - \mu(\omega)\} = 0$$

so that $\widehat{\mu(\omega)}$ is unbiased.

c) Use the preceding exercise to deduce that

$$\text{var}\{\widehat{\mu(\omega)} - \mu(\omega)\} = \widehat{\mu(\omega)}\left(1 + \frac{\xi}{\alpha + m\xi}\right).$$

This result suggests that the estimate $\widehat{\mu(\omega)}$ is unlikely to very accurate.

d) Why is that? [Hint: Ignore the last term in the expression for var$\{\widehat{\mu(\omega)}$ and examine the *relative* eroor.]

Explanatory variables could be the location of a house with respect to floods, storms or earthquackes or descriptions of individuals in terms of sex, age, claim record and other things. The case used for illustration is a simplified one from motor insurance where premium is broken down on **age** (2 categories), **distance limit on policy** (6 categories) and **geographical region** (also 6 categories). There are then $2 \times 6 \times 6 = 72$ different groups. Why can't simply straightforward estimation techniques be applied 72 times, once to each group? What typically happens is illustrated by the following estimates, obtained by applying the elementary estimate (**??**) to the youngest age group of the most densely populated region:

*Distance limit on policy (*10000 *km)*	8	12	16	20	25-30	No limit
Estimatd annual claim intensity (%)	4.5	30.4	18.7	16.5	7.3	91.3

These estimates do not make sense! Random error is enormous despite the portfolio having 100000 policies (exposure two years on average). But they are very unevenly divided among the 72 groups and the smaller ones too thin to return reliable results. What regression techniqes do is to present smoother (and without doubt truer) pictures of the reality.