1 Modelling claim frequency

1.1 Introduction

Actuarial modelling in property insurance may be broken down on claim size (next chapter) and claim frequency (treated here). Section 3.2 introduced the Poisson distribution which was used as model for the numbers of claims. The parameter was $\lambda = \mu T$ (for single policies) and $\lambda = J\mu T$ (for portfolios) where J was the number of policies, μ claim intensity and T the time of exposure. Most claim number models used in practice is related to the Poisson distribution in some way. Such a line has strong theoretical support through the **Poisson point process** outlined in Section 8.2. This leads to the Poisson model as a plausible one under a wide range of circumstances.

It also explores the meaning of the key parameter μ, the vehicle for model extensions. There are here two main viewpoints. The first (with a long tradition in actuarial science) is to regard μ as a random variable, either re-drawn independently *for each* customer or re-drawn each period as common background *for all*. Models of that kind were initiated in Section 6.3, and there will be more below. Then there are situations where variations in μ are linked to **explanatory factors**, such as young drivers causing accidents more often than old or eartquackes or hurricanes striking certaing regions more frequently than others. In a similar vein risk may be growing systematically over time or being influenced by the season of the year, as in Figure 8.1 below. Such situations are best treated through **Poisson regression**, introduced in Section 8.4.

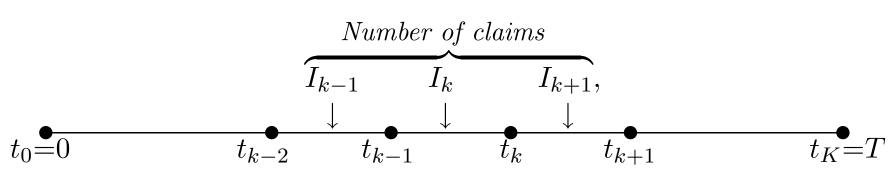
1.2 The world of Poisson

Introduction

Poisson modelling can be understood in terms of the **Poisson point process**, starting from the obvious fact that accidents and incidents occur suddenly and unexpectedly and may take place at any point in time. The mathematical formulation is based on cutting the interval from $t_0 = 0$ to the end of period $t_K = T$ into K pieces of equal length

$$h = T/K \tag{1.1}$$

so that $t_k = kh$ $(k = 0, 1 \dots K)$ are the changepoints; consult the following display:



Eventually the time increment h will approach 0 (making the number K of periods infinite), but to begin with these quantities are kept fixed. Let I_k (for single policies) or $\mathcal{I}_k$ (for portfolios) be the number of claims in the k'th interval from t_{k-1} to t_k. The sums

$$\underbrace{N = I_1 + I_2 + \ldots + I_K}_{\textit{single policy}} \qquad \text{and} \qquad \underbrace{\mathcal{N} = \mathcal{I}_1 + \mathcal{I}_2 + \ldots + \mathcal{I}_K}_{\textit{an entire portfolio}} \tag{1.2}$$

are then the total number of claims up to T. Their probability distribution follows from a theoretical argument covering most of the present section.

Modelling single policies

Consider a single policy over a *short* time interval of length h. More than one claim is highly unlikely, and if that possibility is ruled out (relaxed later), I_k is either 0 or 1 (hence the symbol I, signifying an indicator variable. Suppose that

$$\text{(i) } p = \Pr(I_k = 1) = \mu h \qquad \text{and} \qquad \text{(ii) } I_1, \ldots I_K \text{ are stochastically independent}$$

both plausible conditions. Surely the probability of an incident is proportional to the time increment h whereas accidents typically are consequences of forces and events with no relationship to each other. An incident that has occurred makes further incidents neither more nor less probable which is condition (ii). There may be violations to this, but they are usually better handled by allowing the intensity μ to vary with time; see below.

A sequence of indicator variables $\{I_k\}$ satisfying (i) and (ii) is called a **time-homogeneous** Poisson point process. Formally the definition applies in the limit as $h \to 0$. This yields a mathematical argument which leads to

$$\Pr(N = n) \to \frac{(\mu T)^n}{n!} \exp(-\mu T) \qquad \text{as} \qquad h \to 0,$$

and N is Poisson distributed with parameter $\lambda = \mu T$. The coefficient of proportionality μ is an **intensity** (i.e per time unit). Accepting (i) and (ii) yields the Poisson model as a logical consequence.

This important result follows from the fact that $I_1, \ldots, I_K$ is a Bernoulli series and their sum N a binomially distributed random variable with density function $(n = 0, 1, \ldots, K)$

$$\Pr(N = n) = \frac{K!}{n!(K-n)!} p^n (1-p)^{K-n}, \qquad \text{where} \qquad p = \mu h = \mu T/K.$$

It is easy to verify that

$$\Pr(N = n) = B_1 \cdot B_2 \cdot B_3 \cdot B_4$$

where

$$B_1 = \frac{(\mu T)^n}{n!}, \qquad B_2 = \frac{K(K-1)\cdots(K-n+1)}{K^n}$$
$$B_3 = (1 - \mu T/K)^K, \qquad B_4 = \frac{1}{(1-\mu T/K)^n}.$$

Simply multiply $B_1, \ldots, B_4$ together to convince yourself that the product equals $\Pr(N = n)$. Let $h \to 0$, or, equivalently, $K \to \infty$, keeping n fixed. The first factor B_1 is unchanged whereas the others in the limit become

$$B_2 \to 1, \qquad B_3 \to \exp(-\mu T) \qquad \text{and} \qquad B_4 \to 1.$$

where B_3 is a consequence of the fact that $(1 + a/K)^K \to \exp(a)$, applied with $a = -\mu T$. Collecting the expressions for B_1, B_2, B_3 and B_4 yields the Poisson distribution as claimed.

Many policies jointly

On the portfolio level J independent process (one for each policy) run in parallel. Let $\mu_1, \dots, \mu_J$ be their intensities and $\mathcal{I}_k$ their total number of claims in period k. Note that

$$\Pr(\mathcal{I}_k = 0) = \underbrace{\prod_{j=1}^{J}(1-\mu_j h)}_{no\ claims} \qquad \text{and} \qquad \Pr(\mathcal{I}_k = 1) = \sum_{i=1}^{J} \underbrace{\{\mu_i h \prod_{j \neq i}(1-\mu_j h)\}}_{claim\ policy\ i\ only}.$$

Here $1-\mu_j h$ is the probability of no claims from policy j, and multiplying all those together yields the probability of no claims against the *portfolio*. The probability of exactly one claim is a little more complicated. Such an event may affect the the first, second, third policy and so on. That gives J different probabilities that must be added for the expression on the right.

Both probabilities may be simplified by multiplying their products out and identifying the powers of h. Try to do it for $J = 3$, and the general structure emerges as

$$\Pr(\mathcal{I}_k = 0) = 1 - h\left(\sum_{j=1}^{J}\mu_j\right) + o(h) \qquad \text{and} \qquad \Pr(\mathcal{I}_k = 1) = h\left(\sum_{j=1}^{J}\mu_j\right) + o(h), \qquad (1.3)$$

where terms of order h^2 and higher have been ignored and lumped into the $o(h)$ contributions. If you are unfamiliar with that notation, it signifies a mathematical expression for which

$$\frac{o(h)}{h} \to 0 \qquad \text{as} \qquad h \to 0.$$

These small quantities do not count in the limit as $h \to 0$. A mathematical proof is given in Section 8.7.

It follows that the portfolio sequence $\mathcal{I}_1, \dots, \mathcal{I}_K$ satifies the same model as the policy analogue $I_1, \dots, I_K$ with the sum $\mu_1 + \dots + \mu_J$ taking over from μ. The portfolio number of claims $\mathcal{N}$ are therefore Poisson too with parameter

$$\lambda = (\mu_1 + \dots + \mu_J)T = J\bar{\mu}T \qquad \text{where} \qquad \bar{\mu} = (\mu_1 + \dots + \mu_J)/J,$$

When claim intensities vary over the portfolio, only their average counts.

Time variation

Claim intensities do not necessarily remain constant over time. This is evident in the example in Figure 8.1 which comes from a Scandinavian insurance company. Monthly claim numbers have been converted into *estimated* intensities (method below!) and plotted against time. There are significant fluctuations due to the season of the year (slippery roads causing accident in winter!). Over the period in question there also seems to be systematic growth. The data will be examined more closely in Chapter 11. Here they serve as an example of a seasonal variation that is present in many parts of the world (wet season againt dry, stormy against calm). What is the impact of such factors on how we evaluate risk? Answer (perhaps surprisingly): Not very strong!

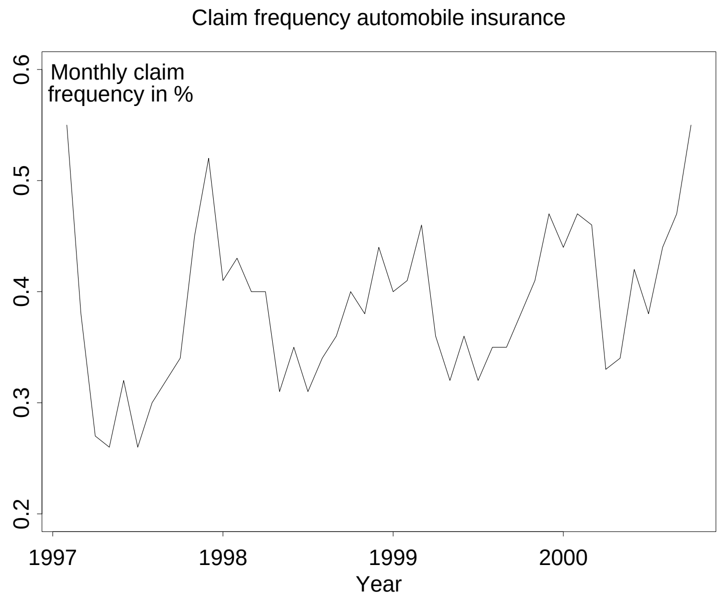


Figure 8.1 *Estimated monthly claim intensities for data from a Norwegian insurance company*

The mathematics is handled through a time-varying function $\mu = \mu(t)$. Now the binary variables $I_1, \ldots, I_K$ used earlier are based on intensities $\mu_1, \ldots, \mu_k$ where

$$\mu_k = \mu(t_k) \qquad \text{for} \qquad k = 1, \ldots, K,$$

and when $I_1, \ldots, I_K$ are added to the total count N, this is the same issue as if K different *policies* apply on an interval of length h. In other words, N must still be Poisson, now with parameter

$$\lambda = h \sum_{k=1}^{K} \mu_k \quad \to \quad \int_0^T \mu(t)\,dt \qquad \text{as} \qquad h \to 0,$$

where the limit simply is how integrals are defined. The Poisson parameter for N can also be written

$$\lambda = T\bar{\mu} \qquad \text{where} \qquad \bar{\mu} = \frac{1}{T}\int_0^T \mu(t)\,dt,$$

and the introduction of the function $\mu(t)$ hasn't changed much. A *time average* $\bar{\mu}$ takes over from a constant μ.

The Poisson model.
In summary, the world of Poisson, governed by the point process of the same name, is a tidy one where claim numbers, N for policies and $\mathcal{N}$ for portfolios, are Poisson distributed with parameters

$$\underset{\textit{policy level}}{\lambda = \mu T} \qquad \textit{and} \qquad \underset{\textit{portfolio level}}{\lambda = J\mu T.}$$

The intensity μ, vehicle for much extended modelling later on, is an average over time and policies.

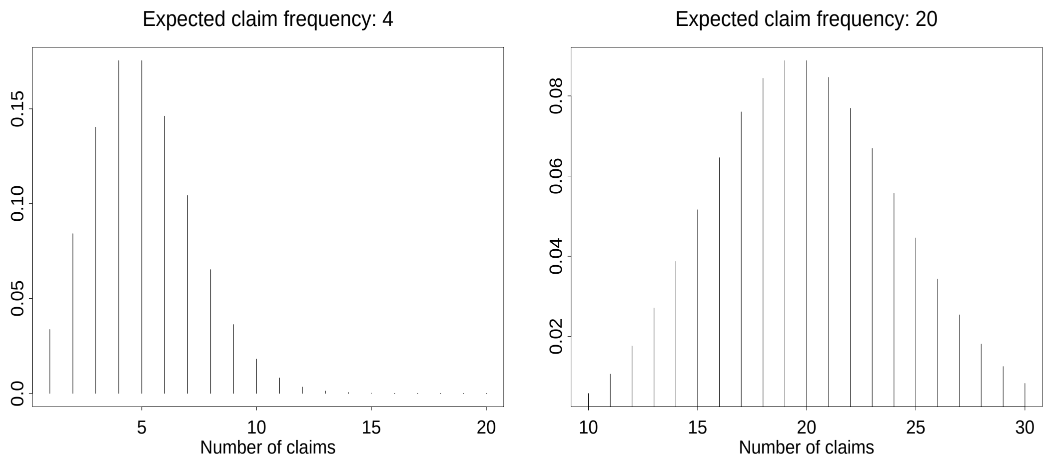


Figure 8.2 *Poisson density functions for $\lambda = 4$ (left) and $\lambda = 20$ (right).*

Poisson models have useful operational properties. The earlier argument connecting portfolio and policies also tells us that sums of independent Poisson variables must remain Poisson; i.e. if $N_1, \ldots, N_J$ are independent and Poisson with parameters $\lambda_1, \ldots, \lambda_J$, then

$$\mathcal{N} = N_1 + \ldots + N_J \sim \text{Poisson}(\lambda_1 + \ldots + \lambda_J).$$

A simple consequence of this **convolution property** is that Poisson variables become Gaussian as $\lambda \to \infty$ (Exercise 8.3.1). Signs of this are evident in Figure 8.2 where the density function has been plotted for two values of λ. The skewness at $\lambda = 4$ has largely disappeared at $\lambda = 20$. Mean, standard deviation, skewness and kurtosis of a Poisson count N are

$$E(N) = \lambda, \qquad \text{sd}(N) = \sqrt{\lambda}, \qquad \text{skew}(N) = 1/\sqrt{\lambda}, \quad \text{and} \quad \text{kurt}(N) = 1/\lambda.$$

This leads to a quick appraisal from historical data. For a Poisson distribution $\text{var}(N) = E(N)$, and their sample versions shouldn't deviate too much.

Using historical data

Poisson models are usually fitted empirically by means of historical claim numbers $n_1, \ldots, n_n$ from n policies. Let T_j be the time policy j has been under risk. The standard estimate of a common intensity μ is then

$$\hat{\mu} = \frac{n_1 + \ldots + n_n}{T_1 + \ldots + T_n}, \tag{1.4}$$

for which

$$E(\hat{\mu}) = \mu \qquad \text{and} \qquad \text{sd}(\hat{\mu}) = \sqrt{\frac{\mu}{T_1 + \ldots + T_n}}. \tag{1.5}$$

see Exercise 8.2.9. The estimate is unbiased with standard deviation determined by the the total time of exposure $T_1 + \ldots + T_n$. This is known to be the most accurate method possible (see Lehmann and Casella (1998), p.121) and can even be used when data are coded so that the same policy holder is hiding under different T_j.

Rates in % annually

	Age groups (years)				
	20-25	26-39	40-55	56-70	71-94
Males	10.5 (0.76)	6.0 (0.27)	5.8 (0.13)	4.7 (0.15)	5.1 (0.22)
Females	8.4 (1.12)	6.3 (0.28)	5.7 (0.23)	5.4 (0.28)	5.8 (0.41)

Table 8.1 *Estimated, annual accident rates of the Scandinavian automobile portfolio (standard error in paranthesis).*

Example: A Norwegian automobile portfolio
The automobile portfolio plotted in Figure 8.1 will serve as example throughout much of this chapter. Seasonal variations and the apparent increase in μ will be not be taken into account (see Section 11.4 for those). The total number of claims is 6978, and the total risk exposure $T_1 + \ldots + T_n = 123069$ automobile years which yields

$$\hat{\mu} = \frac{6978}{123069} = 5.67\% \qquad \text{with estimated standard deviation} \qquad \sqrt{\frac{0.0567}{123069}} \doteq 0.07\%.$$

The estimate is annual and tells us that during one year about one car in twenty is causing an accident.

Considerable variations between groups of individuals are hiding behind this portfolio average. Consider the cross-classification in Table 8.1 where the intensities have been broken down on male/female and five age categories. There are now $2 \times 5 = 10$ different sub-groups to which the estimate (1.4) could be applied. What emerged was strong dependence on age. Also note the lower estimate for women in the youngest group, indicating that it might be fair to charge young women 20 to 25% less that young men[1].

1.3 Random intensity

Introduction
How μ varies over the portfolio will in the next section be refered to observables such as the age or sex of the individual, but there are also personal factors a company can't know much about. In automobile insurance some drivers are reckless, others not, some have excellent power of concentration while others easily lose theirs. Things like that create uncertainty beyond what's coming from the traffic, and is captured by making μ a random variable, re-drawn for each individual. Stochastic intensities have a second rationale too. All insurance processes run against a background that is itself subject to random variation. A striking example is driving conditions during winter months that may be rainy or icy (and perilous) one year and safe and dry the next one. Such uncertainty affects all policy holders simultaneously. Much was in Section 6.3 made of the distinction between individual and collective hidden randomness. Consequences for risk were widely different, but the *mathematics* is very much the same. The models are in either case conditional ones of the form

$$\underset{\textit{policy level}}{N|\mu \sim \text{Poisson}(\mu T)} \qquad \text{and} \qquad \underset{\textit{portfolio level}}{\mathcal{N}|\mu \sim \text{Poisson}(J\mu T).}$$

[1]In Norway with its stern laws on the equality of sexes such price differentiation is considerered illegal! Discrimination according to age is all right.

Most of this section concentrates on the version on the left. When T is replaced by JT, the results apply on portfolio level too.

A first look

Let $\xi = E(\mu)$ and $\sigma = \text{sd}(\mu)$ be the mean and standard deviation of μ. It follows as in Section 6.3 that

$$E(N) = E(\mu T) = \xi T \qquad \text{and that} \qquad \text{var}(N) = E(\mu T) + \text{var}(\mu T) = \xi T + \sigma^2 T^2, \tag{1.6}$$

using the double rules for mean and variance. Note that $E(N) > \text{var}(N)$, and N is no longer Poisson distributed. A lot can be learned without further modelling (see again Section 6.3), and it will below be demonstrated how ξ and σ can be estimated from historical data.

Specific models for μ is of interest too. Their effect on the distribution of N is bestowed through the **mixing** relationship

$$\Pr(N = n) = \int_0^\infty \Pr(N = n|\mu) g(\mu)\, d\mu, \tag{1.7}$$

where $g(\mu)$ is the density function of μ. Gamma models are the traditional choice for $g(\mu)$ and detailed below. Another possibility is log-normal distributions. Closed formulae for the density function of N are now unavailable, but simulation is easy:

Algorithm 8.1 Poisson mixing

0 Input: T and density function $g(\mu)$ for μ
1 Draw $\mu^* \sim g$ *%Many possibilites*
2 Draw $N^* \sim$ Poisson $(\mu^* T)$ *%Algorithm 2.10*
3 Return N^*

Running the algorithm many times enables you to approximate the density function of N. There are several variants. The same Monte Carlo μ^* may apply to an indvidual over time or to an entire portfolio; see exercises. Another possibility is time-varying models for μ which is introduced in Section 11.3.

Estimating the mean and variance of μ

It is possible to estimate ξ and σ from historical data without specifying $g(\mu)$. Suppose there are J policy holders with claims $n_1, \ldots, n_J$ and risk exposures $T_1, \ldots, T_J$. Their intensities $\mu_1, \ldots, \mu_J$ are estimated through the ratios $\hat{\mu}_j = n_j/T_j$ with huge uncertainty, yet these individual estimates may be pooled for the portfolio parameters. The problem is similar to the Bühlman-Straub-Sundt estimate in credibility theory (Section 10.5). One solution is

$$\hat{\xi} = w_1\hat{\mu}_1 + \ldots + w_n\hat{\mu}_n \qquad \text{where} \qquad w_j = \frac{T_j}{T_1 + \ldots + T_n}, \tag{1.8}$$

and

$$\hat{\sigma}^2 = \frac{\sum_{j=1}^n w_j(\hat{\mu}_j - \hat{\xi})^2 - C}{1 - \sum_{j=1}^n w_j^2} \qquad \text{where} \qquad C = \frac{(n-1)\hat{\xi}}{T_1 + \ldots + T_n}. \tag{1.9}$$

$\Gamma(x) \doteq (x+4.5)^{x-0.5}e^{-(x+4.5)} \times \sqrt{2\pi}\left(a_0 + \frac{a_1}{x} + \frac{a_2}{x^2} + \frac{a_3}{x^3} + \frac{a_4}{x^4} + \frac{a_5}{x^5} + \frac{a_6}{x^6}\right)$		
$a_0 = 1.000000000190015$	$a_1 = 76.18009172947146$	$a_2 = 86.50532032941677$
$a_3 = 24.01409824083091$	$a_4 = -1.231739572450155$	$a_5 = 0.1208650973866179\,10^{-2}$
$a_6 = -0.5395239384953\,10^{-5}$		

Table 8.2 *An approximation to the Gamma function (the error is less than* $2\,10^{-10}$*).*

The estimates are examined in Section 8.6 where it is proved that both are unbiased and tend to the correct values ξ and σ as the total exposure time $T_1 + \ldots + T_n \to \infty$ (weak additional assumption needed). The estimated variance does not have to be positive since a correction term C is subtracted (this makes the estimate unbiased). If a negative value appears, take the position that μ variation over the portfolio is unlikely to be important and let $\hat{\sigma} = 0$.

The negative binomial model

The most commonly applied model for μ is the Gamma distribution. It is then assumed that.

$$\mu = \xi G \qquad \text{where} \qquad G \sim \text{Gamma}(\alpha) \tag{1.10}$$

where Gamma(α) has mean one; see Section 2.6. This makes μ fluctuate around ξ with uncertainty controlled by α. Specifically

$$E(\mu) = \xi \qquad \text{and} \qquad \text{sd}(\mu) = \xi/\sqrt{\alpha}. \tag{1.11}$$

Note that sd(μ) $\to 0$ as $\alpha \to \infty$, and the pure Poisson model with fixed intensity emerges in the limit.

One of the reasons for the popularity of this model is without doubt the closed form of the density function of N. It is in Section 8.6 shown that ($n = 0, 1, \ldots$)

$$\Pr(N = n) = \frac{\Gamma(n+\alpha)}{\Gamma(n+1)\Gamma(\alpha)} p^{\alpha}(1-p)^n \qquad \text{where} \qquad p = \frac{\alpha}{\alpha + T\xi}. \tag{1.12}$$

This is known as the **negative binomial** distribution and will be denoted nbin(ξ, α). Computation (needed below) requires the Gamma function $\Gamma(x) = \int_0^\infty y^{x-1}e^{-y}\,dy$. The approximation in Table 8.2, taken from Press, Teukolsky, Vetterling and Flannery (1994), is extremely accurate, but be sure to use it on logarithmic form to avoid overflow in the computer.

Mean and variance are

$$E(N) = T\xi \qquad \text{and} \qquad \text{var}(N) = T\xi(1 + T\xi/\alpha); \tag{1.13}$$

see (1.6) into which you insert $\sigma = \xi/\sqrt{\alpha}$. There is a convolution property similar to one for Poisson models. Suppose $N_1, \ldots, N_J$ are independent with common distribution nbin(ξ, α). Then

$$\mathcal{N} = N_1 + \ldots + N_J \sim \text{nbin}(J\xi, J\alpha); \tag{1.14}$$

see Section 8.6 for the proof.

Fitting the negative binomial

Moment estimation using (1.8) and (1.9) is simplest technically. The estimate of ξ is simply $\hat{\xi}$, and for α invoke (1.11) right which yields

$$\hat{\sigma} = \hat{\xi}/\sqrt{\hat{\alpha}} \qquad \text{which implies} \qquad \hat{\alpha} = \hat{\xi}^2/\hat{\sigma}^2.$$

It was seen above that $\hat{\sigma} = 0$ is a distinct possibility. When that happens, interprete it as an infinite $\hat{\alpha}$ which signals a pure Poisson model.

Likelihood estimates are a little more accurate in theory, but require more work to implement. The log likelihood function follows by inserting n_j for n in (1.12) and adding the logarithm over all j. This leads to the criterion

$$\mathcal{L}(\xi, \alpha) = \sum_{j=1}^{n} \log\{\Gamma(n_j + \alpha)\} - n\{\log\{\Gamma(\alpha)\} - \alpha\log(\alpha)\}$$
$$+ \sum_{j=1}^{n} \{n_j \log(\xi) - (n_j + \alpha)\log(\alpha + T_j\xi)\}$$

where constant factors not depending on ξ and α have been omitted. It takes numerical software to optimize this function. A primitive way is to compute it over a grid of points (ξ, α) and select the maximizing pair (better ways in Appendix C!). An accurate approximation to $\Gamma(x)$ was given in Table 8.2.

Automobile example continued
For the Norwegian automobile portfolio the estimates (1.8) and (1.9) were

$$\hat{\xi} = 5.60\% \qquad \text{and} \qquad \hat{\sigma} = 2.0\%,$$

and the variation in μ over the portfolio is huge. The moment estimate of α in a negative binomial model is $\hat{\alpha} = 0.056^2/0.02^2 = 7.84$, and the likelihood estimate is another possibility. There are now two competing versions:

$$\underbrace{\hat{\xi} = 5.60\%, \ \ \hat{\alpha} = 7.84}_{\textit{moment estimates}} \qquad \text{and} \qquad \underbrace{\hat{\xi} = 5.60\%, \ \ \hat{\alpha} = 2.94}_{\textit{likelihood estimates}}$$

For ξ the results are identical to two decimal places (they are *not* the same estimate!), but for α the descrepancy is huge, the likelihood estimate portraying a much more hetereogeneous phenomenon. Behind the estimates are more than 120000 automobile years, so the deviation is *not* accidental. But what lies behind and which one is to be trusted? Answer: In a sense neither! The Gamma model does *not* describe the underlying variation in μ well. If it had, the estimates would have been much closer. The moment estimate is best, because $\text{sd}(\mu)$ (determined without the Gamma assumption) is captured correctly. An example where the Gamma model does work is given in Exercise 8.5.2.

1.4 Explaining discrepancy: Poisson regression

Introduction
The automobile example in Table 8.1 is a special case of the important problem of linking risk to **explanatory** variables (or **covariates**). Insurance companies do this to understand which

customers are profitable and which are not and to charge differently in different segments of the portfolio. Whether this accords with the principle of solidarity behind insurance might be a matter of opinion, but individual pricing has an any case become more and more widespread. The modern actuary must certainely understand and master the techniques involved.

Credibility theory is a traditional answer, but this historically important method has a scope somewhat different from the present one and is treated in Chapter 10. The issue now is the use of **observable** variables such as age and sex of drivers, geographical location of a house and so forth. In Table 8.1 the portfolio was partitioned into groups of policy holders, examined one by one. This shouldn't be the only approach as there might easily be too many groups. Many relevant factors are missing in Table 8.1, for example the amount of driving[2], geographical region and the type of car. If these variables are classified into categories and cross-classified with sex and age, there can easily be hundreds and even thousands of different categories which would require huge amounts of historical data.

The model

The method most widely used in practice is **Poisson regression** where the claim intensity μ is 'explained' by a set of observable variables $x_1, \ldots, x_v$ through a relationship of the form

$$\log(\mu) = b_0 + b_1 x_1 + \ldots + b_v x_v. \tag{1.15}$$

Here $b_1, \ldots, b_v$ are coefficients that are determined from historical records. The link between μ and the explanatory variables $x_1, \ldots, x_v$ allows us to discriminate customers according to the risk they represent.

Why is $\log(\mu)$ used in (1.15) and not μ itself? It will emerge below that it does lead to a useful interpretation of the model, but the most compelling reason is almost philosophical. Linear functions, as those on the right, extend over the whole real line whereas μ is always positive. It does not really make sense to equate two quantites differing so widely in range, and a log-transformation arguably makes the scales more in line with each other.

Data and likelihood function

Historical data are of the following form:

n_1	T_1	$x_{11} \cdots x_{1v}$
n_2	T_2	$x_{21} \cdots x_{2v}$
.	.	. .... .
.	.	. .... .
n_n	T_n	$x_{n1} \cdots x_{nv}$
Claims	*exposure*	*covariates*

On row j we have the number of claims n_j, the exposure to risk T_j (sometimes called **the offset**) and the values of the explanatory variables $x_{j1} \ldots x_{jv}$. This is known as a **data matrix**. Many software packages work from information stored in that way. How specific situations are read into it will be indicated below.

[2]The companies do not know that, but they have a *proxy*, in the annual distance limit on policies.

The coefficients $b_0, \ldots, b_v$ are usually determined by likelihood estimation. It is then assumed that n_j is Poisson with parameter $\lambda_j = T_j\mu_j$ where μ_j are tied to covariates $x_{j1}, \ldots, x_{jv}$ as in (1.15). The density function of n_j is then

$$f(n_j) = \frac{(T_j\mu_j)^{n_j}}{n_j!}\exp(-T_j\mu_j)$$

or

$$\log\{f(n_j)\} = n_j\log(\mu_j) + n_j\log(T_j) - \log(n_j!) - T_j\mu_j,$$

which is to be added over all j for the likelihood function $\mathcal{L}(b_0, \ldots, b_v)$. We may drop the two middle terms n_jT_j and $\log(n_j!)$ (in this context constants). The likelihood criterion then becomes

$$\mathcal{L}(b_0, \ldots, b_v) = \sum_{j=1}^{n}\{n_j\log(\mu_j) - T_j\mu_j\} \quad \text{where} \quad \log(\mu_j) = b_0 + b_1x_{j1} + \ldots + b_vx_{jv}, \quad (1.16)$$

and there is little point in carrying the mathematics further. Optimization of (1.16) must be done by numerical software. It is rather straightforward, since it can be proved (McCullagh and Nelder, 1989) that $\mathcal{L}(b_0, \ldots, b_v)$ is a convex surface with a single maximum.

Interpretation and coding

Poisson regression is best learned by example, and it will be illustrated on the data in Table 8.1. There are two explanatory variables age (x_1) and sex (x_2). The most obvious way of feeding them into the regression model (1.15) is to write

$$\log(\mu_j) \;=\; b_0 \;+\; \underset{\textit{age effect}}{b_1x_{j1}} \;+\; \underset{\textit{male/female}}{b_2x_{j2}}. \quad (1.17)$$

Here x_{j1} is the age of the owner of car j and

$$x_{j2} = 0, \quad \text{if } j \text{ is male}$$
$$= 1, \quad \text{if } j \text{ is female.}$$

Suppose owners j and i are of the same age x_1, the former being a male and the other a female. Then

$$\frac{\mu_i}{\mu_j} = \frac{e^{b_0+b_1x_1+b_2}}{e^{b_0+b_1x_1}} = e^{b_2},$$

and if $b_2 = 0.037$ (taken from Table 8.3), then $\mu_i/\mu_j = \exp(0.037) \doteq 1.037$; i.e the average Norwegian female driver is causing 3.7% more accidents than the average male.

The coefficient b_1 is likely to be negative. As drivers become more experienced, their accident rate goes down. But $\log(\mu)$ should not necessarily be a strictly *linear* function of age. Indeed, the accident rate could well go up again when people become very old. A more flexible mathematical formulation is to divide into *categories*. The *exact* age x_2 is then replaced by the age *group* to which the policy holder belongs. With g such groups the model is changed to

$$\log(\mu_j) = b_0 + \sum_{i=1}^{g} b_1(i)x_{j1}(i) + b_2x_{j2}, \quad (1.18)$$

Intercept		Male	Female	
-2.315 (0.065)		0 (0)	0.037 (0.027)	
		Age groups (years)		
20-25	26-39	40-55	56-70	71-94
0 (0)	-0.501 (0.068)	-0.541 (0.067)	-0.711 (0.070)	-0.637 (0.073)

Table 8.3 *Poisson regression fitted on the Norwegian automobile portfolio (standard deviations in paranthesis)*

where for an individual j in age group l

$$x_{j1}(i) = 1, \quad \text{if } i = l$$
$$x_{j1}(i) = 0, \quad \text{if } i \neq l.$$

The age component is now represented by g different *binary* variables $x_{j1}(1), \ldots, x_{j1}(g)$. For a given policy holder *exactly one* of them is equal to one; the rest are zero.

Does this look odd? The number of unknown parameters has gone up and more historical data might therefore be needed. But there are advantages too. The relationship between age and risk can be made more flexible (and hence truer), and the partition into groups avoids excessively many different prices. The second variable (sex) is on categorical form by definition, and the model can be written

$$\log(\mu_j) = b_0 + b_1(l) + b_2(s), \qquad \text{for policy holder } j \text{ of age } l \text{ and sex } s. \tag{1.19}$$

Superficially there are now two parameters for the second variable, but that is an illusion. With g categories there are only $g-1$ free parameters (they would be confounded with the intercept b_0 otherwise). Standard in much sofware is to zero the first one, i.e to take $b_1(1) = 0$ and $b_2(1) = 0$, as in Table 8.3.

The multiplicative form

Specifications of the form (1.19) are **multiplicative** and on the original scale reads

$$\mu_j = \underset{baseline}{\mu_0} \cdot \underset{age}{e^{b_1(l)}} \cdot \underset{sex}{e^{b_2(s)}}. \qquad \text{where} \qquad \mu_0 = e^{b_0},$$

The baseline intensity μ_0 (here for men of the youngest group) are modified by two factors working independently of each other. This is a common (and useful!) approach, but there are also pitfalls. From the estimates in Table 8.3 we obtain the annual claim rates Table 8.4. The results follow those in Table 8.1 with the notable exception that the much higher accident rate for men in the youngest group now has disappeared!

Why is that? The reason is the multiplicative form which fixes the intensity ratio between males and females as a constant, the same at any age. This washes away the *fine structure* in Table 8.1 and a feature of some importance has been lost. Yet muliplicative model specifications (or equivalently additions on logarithmic scale) *is* of high interest. Simplifications that cut down on the number of parameters and groups are often necessary, but it must be done with care, and larger problems are harder to read than this simple example. There is more on Poisson regression in Section 10.4; see also Section 11.4.

Intensities in % annually

	Age groups (years)				
	20-25	26-39	40-55	56-70	71-94
Male	9.9	5.9	5.8	4.9	5.2
Female	10.3	6.1	6.1	5.1	5.4

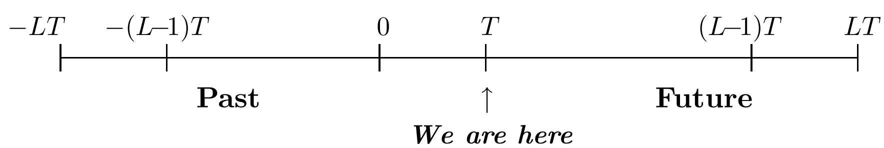
Table 8.4 *Annual claim intensities under Poission regression*

1.5 Modelling delay

Introduction

Claims are never settled immediately, and for some types of injuries or damages the delay is rather long. A typical case is back or neck ailments following a car crash; it may take years before their symptoms arrive. Other examples originate with environmental factors like asbesto or radioactivity. Their severe effect on health was not understood at the time, and the repercussions for the insurance industry have been dire indeed. Claims that are only discovered years after the insurable event are called **IBNR** (**i**ncurred, **b**ut **n**ot **r**eported). Companies have to set aside funds to cover them (even though they may not yet have been identified). Claims may also be of **RBNS** status (**r**eported, **b**ut **n**ot **s**ettled) which means that they are in the process of being liquidated, but not finished.

We shall concentrate on IBNR claims which require a longer look than the ordinary period of accountacy T. The situation is as follows:



At the end of year zero (where we are at the moment) there are outstanding claims that will surface later. Companies are responsible. Their balance sheets will be affected up to L periods ahead by claims not yet known, and must look back equally long to include all contributions. A natural approach (though currently not the most popular one) is to see delay as a random phenomenon. Let q_l be the probability that it takes l periods to liquidate the claim from the time the incident occurred. IBNR modelling is in this section developed through these quantities; for other approaches see Section 8.7.

Multinomial delays

We are assuming that the time to settlement of a IBNR claim is random with distribution $q_0, \dots, q_L$. Here $q_0 + \dots + q_L = 1$ and L the maximum delay; i.e. all claims have been finished L periods after the incident took place. Different events are reasonably seen as independent. We are then dealing with **multinomial** processes, and if $\mathcal{N}_l$ is the number of claims liquidated l periods late, then $(\mathcal{N}_0, \dots, \mathcal{N}_L)$ is a multinomially distributed random vector. More precisely, let $\mathcal{N}$ be the total number of claims from a period of length T with J policies under risk. Then

(i) $\mathcal{N}_0, \dots, \mathcal{N}_L | \mathcal{N} = n$ is multinomial$(n, q_0, \dots, q_L)$
(ii) $\mathcal{N} \sim$ Poission(λ) where $\lambda = J\mu T$.

It follows from this that

$$\mathcal{N}_l \sim \text{Poisson}(q_l J\mu T) \qquad \text{with} \qquad \mathcal{N}_0, \dots, \mathcal{N}_L \quad \text{stochastically independent;} \tag{1.20}$$

see Section 8.6 for a formal verification. An intuitive point process argument runs as follows. Divide the period T under risk into K intervals of length $h = T/K$ as in Section 8.2. The probability of an incident against the portfolio is $J\mu h$ per interval, but only a fraction q_l affects $\mathcal{N}_l$. Hence $q_l J\mu T$ is the intensity of the point process underlying $\mathcal{N}_l$ which becomes Poisson distributed as stated. There are $L+1$ such point processes (one for each delay l). They are dependent, but only to the of order $o(h)$ which vanishes as $h \to 0$ so that $\mathcal{N}_0, \dots, \mathcal{N}_L$ become independent.

IBNR claim numbers

We are dealing with a sequence of periods (indexed $-s$) from year zero and back; see the scheme above. Claims from that period haven't necessarily come to light at the end of year zero. Let J_{-s} be the number of policies at risk in year $-s$ (known), μ_{-s} the average claim intensity during the same year (unknown) and let $\mathcal{N}_{-s,l}$ be the number of claims originating that year and settled l years later. By (1.20) the entire set of $\{\mathcal{N}_{-s,l}\}$ are stochastically independent with

$$\mathcal{N}_{-s,l} \sim \text{Poisson}(\lambda_{-s,l}) \qquad \text{where} \qquad \lambda_{-s,l} = J_{-s} q_l \mu_{-s}. \tag{1.21}$$

What affects our balance sheet k years ahead is the total number of claims disposed of that year; i.e.

$$\mathcal{N}_k = \sum_{s=0}^{L-k} \mathcal{N}_{-s,k+s}. \tag{1.22}$$

As a sum of independent Poisson counts these quantities become Poisson themselves, the parameter of $\mathcal{N}_k$ being

$$\lambda_k = \sum_{s=0}^{L-k} \lambda_{-s,k+s} = \sum_{s=0}^{L-k} q_{-s+k} J_{-s} \mu_{-s}, \tag{1.23}$$

and, moreover, all of $\mathcal{N}_1, \dots, \mathcal{N}_L$ are stochastically independent. This useful observation permits us to use the same methods for IBNR as for ordinary reserving.

Fitting delay models

A convenient way to fit model with delay probabiltities is to utilize that $\lambda_{-s,l}$ in (1.21) is on multiplicative form so that

$$\log(\lambda_{-s,l}) = \log(J_{-s}) + \alpha_l + \beta_s, \tag{1.24}$$

where

$$\alpha_l = \log(q_l) \qquad \text{and} \qquad \beta_s = \log(\mu_{-s}). \tag{1.25}$$

This is a Poission log-linear regression of the same type as in Section 8.4. There are now *two* indexing variables l and s (known in statistics as a **two-way design**), but it can still be handled by ordinary Poisson regression software if special programs are unavailable. The data matrix $\{n_{-s,l}\}$ must then be concatenated into a single vector; see Section 8.6 for the details.

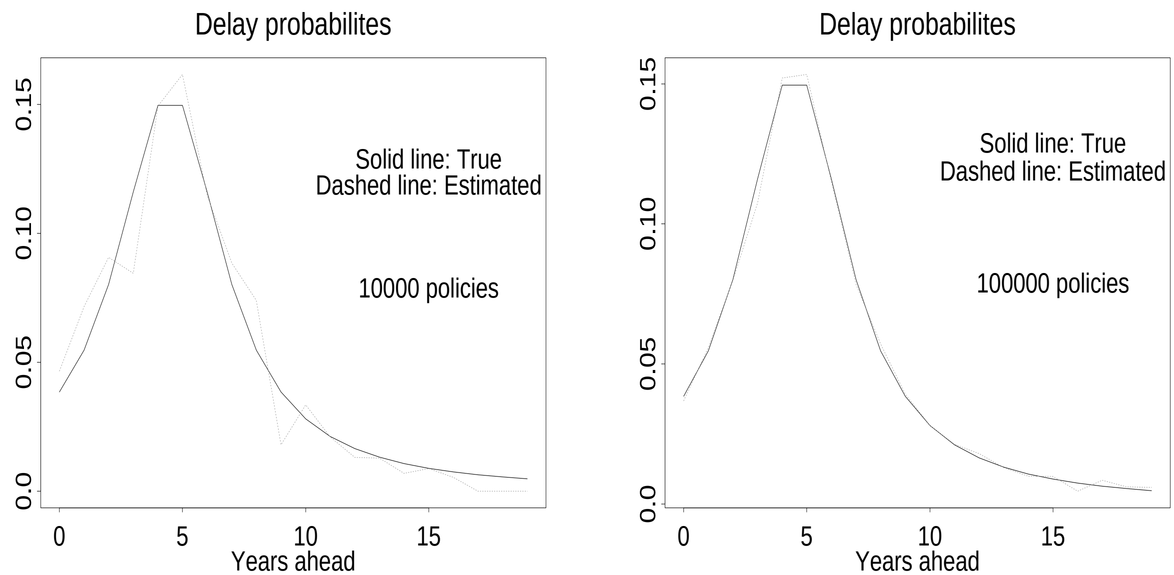


Figure 8.3 *Estimates of delay probabilites under circumstances described in the text*

The estimates $\hat{\alpha}_l$ and $\hat{\beta}_s$ so obtained are likely to be wrongly scaled; i.e. when they are converted to the orginal parameters we must ensure that $\{\hat{q}_l\}$ is a probability distribution adding to one. That is achieved by taking

$$\hat{q}_l = \exp(\hat{\alpha}_l)/C, \quad \hat{\mu}_{-s} = C\hat{\beta}_s, \qquad \text{where} \qquad C = \exp(\hat{\alpha}_0) + \ldots + \exp(\hat{\alpha}_K). \tag{1.26}$$

The resulting estimates are likelihood ones.

Syntetic example: Car crash injury

The example shown in Figures 8.3 and 8.4 is patterned on a real automobile portfolio of an Scandinavian insurance company. We are considering personal injuries following road accidents. A typical claim rate could be around 0.5% annually. The true model generated in the computer is based on the annual frequency

$$\mu = \xi G, \quad G \sim \text{Gamma}(\alpha) \qquad \text{where} \qquad \xi = 0.5\%, \quad \alpha = 7.85.$$

Note that the true frequency of personal injuries varies randomly from one year to another in a manner reflecting the automobile portfolio examined earlier in this chapter. The delay probabilities were

$$q_l = C\exp(-\beta|l - l_0|), \quad l = 0, \ldots, L,$$

where the constant C ensures that $q_0 + \ldots q_L = 1$. Parameters were

$$L = 20, \quad \beta = 0.15, \quad l_0 = 4.5, \quad K = 20$$

which means that the distribution $\{q_l\}$ has a top after four and five years ; see Figure 8.3. All claims have been settled after twenty years.

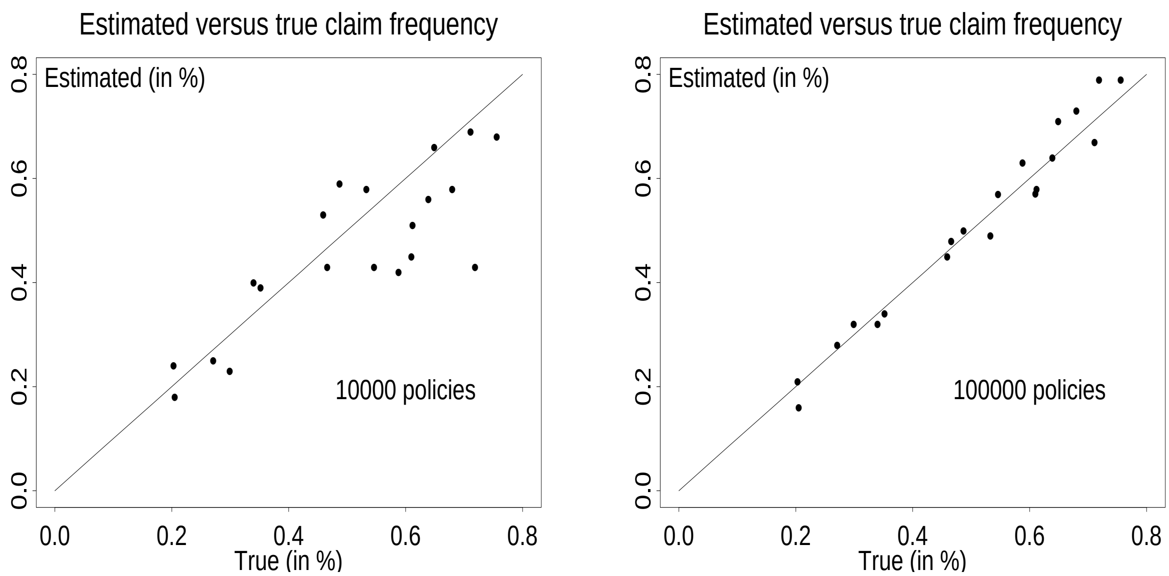


Figure 8.4 *Estimates of claim frequencies under circumstances described in the text*

Historical claim data were created for portfolios of $J = 10000$ and $J = 100000$ policies by means of the following commands:

Algorithm 8.2 Simulating IBNR counts
0 Input: J,$\{q_l\}$, ξ, α
1 For $s = 0, 1 \dots, L-1$ do
2 Draw $Z^* \sim$ Gamma(α) and $\mu^*_{-s} \leftarrow \xi Z^*$
3 For $s = 0, 1 \dots, L$ do
Draw $\mathcal{N}^*_{-sl} \sim$ Poisson($Jq_l\mu^*_{-s}$)
4 For all s and l return $\mathcal{N}^*_{-sl}$

The collection $\{\mathcal{N}^*_{-s,l}\}$ (one single round of simulations) were used as historical material and parameter estimates extracted from them.

Figures 8.3 and 8.4 suggest that delay probabilites $\{q_l\}$ and actual claim intensities $\{\mu^*_{-s}\}$ can be reconstructed. The pattern in the true delay probabiltites are certainely picked up (Figure 8.3), and the match between the true and fitted claim intensitites (Figure 8.4) is a good one.

1.6 Mathematical arguments

Section 8.2
Deriving the Poisson distribution *Preliminary* The general form of the Poisson point process assumes stochastically independent variables $I_1, \dots, I_K$ under the the model

$$\Pr(I_k = 0) = 1 - \mu h + o(h), \qquad \Pr(I_k = 1) = \mu h + o(h), \qquad \Pr(I_k > 1) = o(h).$$

where $o(h)/h \to 0$ as $h \to 0$. We are to prove that $N = I_1 + \ldots + I_K$ becomes Poisson in the limit. Introduce the events

$$A = \max_{1 \le k \le K} I_k \le 1 \qquad \text{and} \qquad A^c = \max_{1 \le k \le K} I_k > 1$$

and note that

$$\Pr(A^c) = \Pr(I_1 > 1 \text{ or} \ldots \text{ or } I_K > 1) \le \sum_{k=1}^{K} \Pr(I_k > 1) = Ko(h) = T\frac{o(h)}{h} \to 0.$$

Moreover,

$$\Pr(N = n) = \Pr(N = n|A)\Pr(A) + \Pr(N = n|A^c)\Pr(A^c)$$

and therefore

$$|\Pr(N = n) - \Pr(N = n|A)| = |\Pr(N = n|A) - \Pr(N = n|A^c|\Pr(A^c) \le 2\Pr(A^c) \to 0$$

so that $\Pr(N = n)$ and $\Pr(N = n|A)$ have the same limit.

Main argument It follows that the density function of N may be studied given A. All I_k are then zero or one so that $N|A$ is binomial with 'success' probability

$$p = \frac{\Pr(I_k = 1)}{\Pr(I_k = 0) + \Pr(I_k = 1)} = \frac{\mu h + o(h)}{1 - \mu h + o(h) + \mu h + o(h)} = \mu h + o(h) = \frac{\mu T}{K} + o(1/K)$$

which must be inserted the factorization used in Section 8.2. The details are almost the same; i.e.

$$\Pr(N = n|A) = B_1 \cdot B_2 \cdot B_3 \cdot B_4$$

where

$$B_1 = \frac{\{\mu T + Ko(1/K))\}^n}{n!}, \qquad B_2 = \frac{K(K-1)\cdots(K-n+1)}{K^n},$$
$$B_3 = \{1 - \mu T/K + o(1/K)\}^K, \qquad B_4 = \frac{1}{\{1 - \mu T/K + o(1/K)\}^n}.$$

Limits of the these quantities are the same as before, and only B_3 requires an argument. Utilize that $\log(1 + x) = x - x^2/2 + \ldots$, taking $x = -\mu T/K + o(1/K)$. This yields

$$\log(B_3) = K \log\{1 - \mu T/K + o(1/K)\} = -\mu T + Ko(1/K) \to -\mu T$$

and $B_3 \to \exp(-\mu T)$ as in Section 8.2.

Section 8.3

The moment estimates of $\boldsymbol{\xi}$ and $\boldsymbol{\sigma}$ *The individual estimates.* Let $\hat{\mu} = N_j/T_j$. Then

$$E(\hat{\mu}_j|\mu_j) = \mu_j \qquad \text{and} \qquad \text{var}(\hat{\mu}_j|\mu_j) = \frac{\mu_j}{T_j},$$

and the double ruls for mean and variance yield

$$E(\hat{\mu}_j) = E(\mu_j) = \xi \qquad \text{and that} \qquad \text{var}(\hat{\mu}_j) = \text{var}(\mu_j) + E\left(\frac{\mu_j}{T}\right) = \sigma^2 + \frac{\xi}{T_j}.$$

Estimate of ξ It was suggested in Section 8.3 that ξ can be estimated through

$$\hat{\xi} = w_1\hat{\mu}_1 + \ldots + w_n\hat{\mu}_n \qquad \text{where} \qquad w_j = \frac{T_j}{T_1 + \ldots + T_n},$$

which is (1.8). Recall that $w_1 + \ldots + w_n = 1$. Hence

$$E(\hat{\xi}) = E\left(\sum_{j=1}^{n} w_j\hat{\mu}_j\right) = \sum_{j=1}^{n} w_j E(\hat{\mu}_j) = \sum_{j=1}^{n} w_j\xi = \xi,$$

and $\hat{\xi}$ is unbiased. Moreover, since $\hat{\mu}_1, \ldots, \hat{\mu}_n$ are independent

$$\text{var}(\hat{\xi}) = \sum_{j=1}^{n} w_j^2 \text{var}(\hat{\mu}_j) = \sum_{j=1}^{n} w_j^2(\sigma^2 + \frac{\xi}{T_j}) = \sigma^2(w_1^2 + \ldots + w_n^2) + \frac{\xi}{T_1 + \ldots + T_n}.$$

This tends to zero as $T_1 + \ldots + T_n \to \infty$ if at the same time all $T_j \leq M$. Indeed

$$w_1^2 + \ldots + w_n^2 = \frac{T_1^2 + \ldots + T_n^2}{(T_1 + \ldots + T_n)^2} \leq \frac{MT_1 + \ldots + MT_n}{(T_1 + \ldots + T_n)^2} = \frac{M}{T_1 + \ldots + T_n} \to 0,$$

which covers the first term of the variance formula. The second term is obvious.

Estimate for σ Consider

$$Q = \sum_{j=1}^{n} w_j(\hat{\mu}_j - \hat{\xi})^2 = \sum_{j=1}^{n} w_j\hat{\mu}_j^2 - \hat{\xi}^2 \qquad \text{which yields} \qquad E(Q) = \sum_{j=1}^{n} w_j E(\hat{\mu}_j^2) - E(\hat{\xi}^2).$$

Here

$$E(\hat{\mu}_j^2) = \text{var}(\hat{\mu}_j) + (E\hat{\mu}_j)^2 = \sigma^2 + \frac{\xi}{T_j} + \xi^2$$

and

$$E(\hat{\xi}^2) = \text{var}(\hat{\xi}) + (E\hat{\xi})^2 = \sigma^2 \sum_{j=1}^{n} w_j^2 + \frac{\xi}{\sum_{j=1}^{n} T_j} + \xi^2.$$

Inserting this yields

$$E(Q) = \sum_{j=1}^{n} w_j\left(\sigma^2 + \frac{\xi}{T_j} + \xi^2\right) - \left(\sigma^2\sum_{j=1}^{n} w_j^2 + \frac{\xi}{\sum_{j=1}^{n} T_j} + \xi^2\right),$$

which simplifies to

$$E(Q) = \sigma^2(1 - \sum_{j=1}^{n} w_j^2) + \frac{(n-1)\xi}{\sum_{j=1}^{n} T_j}.$$

A moment estimate for σ is the solution of the equation

$$Q = \hat{\sigma}^2\left(1 - \sum_{j=1}^{n} w_j^2\right) + C \qquad \text{where} \qquad C = \frac{(n-1)\hat{\xi}}{\sum_j T_j},$$

which is (1.9). The argument also shows that the estimate must be unbiased. The variance is complicated and is omitted. It tends to zero under the same conditions as for the mean.

The negative binomial*Density function* The negative binomial distribution is defined as

$$\Pr(N = n|\mu) = \frac{(\mu T)^n}{n!} e^{-\mu T} \qquad \text{and} \qquad g(\mu) = \frac{(\alpha/\xi)^\alpha}{\Gamma(\alpha)} \mu^{\alpha-1} e^{-\mu\alpha/\xi}.$$

By the mixing formula (1.7)

$$\Pr(N = n) = \int_0^\infty \frac{(T\mu)^n}{n!} e^{-T\mu} \times \frac{(\alpha/\xi)^\alpha}{\Gamma(\alpha)} \mu^{\alpha-1} e^{-\mu\alpha/\xi}\, d\mu$$

or when reorganized,

$$\Pr(N = n) = \frac{T^n (\alpha/\xi)^\alpha}{n!\Gamma(\alpha)} \int_0^\infty \mu^{n+\alpha-1} e^{-\mu(T+\alpha/\xi)}\, d\mu.$$

Substituting $z = \mu(T + \alpha/\xi)$ in the integrand yields

$$\Pr(N = n) = \frac{T^n (\alpha/\xi)^\alpha}{n!\Gamma(\alpha)(T + \alpha/\xi)^{n+\alpha}} \int_0^\infty z^{n+\alpha-1} e^{-z}\, dz,$$

where the integral is $\Gamma(n + \alpha)$. Hence

$$\Pr(N = n) = \frac{\Gamma(n + \alpha)}{n!\Gamma(\alpha)} \cdot \frac{T^n (\alpha/\xi)^\alpha}{(T + \alpha/\xi)^{n+\alpha}} = \frac{\Gamma(n + \alpha)}{n!\Gamma(\alpha)} p^\alpha (1 - p)^n$$

where $p = \alpha/(\alpha + \xi T)$. This is the expression (1.12) since $\Gamma(n + 1) = n!$.

Convolution Let $N_1, \ldots, N_J$ be independent and indentically distributed with $N_j \sim \text{nbin}(\xi, \alpha)$. Each N_j can then be represented as

$$N_j|\mu_j \sim \text{Poisson}(\mu_j T) \qquad \text{where} \qquad \mu_j = \xi Z_j, \quad Z_j \sim \text{Gamma}(\alpha).$$

Consider the conditional distribution of $\mathcal{N} = N_1 + \ldots + N_J$ given $\mu_1, \ldots, \mu_J$. This is Poisson with parameter $\zeta = \mu_1 + \ldots + \mu_J$ which can be represented as

$$\zeta = \mu_1 + \ldots + \mu_J = \xi(Z_1 + \ldots + Z_j) = J\xi\bar{Z} \qquad \text{where} \qquad \bar{Z} = (Z_1 + \ldots + Z_J)/J.$$

Here $\bar{Z}$ is the average of J variables that are Gamma(α) and is therefore Gamma($J\alpha$) itself; see Section 9.3, and it has been established that

$$\mathcal{N}|\zeta \sim \text{Poisson}(\zeta T) \qquad \text{where} \qquad \zeta = (J\xi)\bar{Z}, \quad \bar{Z} \sim \text{Gamma}(J\alpha),$$

and is nbin($J\xi, J\alpha$) by definition.

Section 8.5

IBNR: The delay model Let $\mathcal{N}$ be the total number of claims that arise in a given period (of length T) and $\mathcal{N}_l$ those among them settled l periods later. Then $\mathcal{N} = \mathcal{N}_0 + \ldots + \mathcal{N}_L$ and with $n = n_0 + \ldots + n_L$

$$\Pr(\mathcal{N}_0 = n_0, \ldots, \mathcal{N}_L = n_L) = \Pr(\mathcal{N}_0 = n_0, \ldots, \mathcal{N}_L = n_L|\mathcal{N} = n)\Pr(\mathcal{N} = n).$$

The first factor is a multinomial probability, i.e

$$\Pr(\mathcal{N}_0 = n_0, \ldots, \mathcal{N}_L = n_L | \mathcal{N} = n) = \frac{n!}{n_0! \cdots n_L!}\; q_0^{n_0} \cdots q_L^{n_L}$$

whereas the second one is the Poisson probability

$$\Pr(\mathcal{N} = n) = \frac{\lambda^n}{n!} e^{-\lambda} = \frac{\lambda^{n_0 + \ldots + n_L}}{n!} e^{-\lambda(q_0 + \ldots + q_L)} = \frac{1}{n!} \left(\lambda^{n_0} e^{-q_0 \lambda}\right) \cdots \left(\lambda^{n_L} e^{-q_L \lambda}\right)$$

since $q_0 + \ldots + q_L = 1$. Combining these probabiltities yields

$$\Pr(\mathcal{N}_0 = n_0, \ldots, \mathcal{N}_L = n_L) = \frac{q_0^{n_0} \cdots q_L^{n_L}}{n_0! \cdots n_L!}\; \left(\lambda^{n_0} e^{-q_0 \lambda}\right) \cdots \left(\lambda^{n_L} e^{-q_L \lambda}\right) = \prod_{l=0}^{L} \frac{(q_l \lambda)^{n_l}}{n_l!} e^{-q_l \lambda}$$

as claimed in (1.20).

IBNR: Implementation You may have trouble hitting an implementation of Poisson regression that handles the special two-way structure in Section 8.6. Here is how it is implemented by means of standard software.

The parameters are $\alpha_l = \log(q_l)$ and $\beta_s = \log(\mu_{-s})$ which yields the column vector

$$\mathbf{b} = (\alpha_0, \ldots, \alpha_{K-1}, \beta_0, \ldots, \beta_n)' \qquad \text{of length} \qquad n_c = K + n$$

where the historical record extends n years back. Let $n_{-s,l}$ be the observed count from year $-s$ that took l years to settle and concanate them in any order you like into a single vector $\mathbf{y}$. The length of this vector is

$$n_r = \sum_{s=0}^{n} \{1 + \min(s, K)\}$$

since l runs from 0 to the minimum of s and K. This equals

$$\begin{aligned} n_r &= (n+1)(n+2)/2, && \text{if } n \le K \\ &= n + 1 + K(n - (K-1)/2), && \text{if } n > K. \end{aligned}$$

We need a $(n_r \times n_c)$ design matrix $\mathbf{X} = (x_{ij})$ where the row i runs over all $(-s, l)$ and for which $x_{ij} = 1$ if $i = (-s, l)$ and $j = l$ or $j = L + 1 + s$ and $x_{ij} = 0$ otherwise. The order of the rows must match that of the vector $\mathbf{y}$. If $\boldsymbol{\lambda}$ is the vector of the expdectations of $\mathbf{y}$, then $\boldsymbol{\lambda} = \mathbf{X}\mathbf{b}$, and $\mathbf{b}$ is estimated by a Poisson regression program. It might in practice well be that $n < K$. If so, it is impossible to estimate the entire sequence $\{q_l\}$ and simplifying formulations must be introduced.

1.7 Bibliographic notes

Poisson processes with or without random intensities is the usual mathematical model construction for of claim numbers, and there are presentations in standard textbooks on general insurance, for example in Daykin, Pentikäinen and Pesonen, M. (1994) and Mikosch (2004); see also Grandell (2004) and Jacobsen (2006) for more advanced treatments. Extended modelling with claim intensities as stochastic process is discussed in Chapter 11. Poisson regression in Section 8.4 is a special case of the class of **generalized linear models** where expectation is linked to explanatory variables for many of the most common distributions. These techniques are also relevant for claim size (Section 9.3) and in life insurance. How they are put to use has nuch in common with the account in Section 8.4. The classical, general reference on generalized linear models is McCullagh and Nelder (1989), but you may well find Dobson and Barnett (2008) or perhaps Dunteman and Ho (2006) easier to read. Lee, Nelder, and Pawitan (2006) allows random coefficients which is of interest in insurance, see Wütrich and Merz (2008). An extension in another direction is Wood (2006). Reviews in an actuarial context have been provided by Haberman and Renshaw (1996) and Antiono and Beirlant (2007). For specific applications covering a wide range of topics in insurance consult Yau, Yip and Yuen (2003), Butt and Haberman (2004), Hoedemakers, Beirlant, Goovaerts and Dhaene (2005) and England and Verall (2007).

The delay model in Section 8.5 goes back at least to Kaminsky (1987). It makes reserving for claims that surface late much the same problem as for ordinary claims. Other methods have (perhaps surprisingly) been more popular. Taylor (2000) and Wütrich and Merz (2008) are thorough reviews, containing much more advanced and complex modelling, for example of a type related to the stochastic intensity processes in Section 11.3.

Antiono, K. and Beirlant, J. (2007). Actuarial Statistics with Generalized Linear Mixed Models. *Insurance: Mathematics and Economics*,40, 58-76.
Boland, P.J. (2007). *Statistical and Probabilistic Methods in Actuarial Science.* John Wiley & Sons, Hoboken, New Jersey.
Butt, Z. and Haberman, S. (2004). Application of Frailty-Based Mortality Models Using Generalized Linear Models. *Astin Bulletin*, 34, 175-197.
Daykin, C.D., Pentikäinen, T. and Pesonen, M. (1994). *Practical Risk theory for Actuaries.* Chapman & Hall/CRC, London.
Dobson, A.J. and Barnett, A.G. (2008). *An Introduction to Generalized Linear Models.* Chapman & Hall, Boca Raton, Florida.
Dunteman, G.H. and Ho, M-H., R. (2006). *An Introduction to Generalized Linear Models.* Sage Publications, Thousand Oaks, California.
England, P.D. and Verall, R.J. (2007). Predictive Distributions of Outstanding Liabilities in General Insurance. *Annals of Actuarial Science*, 1/2, 221-270.
Grandell, J. (2004). Poisson Processes. In *Encyclopedia of Actuarial Science*, Teugels, J, and Sundt, B. (eds), John Wiley & Sons, Chichester, 1296-1301.
Haberman, S. and Renshaw, A.E. (1996). Generalized Linear Models and Actuarial Science. *The Statistician*, 45, 407-436.
Hoedemakers, T, Beirlant, J., Goovaerts, M.J. and Dhaene, J. (2005). On the Distibution of Discounted Loss Reserves Using Generalized Linear Models. *Scandinavian Actuarial Journal*, 1, 25-45.
Jacobsen, M. (2006). *Point Process Theory and Applications.* Birkhäuser, Boston.
Kaminsky, K. (1987). Prediction of IBNR claim counts by modeling the distribution of report lags. *Insurance: Mathematics and Economics*, 6, 151-159.
Lee, Y., Nelder, J.A. and Pawitan, Y. (2006). *Generalized Linear Models with Random Effects: Unified Analysis via H-Likelihood.* Chapman & Hall, BoJacobsen, M. (2006). *Point Process Theory and Applications.* Birkhäuser, Boston.ca Raton, Florida.
Lehmann, E. and Casella, G. (1998), second ed. *Theory of Point estimation.* Springer Verlag, New York.

McCullagh, P., and Nelder, J.A. (1989), second ed. *Generalized Linear Models.* Chapman & Hall, London.
Press, W.H:, Teukolsky, S.A, Vetterling, W.T and Flannery, B. (2007). Third ed. *Numerical Recipes in C.* Cambridge University Press, Cambridge.
Taylor, G. (2000). *Claim Reserving: an Actuarial Perspective.* Elsvier, New York.
Wood, S.N. (2006). *Generalized Additive Models: an Introduction with R.* Chapman & Hall, Boca Raton, Florida
Wütrich, M.V. and Merz, M. (2008). *Stochastic Claim Reserving Methods in Insurance.* John Wiley & Sons, Chichester.
Yau, K.K.W., Yip, K.C.H. and Yuen, H.K. (2003). Modelling Repeated Insurance Claim Data Using the Generalized Linear Mixed Model. *Journal of Applied Statistics*, 30, 857-865.

1.8 Exercises

Section 8.2

Exercise 8.2.1 Let N_t be the number of events up to time t in a Poisson point process with constant intensity μ and let X be the time the *first* event appears. **a)** Explain that

$$\Pr(X > t) = \Pr(N_t = 0) = \exp(-\mu t).$$

b) Identify the distribution of X. What is its expectation?

Exercise 8.2.2 The random variable X defined in the preceding exercise is known as a **waiting time**. Let $X_1 = X$ be the time to the first Poisson event, and more generally let X_i be the time between events $i-1$ and i. Suppose that the claim intensity μ is constant. **a)** Explain that $X_1, X_2, \dots$ are independent and exponentially distributed. Let $S_i = X_1 + \dots + X_i$ be the time of the i'th event with $S_0 = 0$. **b)** Argue that

$$\Pr(S_i \le t < S_{i+1}) = \Pr(N_t = i)$$

where N_t is the number of events up to t. **c)** Explain that this result leads to Algorithm 2.10; i.e. that N_t can be sampled by selecting it is as the largest integer i^* for which $X_1^* + \dots + X_{i^*}^* \le t$.

Exercise 8.2.3 Consider an insurance portfolio with J policies that all generate claims according to a Poisson point process with fixed intensity μ. **a)** Use the preceding exercise to explain how you simulate the time S_i of the i'th claim of the portfolio. Let $\mu = 0.5\%$ annually and $J = 400$. **b)** Simulate the pair (S_1, S_5) and display $m = 1000$ replications in a scatter plot. **c)** Use the scatter plot to argue that these expenses would vary enormously from one year to another if we are dealing with big-claim insurance with possible huge pay-offs each time an incident occurs.

Exercise 8.2.4 Consider a time-hetereogenous Poisson point process and let N_t be the number of events up to t. **a)** With X as the waiting time as in Exercise 8.2.2 argue as in Exercise 8.2.1 to deduce that

$$\Pr(X > t) = \Pr(N_t = 0) = \exp\left(-\int_0^t \mu_s\, ds\right).$$

Let $F(x)$ and $f(x) = F'(x)$ be the distribution and density function of X **b)** By differentiating $\Pr(X > x)$ show that

$$f(x) = \mu_x \exp\left(-\int_0^x \mu_s\, ds\right) \qquad \text{which implies that} \qquad \mu_x = \frac{f(x)}{1 - F(x)}.$$

These relationships are also relevant with survival modelling in Section 12.3.

Exercise 8.2.5 This is a continuation of the preceding exercise. **a)** Use inversion (Algorithm 2.6) to show that a Monte Carlo realisation of the waiting time X is generated by solving the equation

$$\int_0^{X^*} \mu_s \, ds = -\log(U^*) \qquad \text{for} \qquad U^* \sim \text{uniform}.$$

Suppose $\mu_s = \mu_0 \exp(\gamma s)$ where $\gamma \neq 0$ is a growth (or decline) parameter. **b)** Show that a Monte Carlo realization of the waiting time is generated by taking

$$X^* = \frac{1}{\gamma} \log\left(1 - \frac{\gamma}{\mu_0} \log(U^*)\right)$$

c) Explain how you sample the time S_i of the i'th event in this Poisson growth process.

Exercise 8.2.6 This exercise treats the same intensity function $\mu_s = \mu_0 \exp(\gamma s)$ as in the preceding exercise, but now we introduce N_k as the number of incidents between t_{k-1} and t_k where $t_k = kh$. **a)** Identify the distribution of N_k. **b)** Explain how the sequence $N_1, N_2, \ldots$ are simulated. In practice this would often be a more natural approach than simulating waiting times.

Exercise 8.2.7 Suppose the premium (paid up-front) for an insurance lasting up to time T is

$$\pi = (1+\gamma)\xi \int_0^T \mu_s \, ds,$$

where γ is the loading and ξ is mean payment per claim. If the insured leaves the contract at time $T_1 < T$, how much would it be fair to repay out of the original premium?

Exercise 8.2.8 Let $N_1, N_2 \ldots$ be stochastically independent and Poisson distributed with common parameter η. **a)** Argue that $X = N_1 + \ldots + N_K$ is Poisson with parameter $\lambda = K\eta$. **b)** Why does X tend to the normal distribution as $K \to \infty$? **c)** Use a) and b) to deduce that the Poisson distribution with parameter λ becomes normal as $\lambda \to \infty$.

Exercise 8.2.9 Consider the estimate $\hat{\mu} = (n_1 + \ldots + n_n)/(T_1 + \ldots + T_n)$ in (1.4) where $n_1, \ldots n_n$ are stochastically independent observations, the $j'th$ being Poisson distributed with parameter μT_j. **a)** Show that $\hat{\mu}$ is unbiased with standard deviation (1.4). Suppose the intensity for n_j is μ_j, depending on j. (**b)** Recalculate $E(\hat{\mu})$ and argue that the estimate $\hat{\mu}$ has little practical interest if the times of exposure $T_1, \ldots, T_n$ vary a lot. **c)** However, suppose you arrange things so that $T_1 = \ldots = T_n$. Which parameter of practical interest does $\hat{\mu}$ estimate now?

Exercise 8.2.10 A classical set of historical data is due to Greenwood (1927) and shows accidents among 648 female munition workers in Britain during The First World War (the men were at war!). Many among them experienced more than one accident during a period of almost one year. The data were as follows:

Number of accidents	0	1	2	3	4	≥ 5
Number of cases	448	132	42	21	3	2,

which shows that 448 hadn't had any accidents, 132 had have one, 42 two and so on. **a)** Compute the mean and standard deviation $\bar{n}$ and s of the 648 observations. **b)** Argue that $\hat{\lambda} = \bar{n}$ is a natural estimate of λ if we postulate that the number of accidents is Poisson distributed with parameter λ. [Hint: Answer: $\hat{\lambda} = 0.465$]. **c)** Compute the **coefficient of dispersion** $D = s^2/\bar{n}$. What kind of values would we expect to find if the Poisson model is true? **d)** Simulate $n = 648$ Poisson variables from the parameter $\lambda = 0.465$ and compute D from the simulations. Repeat 9 (or 99) times. Does the Poisson distribution look like a plausible one? This story is followed up in Exercise 8.5.4.

Exercise 8.2.11 a) Implement Algorithm 8.1 due to Atkinson in a way that you can keep track on how many repetitions are needed in the outer loop. **b)** Run it $m = 1000$ times for $\lambda = 10$ and $\lambda = 50$ and determine the average number of trials required.

Section 8.3
Exercise 8.3.1 let $\mathcal{N}$ be the total number of claims against a portfolio of J policies with identical claim intensity μ, and suppose $\mathcal{N}|\mu$ is Poisson with parameter $J\mu T$. **a)** What is the distribution of $\mathcal{N}$ if $\mu = \xi Z$, $Z \sim$ Gamma(α)? **b)** Determine the mean and standard deviation of $\mathcal{N}$. **c)** Calculate $\text{sd}(\mathcal{N})/E(\mathcal{N})$ and comment on how it depends on J. What happens as $J \to \infty$?

Exercise 8.3.2 This is a continuation of Exercise 8.3.3. A reasonable addition to the Poisson model presented there is to assume that the parameter λ is drawn randomly and independently for each woman. Each of the the $n = 648$ observed counts N is then Poisson distributed given λ, and a distribution must be supplied for λ. Assume that $\lambda = \xi Z$ and Z is Gamma(α). **a)** What is the interpretation of the parameters ξ and α for the situation described in Exercise 8.3.3? **b)** What's the distribution of N now? **c)** Use the moment method to fit it [Answer: You get $\hat{\xi} = 0.465$ and $\hat{\alpha} = 0.976$]. One way to investigate the fit is to compare $E_n = 648 \times \Pr(N = n)$ (*expected* number) with the *observed* number O_n having had n accidents; see the table in Exercise 8.3.3. **d)** For $n = 0, 1, 2, 3, 4$ and $n \geq 5$ compare E_n and O_n, both for the pure Poisson model and for the present extension of it. Comments? We should go for the negative binomial! **e)** What is the likelihood that a given female worker carries more than twice as much risk as the average? What about four times a much or at most one one half or one fourth? [Hint: To answer the questions use (for example) the exponential distribution.]

Exercise 8.3.3 Suppose $N|\mu$ is Poisson with parameter μT with two models under consideration for μ. Either take $\mu = \xi Z$ with $Z \sim$ Gamma(α) or $\mu = \beta \exp(\tau\varepsilon)$ with $\varepsilon \sim N(0,1)$. **a)** Show that $E(\mu)$ and $\text{sd}(\mu)$ is the same under both models if

$$\beta = \frac{\xi}{\sqrt{1 + 1/\alpha}} \qquad \text{and} \qquad \tau = \sqrt{\log(1 + 1/\alpha)}.$$

b) Argue that $E(N)$ and $\text{sd}(N)$ then are the same too. **c)** Determine β and τ when $\xi = 5\%$ and $\alpha = 4$ and simulate N under both models when $T = 1$. **d)** Generate (for both models!) $m = 1000$ simulations, store and sort each set and plot them against each other in a Q-Q plot. Comments on their degree of discrepancy.

Exercise 8.3.4 Let $\hat{\mu}_k = \mathcal{N}_k/\mathcal{T}_k$ be the estimate of the intensity μ_k in year k where $\mathcal{N}_k$ is the number of observed claims and $\mathcal{T}_k$ the total time of exposure; see (1.4). The estimates are available for K years. Suppose all μ_k have been randomly drawn independently of each other with $E(\mu_k) = \xi$ and $\text{sd}(\mu_k) = \sigma_\mu$ (same for all k). If we ignore randomness in μ_k, our natural estimate of a common μ is

$$\hat{\mu} = \frac{\mathcal{N}_1 + \ldots + \mathcal{N}_K}{\mathcal{T}_1 + \ldots + \mathcal{T}_K} \qquad \text{or} \qquad \hat{\mu} = w_1\hat{\mu}_1 + \ldots + w_K\hat{\mu}_K \quad \text{where} \quad w_k = \frac{\mathcal{T}_k}{\mathcal{T}_1 + \ldots + \mathcal{T}_K}$$

a) Justify this estimate. **b)** Show that $E(\hat{\mu}_k) = \xi$ so that $E(\hat{\mu}) = \xi$. **c)** Also show that

$$\text{var}(\hat{\mu}) = \frac{\xi}{\mathcal{T}_1 + \ldots \mathcal{T}_K} \quad + \quad \sigma_\mu^2(w_1^2 + \ldots + w_K^2)$$

[Hint: Utlize that $E(\hat{\mu}_k|\mu_k) = \mu_k$ and $\text{var}(\hat{\mu}_k|\mu_k) = \mu_k/\mathcal{T}_k$, the rule of double variance and the formula for the variance of sums formula.] In practice the number of years K is moderate, perhaps no more than a few. **c)** Use the results above to argue that the uncertainty in an estimate $\hat{\mu}$ could be substantial despite portfolios being large.

Section 8.4
Exercise 8.4.1 Consider a Poisson regression with v explanatory, *categorical* variables with c_i categories for

variable i. **a)** Explain that if you cross all variables with all others, then the number of different groups will be $c_1 c_2 \cdots c_v$. **b)** What is this number when the variables (automobile insurance) are sex, age (6 categories), driving limit (8 categories), car type (10 categories) and geographical location (6 categories). **c)** Explain how the the simple estimator (1.4) can be applied in this example and suggest an obstacle to its use.

Exercise 8.4.2 Consider Table 8.2. **a)** What would have happened to the estimated annual claim intensities in Table 8.3 if the intercept term was -2.815 rather than -2.315? **b)** The same question if the female parameter was 0.074 instead of 0.037. **c)** Suggest a maximum and a minimum difference between the youngest and the next youngest age group by using the standard deviation recorded.

Exercise 8.4.3 Consider two individuals in a Poission regression with explanatory variables $x_1, \ldots, x_v$ and $x_1', \ldots, x_v'$ respectively. **a)** Show that the ratio of intensities are

$$\frac{\mu}{\mu'} = e^{b_1(x_1 - x_1') + \ldots + b_v(x_v - x_v')}, \qquad \text{estimated as} \qquad \frac{\hat{\mu}}{\hat{\mu}'} = e^{\hat{b}_1(x_1 - x_1') + \ldots + \hat{b}_v(x_v - x_v')},$$

Suppose the estimates $\hat{b}_1, \ldots, \hat{b}_v$ are approximately normally distributed with means $b_1, \ldots, b_v$ (often reasonable in practice). **b)** What is the approximate distribution of $\hat{\mu}/\hat{\mu}'$? Is it unbiased? **c)** How do you determine the variance? [Hint: Look up log-normal distributions.]

Exercise 8.4.4 This is a continuation of the preceding exercise. Consider intensities μ and μ' on two individuals of the same sex, one belonging to the oldest age group in Table 8.2 and the other to the group $56 - 70$. Estimates of the coefficients are then $= -0.637$ (0.073) (oldest group) and -0.711 (0.070) (the next oldest one). Estimated standard deviations are in parantheses. The estimated covariance (not shown in Table 8.2) was 0.0025 and we shall assume that estimates are normally distributed. **a)** Estimate the ratio of the two intensities. **b)** Compute the estimated standard deviation of $\log(\hat{\mu}/\hat{\mu}')$. **c)** Adjust the estimate of the ratio so that it becomes approximately unbiased. **d)** Determine a 95% confidence interval for the ratio.

Section 8.5

Exercise 8.5.1 Suppose the delay probabiltities are of the form $q_0 = 0$ and $q_l = C\exp(-\gamma l)$ for $l = 1, \ldots K$ where C ensures that their sum is one. **a)** Determine C. **b)** Plot the delay probabilities when $\gamma = 0.1$ and $\gamma = 0.2$. Suppose all $J_{-s} = J$ and all $\mathcal{N}_{-s} \sim \text{Poisson}(J\mu T)$. c) Determine the distribution of the claims $\mathcal{N}_k$ that will be settled at time k.

Exercise 8.5.2 In the situation in Section 8.5 suppose $\mu_{-s} = \xi Z_{-s}$ where $Z_{-s} \sim \text{Gamma}(\alpha)$ and all Z_{-s} are independent of each other. **a)** Argue that each $\mathcal{N}_{-s,k+s}$ is negatively binomial. Suppose $J_{-s} = J$ is the same for all s; i.e that the portfolios of the past have been equally large at all times. **b)** Show that the conclusion in a) carries over to $\mathcal{N}_k$ which is the sum over all s. **c)** Can it be concluded that $\mathcal{N}_k$ are independent for different k? [Hint: Circumstances must be very special.]. Note that the conclusions have consequences for how we simulate.