Non-Life Insurance: Mathematics and Statistics

Solution sheet 8

Solution 8.1 Panjer Algorithm

For the expected yearly claim amount π_0 we have

$$\pi_0 = \mathbb{E}[S] = \mathbb{E}[N]\,\mathbb{E}[Y_1] = 1 \cdot \mathbb{E}[k + Z] = k + \mathbb{E}[Z] = k + \exp\left\{\mu + \frac{\sigma^2}{2}\right\} \approx 4123.872.$$

Let Y_i^+ denote the discretized claim sizes using a span of $s = 10$, where we put all the probability mass to the upper end of the intervals. If we write $g_m = \mathbb{P}[Y_1^+ = sm]$ for $m \in \mathbb{N}$, then we have

$$g_1 = g_2 = \cdots = g_{10} = 0,$$

since $\mathbb{P}[Y_1^+ \le k] = \mathbb{P}[Z \le 0] = 0$ and $k = 10s$. For all $l \ge 11$, we get

$$\begin{aligned} g_l &= \mathbb{P}[Y_1^+ = sl] \\ &= \mathbb{P}[Y_1^+ = k + s(l-10)] \\ &= \mathbb{P}[k + s(l-11) < Y_1 \le k + s(l-10)] \\ &= \mathbb{P}[Y_1 \le k + s(l-10)] - \mathbb{P}[Y_1 \le k + s(l-11)] \\ &= \mathbb{P}[Z \le s(l-10)] - \mathbb{P}[Z \le s(l-11)] \\ &= \mathbb{P}\left[\log Z \le \log(s[l-10])\right] - \mathbb{P}\left[\log Z \le \log(s[l-11])\right] \\ &= \Phi\left(\frac{\log[s(l-10)] - \mu}{\sigma}\right) - \Phi\left(\frac{\log[s(l-11)] - \mu}{\sigma}\right), \end{aligned}$$

where Φ is the distribution function of the standard Gaussian distribution and where we define $\log 0 = -\infty$. From now on we will replace the claim sizes Y_i with the discretized claim sizes Y_i^+. In particular, we will still write S for the yearly claim amount that changed to

$$S = \sum_{i=1}^{N} Y_i^+.$$

Note that $N \sim \text{Poi}(1)$ has a Panjer distribution with parameters $a = 0$ and $b = 1$, see the proof of Lemma 4.7 of the lecture notes. Applying the Panjer algorithm given in Theorem 4.9 of the lecture notes, we have for $r \in \mathbb{N}_0$

$$f_r \stackrel{\text{def.}}{=} \mathbb{P}[S = sr] = \begin{cases} \mathbb{P}[N = 0] & \text{for } r = 0, \\ \sum_{l=1}^{r} \frac{l}{r} g_l f_{r-l} & \text{for } r > 0. \end{cases}$$

Since the yearly amount that the client has to pay by himself is given by

$$S_{\text{ins}} = \min\{S, d\} + \min\{\alpha \cdot (S-d)_+, M\} = \min\{S, d\} + \alpha \cdot \min\left\{(S-d)_+, \frac{M}{\alpha}\right\},$$

$M/\alpha = 7\text{'}000$ and the maximal possible franchise is 2'500, we have to apply the Panjer algorithm until we reach $\mathbb{P}[S = 9\text{'}500] = f_{950}$. Here we limit ourselves to determine the values of $f_0, \ldots, f_{12}$ to illustrate how the algorithm works. In particular, we have

$$f_0 = \mathbb{P}[N = 0] = e^{-1} \approx 0.36$$

and

$$f_1 = f_2 = \cdots = f_{10} = 0,$$

since $g_1 = g_2 = \cdots = g_{10} = 0$. For $r = 11$ and $r = 12$, we get

$$f_{11} = \sum_{l=1}^{11} \frac{l}{11} g_l f_{11-l} = g_{11} f_0 = \left[\Phi\left(\frac{\log s - \mu}{\sigma}\right) - \Phi\left(\frac{\log 0 - \mu}{\sigma}\right)\right] e^{-1} \approx 7.089 \cdot 10^{-9}$$

and

$$f_{12} = \sum_{l=1}^{12} \frac{l}{12} g_l f_{12-l} = g_{12} f_0 = \left[\Phi\left(\frac{\log 2s - \mu}{\sigma}\right) - \Phi\left(\frac{\log s - \mu}{\sigma}\right)\right] e^{-1} \approx 2.786 \cdot 10^{-7}.$$

Using the discretized claim sizes, the yearly expected amount π_{ins} paid by the client is given by

$$\pi_{\text{ins}} = \mathbb{E}[S_{\text{ins}}] = \mathbb{E}\left[\min\{S, d\}\right] + \alpha\, \mathbb{E}\left[\min\left\{(S-d)_+, \frac{M}{\alpha}\right\}\right],$$

where we have

$$\mathbb{E}\left[\min\{S, d\}\right] = \sum_{r=0}^{d/s} f_r s r + d\left(1 - \sum_{r=0}^{d/s} f_r\right) = d + \sum_{r=0}^{d/s} f_r (sr - d)$$

and

$$\begin{aligned}\mathbb{E}\left[\min\left\{(S-d)_+, \frac{M}{\alpha}\right\}\right] &= \sum_{r=d/s+1}^{d/s+M/s\alpha} f_r (sr - d) + \frac{M}{\alpha}\left(1 - \sum_{r=0}^{d/s+M/\alpha} f_r\right) \\ &= \frac{M}{\alpha} + \sum_{r=d/s+1}^{d/s+M/s\alpha} f_r\left(sr - d - \frac{M}{\alpha}\right) - \frac{M}{\alpha}\sum_{r=0}^{d/s} f_r.\end{aligned}$$

Therefore, we get

$$\begin{aligned}\pi_{\text{ins}} &= d + \sum_{r=0}^{d/s} f_r (sr - d) + \alpha\left[\frac{M}{\alpha} + \sum_{r=d/s+1}^{d/s+M/s\alpha} f_r\left(sr - d - \frac{M}{\alpha}\right) - \frac{M}{\alpha}\sum_{r=0}^{d/s} f_r\right] \\ &= d + M + \sum_{r=0}^{d/s} f_r (sr - d - M) + \sum_{r=d/s+1}^{d/s+M/s\alpha} \alpha f_r\left(sr - d - \frac{M}{\alpha}\right).\end{aligned}$$

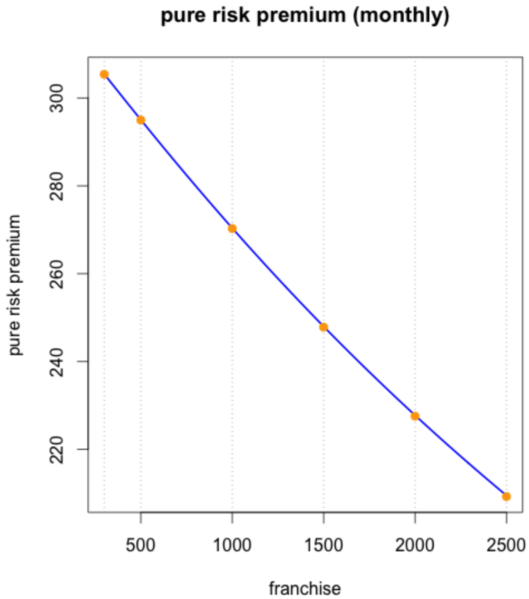
Finally, if the client has chosen franchise d, then the monthly pure risk premium π is given by

$$\begin{aligned}\pi &= \frac{\pi_0 - \pi_{\text{ins}}}{12} \\ &= \frac{1}{12}\left[k + \exp\left\{\mu + \frac{\sigma^2}{2}\right\} - d - M - \sum_{r=0}^{d/s} f_r (sr - d - M) - \sum_{r=d/s+1}^{d/s+M/s\alpha} \alpha f_r\left(sr - d - \frac{M}{\alpha}\right)\right].\end{aligned}$$

In the end, we get the following monthly pure risk premiums for the different franchises:

d	300	500	1’000	1’500	2’000	2’500
π	307	297	274	253	233	216

More generally, the monthly pure risk premium as a function of the franchise, which is allowed to vary between 300 CHF and 2’500 CHF, looks as follows:



Note that the above values only represent the pure risk premiums. In order get the premiums that the customer has to pay in the end, we would need to add an appropriate risk-loading, which may vary between different health insurance companies. The above plot can be created by the R-code given below, where we calculated the premiums using two different discretizations of the claim sizes: in one we put the probability mass to the upper end of the intervals and in the other to the lower end of the intervals. However, the resulting premiums for these two versions are basically the same.

```
### Define the function KK_premium with the variables:
### lambda = mean number of claims
### mu = mean parameter of log-normal distribution
### sigma2 = variance parameter of log-normal distribution
### span = span size used in the Panjer algorithm
### shift = shift of the translated log-normal distribution
KK_premium <- function(lambda, mu, sigma2, span, shift){
  ### we will calculate the distribution of S until M (M = 2500 + 7000)
  M <- 9500

  ### number of steps
  m <- M/span

  ### we won't have any mass until we reach shift, which happens at the k0-th step
  k0 <- shift/span

  ### initialize array where mass is put to the lower end of the interval
  g_min <- array(0, dim=c(m+1,1))

  ### initialize array where mass is put to the upper end of the interval
  g_max <- array(0, dim=c(m+1,1))
```

```

### discretize the log-normal distribution putting the mass to the lower end of the interval
for (k in (k0+1):(m+1)){g_min[k,1] <- pnorm(log((k-k0)*span), mean=mu, sd=sqrt(sigma2))-pnorm(log((k-k0-1)*span), mean=mu, sd=sqrt(sigma2))}

### discretize the log-normal distribution putting the mass to the upper end of the interval
g_max[2:(m+1),1] <- g_min[1:m,1]

### initialize matrix, where we will store the probability distribution of S
f1 <- matrix(0, nrow=m+1, ncol=3)

### store the probability of getting zero claims (in both lower bound and upper bound)
f1[1,1] <- exp(-lambda*(1-g_min[1,1]))
f1[1,2] <- exp(-lambda*(1-g_max[1,1]))

### calculate the values "l * g_{l}" of the discretized claim sizes (lower bound and upper bound), we need these values in the Panjer algorithm
h1 <- matrix(0, nrow=m, ncol=3)
for (i in 1:m){
  h1[i,1] <- g_min[i+1,1]*(i+1)
  h1[i,2] <- g_max[i+1,1]*(i+1)
}

### Panjer algorithm (note that in the Poisson case we have a = 0 and b = lambda*v, which is just lambda here)
for (r in 1:m){
  f1[r+1,1] <- lambda/r*(t(f1[1:r,1])%*%h1[r:1,1])
  f1[r+1,2] <- lambda/r*(t(f1[1:r,2])%*%h1[r:1,2])
  f1[r+1,3] <- r * span
}

### maximal and minimal franchise
m1 <- 2500
m0 <- 300

### number of iterations needed to get to m1 and m0
i1 <- m1/span+1
i0 <- m0/span+1

### calculate the part that the insured pays by himself
franchise <- array(NA, c(i1, 3))
for (i in i0:i1){
  franchise[i,1] <- f1[i,3]   ### this represents the franchise
  franchise[i,2] <- sum(f1[1:i,1]*f1[1:i,3]) + f1[i,3] * (1-sum(f1[1:i,1]))
  franchise[i,2] <- franchise[i,2] + sum(f1[(i+1):(i+7000/span)
```

```
      ,1]*f1[2:(7000/span+1),3])*0.1 + 700 * (1-sum(f1[1:(i+7000/
      span),1]))
    franchise[i,3] <- sum(f1[1:i,2]*f1[1:i,3]) + f1[i,3] * (1-sum(
      f1[1:i,2]))
    franchise[i,3] <- franchise[i,3] + sum(f1[(i+1):(i+7000/span)
      ,2]*f1[2:(7000/span+1),3])*0.1 + 700 * (1-sum(f1[1:(i+7000/
      span),2]))
  }

  ### calculate the price of the monthly premium
  price <- array(NA, c(i1, 3))
  price[,1] <- franchise[,1]    ### this represents the franchise
  price[,2:3] <- (lambda*(exp(mu+sigma2/2)+shift) - franchise
     [,2:3])/12
  price
}

### Load the add-on packages stats and MASS
require(stats)
require(MASS)

### Determine values for the input parameters of the function KK_
    premium
lambda <- 1
mu <- 7.8
sigma2 <- 1
span <- 10
shift <- 100

### The coefficient of variation of the translated log-normal
    distribution is given by
exp(mu+sigma2/2)*sqrt(exp(sigma2)-1)/(shift+exp(mu+sigma2/2))

### Run the function KK_premium
price <- KK_premium(lambda, mu, sigma2, span, shift)

### Plot the monthly pure risk premium as a function of the
    franchise
plot(x=price[,1], y=price[,2], lwd=2, col="blue", type='l', ylab="
    pure risk premium", xlab="franchise", main="pure risk premium (
    monthly)")
lines(x=price[,1], y=price[,2], lwd=1, col="blue")
points(x=c(300,500, 1000, 1500, 2000, 2500), y=price[c(300,500,
    1000, 1500, 2000, 2500)/span+1,3], pch=19, col="orange")
abline(v=c(300, 500, 1000, 1500, 2000, 2500), col="darkgray", lty
    =3)

### Give the monthly pure risk premiums for the six franchises
    listed on the exercise sheet
round(price[c(300,500, 1000, 1500, 2000, 2500)/span+1,2])
round(price[c(300,500, 1000, 1500, 2000, 2500)/span+1,3])
```

Solution 8.2 Variance Loading Principle

(a) Let S_1, S_2, S_3 be the total claim amounts of the passenger cars, delivery vans and trucks, respectively. Then, according Proposition 2.11 of the lecture notes, for the expected total claim amounts we have

$$\mathbb{E}[S_i] = \lambda_i v_i \, \mathbb{E}\left[Y_1^{(i)}\right],$$

for all $i \in \{1, 2, 3\}$. Using the data given in the table on the exercise sheet, we get

$$\mathbb{E}[S_1] = 0.25 \cdot 40 \cdot 2'000 = 20'000,$$
$$\mathbb{E}[S_2] = 0.23 \cdot 30 \cdot 1'700 = 11'730 \quad \text{and}$$
$$\mathbb{E}[S_3] = 0.19 \cdot 10 \cdot 4'000 = 7'600.$$

If we write S for the total claim amount of the car fleet, we can conclude that

$$\mathbb{E}[S] = \mathbb{E}[S_1 + S_2 + S_3] = \mathbb{E}[S_1] + \mathbb{E}[S_2] + \mathbb{E}[S_3] = 39'330.$$

(b) Again using Proposition 2.11 of the lectures notes, we get

$$\mathrm{Var}[S_i] = \lambda_i v_i \, \mathbb{E}\left[\left(Y_1^{(i)}\right)^2\right] = \lambda_i v_i \left(\mathrm{Var}\left(Y_1^{(i)}\right) + \mathbb{E}\left[Y_1^{(i)}\right]^2\right) = \lambda_i v_i \, \mathbb{E}\left[Y_1^{(i)}\right]^2 \left(\mathrm{Vco}(Y_1^{(i)})^2 + 1\right),$$

for all $i \in \{1, 2, 3\}$. Using the data given in the table on the exercise sheet, we find

$$\mathrm{Var}(S_1) = 0.25 \cdot 40 \cdot 2'000^2(2.5^2 + 1) = 290'000'000,$$
$$\mathrm{Var}(S_2) = 0.23 \cdot 30 \cdot 1'700^2(2^2 + 1) = 99'705'000 \quad \text{and}$$
$$\mathrm{Var}(S_3) = 0.19 \cdot 10 \cdot 4'000^2(3^2 + 1) = 304'000'000.$$

Since S_1, S_2 and S_3 are independent by assumption, we get for the variance of the total claim amount S of the car fleet

$$\mathrm{Var}(S) = \mathrm{Var}(S_1) + \mathrm{Var}(S_2) + \mathrm{Var}(S_3) = 693'705'000.$$

Using the variance loading principle with $\alpha = 3 \cdot 10^{-6}$, we get for the premium π of the car fleet

$$\pi = \mathbb{E}[S] + \alpha \mathrm{Var}(S) = 39'330 + 3 \cdot 10^{-6} \cdot 693'705'000 \approx 39'330 + 2'081 = 41'411.$$

Note that we have

$$\frac{\pi - \mathbb{E}[S]}{\mathbb{E}[S]} = \frac{\alpha \mathrm{Var}(S)}{\mathbb{E}[S]} \approx \frac{2'081}{39'330} \approx 5.3\%.$$

Thus, the loading $\pi - \mathbb{E}[S]$ is given by 5.3% of the pure risk premium.

Solution 8.3 Panjer Distribution

If we write

$$p_k = \mathbb{P}[N = k]$$

for all $k \in \mathbb{N}$, then, by definition of the Panjer distribution, we have

$$p_k = p_{k-1}\left(a + \frac{b}{k}\right),$$

for all k in the range of N. We can use this recursion to calculate $\mathbb{E}[N]$ and $\mathrm{Var}(N)$. Note that the range of N is $\mathbb{N}$ if $a \geq 0$ and it is $\{0, 1, \ldots, n\}$ for some $n \in \mathbb{N}_{\geq 1}$ if $a < 0$.

First, we consider the case where $a < 0$, i.e. where the range of N is $\{0, 1, \ldots, n\}$. According to the proof of Lemma 4.7 of the lecture notes, we have

$$n = -\frac{a+b}{a}. \tag{1}$$

For the expectation of N, we get

$$\begin{aligned}
\mathbb{E}[N] &= \sum_{k=0}^{n} k\, p_k \\
&= \sum_{k=1}^{n} k\, p_k \\
&= \sum_{k=1}^{n} k\, p_{k-1} \left(a + \frac{b}{k}\right) \\
&= a \sum_{k=1}^{n} k\, p_{k-1} + b \sum_{k=1}^{n} p_{k-1} \\
&= a \sum_{k=0}^{n-1} (k+1)\, p_k + b \sum_{k=0}^{n-1} p_k \\
&= a \sum_{k=0}^{n-1} k\, p_k + (a+b) \sum_{k=0}^{n-1} p_k \\
&= a\, (\mathbb{E}[N] - np_n) + (a+b)(1-p_n) \\
&= a\, \mathbb{E}[N] + a + b + p_n(-an - a - b).
\end{aligned}$$

Using (1), we get

$$-an - a - b = a\,\frac{a+b}{a} - a - b = 0. \tag{2}$$

Hence, the above expression for $\mathbb{E}[N]$ simplifies to

$$\mathbb{E}[N] = a\mathbb{E}[N] + a + b,$$

from which we can conclude that

$$\mathbb{E}[N] = \frac{a+b}{1-a}.$$

In order to get the variance of N, we first calculate the second moment of N:

$$\begin{aligned}
\mathbb{E}[N^2] &= \sum_{k=0}^{n} k^2\, p_k \\
&= \sum_{k=1}^{n} k^2\, p_k \\
&= \sum_{k=1}^{n} k^2\, p_{k-1} \left(a + \frac{b}{k}\right) \\
&= a \sum_{k=1}^{n} k^2\, p_{k-1} + b \sum_{k=1}^{n} k\, p_{k-1} \\
&= a \sum_{k=0}^{n-1} (k+1)^2\, p_k + b \sum_{k=0}^{n-1} (k+1)\, p_k \\
&= a \sum_{k=0}^{n-1} k^2\, p_k + (2a+b) \sum_{k=0}^{n-1} k\, p_k + (a+b) \sum_{k=0}^{n-1} p_k \\
&= a\,(\mathbb{E}[N^2] - n^2 p_n) + (2a+b)(\mathbb{E}[N] - np_n) + (a+b)(1-p_n) \\
&= a\,\mathbb{E}[N^2] + (2a+b)\,\mathbb{E}[N] + a + b + p_n[-an^2 - (2a+b)n - a - b].
\end{aligned}$$

Using (1), we get

$$\begin{aligned}
-an^2 - (2a+b)n - a - b &= -a\left(\frac{a+b}{a}\right)^2 + (2a+b)\frac{a+b}{a} - a - b \\
&= -\frac{a^2 + 2ab + b^2}{a} + \frac{2a^2 + 3ab + b^2}{a} - \frac{a^2 + ab}{a} \\
&= 0.
\end{aligned} \tag{3}$$

Hence, the above expression for $\mathbb{E}[N^2]$ simplifies to

$$\mathbb{E}[N^2] = a\,\mathbb{E}[N^2] + (2a+b)\,\mathbb{E}[N] + a + b,$$

from which we get

$$\begin{aligned}
\mathbb{E}[N^2] &= \frac{(2a+b)\,\mathbb{E}[N] + a + b}{1-a} \\
&= \frac{(2a+b)\,(a+b) + (a+b)(1-a)}{(1-a)^2} \\
&= \frac{2a^2 + 3ab + b^2 + a - a^2 + b - ab}{(1-a)^2} \\
&= \frac{(a+b)^2 + a + b}{(1-a)^2}.
\end{aligned}$$

Finally, the variance of N then is

$$\mathrm{Var}(N) = \mathbb{E}[N^2] - \mathbb{E}[N]^2 = \frac{(a+b)^2 + a + b}{(1-a)^2} - \frac{(a+b)^2}{(1-a)^2} = \frac{a+b}{(1-a)^2}.$$

In the case where $a \geq 0$, i.e. where the range of N is $\mathbb{N}$, we can perform analogous calculations with the only difference that the index of summation in all the sums involved goes up to ∞ instead of stopping at n. As a consequence, the calculations in (2) and in (3) aren't necessary anymore. The formulas for $\mathbb{E}[N]$ and $\mathrm{Var}(N)$, however, remain the same.

The ratio of $\mathrm{Var}(N)$ to $\mathbb{E}[N]$ is given by

$$\frac{\mathrm{Var}(N)}{\mathbb{E}[N]} = \frac{a+b}{(1-a)^2}\,\frac{1-a}{a+b} = \frac{1}{1-a}.$$

Note that if $a < 0$, i.e. if N has a binomial distribution, we have $\mathrm{Var}(N) < \mathbb{E}[N]$. If $a = 0$, i.e. if N has a a Poisson distribution, we have $\mathrm{Var}(N) = \mathbb{E}[N]$. Finally, in the case of $a > 0$, i.e. for a negative-binomial distribution, we have $\mathrm{Var}(N) > \mathbb{E}[N]$.