# 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). The Poisson distribution is the central model of the chapter. Its parameter is $\lambda = \mu T$ (for single policies) and $\lambda = J\mu T$ (for portfolios), where $J$ is the number of policies, $T$ the time of exposure and $\mu$ claim intensity; see Section 3.2. These probabilistic descriptions rest on the **Poisson point process** presented in the next section. This construction provides a strong theoretical basis and enables us to explore the meaning of $\mu$ and how it comes about that the same familiy of distributions is used at both policy and portfolio level. It also indicates how the default Poisson model should be extended to cover different forms of risk hetereogeneity.

The issue of indvidual risk was raised in Chapter 3, but the discussion will now be taken much further. One approach is through random intensities. Each policy holder $j$ is then assigned an intensity $\mu_j$ through a drawing from some probabiltity distribution. This viewpoint leads to the theory of **credibility** in Chapter 10. Another way is to link $\mu_j$ to **explanatory** variables. An example is automobile insurance where (for instance) age and sex of the drivers typically influence accident frequency. The method to use is **Poisson regression** presented in Section 8.4. But variations in risk may also have to do with time. One feature is seasonal effects due to climate and other things, but there are also systematic trends upwards or downwards (see Figure 8.1). Such phenomena make certain periods collectively more risky than others with an effect on portfolio risk that is very different from hetereogeneity at individual level. A time effect of a different kind is delays (up to decades) from the insured incidence to the actual compensation. That is treated at the end.

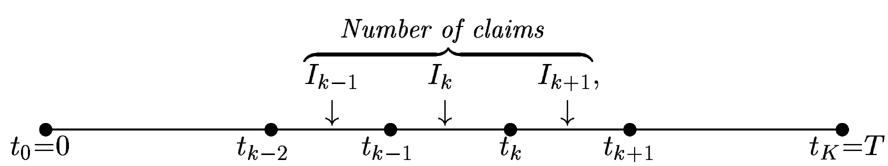
## 1.2 The Poisson point process

**Introduction**

The Poisson point process provides the genesis for most models of claim size used in practice, and its construction is the natural way to elucidate the meaning of the key parameter $\mu$. Start 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 $t_k = kh$ $(k = 0, 1 \ldots K)$ are the changepoints; see 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 $T$. Their probability distributions are needed for most evaluations of risk in property insurance. What they are will in this section be deduced through a theoretical argument.

**Modelling single policies**

It is overwhelmingly likely that *no* claims is received from a policy holder during a single, *short* time increment $h$, and the risk of more than one claim is remote. Let us go to the extreme and assume that each $I_k$ is either 0 or 1 (hence the symbol $I$, signifying an *indicator* variable). These assumptions will be relaxed later without this changing anything substantial. But for now a sequence of indicator variables $\{I_k\}$ is called a **time-homogeneous Poisson point process** if

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

Formally the definition applies in the limit as $h \to 0$; see below. There is a sensible rationale behind both assumptions. Surely an insurance incident occurs with probability proportional to the time increment $h$. The coefficient of proportionality $\mu$ is an **intensity** and defines the risk of the individual. Independence is also plausible. Typically accidents are consequences of forces, behaviour or events that bear no relationship to each other. An incident that has occurred simply makes further incidents in the aftermatch neither more nor less probable. There may be violations to this, but they are usually better handled by allowing the intensity $\mu$ to vary with time, see below.

The assumptions make $\{I_k\}$ a *Bernoulli* series and their sum $N$ a *binomially* distributed random variable. This chance experiment is about the first you are taught in courses in probability and statistics; look it up in any elementary textbook if you are unfamiliar with it. The probability distribution for $N$ is then $(n = 0, 1, \ldots, K)$

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

It is easy to verify that the expression for $\Pr(N = n)$ may be rewritten

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

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

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

The relationship for $B_3$ is a consequence of the fact that $(1 + a/K)^K \to \exp(a)$, applied with $a = -\mu T$. By collecting the terms it follows that

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

and the limit is the Poisson distribution with parameter $\lambda = \mu T$. The Poisson distribution is the logical consequence of the Poisson point process and is accurate if the conditions (i) and (ii) are reasonable descriptions of the reality.

**A more general viewpoint**

On the portfolio level $J$ independent process based on intensities $\mu_1, \ldots, \mu_J$ run in parallel. Together they produce each period $k$ a total number of claims $\mathcal{I}_k$ against the portfolio. Its distribution is derived by noting that

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

Here $1 - h\mu_j$ is the probability of no claim from policy $j$. On the left all of those are multiplied together for the probability of no claims against the *portfolio*. The probability of exactly one claim is a little more complicated. We must add the probabilitites that it is due to the the first, second, third policy and so on. This yields 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 $\mathcal{I}_1, \ldots, \mathcal{I}_K$ has the same type of probabilistic description as their analogue $I_1, \ldots, I_K$ for policies. The only difference is that the former intensity $\mu$ now has become the sum $\mu_1 + \ldots + \mu_J$. Note that we could have *started* with a construction that allows these small $o(h)$ terms. The earlier definition is then replaced by

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

Again, what happens in the limit as $h \to 0$ doesn't change, and (1.3) appears as a consequence. This less rigid formulation is a more satisfactory one from a theoretical point of view.

**Summary of conclusions**

The preceding argument has established several useful results. Under the Poisson point process claim frequencies $N$ (policy level) and $\mathcal{N}$ (for portfolios) are both Poisson distributed. Their parameters are

$$\underset{\textit{policies}}{\lambda = \mu T} \qquad \textit{and} \qquad \underset{\textit{portfolios}}{\lambda = (\mu_1 + \ldots + \mu_J)T = J\bar{\mu}T}$$

where $\bar{\mu} = (\mu_1 + \ldots + \mu_J)/J$ is the average intensity over all policy holders. It doesn't for $\mathcal{N}$ matter that $\mu_1, \ldots, \mu_J$ vary over the portfolio; we simply insert their average. That seems tidy and easy, and yet isn't the entire story as new effects emerge when the intensities are randomly drawn; see Section 8.5 below.

An alternative way of stating the conclusion is to start with $J$ independent Poisson variables $N_1, \ldots, N_J$ with parameters $\lambda_1, \ldots, \lambda_J$. Their sum $\mathcal{N} = N_1 + \ldots + N_J$ is then Poisson too; the parameter being $\lambda_1 + \ldots + \lambda_J$.

**Hetereogeneity over time**

In practice claim intensities often fluctuate over time. An example is shown in Figure 8.1 where monthly accident frequencies for automobiles are plotted. The data comes from a Norwegian insurance company and have been converted into estimated intensities on monthly time scale through the method introduced in Section 8.3. One feature is that claim frequency seems to rise over the period in question, perhaps by about 0.1% monthly (or 1.2% annually). There are also systematic variations due to the season of the year. In cold countries (like Norway) a main cause is slippery roads during winter; elsewhere it could be strong winds or torrents of rain during the stormy season. Peaks in summer (when people drive a lot) is still another example. Are such time factors relevant for actuarial calculations? Systematic trends upwards surely must be, but for seasonal effects the first answer is negative. Calulations are usually annual which means that variations over the year average out and can be ignored.

To analyse what goes on introduce $\mu = \mu_t$ as an intensity fluctuating in time. We may envisage $\mu_t$ as a continuous function which can be approximated by a step function on the $K$ sub-intervals defining the Poisson process. This means that $\mu = \mu_k$ applies between $t_{k-1}$ and $t_k$, and the number of claims $I_k$ for this period must be Poisson with parameter $\mu_k h$. Hence $N = I_1 + \ldots + I_K$ is Poisson too. Its parameter is

$$\lambda = \sum_{k=1}^{K} h\mu_k \quad \to \quad \int_0^T \mu_t \, dt \qquad \text{as} \qquad h \to 0.$$

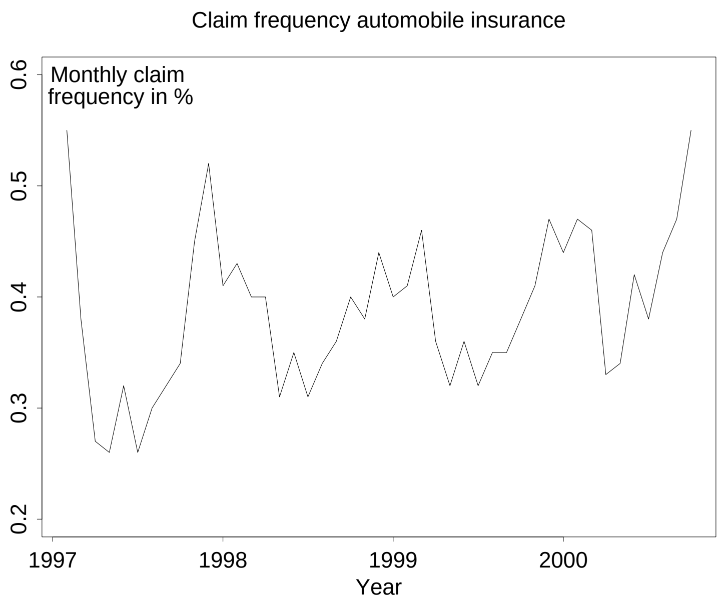



*Figure 8.1. Estimated* **monthly** *claim intensities for data from a Norwegian insurance company*

The limit is simply how integrals are defined. In other words we have proved that the Poisson parameter for $N$ has become

$$\lambda = T\tilde{\mu} \qquad \text{where} \qquad \tilde{\mu} = \frac{1}{T}\int_0^T \mu_t\, dt. \tag{1.4}$$

Here $\tilde{\mu}$, the *time average* of $\mu_t$, takes over from the former constant $\mu$. It in this sense time variations over the year can be ignored.

## 1.3 The Poisson model

### Introduction

The preceding section established the Poisson distribution, introduced in Section 2.6, as the default model for claim frequencies. Mean, standard deviation, skewness and kurtosis for a Poisson variable with parameter $\lambda$ are

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

Note that variance and mean are equal; i.e. $\text{var}(N) = E(N)$ which is often used for a quick appraisal from the historical data $N_1, \ldots, N_n$. If their average and variance do not deviate too much, the Poisson model becomes a strong contender.

The important property that sums of independent Poisson variables are Poisson distributed themselves was pointed out in the preceding section. It has the interesting consequence that $N$ becomes *normally* distributed as $\lambda \to \infty$ (Exercise 8.3.1). There are clear signs of this in Figure 8.2 where the probability function has been plotted for two values of $\lambda$. The skewness to the right when $\lambda = 4$

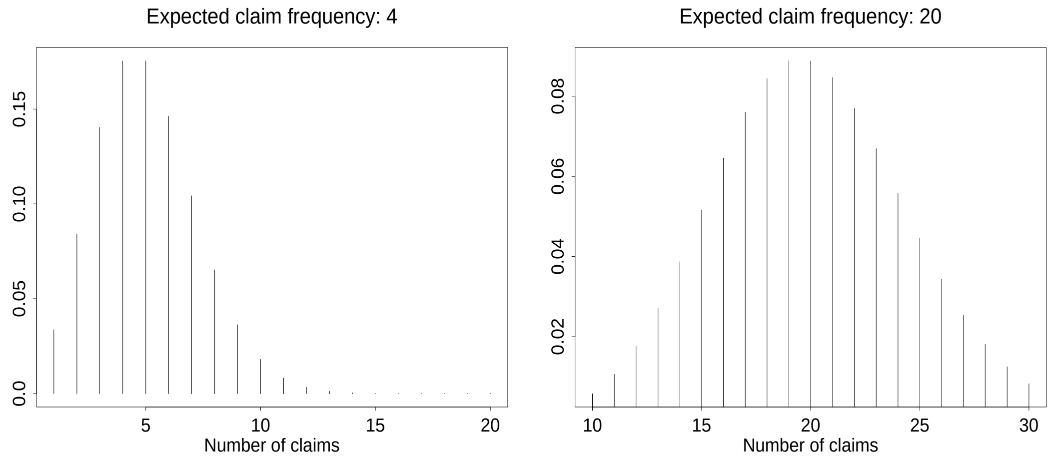



*Figure 8.2 Poisson density functions for* $\lambda = 4$ *(left) and* $\lambda = 20$ *(right).*

has largely disappeared at $\lambda = 20$, the shape now being approximately symmetrical.

**Fitting**

Let $n_1, \ldots, n_n$ be the number of claims from $n$ policies with common intensity $\mu$. Suppose they have been under exposure $T_1, \ldots, T_n$. The standard estimate of $\mu$ is then

$$\hat{\mu} = \frac{n_1 + \ldots + n_n}{T_1 + \ldots + T_n}. \tag{1.5}$$

This is a likelihood method. It is *unbiased*, i.e. $E(\hat{\mu}) = \mu$ and the standard deviation is

$$\mathrm{sd}(\hat{\mu}) = \sqrt{\frac{\mu}{T_1 + \ldots + T_n}} \tag{1.6}$$

see Exercise 8.3.2. Estimation error is determined by the size $n$ of the historical record *and* the exposure to risk. The sum $T_1 + \ldots + T_n$ is the number of automobile years behind the estimate.

Sometimes claims are recorded on an annual basis. Each $T_j$ is then either one or some fraction of a year (if the policy wasn't there the entire year). The same policy will then occur several times in the data record. That doesn't matter here (and neither in the next section), but for the models with random intensities later in this chapter, it *does* matter. In such situations each individual must be recorded a single time only. Of course, isn't necessary that a client has remained with the company. The experience with former ones are relevant too. How far back we should look is a matter of practical judgement.

**Numerical example**

Poisson modelling will be illustrated on the Norwegian automobile portfolio in Figure 8.1. The number of claims where 6555 over 115000 automobile years which yield the estimate

$$\hat{\mu} = \frac{6555}{115000} = 5.7\% \qquad \text{with estimated sd} \qquad \sqrt{\frac{0.057}{115000}} \doteq 0.07\%,$$

*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 (standard error in paranthesis)*

where the accuracy is less than it appears; see Section 10.3 where a trap is elucidated. About one car in twenty causes accidents each year, but underneath this average value there must be considerable variations from one individual to another. The annual intensity estimates have in Table 8.1 been broken down on male/female and five age categories. This is known as a **cross-classification** between the two variables and leads to $2 \times 5 = 10$ sub-groups to which the earlier estimation method could be applied. The analysis has ignored other explanatory variables (such as how much people drive) in order to keep things simple. Let us nevertheless see what we can learn.

There is a strong age effect. Accidents rates go consistently down as people gain in experience. (people over 70 being an exception). In the youngest group $(20-25)$ men cause distinctly more incidents than women. Could the difference of 2.1% be due to chance? The estimated standard error of the *difference* is from the table

$$(0.76^2 + 1.12^2)^{1/2} \doteq 1.35,$$

and the ratio $2.1/1.35 \doteq 1.5$ could be compared against a normal distribution. In economics that is known as the **Wald test**. Here 1.5 is close to being statistical significant at the 5% level which might lead us to the position that it is 'fair' to charge young women less that young men. A difference of about 25% would not be unreasonable since $10.5/8.4 \doteq 1.25$[1]. The advantage of women disappears with age, and they are more and not less risky in the highest age groups.

**How Poisson variables are sampled**

The method of choice depends on the circumstances. If we are dealing the aggregated claim against a portfolio, computing time will be dominated by time spent on claim size. With $m$ simulations, there will be about $mE(N)$ claim size drawings against only $m$ of $N$. This suggests that the speed of the Poisson sampler does not matter too much. Often the simple and elementary Algorithm 2.10 is good enough. This applies equally with the mixture models capturing risk hetereogeneity in later sections of this chapter.

If speed *is* critical, we might fall back on the methods of guide tables in Section 4.2, but this approach is useless with mixture models (because the time-consuming set-up phase has to be redone many times). The problem with Algorithm 2.10 is that it slows down for large values of $\lambda$, but under such circumstances we may turn to the method of Atkinson (1979):

**Algorithm 8.1 Atkinson's Poisson generator**

[1] In Norway, with its stern laws on equal treatment of sexes, such differentiated pricing isn't legal! (charging unequally according to *age* is all right!).

0 Input: $c \leftarrow 0.767 - 3.36/\lambda, \quad a \leftarrow \pi/\sqrt{3\lambda}, \quad b \leftarrow \lambda a, \quad d \leftarrow \log(c/a) - \lambda$
1 Repeat

2 Repeat
3 Draw $U^* \sim$ uniform and $X^* \leftarrow \{b - \log(1/U^* - 1)\}/a$
until $X^* > -0.5$

4 $N^* \leftarrow [X^* + 0.5]$ and draw $U^* \sim$ uniform
5 If $b - aX^* - \log\{\{1 + \exp(b - aX^*)\}^2/U^*\} < d + N^* \log(\lambda) - \log(N^*!)$
stop and return $N^*$

Before running the algorithm it is necessary compute (recursively!) and store the sequence $\log(n!)$ up to some number the Poisson variable has microsopic chances to exceed ($5\lambda$ could be a sensible choice). The method is derived through rejection sampling; see Cassela and Robert (1998). Atkinson recommends that $\lambda > 30$ for his procedure to be used. Devroye (1986) contains other possibilities.

## 1.4 Poisson regression

### Introduction

The automobile example in Table 8.1 is a special case of the important problem of linking variations in risk to *explanatory* variables. Insurance companies run such studies to understand which customers are profitable and which are not and to charge differently in different segments of the portfolio. Some will say the latter contradicts the the principle of solidarity which lies behind the very idea of insurance itself. Whatever the merit of this view there has through the last decades certainely been a growing trend towards individual pricing. It can be little doubt that the modern actuary must understand and master the techniques involved.

Credibility theory is one of the answers provided by actuarial science, 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 relevance of **observable** variables such as age and sex of drivers, geographical location of a house and so forth. In the preceding section the portfolio was partitioned into groups of policy holders, and the estimate (1.5) applied separately to each group. Although there are many situations where this is a sensible approach, its range of applications is too restrictive for the method to be the single one. One reason is that there might be too many groups. Take the automobile example in Section 8.3. Among relevant factors neglected are 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 1000 different categories or more. This would require huge amount of data for estimation. Another problem is that we need an approach that can handle *numerical* variables, not only categorical ones.

### The model

The approach most widely used in practice is that of **Poisson regression** where the claim intensity

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

$\mu$ is 'explained' by a set of observable variables $x_1, \ldots, x_v$ through a relationship of the form

$$\log(\mu) = b_0 + b_1 x_1 + \ldots + b_v x_v. \tag{1.7}$$

Here $b_1, \ldots, b_v$ are coefficients that are to be determined from historical records. When that has been accomplished, the link between $\mu$ and the explanatory variables $x_1, \ldots, x_v$ (also called **covariates**) allows us to discriminate customers according to the risk they represent.

Why is $\log(\mu)$ used in (1.7) and not $\mu$ 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 $\mu$ 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 will be 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 $T_j$ 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.

It is assumed that the random variable $N_j$ corresponding to $n_j$ is Poisson distributed with parameter $\lambda_j = T_j \mu_j$, where

$$\log(\mu_j) = b_0 + b_1 x_{j1} + \ldots + b_v x_{jv}. \tag{1.8}$$

The likelihood function (see Section 7.3) is derived by noting that the probability function of $N_j$ is

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

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 $j$ to produce the likelihood function $\mathcal{L}(b_0, \ldots, b_v)$. We may drop the two middle terms $n_j T_j$ and $\log(n_j!)$ (*constants* in this context). The likelihood criterion then becomes

$$\mathcal{L}(b_0, \ldots, b_v) = \sum_{j=1}^{m} \{n_j \log(\mu_j) - T_j \mu_j\}, \tag{1.9}$$

where $\mu_j$ is given in (1.8). The exposure to risk $T_j$ is sometimes called the **offset**.

There is little point in carrying the mathematics further. Together (1.8) and (1.9) define a criterion, the maximimum $\hat{b}_0, \ldots, \hat{b}_v$ of which being the estimates. Optimization is carried out by commercial software (and you do not have to know how it is done). In fact, it is easy, since it can be proved (McCullagh and Nelder, 1994) that $\mathcal{L}(b_0, \ldots, b_v)$ is a convex surface with a single maximum. Approximate standard errors are also computed.

**Interpretation and coding**

The method will applied to the example in Table 8.1 for illustration. Perhaps the most obvious way of feeding the two explanatory variables age ($x_1$) and sex ($x_2$) into the regression model (1.8) is to write

$$\log(\mu_j) \quad = \quad b_0 \quad + \quad \underset{\textit{age effect}}{b_1 x_{j1}} \quad + \quad \underset{\textit{male/female}}{b_2 x_{j2}}. \tag{1.10}$$

Here $x_{1j}$ is the age of the owner of car $j$ and

$$\begin{aligned} x_{2j} &= 0, \quad \text{if } j \text{ is male} \\ &= 1, \quad \text{if } j \text{ is female.} \end{aligned}$$

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

$$\frac{\mu_i}{\mu_j} = \exp(b_2). \qquad \text{Example: } b_2 = 0.037 \text{ yields} \quad \frac{\mu_i}{\mu_j} = \exp(0.037) \doteq 1.037,$$

where the value on $b_2$ is taken from Table 8.2 below. According to the estimate an average Norwegain female driver causes 3.7% more accidents than the average male.

The coefficient $b_1$ in (1.8) is likely to be negative. As drivers become more experienced, so their accident rate is likely to go down. But $\log(\mu)$ should not nevessarily 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 $c$ such groups (1.8) is changed to

$$\log(\mu_j) = b_0 + \sum_{i=2}^{c} b_1(i) x_{j1}(i) + b_2 x_{j2}, \tag{1.11}$$

where for an individual $j$ in age group $l$

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

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

$$\log(\mu_j) = b_0 + b_1(l) + b_2 x_{j2}, \qquad \text{for policy holder } j \text{ at age } l,$$

| | | | | |
|---|---|---|---|---|
| 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 8.2 Fitted Poisson regression (standard deviations in paranthesis)*

but (1.11) may be the way the model is entered Poisson regression software.

Does this look contrived? The number of unknown parameters has gone up, and longer records historical data are needed. However, the modelling now permits more complicated (and hence truer) relationships between age and risk, and the partition into groups may also be the natural way to avoid excessively many different premia when charging people. Note that $b_1(1) = 0$ in (1.11). A restriction of that kind is necessary to avoid confounding with the intercept parameter $b_0$. Such conventions vary with the software you are using; more on the interpretation of these coefficients among the exercises.

**Automobile example continued**
The automobile data has been fitted in Table 8.2. Note the huge jump from the lowest age group to the next one, and for the oldest age group claim frequency goes up. Statistical significance can be evaluated by dividing the estimates (or their differences) on the estimated standard deviations as explained in Section 8.3. Usually actuaries are not overly concerned with statistical significance.

The estimated coeficients plugged into (1.10) yield estimates for the claim intensities in the various groups. The results in Table 8.3 follow those in Table 8.1 with the notable exception that the much higher accident rate for men in the youngest group has disappeared! This is caused by the model imposed. Since the contributions of age and sex are added on log-scale, the female intensity is forced to be proportional to the male on ordinary scale, and the *fine structure* found in Table 8.1 has been lost. Presumably the former version obtained without the additional mathematical restrictions is closest to the truth. So why bother with the other? That's because such simplifications frequently are necessary in practice to reduce the number of parameters. Often there are many explanatory variables present, as is in this example when all of them are included. Crossing four or five variables with all others would lead to hundreds or even thousands of different groups with the same number of parameters that would demand huge amount of data to be fitted.

## 1.5 Random risk parameters

**Introduction**
Variation in $\mu$ over the portfolio was in the preceding section attributed to *observable* explanatory variables. But there are also personal factors about which we can not possibly possess detailed knowledge. In automobile insurance, how are we to know that some drivers are reckless and some are not, that some have excellent power of concentration while others easily lose theirs. Even within a group that has been made as homogeneous as possible uncertainty due to such factors

*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 8.3 Annual claim intensities under Poission regression*

will remain. The natural mathematical formulation is to work with a *stochastic* $\mu$, and that has 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. That contributes a random component too, but now one affecting all policy holders jointly. Much was made of this distinction between *individual* and *collective* hidden randomness in Section 6.3. The basic models were

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

Their impact on portfolio risk differed violently, yet their basic methodology is very much the same and will only be written out for the model version on the left. Modifications required for the other case are treated among the exercises.

The density function for $N$ is given by the so-called *mixing relationship*

$$\Pr(N=n) = \int_0^\infty \Pr(N=n|\mu)g(\mu)\,d\mu. \tag{1.12}$$

Here $g(\mu)$ is the density function of $\mu$. The Gamma family will be used during the second half of this section, but first two issues that are best discussed at a general level.

**Estimation of mean and standard deviation**

It is possible to estimate $\xi = E(\mu)$ and $\sigma_\mu = \text{sd}(\mu)$ from historical data without imposing further conditions on the model. The likelihood method is not available, since $g(\mu)$ has not been specified, but moment estimation (Section 7.3) is possible. Suppose policy holder $j$ has reported $n_j$ claims over a period of length $T_j$ for $j = 1,\dots,n$. Estimated claim intensities are then

$$\hat{\mu_j} = \frac{n_j}{T_j} \qquad \text{with mean and variance} \qquad E(\hat{\mu_j}|\mu_j) = \mu_j \quad \text{and} \quad \text{var}(\hat{\mu}_j|\mu_j) = \frac{\mu_j}{T_J};$$

see (1.6). Note that the mean and variance is conditional. By the double rules of Section 6.3

$$E(\hat{\mu_j}) = \xi \qquad \text{and} \qquad \text{var}(\hat{\mu}_j) = \sigma_\mu^2 + \frac{\xi}{T_j},$$

and the question is how the information in $\hat{\mu}_1,\dots,\hat{\mu}_n$ should be pooled. The problem is much the same as with Anova II models in biostatistics; see Sokal and Rholf (1973) for a classical treatise. Their approach is based on assigning each estimate $\hat{\mu}_j$ a weight proportional to $T_j$. For the estimate of $\xi$ this leads to

$$\hat{\xi} = w_1\hat{\mu}_1 + \dots + w_n\hat{\mu}_n \qquad \text{where} \qquad w_j = \frac{T_j}{T_1 + \dots + T_n}, \tag{1.13}$$

which is the same as (1.5). It is proved in Section 8.7 that

$$E(\hat{\xi}) = \xi \qquad \text{and} \qquad \text{sd}(\hat{\xi}) = \left(\sigma_\mu^2(w_1^2 + \ldots + w_n^2) + \frac{\xi}{T_1 + \ldots + T_n}\right)^{1/2},$$

and $\hat{\xi}$ is unbiased. The error normally vanishes if $T_1 + \ldots + T_n \to \infty$. A precise statement requires an additional (weak) condition on the sequence $\{T_j\}$.

The estimate for the variance is less intuitive. One possibility, justified in Section 8.7, is

$$\hat{\sigma}_\mu^2 = \frac{\sum_{j=1}^n w_j(\hat{\mu}_j - \hat{\xi})^2 - C}{1 - \sum_{j=1}^n w_j^2} \qquad \text{where} \qquad C = \frac{(n-1)\hat{\xi}}{T_1 + \ldots + T_n}. \tag{1.14}$$

In the numerator $C$ is a correction term that makes the estimate unbiased, but this adjustment doesn't guarantee a positive estimate. With negative values take the obvious position that the variation in $\mu$ over the portfolio is unlikely to be very important and let $\hat{\sigma}_\mu = 0$. The expression for $\text{sd}(\hat{\sigma}_\mu^2)$ is complicated and is omitted.

Application to the automobile portfolio described earlier gave

$$\hat{\xi} = 5.60\% \qquad \text{and} \qquad \hat{\sigma_\mu} = 2.0\%.$$

That tells us that the variability in claim intensity among the policies is huge. We got the same message in Section 8.4 from another perspective. The present model would be used for purposes quite different from those of Poisson regression.

**Simulation**

Monte Carlo sampling requires a specific model for $\mu$ (such as the one in the next section). The commands reflect the model definition:

**Algorithm 8.2 Poisson mixing (single policy)**

0 Input: Density $g(\mu)$ for $\mu$

1 Draw $\mu^* \sim g$      % *Many possibilites*

2 For $k = 1, \ldots, K$ do

3     Draw $N_k^* \sim$ Poisson $(T\mu^*)$      % *Algorithm 2.10*

4 Return $N_1^*, \ldots, N_K^*$

On output claim frequencies over $K$ different periods of length $T$ has been generated. Note that the intensity $\mu^*$ (linked to the policy), is only generated once. That makes all simulations stochastically dependent, since they originate with the same $\mu^*$. For $J$ policies Algorithm 8.2 is run $J$ times. Alternatively when $\mu$ is a common factor affecting the entire portfolio, the intensity $\mu^*$ is only drawn once in the beginning and the loop repeated for all $J$ policies.

**The negative binomial model**

The most commonly applied model for $\mu$ is the Gamma family; i.e. taking

$$\mu = \xi Z \qquad \text{where} \qquad Z \sim \text{Gamma}(\alpha). \tag{1.15}$$

This allows $\mu$ to vary around a mean $\xi$, the uncertainty being controlled by $\alpha$. Specifically

$$E(\mu) = \xi, \qquad \text{and} \qquad \text{sd}(\mu) = \xi/\sqrt{\alpha}; \tag{1.16}$$

see Section 2.6. Note that the variation in $\mu$ is now captured by $\alpha$. Clearly $\text{sd}(\mu) \to 0$ as $\alpha \to \infty$. The pure Poisson case with *fixed* intensity appears in the limit.

One of the reasons for the popularity of this model is without doubt that (1.12) can be evaluated in closed form. The detailed mathematical argument which is shown in Section 8.7, yields ($n = 0, 1, \ldots$)

$$\Pr(N = n) = \frac{\Gamma(n+\alpha)}{\Gamma(n+1)\Gamma(\alpha)} p^{\alpha}(1-p)^{n} \qquad \text{where} \qquad p = \frac{\alpha}{\alpha + T\xi}. \tag{1.17}$$

This is known as the **negative binomial** distribution and will be denoted $\text{nbin}(\xi, \alpha)$.

**Properties of the negative binomial**
Mean and variance are

$$E(N) = T\xi \qquad \text{and} \qquad \text{var}(N) = T\xi(1 + T\xi/\alpha); \tag{1.18}$$

see Section 8.7. Unlike pure Poisson models $\text{var}(N) > E(N)$, but they become equal in the limit as $\alpha \to \infty$. It had to like that since an infinite $\alpha$ signifies a pure Poisson distribution.

The model has a convolution property which is similar, but less general than for the Poisson familiy. Suppose $N_1, \ldots, N_J$ are independent with common distribution $\text{nbin}(\xi, \alpha)$. Each $N_j$ can then be represented as

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

If $\zeta = \mu_1 + \ldots + \mu_J$ and $\bar{Z} = (Z_1 + \ldots + Z_J)/J$, we have for $\mathcal{N} = N_1 + \ldots + N_J$ the analogous representation

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

and where $Z_1, \ldots, Z_J$ are independent. What lies behind this? Firstly that the sum of independent Poisson variables are Poisson itself, secondly that the new parameter

$$\zeta = \mu_1 + \ldots + \mu_J = \xi(Z_1 + \ldots + Z_j) = J\xi\bar{Z}$$

and finally that the average $\bar{Z}$ is $\text{Gamma}(J\alpha)$; see Section 9.3. But the representation established for $\mathcal{N}$ is the defining one for the negative binomial model, and so $\mathcal{N} \sim \text{nbin}(J\xi, J\alpha)$. Note that $\text{sd}(\mathcal{N})/E(\mathcal{N}) = 1/\sqrt{J\alpha}$, and random variation is strongly diminished when moving to the portfolio level. This was also established in Section 6.3 through a more general route.

**Fitting the negative binomial**
Moment estimation through $\hat{\xi}$ and $\hat{\sigma_\mu}$ in (1.13) and (1.14) is easiest technically. Since $\xi$ is the mean, its estimate is simply (1.13) whereas $\hat{\alpha}$ is determined by

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

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

Likelihood estimates are more accurate in theory, but require more work to implement. To derive the log likelihood function the obervations $n_j$ must be inserted for $n$ in (1.17) and the logarithm added 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 $\xi$ and $\alpha$ have been omitted. It takes numerical software to optimize this function. A primitive way is to compute it over a grid of points $(\xi, \alpha)$ and select the maximizing pair. If not available in your software, you will need approximations for $\log\{\Gamma(\alpha)\}$, see Appendix C.

**Automobile example continued**
Moment estimates for $\xi$ and $\tau$ were $\hat{\xi} = 5.60\%$ and $\hat{\sigma_\mu} = 2.0\%$, which leads to $\hat{\alpha} = 0.056^2/0.02^2 = 7.84$ as assessment of $\alpha$ in the negative binomial distribution. The likelihood method was applied as well, and together the two methods gave the following sets of estimates:

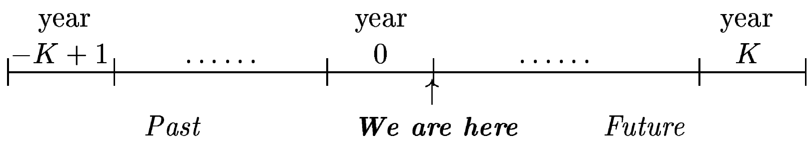
| | | |
|---|---|---|
| $\hat{\xi} = 5.60\%$, $\hat{\alpha} = 7.84$ | and | $\hat{\xi} = 5.60\%$, $\hat{\alpha} = 2.94$ |
| *moment estimates* | | *likelihood estimates* |

For $\xi$ the results are identical to two decimal places (they are *not* the same estimate!), but for $\alpha$ the descrepancy is huge, the likelihood estimate potraying a much more variable claim frequency. Behind the estimates are about half a million automobile years, so there is no question of the deviation being accidental. What lies behing and which one is to be trusted? Answer: In a sense neither! What it tells us is that the Gamma family doesn't (in this case) provide a good description of the variability of $\mu$. If it had, the estimates would have been much closer together. The moment estimate is best (because the standard deviation is captured correctly). An example where the model does fit is given in Exercise 8.5.2.

## 1.6 Delayed settlements of claims

**Introduction**
Claims are never settled immediately, and for some types of injuries or damages 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 took place are called **IBNR** (**i**ncurred, **b**ut **n**ot **r**eported). Companies have to set aside funds to cover them (even though they are not yet identified).

The situation may be depicted as follows:




At the end of year zero (where we are at the moment) there are outstanding claims that will surface later. They have occurred during periods insured, and companies are responsible. If the maximum displacement between incident and settlement is $K$ years, we have to look $K$ years back (to include everything undetected) and also $K$ years ahead (the period our balance sheet will be affected). This section is concerned with the modelling of claim frequencies of this nature.

**The delay model**

A natural approach (though not currently the most popular one) is to regard IBNR claims as a random phenomenon. If $D$ is the number of periods of postponement until a claim is settled, we may introduce *delay probabilities* $q_0, \ldots, q_K$ through

$$q_l = \Pr(D = l) \qquad \text{for which} \qquad q_0 + \ldots + q_K = 1. \tag{1.19}$$

Their meaning is straightforward; $q_0$, for example, is the chance that the claim is disposed of the same year the incident took place. It is reasonable to assume that $D$ follows the same probability distribution for all events and that there is stochastic independence between events. We are then in a **multinomial** situation (Section 6.5). Let $\mathcal{N}$ be the total (and Poisson) number of claims orginating a given year (some are undetected) and $\mathcal{N}_l$ those settled $l$ years later, $l = 0, 1, \ldots, K$. We are then assuming that

(i) $\mathcal{N}_0, \ldots, \mathcal{N}_K | \mathcal{N} = n$ is multinomial$(n, q_0, \ldots, q_K)$
(ii) $\mathcal{N} \sim$ Poission$(\lambda)$,

which yield the simple result that

$$\mathcal{N}_k \sim \text{Poisson}(q_k \lambda) \qquad \text{with} \qquad \mathcal{N}_0, \ldots, \mathcal{N}_K \quad \text{stochastically independent;} \tag{1.20}$$

see Section 8.7 for the proof. The delayed claims originate with the *same* basic $\mathcal{N}$; still they are independent. This allows us to keep track on many years simultaneously through the same machinery as elsewhere; even the fitting of the model follows earlier leads. The rest of the present section is devoted to these issues.

**IBNR claim frequencies**

We are dealing with a sequence of years (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), $\mu_{-s}$ the average claim intensity during the same year (unknown) and let $N_{-s,l}$ 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.

$$\mathcal{N}_k = \sum_{s=0}^{K-k} \mathcal{N}_{-s,k+s}. \qquad (1.22)$$

As a sum of independent Poisson counts these quantities become Poisson themselves, with

$$\lambda_k = \sum_{s=0}^{K-k} \lambda_{-s,k+s} = \sum_{s=0}^{K-k} q_{-s+k} J_{-s} \mu_{-s}. \qquad (1.23)$$

being the Poisson parameter of $\mathcal{N}_k$. The pay-outs $\mathcal{N}_1, \ldots, \mathcal{N}_K$ over the entire time horizon ahead are still stochastically independent. This useful observation enables us to employ the same techniques for IBNR reserving as those introduced in Chapter 3.

**Fitting delay models**

A convenient way to fit model with delay probabiltities is to utilize that $\lambda_{-s,l}$ in (1.20) is on multiplicative form so that

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

where

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

This is a Poisson 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. Details in how that is done are given in Section 8.7.

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

$$\hat{q}_l = \exp(\hat{\alpha}_l)/C, \quad \hat{\mu}_{-s} = C\hat{\beta}_s, \qquad \text{where} \qquad C = \exp(\hat{\alpha_0}) + \ldots + \exp(\hat{\alpha_K}). \qquad (1.26)$$

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

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

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

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

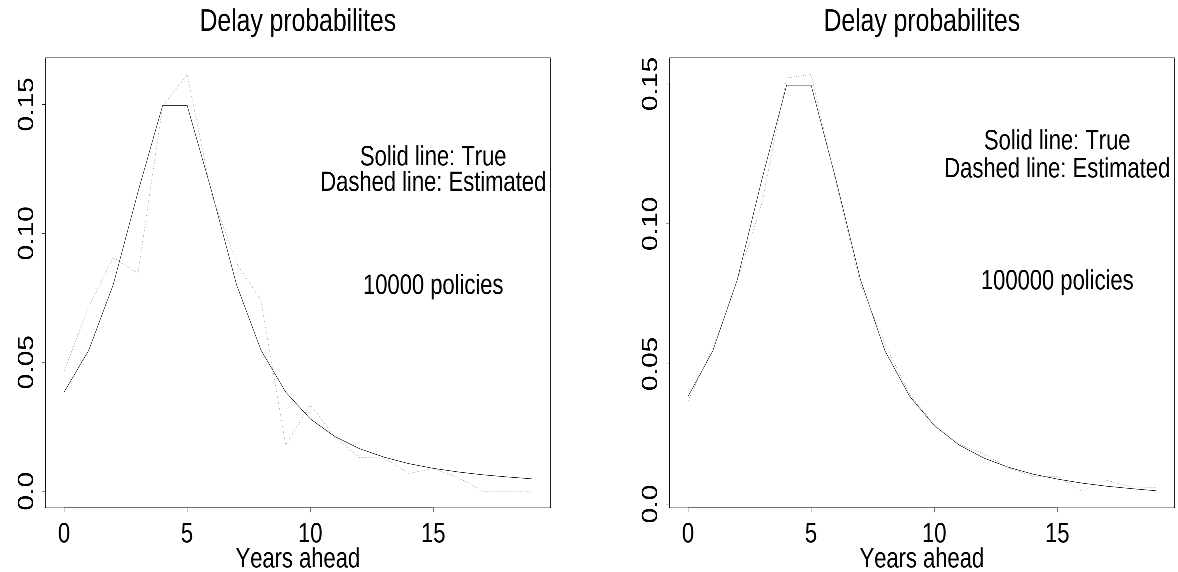



*Figure 8.3 Estimates of delay probabilites under circumstances described in the text*

where the constant ensured that $\{q_l\}$ adds to one. Parameters were

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

which means that the distribution $\{q_l\}$ has a top after four and five years ; it is plotted in Figure 8.3. All claims have been settled after $K = 20$ years.

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

For $s = 0, 1 \ldots, K-1$ do
  Draw $Y^* \sim \text{Gamma}(\alpha)$ and $\mu^*_{-s} \leftarrow \xi Y^*$
  For $s = 0, 1 \ldots, K$ do
    Draw $N^*_{-sl} \sim \text{Poisson}(Jq_l\mu^*_{-s})$.

The collection $\{N^*_{-s,l}\}$ (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 there is also good correspondence between the true and fitted claim intensitites (Figure 8.4).

## 1.7 Mathematical arguments

**Section 8.2**

**Derivation the Poisson distribution** Let $N = I_1 + \ldots + I_K$ where $I_1 \ldots I_K$ are stochastically

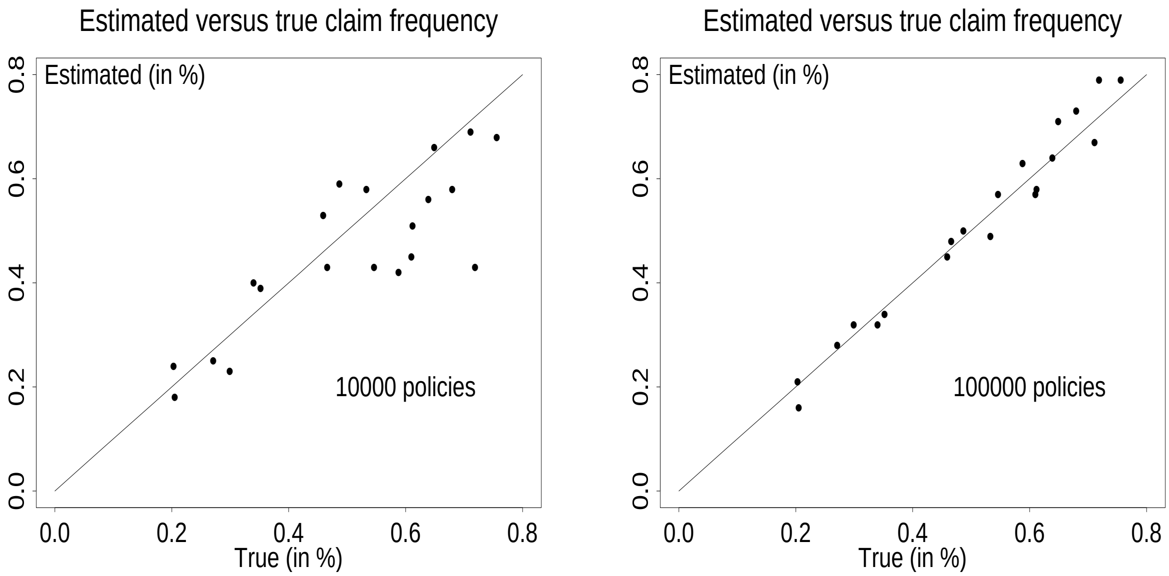



*Figure 8.4 Estimates of claim frequencies under circumstances described in the text*

independent with distribution

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

We shall let $h \to 0$ and $K \to \infty$ while keeping their product $T = Kh$ fixed. Introduce

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

for which

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

This is due to the Bonferroni inequality; see Appendix A. It follows that $\Pr(A^c) \to 0$ as $h \to 0$. Moreover,

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

so that

$$\Pr(N = n) - \Pr(N = n|A) = \{\Pr(N = n|A) - \Pr(N = n|A\}\Pr(A^c),$$

and it follows that

$$|\Pr(N = n) - \Pr(N = n|A)| \le 2\Pr(A^c) \to 0 \quad \text{as} \quad h \to 0.$$

This tells us that $\Pr(N = n)$ and $\Pr(N = n|A)$ have the same limit, and we may calculate the latter.

Under the event $A$ all $I_k$ are either zero or one, and $N$ becomes binomial with 'success' probability

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

so that

$$p = \frac{T\mu}{K} + o(\frac{1}{K}).$$

The only difference from the argument in Section 8.2 is the presence of the error terms $o(1/K)$. Thus

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

where

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

When $K \to \infty$ (and $n$ is kept fixed), we obtain the limits

$$B_1 \to (T\mu)^n/n!, \qquad B_2 \to 1,$$
$$B_3 \to \exp(-\mu T), \qquad B_4 \to 1\ .$$

The limits for $B_1$, $B_2$ and $B_4$ are obvious since $o(1/K)$ and $Ko(1/K)$ both tends to 0 whereas the one for $B_3$ is proved in Appendix C. Collecting the terms we have proved that

$$\Pr(N = n|A) \to \frac{(T\mu)^n}{n!}\exp(-T\mu),$$

which completes the derivation.

**Section 8.5**
**Justification of the estimates (1.13) and (1.14)** The argument is based on

$$E(\hat{\mu}_j) = \xi \qquad \text{and} \qquad \text{var}(\hat{\mu}_j) = \sigma_\mu^2 + \frac{\xi}{T_j},$$

which was established in Section 8.5. Since the sum of all the weights $w_j = T_j/(T_1 + \ldots + T_n)$ in (1.13) is one, it follows that

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

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

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

which is the expression for $\mathrm{sd}(\hat{\xi})$ in Section 8.5.

The estimate for $\sigma_\mu$ is based on

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

Here

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

and

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

Inserting this yields

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

which simplifies to

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

A moment estimate for $\sigma_\mu$ is the solution of the equation

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

which is the estimate (1.14), and the argument also shows that it must be unbiased.

**Negative binomial: Density** The definition was based on

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

By (1.12)

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

or when reorganized,

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

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

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

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 (1.17) since $\Gamma(n + 1) = n!$.

**Negative binomial: Mean and variance** By definition $\mu$ is Gamma distributed and

$$E(N|\mu) = \mu T \qquad \text{and} \qquad \text{var}(N|\mu) = \mu T.$$

Since $E(\mu) = \xi$ and $\text{var}(\mu) = \xi^2/\alpha$ it follows by the double rules of Section 6.3 that

$$E(N) = \xi T \qquad \text{and} \qquad \text{var}(N) = \xi T + \tau^2 T^2 = \xi T + \xi^2 T/\alpha,$$

as claimed in (1.15).

**Section 8.6**
**Poisson with multinomial delay**. Let $\mathcal{N}$ be Poisson($\lambda$) and suppose $(\mathcal{N}_0, \dots, \mathcal{N}_K)$ given $\mathcal{N} = n$ is multinomial $(n, q_0, \dots, q_K)$. Since $\mathcal{N}_0 + \dots + \mathcal{N}_K = \mathcal{N}$, we have

$$\Pr(\mathcal{N}_0 = n_0, \dots, \mathcal{N}_K = n_K) = \Pr(\mathcal{N}_0 = n_0, \dots, \mathcal{N}_K = n_K | \mathcal{N} = n)\Pr(\mathcal{N} = n)$$

where $n = n_0 + \dots + n_k$. Inserting the expressions for the probabilites on the right yields

$$\Pr(\mathcal{N}_0 = n_0, \dots, \mathcal{N}_K = n_K) = \left(\frac{n!}{n_0! \cdots n_K!} q_0^{n_0} \cdots q_K^{n_K}\right) \times \left(\frac{\lambda^n}{n!}e^{-\lambda}\right).$$

or

$$\Pr(\mathcal{N}_0 = n_0, \dots, \mathcal{N}_K = n_K) = \frac{q_0^{n_0} \cdots q_K^{n_K}}{n_0! \cdots n_K!} \times (\lambda^n e^{-\lambda}),$$

where

$$\lambda^n e^{-\lambda} = (\lambda^{n_0} e^{-\lambda q_0}) \cdots (\lambda^{n_K} e^{-\lambda q_K})$$

since $n_0 + \dots + n_K = n$ and $q_0 + \dots + q_K = 1$. Hence

$$\Pr(\mathcal{N}_0 = n_0, \dots, \mathcal{N}_K = n_K) = \prod_{k=0}^{K} \frac{(q_k\lambda)^{n_k}}{n_k!} e^{-q_k\lambda}$$

which is a product of independent Poisson densities, as claimed in Section 8.6.

**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 you can implement it yourself by means of standard software.

Start by noting that there are $K$ probailities $q_l$ estimated through their logaritms $\alpha_l$. With $n$ years back there are $n$ parameters $\mu_{-s}$, again handled through their logarithms $\beta_s$. Concatenating yields the parameter vector

$$\mathbf{b} = (\alpha_0, \dots, \alpha_{K-1}, \beta_0, \dots, \beta_n)'.$$

of length

$$n_c = K + n.$$

Note that $\mathbf{b}$ is a column vector. It is operated by a so-called design matrix $X = (x_{ij})$ with $n_c$ columns and as many rows as there are observations; i.e.

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

which equals[3]

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

To define $X$ we must run the row index $i = l + s$ over both $l$ and $s$ in some order. At $i = s + l$

$$\begin{aligned} x_{ij} &= 1, \quad &\text{for } i = l, K + s, \\ x_{ij} &= 0, \quad &\text{otherwise.} \end{aligned}$$

Finally, the observed counts must be concatenated into a single vector $\mathbf{y}$ with $N_{-sl}$ at row $i = s + l$ in the order consistent with that used for $X$.

When $\mathbf{y}$ and $X$ is entered a Poisson regression program, likelihood estimates $\hat{\mathbf{b}}$ of $\mathbf{b}$ are produced. It might in practice well be that $n < K$. If so, it is imposiible to estimate the entire sequence $\{q_l\}$ and simplifying formulations must be introduced.

## 1.8 Further reading

## 1.9 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 $\mu$ 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**.

[3] There is a single observation for the last year ($s = 1$), two for the next to-last-year ($s = 2$), three for the year before that and so on and no more than $K$ observations for a single $s$.

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 $\mu$ is constant. **a)** Explain that $X_1, X_2, \ldots$ are independent and exponentially distributed. Let $S_i = X_1 + \ldots + X_i$ be the time of the $i$'th event with $S_0 = 0$. **b)** Argue that

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

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 $X_1^* + \ldots + X_{i^*}^* \leq t$.

**Exercise 8.2.3** Consider an insurance portfolio with $J$ policies that all generate claims according to a Poisson point process with fixed intensity $\mu$. **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 $\gamma$ is the loading and $\xi$ 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?

**Section 8.3**

**Exercise 8.3.1** Let $N_1, N_2 \ldots$ be stochastically independent and Poisson distributed with common parameter $\eta$. **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 $\lambda$ becomes normal as $\lambda \to \infty$.

**Exercise 8.3.2** Consider the estimate $\hat{\mu} = (n_1 + \ldots + n_n)/(T_1 + \ldots + T_n)$ in (1.5) where $n_1, \ldots n_n$ are stochastically independent observations, the $j'th$ being Poisson distributed with parameter $\mu T_j$. **a)** Show that $\hat{\mu}$ is unbiased with standard deviation (1.5). Suppose the intensity for $n_j$ is $\mu_j$, 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 8.3.3** 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 | $\geq 5$ |
| 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 $\lambda$ if we postulate that the number of accidents is Poisson distributed with parameter $\lambda$. [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 8.3.4 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 8.4**

**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 (1.5) 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 $\mu$ and $\mu'$ 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** let $\mathcal{N}$ be the total number of claims against a portfolio of $J$ policies with identical claim intensity $\mu$, 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($\alpha$)? **b)** Determine the mean and standard deviation of $\mathcal{N}$. **c)** Calculate $\mathrm{sd}(\mathcal{N})/E(\mathcal{N})$ and comment on how it depends on $J$. What happens as $J \to \infty$?

**Exercise 8.5.2** This is a continuation of Exercise 8.3.3. A reasonable addition to the Poisson model presented there is to assume that the parameter $\lambda$ is drawn randomly and independently for each woman. Assume the underlying distribution to be Gamma. Each of the the $n = 648$ observed counts $N$ is then Poisson distributed given $\lambda$ where $\lambda = \xi Z$ and $Z$ is Gamma($\alpha$). **a)** What is the interpretation of the parameters $\xi$ and $\alpha$ 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 8.5.3** Suppose $N|\mu$ is Poisson with parameter $\mu T$ with two models under consideration for $\mu$. Either take $\mu = \xi Z$ with $Z \sim$ Gamma($\alpha$) or $\mu = \beta \exp(\tau\varepsilon)$ with $\varepsilon \sim N(0,1)$. **a)** Show that $E(\mu)$ and $\mathrm{sd}(\mu)$ 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 $\mathrm{sd}(N)$ then are the same too. **c)** Determine $\beta$ and $\tau$ 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 8.5.4** Let $\hat{\mu}_k = \mathcal{N}_k/\mathcal{T}_k$ be the estimate of the intensity $\mu_k$ in year $k$ where $\mathcal{N}_k$ is the number of observed claims and $\mathcal{T}_k$ the total time of exposure; see (1.5). The estimates are available for $K$ years. Suppose all $\mu_k$ have been randomly drawn independently of each other with $E(\mu_k) = \xi$ and $\mathrm{sd}(\mu_k) = \sigma_\mu$ (same for all $k$). If we ignore randomness in $\mu_k$, our natural estimate of a common $\mu$ 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

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

[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 8.6**

**Exercise 8.6.1** 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 8.6.2** In the situation in Section 8.6 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.