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 liability $\mathcal{X}$, known as the **reserve**. Modelling and computation are required. Examples have been spread over 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*. What counts now is experience with the group which could be defined according to age, what kind of car you own, where your residence are and so on. The natural method is regression, but credibility may be used as well. 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 are looming underneath. Liability over longer time horizons is taken up in the next chapter.

1.2 Portfolio liabilities by simple approximation

Introduction

The portfolio loss $\mathcal{X}$ becomes Gaussian when the number of policies $J \to \infty$. This is a consequence of the central limit theorem and leads to 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, the 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 be avoided by using the sample mean and the sample standard deviation as estimates $\hat{\xi}_z$ and $\hat{\sigma}_z$. Another way is to fit parametric distributions and use their mean and standard deviation.

The approximation (1.2) is nearly always valid for large portfolios even when μ, ξ_z and σ_z depend on j. This is due to the Lindeberg extension of the cental limit theorem; see Appendix A.4. The coefficients a_1 and a_1 are now 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 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 the policy holders of the portfolio are regarded as an independent random sample. Their parameters are then random too, and the coefficients a_0 and a_1 are no longer valid in their previous form. The most important special case is when claim frequencies $\mu_1, \ldots, \mu_J$ are drawn (independently of each other) from a distribution with common mean and standard deviation ξ_μ and σ_μ. If the mean and standard deviation ξ_z and σ_z of the size of claims are fixed, the coefficients (1.1) 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. The following example examines the numerical impact.

Example: Motor insurance
The Norwegian autmobile portfolio was introduced in Chapter 8. Its parameters are

$$\underbrace{\hat{\xi}_\mu = 5.6\%, \qquad \hat{\sigma}_\mu = 2.0\%}_{\textit{annual parameters}} \qquad \text{and} \qquad \underbrace{\hat{\xi}_z = 0.30, \qquad \hat{\sigma}_z = 0.35,}_{\textit{unit: 1000 euro}}$$

where the model for claim intensity was identified in Section 8.3. Claim size excludes personal injuries, and the parameters were obtained from almost 7000 incidents; see also Section 10.4. This 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 (1.4). After having looked up the Gaussian percentiles this leads to the following assessments (in 1000 euro):

Fixed claim frequency			*Random claim frequency*	
1860,	1934	and	1860,	1935.
95% *reserve*	99% *reserve*		95% *reserve*	99% *reserve*

Note how little hetereogeneity among policy holders matters. The message was the same in Section 6.3. Even a quite substantial variation among individuals (as in the present example) is of no more than minor importance 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, usually 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

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

see Section 10.7 for justification. When (1.1) replaces the normal approximation $q_\epsilon^{\rm No}$, this yields

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

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

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

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 of the reserve becomes those in Table 10.1. On the the small portfolio on the left the NP correction has considerable impact, raising the 99% the reserve by as much as 60%. What lies behind is principally losses being strongly skewed towards the right (with skewness coefficient exceeding 18). But when the number of policies is higher, the relative effect is smaller. With 100000 policies the

Money unit: Million DKK

	Portfolio size: $J = 1000$		Portfolio: $J = 100000$	
	95% *reserve*	99% *reserve*	95% *reserve*	99% *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.

difference between the two methods is of minor importance and their almost common assessment one to be trusted.

However, how about the small portfolio? 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. The approximations of this section is likely to work best when the NP term isn't a dominating one.

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 for 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 (apart from mean and variance) doesn't matter anyhow.

What about computational speed? Two methods 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 might appear slower than Algorithm 3.1. An experiment to measure performance is reported in Table 10.2 using a Fortran90 implementation, combining Algorithm 2.10 (Poisson) with Algorithm 4.1 (the empirical distribution function) and Algorithm 9.1a (Gamma). Detailed conditions were

$\mu T = 5\%$	and for losses	empirical 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

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.*

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

generating the cost of claims is the same within both algorithms, and when this component dominates, much of their differences are wiped out. Among the distributions used in this book the Gamma distribution is the most labourious one to sample.

A skeleton algorithm

The portfolio liabiltity is a central issue in general insurance, and its seems worthwile to sketch a general method that collect algorithms spread over several chapters. Suppose claim intensities $\mu_1, \ldots, \mu_J$ are stored on file along with J different 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.

Usually individual losses require most of the computer time. It is therefore little point in faster Poisson samplers such as the guide tables (Section 4.2) and the Atkinson method (Section 2.6) which would *not* bring worthwile improvements.

Danish fire data: The impact of the claim size model

The Danish fire data was examined in Section 9.6 and a number of models were tried. Some worked better than others, and 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, pure 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 $\hat{\mu} = 1\%$,

[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
95%	72	100	104	105	100	71	94	49
99%	173	200	217	230	225	137	214	61
99.97%	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).

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. On its own the empirical distribution function underestimates risk, but it seems to work well when mixed with the Pareto distribution, and the results are not overly dependent on where the threshold b is placed. Another well-fitting model in Section 9.5 was the Gamma distribution on log-scale, and the reserve calculated under it does not deviate much from Pareto mixing. Other models in Section 9.5 were grossly in error, and produce strongly deviating results here. If you compare with the normal power method in Table 10.1 you will discover that it over-shoots at 95% and under-shoots at 99%.

Reserves at level 99.97% have been added. Luckily those figures are not in demand! The results are a mess of unstability, an example of the extreme difficulty of evaluations very far out into the tails of a distribution where they become sensitive to modelling details. Although percentiles so close to one are rarely needed with insurance liabilities, they are used by rating bureaus in finance.

1.4 Differentiated pricing through regression

Introduction

Very young male drivers or owners of fast cars are groups of clients notoriously more risky than others, and it may not be unfair to charge them more. The technological development which makes it easier to collect and store information with bearing on risk, can only further such practice. A picture of how insurance incidents and their cost are connected to circumstances, conditions and the people causing them must be built up from experience, and the principal tool is regression, typically on log-linear form. The purpose of this section is to indicate how Poisson, Gamma and log-normal regression from earlier chapters are put to work.

Explanatory variables (registrations and measurements) $x_1 \dots, x_v$ are then linked to claim intensity μ and mean loss per event ξ_z through

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

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 do not have to be the same for μ and ξ_z, 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_z. In motor insurance (the example below) regression relationships are typically stronger for μ than for ξ_z. Inserting the defining equations for μ and ξ_z into the pure premium $\pi = \mu T \xi_z$ yields

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

and estimates for the coefficients must be supplied.

Estimates of the pure premium

The pure premium of a policy holder with $x_1, \ldots, x_v$ as explanatory variables is estimated as

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

Here $\hat{b}_{\mu i}$ and $\hat{b}_{\xi i}$ are obtained from historical data, usually through statistical software. Assessment of their error is provided too, and we must learn how this is passed on to $\hat{\pi}$ itself. Bootstrapping (Section 7.4) can be used (as always), but there is also a simpler normal technique. The estimated regression coefficients are often approximately normal and therefore their sum $\hat{\eta}$ as well. It follows that $\hat{\pi}$ is log-normal. This is a large-sample result which requires (in principle) much historical data, but a robust attitude is here in order. High accuracy in *error* estimates 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 so that $(\hat{b}_{\mu i}, \hat{b}_{\xi j})$ is uncorrelated for all (i, j). If $\sigma_{\mu ij} = \text{cov}(\hat{b}_{\mu_i} \hat{b}_{\mu_j})$ and $\sigma_{\xi ij} = \text{cov}(\hat{b}_{\xi_i} \hat{b}_{\xi_j})$ are the covariances *within* sets, then

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

where the relationship on the right follows from the general variance formula for sums (rule (A20) in Table A.2). These results are passed on to $\hat{\pi}$ through the usual formulae for the log-normal which yield

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

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

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

Here $\hat{\sigma}_{\mu ij}$ and $\hat{\sigma}_{\xi ij}$ are estimates of their variances/covariances (provided by standard software). In the formula for $\hat{\tau}^2$ take $\hat{\sigma}_{\xi ij} = 0$ or $\hat{\sigma}_{\xi ij} = 0$ if variable i or j isn't included in the regression.

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 multiplicative; i.e.

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

[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.

Here μ_0 is claim intensity when $x_1 = \ldots = x_v = 0$, and the explanatory variables drive intensities up and down compared to this baseline. As an example suppose x_1 is binary, (0 for males and 1 for females). Then

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

and $\mu_f/\mu_m = e^{b_{\mu 1}+b_{\xi 1}}$, is fixed and *independent of all other covariates.*

The female drivers of Section 8.3 who were more reliable than men when young and less when old is not captured by this, but modifications are possible. One way is to design crossed categories. An example was given in Section 8.4. The problem with such procedures is that the number of parameters grows rapidly. Suppose there are three variables consisting of 2, 6 and 6 categories (the example below). The total number of combinations is then $2 \times 6 \times 6 = 72$, and the cross-classification comprises 72 groups which may not appear much when the historical material is over 200000 policy years. On average there would be around 2500 policy years for each group, 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.

Example: 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)

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)

- driving limit (6 categories)
- geographical region (6 categories).

Driving limit is a proxy for how much people drive. Age is simplified drastically compared to what would be done in practice. The regression equation for μ now 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_z. 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. Regression methods used were Poisson (claim frequency) and Gamma (claim size).

The estimated parameters in Table 10.5 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 (see above) varied from 0.2 to 0.5, much smaller than the standard deviation in parenthesis in Table 10.6.

1.5 Differentiated pricing through credibility

Introduction

The preceding section differentiated premium according to observable attributes such as age, sex, geographical location and so on. Other factors with impact on risk could be personal ones that are not easily measured or observed. Drivers of automobiles may be able and concentrated or reckless and inexperienced. Such things influence driving and the accidents caused. Is it possible to deduce the risk they lead to from people's own track record? If so charge unequally! Similar example are shops robbed repeatedly or houses and building frequently damaged. Is there a formal basis for raising the premium?

The general problem is how to rate risks from past experience. In a sense that has been been done that all along by fitting models for claim numbers and size to historical data, but focus is now on the individual. The approach has much in common with the Bayesian ideas of Section 7.6. Policy holders are seen as randomly selected and carriers of pure premia $\pi = \pi^{\text{pu}}$ that vary from one person to another. **Credibility** theory provides a way of estimating them. This is a class of estimation techniques where prior knowledge of how π varies over the portfolio is combined with individual claim records. Such methods can be administered to groups of policy holders too which demands little new. Both viewpoints are pursued below.

Credibility: Approach and modelling

Credibility estimation assumes that each policy holder carries a list of attributes ω with impact on risk. What they are is immaterial. The important thing is that they exist and have been drawn randomly for each individual. Let X be the sum of claims from a certain future period, say a year. Expectation and standard deviation are

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

where the notation reflects that both quantities depend on the underlying ω. We seek $\pi = \pi(\omega)$, the **conditional** pure premium of the policy holder as basis for how much he is charged. The same problem may be posed on group or portfolio level. Now the target is $\Pi = E(\mathcal{X}|\omega)$ where $\mathcal{X}$ is the sum of claims from many individuals and ω applies to all risks jointly, for example as common, uncertain background conditions.

Let $X_1, \ldots, X_K$ (policy level) or $\mathcal{X}_1, \ldots, \mathcal{X}_K$ (group level) be past claims dating up to K years back. The most accurate estimate of π and Π from such records are the conditional means

$$\underbrace{\hat{\pi} = E(X|x_1, \ldots, x_K)}_{\textit{policy level}} \qquad \text{and} \qquad \underbrace{\hat{\Pi} = E(\mathcal{X}|x_1, \ldots, x_K)}_{\textit{group level}} \tag{1.8}$$

where $x_1, \ldots x_K$ are the values of $X_1, \ldots X_K$ or $\mathcal{X}_1, \ldots, \mathcal{X}_K$; see Section 6.4. We may also break claims down on frequencies and size (this viewpoint will be introduced later), but for the moment stay with the estimates (1.8). The issue is essentially the same on either level, and the argument will be written out for single policies. As basic framework introduce the common factor model of Section 6.2 where $X_1 \ldots, X_K, X$ are identically and independently distributed given ω. Surely this is plausible? It won't be true when underlying conditions change systematically during the K periods in question; for credibility methods under such circumstances consult some of the references in Section 10.8.

Complicated modelling can be avoided by leaning on the so-called **structural parameters**. There are three of them; i.e.

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

where ζ is the *average* pure premium over the entire population and υ and τ both represent variation. The former is caused by diversity between individuals and the latter by the physical processes leading to the incidents. These parameters determine mean and standard deviation of X through

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

which are verified by the rules of double expectation and double variance. Indeed,

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

and

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

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

Linear credibility

Let $\hat{\pi}_K$ be an estimate of π based on the claim record $X_1, \ldots, X_K$. The simplest procedure would be to go linear. This means that the estimate is of the form

$$\hat{\pi}_K = b_0 + b_1 X_1 + \ldots + b_K X_K,$$

where $b_0, b_1, \ldots, b_K$ are carefully selected coefficients. The fact that $X_1, \ldots, X_K$ are condionally independent with the same distribution forces $b_1 = \ldots = b_K = b$, and the estimate becomes

$$\hat{\pi}_K = b_0 + b\bar{X}_K \qquad \text{where} \qquad \bar{X}_K = (X_1 + \ldots + X_K)/K. \tag{1.11}$$

A natural way to proceed is to demand that b_0 and b minimize the mean squared error $E(\hat{\pi}_K - \pi)^2$. This sets up a mathematical problem which yields the solution

$$\hat{\pi}_K = (1-w)\zeta + w\bar{X}_K, \qquad \text{where} \qquad w = \frac{\upsilon^2}{\upsilon^2 + \tau^2/K}; \tag{1.12}$$

see Section 10.7 where the argument is given. There is a close resemblance with the Bayes estimate of the normal mean in Section 7.6. The weight w defines a compromise between the average pure premium ζ of the population and the individual record of the policy holder. Note that $w = 0$ if $K = 0$; i.e. with no claim information available the best estimate is the population average. Other interpretations are given among the exercises.

It is also proved in Section 10.7 that

$$E(\hat{\pi}_K - \pi) = 0 \qquad \text{and that} \qquad \text{sd}(\hat{\pi}_K - \pi) = \frac{\upsilon}{\sqrt{1 + K\upsilon^2/\tau^2}}. \tag{1.13}$$

The linear credibility estimate is unbiased, and its standard deviation decreases with K.

Optimal credibility

The preceding estimate is the best *linear* method, but the Bayesian estimate (1.8) offers an improvement since it is optimal among *all* methods; see Section 6.4. Now the aggregate claims $x_1, \ldots, x_K$ are broken down on annual frequencies $n_1, \ldots, n_K$ and individual losses $z_1, \ldots, z_n$ where $n = n_1 + \ldots + n_K$. The Bayes estimate of $\pi = E(X) = E(N)E(Z)$ is

$$\hat{\pi} = E(X|n_1, \ldots, n_K, z_1, \ldots, z_n) = E(N|n_1, \ldots, n_K)E(Z|z_1, \ldots, z_n)$$

if claim numbers and losses are stochastically independent, and the estimation problem has been decoupled into two separate ones. Both the claim intensity μ and the mean claim size $\xi_z = E(Z)$ may vary between individuals, but the diversity is often stronger for the former, and a possible simplification is to fix ξ_z, the same for everybody. Then $\xi_z = E(Z|z_1, \ldots, z_n)$, and the preceding estimate becomes

$$\hat{\pi} = \xi_z E(N|n_1, \ldots, n_K), \tag{1.14}$$

Credibility estimation for ξ_z is discussed in Exercise ??

We need a model for past and future claim numbers $N_1, \ldots, N_K, N$. The natural one is of the common factor type with the sequence conditionally independent given μ and each count Poisson(μT). As model for μ the customary choice is

$$\mu = \xi_\mu G \qquad \text{and} \qquad G \sim \text{Gamma}(\alpha)$$

where G is a standard gamma variable with expectation one. It is verified in Section 10.7 that the estimate (1.14) now becomes

$$\hat{\pi}_K = \zeta \frac{\bar{n} + \alpha/K}{\xi_\mu T + \alpha/K} \qquad \text{where} \qquad \bar{n} = (n_1 + \ldots + n_K)/K, \tag{1.15}$$

and the population average ζ is adjusted up or down according to whether the average claim number $\bar{n}$ is larger or smaller than its expectation $\xi_\mu T$. The error is

$$E(\hat{\pi}_K - \pi) = 0 \qquad \text{and} \qquad \text{sd}(\hat{\pi}_K - \pi) = \frac{\zeta}{\sqrt{\alpha + \xi_\mu K T}} \tag{1.16}$$

which is also proved in Section 10.7.

Credibility on group level

The preceding estimates apply to groups of policies as well. Suppose we seek $\Pi(\omega) = E(\mathcal{X}|\omega)$ where $\mathcal{X}$ is the sum of claims from a group of policy holders. Now ω is common bacground uncertainty, and the linear credibility estimate (1.12) is applied to the claim record $\mathcal{X}_1, \ldots, \mathcal{X}_K$ of the entire group. The structural parameters differ from what they were above. A reasonable assumption is that individual risks are independent given ω. Then

$$E(\mathcal{X}|\omega) = J\pi(\omega) \qquad \text{and} \qquad \text{sd}(\mathcal{X}|\omega) = \sqrt{J}\,\sigma(\omega),$$

and the structural parameters (1.9) become $J\zeta$, $J^2\upsilon^2$ and $J\tau^2$ instead of ζ, υ^2 and τ^2. It follows from (1.12) that the best linear estimate is

$$\hat{\Pi}_K = (1-w)J\zeta + w\bar{\mathcal{X}}_K, \qquad \text{where} \qquad w = \frac{\upsilon^2}{\upsilon^2 + \tau^2/(JK)}, \tag{1.17}$$

Here $\bar{\mathcal{X}}_K = (\mathcal{X}_1 + \ldots + \mathcal{X}_K)/K$ is the average claim on group level. Its weight is much larger than for individual policies and increases with the group size J.

Estimation error is from (1.13)

$$E(\hat{\Pi}_K - \Pi) = 0 \qquad \text{and} \qquad \text{sd}(\hat{\Pi}_K - \Pi) = \frac{J\upsilon}{\sqrt{(1 + KJ\upsilon^2/\tau^2}}, \tag{1.18}$$

and the method is unbiased as before. Note that

$$\frac{\text{sd}(\hat{\Pi}_K - \Pi)}{\text{sd}(\hat{\Pi}_0 - \Pi)} = \frac{1}{\sqrt{(1 + KJ\upsilon^2/\tau^2}},$$

Optimal			Linear, $\sigma_z = 0.1\xi_z$			Linear, $\sigma_z = \xi_z$		
$K = 0$	$K = 10$	$K = 20$	$K = 0$	$K = 10$	$K = 20$	$K = 0$	$K = 10$	$K = 20$
200.0	193.2	187.1	200.0	193.3	187.2	200.0	196.5	193.2

Table 10.7 *Standard deviations of credibility estimates under conditons in the text.*

which decreases with KJ, and the gain from the claim record is much higher when J is large. This is an important observation. It will be suggested below that the accuracy on individual level might be poor, but it could be different for groups.

Similar results apply to the optimal credibility estimates (1.14). On group level the historical claim numbers $n_1, \dots, n_K$ are aggregates from J policies. Their distribution is now Poisson($J\mu T$), and the only thing we have to do is to replace T with JT in (1.14) and (1.15); see Exercise ??.

How accurate is credibility estimation?

Consider esimation error when ξ_z is fixed for all policy holders and μ is random. With $\xi_\mu = E(\mu)$, $\sigma_\mu = \mathrm{sd}(\mu)$, $\xi_z = E(Z)$ and $\sigma_z = \mathrm{sd}(Z)$ we have

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

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 in (1.13) right, we obtain for the linear credibility estimate

$$\mathrm{sd}(\hat{\pi}_K - \pi) = \frac{\sigma_\mu T \xi_z}{\sqrt{1 + K\theta_z T \sigma_\mu^2/\xi_\mu}} \qquad \text{where} \qquad \theta_z = \xi_z^2/(\xi_z^2 + \sigma_z^2).$$

Accurate estimation requires the standard deviation to go down fast as K is raised, and much hinges on the ratio σ_μ^2/ξ_μ. Unfortunately the variability in μ has to be huge for this quantity to be more than a small number.

Standard deviation for the optimal credibility estimate is (1.16) and is in Table 10.7 compared with the linear version using the earlier automobile portfolio as test case. Then $\xi_\mu = 5.6\%$ and $\sigma_\mu = 2\%$ annually ($T = 1$), and additional assumptions are $\xi_z = 10000$ and $\sigma_z = 0.1\xi_z$ or $\sigma_z = \xi_z$. The errors are huge when the mean annual claim $\zeta = 10000 \cdot 0.056 = 560$, and even 20 years of experience with the same client hasn't reduced uncertainty a lot. Nor does the optimal method beat the linear one by much. The picture might not be the same on group level; see Exercise ??.

Finding the parameters

It remains to determine the parameters underlying the credibility estimates. For claim numbers that was discussed in Section 8.3, and only the linear method need be considered here. Historical data are then of the form

$$\begin{array}{cccc|cc}
1 & x_{11} & \dots & x_{1K_1} & \bar{x_1} & s_1 \\
\cdot & \cdot & \cdots & \cdot & \cdot & \cdot \\
\cdot & \cdot & \cdots & \cdot & \cdot & \cdot \\
J & x_{J1} & \dots & x_{JK_J} & \bar{x_J} & s_J, \\
\textit{Policies} & & \textit{Annual claims} & & \textit{mean} & \textit{sd}
\end{array}$$

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

$$\hat{\zeta} = \frac{1}{\mathcal{K}} \sum_{j=1}^{J} K_j \bar{x}_j, \qquad \hat{\tau}^2 = \frac{1}{\mathcal{K} - J} \sum_{j=1}^{J} (K_j - 1) s_j^2 \tag{1.19}$$

and

$$\hat{\upsilon}^2 = \frac{\sum_{j=1}^{J} (K_j/\mathcal{K})(\bar{x}_j - \hat{\xi})^2 - \hat{\tau}^2 (J-1)/\mathcal{K}}{1 - \sum_{j=1}^{J} (K_j/\mathcal{K})^2}; \tag{1.20}$$

for verfication see 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 was introduced in Section 3.2. Parts of **primary** risks placed with a **cedent** are now passed on to **re-insurers** who may in turn go to other re-insurers leading to a global network of risk sharers. Re-insurers may 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 this is irrelevant. For them re-insurance instruments used higher up are without importance as long as the companies involved are solvent. These arrangements are ways to spread risk and may enable small or medium-sized companies to take on heavier responsibilites than its capital base might permit on its own.

Actuarial method don't change much from ordinary insurance. The primary risks rest with cedents, and the stochastic modelling is the same as before. Cash flows differ, but those are merely modifications handled through fixed functions $H(z)$ defining the payment rules and are easily taken care of by Monte Carlo (Section 3.3). The economic impact may be huge, the methodological not. This section outlines some of the most common contracts and indicate consequences for pricing and solvency.

Types of contracts

Re-insurance is expenses shared between two or more parties. Contracts may apply to single events or to sums of claims affecting the entire portfolio during a certain period of time. These losses (denoted Z and $\mathcal{X}$) are then (in obvious mathematical notation) divided between re-insurer and cedent according to

$$\underbrace{Z^{\text{re}} = H(Z), \quad Z^{\text{ce}} = Z - H(Z)}_{\textit{single events}} \qquad \text{and} \qquad \underbrace{\mathcal{X}^{\text{re}} = H(\mathcal{X}), \quad \mathcal{X}^{\text{ce}} = \mathcal{X} - H(\mathcal{X}),}_{\textit{on portfolio level}} \tag{1.21}$$

where $0 \leq H(z) \leq 1$. Here Z^{ce} and $\mathcal{X}^{\text{ce}}$ are the *net* cendent responsibility after accounting for the re-insurance.

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

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

where $Z^{\text{re}} + Z^{\text{ce}}$=Z. The lower bound a is the **retention** limit of the cedent who must cover all claims below. Responsibility (i.e. Z^{ce}) appears unlimited, but in practice there is usually a maximum insured sum S that makes $Z \leq S$. If $b - a = S$, 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**. This type of arrangement is also used with $\mathcal{X}$. Now $\mathcal{X}^{\text{re}}$ and $\mathcal{X}^{\text{ce}}$ are related to $\mathcal{X}$ in a manner similar to the previous relationships for $Z^{\text{re}}$and Z^{ce}, and if b is infinite, the treaty is known as **stopp loss**.

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

$$\underset{\textit{single events}}{Z^{\text{re}} = cZ, \quad Z^{\text{ce}} = (1 - Z)} \qquad \text{and} \qquad \underset{\textit{on portfolio level}}{\mathcal{X}^{\text{re}} = c\mathcal{X}, \quad \mathcal{X}^{\text{ce}} = (1 - c)\mathcal{X}} \tag{1.22}$$

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. Consider the opposite case where $c = c_j$ depends on the contract. Specifically, suppose that

$$c_j = \max(0, 1 - \frac{a}{S_j}) \qquad \text{so that} \qquad Z_j^{\text{re}} = \begin{matrix} 0 & \text{if } a \geq S_j \\ (1 - a/S_j)Z_j & \text{if } a < S_j, \end{matrix} \tag{1.23}$$

where S_j is the maximum insured sum of the j'th primary risk. This is known as **surplus** re-insurance. 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

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

with Monte Carlo approximations

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

Simulation is usually the simplest way if you know the ropes and often takes *less* time to implement than to work out exact formulae (and the latter may not be possible at all). 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 $H(x)$

There is a useful formula for $a \times b$ contracts in terms of *single* events. If $f(z)$ and $F(z)$ are density and distribution function of Z, then the mean re-insurance claim is

$$\xi^{\text{re}} = \int_a^{a+b} (z-a) f(z)\, dz + \int_{a+b}^{\infty} b f(z)\, dz$$
$$= -(z-a)\{1-F(z)\}\Big|_a^{a+b} + \int_a^{a+b} \{1-F(z)\}\, dz + b\{1-F(a+b)\} = \int_a^{a+b} \{1-F(z)\}\, dz$$

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

$$\pi^{\text{re}} = \mu T \int_a^{a+b} \{1 - F_0(z/\beta)\} dz, \tag{1.24}$$

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

$$\pi^{\text{re}} = \mu T \frac{\beta}{\alpha - 1} \left(\frac{1}{(1+a/\beta)^{\alpha-1}} - \frac{1}{(1+(a+b)/\beta)^{\alpha-1}} \right) \qquad \text{for} \qquad \alpha > 0, \tag{1.25}$$

with special treatment being needed for $\alpha = 1$ (Execrise ?).

The example

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

which was used to test Monte Carlo accuracy. With $m = 100000$ the answer was reproduced to two decimal places. Three decimals would take one hundred times more; i.e. $m = 10$ million.

The 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 which changes the parameter of scale from $\beta = \beta_0$ to $\beta_I = (1+I)\beta_0$, see Section 9.2, but the rest of the model is unchanged. Fot $a \times b$ contracts in terms of single events (1.24) shows that that the pure premium π_I^{re} under inflation is related to the original one through

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

How other types of contracts react to inflation is studied among the exercises.

Consider, in particular, the case of infinite b with Pareto distributed claims. Then

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

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

which is not negligable for values of α of some size; try some suitable values if $I = 5\%$, for example. The ratio 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. A second example is

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

and the pure premia π_0^{re} and π_I^{re} can be computed by Monte Carlo. When the upper limit b is infinite and $I = 5\%$, the relative change $(\pi_I^{\text{re}} - \pi_0^{\text{re}})/\pi_0^{\text{re}}$ was found to be

$$\underbrace{\begin{matrix} 9\% & 23\% & 76\% \\ \alpha = 1 & \alpha = 10 & \alpha = 100 \end{matrix}}_{a \text{ median of } Z_0} \qquad \text{and} \qquad \underbrace{\begin{matrix} 17\% & 46\% & 169\% \\ \alpha = 1 & \alpha = 10 & \alpha = 100 \end{matrix}}_{a \text{ upper 10\% percentile of } Z_0}$$

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

The effect of re-insurance on the reserve

Re-insurance may lead to substantial reduction in capital requirements. The cedent company loses money on average, but it can get around on less own capital, and its value per share could be higher. A re-insurance strategy must balance extra cost against capital saved. An illustration is given in Table 10.8. Losses were those of the Norwegain pool of natural disasters in Chapter 7 for which a possible claim size distribution is

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

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

Table 10.8 shows cedent net reserve against the pure re-insurance premium. With claim frequency 1.05 annually 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 per cent, 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 $(1+\gamma^{\text{re}})\pi^{\text{re}}$ where π^{re} is pure premium and γ^{re} the loading, the average loss due to re-insurance is

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

In practice $\gamma^{\rm re}$ is determined by market conditions and may vary enormously in certain branches of insurance. Going from barely zero to 100% and even 200% during a year or two are not unheard of!

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

$$q_\epsilon \doteq \underbrace{E(\mathcal{X})+\mathrm{sd}(\mathcal{X})\phi_\epsilon}_{\textit{normal approximation}} + \underbrace{\mathrm{sd}(\mathcal{X})\frac{1}{6}(\phi_\epsilon^2-1)\mathrm{skew}(\mathcal{X})}_{\textit{skewness correction}}. \tag{1.26}$$

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 (1.4) follows when the formulae for $\mathrm{sd}(\mathcal{X})$ and $\mathrm{skew}(\mathcal{X})$ are inserted into (1.26).

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
Statistical properties of $\bar{X}$. The first part of this section derives the linear credibility estimate and verifies its statistical properties. Three auxiliary results are

$$E(\bar{X}) = \zeta, \qquad \text{var}(\bar{X}) = \upsilon^2 + \tau^2/K \qquad \text{cov}\{\bar{X}, \pi(\omega\} = \upsilon^2. \tag{1.27}$$

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 $\text{cov}\{\bar{X}, \pi(\omega)\}$. In the following we shall write $\pi = \pi(\omega)$.

Linear credibility Let $\hat{\pi}_K$ be the estimate in (1.11). Then

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

after a little reorganization. Hence

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

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

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

and after inserting (1.27) (middle and right) and $\upsilon^2 = \text{var}(\pi)$, 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 $\hat{\pi}_K$ defined in (1.12).

The statistical properties Unbiasedness is a consequence of

$$E(\hat{\pi}_K) = E\{(1 - w)\zeta + w\bar{X}\} = (1 - w)\zeta + wE(\bar{X}) = (1 - w)\zeta + w\zeta = \zeta = E(\pi).$$

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

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

as asserted in (1.13).

Optimal credibility We must determine the distribution of μ given $N_1, \ldots, N_K$. The prior density function assumed for μ is

$$f(\mu) = C\mu^{\alpha-1}e^{-\mu\alpha/\xi_\mu}$$

where C is a constant whereas the the claim numbers are conditionally independent and Poisson given μ. Their joint density function is

$$f(n_1, \ldots, n_K|\mu) = \prod_{k=1}^{K}\left(\frac{(\mu T)^{n_k}}{n_k!}e^{-\mu T}\right) = C\mu^{n_1+\ldots+n_K}e^{-\mu KT}$$

where C (another constant) is an expression not depending on μ. Multiplying the pair of density functions together yields the posterior density function $p(\mu|n_1, \ldots, n_K)$ up to a constant. In other words,

$$p(\mu|n_1, \ldots, n_K) = C\left(\mu^{\alpha-1}e^{-\mu\alpha/\xi_\mu}\right) \cdot \left(\mu^{n_1+\ldots+n_K}e^{-\mu KT}\right) = Ce^{\alpha+K\bar{n}-1}e^{-\mu(\alpha/\xi+KT)}$$

where $\bar{n} = (n_1 + \ldots + n_K)/K$. This is another Gamma density function with expectation

$$E(\mu|n_1, \ldots, n_K) = \frac{\alpha + K\bar{n}}{\alpha/\xi_\mu + KT} = \xi_\mu \frac{\bar{n} + \alpha/K}{\xi_\mu T + \alpha/K}.$$

Multiply with $T\xi_z$, and you get

$$\hat{\pi}_K = \zeta \frac{\bar{n} + \alpha/K}{\xi_\mu T + \alpha/K}.$$

which is the credibility estimate (1.15).

Optimal credibility: Error
Note that

$$\hat{\pi}_K - \pi = \zeta \frac{\bar{n} + \alpha/K}{\xi_\mu T + \alpha/K} - \mu T \xi_z = \zeta \left(\frac{\bar{n} + \alpha/K}{\xi_\mu T + \alpha/K} - \frac{\mu}{\xi_\mu} \right).$$

Since $E(\bar{n}|\mu) = \mu T$ and $\text{var}(\bar{n}|\mu) = \mu T/K$, this implies that

$$E(\hat{\pi}_K - \pi|\mu) = \zeta \left(\frac{\mu T + \alpha/K}{\xi_\mu T + \alpha/K} - \frac{\mu}{\xi_\mu} \right) \qquad \text{and} \qquad \text{var}(\hat{\pi}_K - \pi|\mu) = \zeta^2 \frac{\mu T/K}{(\xi_\mu T + \alpha/K)^2},$$

and by the rule of double variance

$$\text{var}(\hat{\pi}_K - \pi) = \zeta^2 \left(\frac{T}{\xi_\mu T + \alpha/K} - \frac{1}{\xi_\mu} \right)^2 \sigma_\mu^2 + \zeta^2 \frac{\xi_\mu T/K}{(\xi_\mu T + \alpha/K)^2}.$$

Under the model assumed $\sigma_\mu = \xi_\mu/\sqrt{\alpha}$, and when this is inserted, the preceding expression reduces to

$$\text{var}(\hat{\pi}_K - \pi) = \frac{\zeta^2}{\alpha + K\xi_\mu T}.$$

which is (1.16).

The estimates of ξ, τ and υ.
We shall examine the estimates (1.19) and (1.20). The principal part of the argument is to verify unbiasedness. Firstly, since $E(\bar{x}_j) = \zeta$ and $K_1 + \ldots + K_J = \mathcal{K}$ we have

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

For τ we must utilize that s_j^2 is the ordinary empirical variance. Thus

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

and by the rule of double expectation

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

Now (1.19) right yields

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

Finally, note that

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

so that

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

and from (1.27) middle

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

Since

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

it now 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 the estimate (1.20) for $\hat{\upsilon}$, and 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.