8

Poisson Random Measures in Collective Risk Theory

In Chapter 7 we collected the basic notions of point process theory. We have focused on Poisson random measures (PRMs) and their properties. In the present chapter we would like to apply the theory developed there, to models from collective risk theory. In particular, we will make considerable use of the marking and transformation techniques of PRMs introduced in Section 7.3, and we will intensively exploit the independence of Poisson claim numbers and Poisson integrals on disjoint parts of the time-claim size space. In Section 8.1, we consider different decompositions of the time-claim size space, such as decomposition by claim size, year of occurrence, year of reporting, etc. In Section 8.2, we study a major generalization of the Cramér-Lundberg model, called the *basic model*, which accounts for delays in reporting, claim settlements, as well as the payment process in the settlement period of the claim. We also decompose the time-claim size space into its basic ingredients, resulting in settled, incurred but not reported, and reported but not settled claims. We study the distributions of the corresponding claim numbers and total claim amounts.

This chapter was inspired by the ideas in Norberg's [114] article on point process techniques for non-life insurance.

8.1 Decomposition of the Time-Claim Size Space

The aim of this section is to decompose the time-claim size space in various ways into disjoint subsets. The resulting claim numbers and total claim amounts on the subspaces will be independent due to the underlying PRM structure. We will determine the distributions of these independent quantities.

8.1.1 Decomposition by Claim Size

Assume that claims arrive at times T_i according to PRM(μ) on the state space $(0,\infty)$, independently of the iid one-dimensional claim sizes X_i with common

T. Mikosch, *Non-Life Insurance Mathematics*, Universitext
DOI 10.1007/978-3-540-88233-6_8,

distribution F on $(0,\infty)$. We know from Section 7.3.2 that the points (T_i, X_i) constitute PRM$(\mu \times F)$, denoted by N, on the state space $E = (0,\infty)^2$.

For $0 = c_0 < c_1 < \cdots < c_m < \infty$ and $m \geq 1$, the claim size layers

$$A_i = (c_{i-1}, c_i]\,, \quad i = 1, \ldots, m\,, \quad \text{and} \quad A_{m+1} = (c_m, \infty)\,,$$

are disjoint. For a given period of time $(a,b]$ for some $0 \leq a < b < \infty$, such as a year, a quarter, a month, etc., the claim numbers

$$N_i = N((a,b] \times A_i)\,, \qquad i = 1, \ldots, m+1\,,$$

are mutually independent Poisson random variables. In particular, N_i has a Pois$(\mu(a,b]\, F(A_i))$ distribution. This follows from the defining properties of a PRM.

By Lemma 7.2.10, the corresponding total claim amounts

$$\begin{aligned} S_i &= \int_{(a,b]\times A_i} x\, N(dt, dx) \\ &= \int_E x\, I_{(a,b]\times A_i}((t,x))\, N(dt,dx) \\ &= \sum_{j: a < T_j \leq b\,, X_j \in A_i} X_j\,, \qquad i = 1, \ldots, m+1\,, \end{aligned}$$

are mutually independent, since the integrands $f_i(t,x) = x\, I_{(a,b]\times A_i}((t,x))$ have disjoint support. If we assume that $\mu(a,b] < \infty$, then

$$(\mu \times F)((a,b] \times A_i) = \mu(a,b]\, F(A_i) < \infty\,,$$

and therefore every S_i has compound Poisson representation

$$\mathrm{CP}(\mu(a,b]\, F(A_i), P(X_1 \leq \cdot \mid X_1 \in A_i))\,;$$

see Corollary 7.2.8 or Example 7.3.10.

An important special case corresponds to $m = 1$. Then the claim size space is divided into two layers $A_1 = (0,c]$ and $A_2 = (c,\infty)$, i.e., the portfolio splits into small and large claims. The quantity c can be interpreted as the *deductible by minimum franchise* or *first risk* in direct insurance, or as *retention level* in the context of excess-of-loss reinsurance; see Section 3.4 for terminology on reinsurance treaties. In this case, the primary insurer covers the amount

$$\begin{aligned} \sum_{j: T_j \in (a,b]} \min(X_j, c) &= \int_{(a,b]\times(0,c]} x\, N(dt,dx) + c \int_{(a,b]\times(c,\infty)} N(dt,dx) \\ &= S_1 + S_2\,, \end{aligned}$$

and the reinsurer covers

$$\sum_{j:T_j\in(a,b]} (X_j - c)_+ = \int_{(a,b]\times(c,\infty)} (x-c)\, N(dt,dx) = S_3\,.$$

The claim amounts $S_1 + S_2$ and S_3 are *not* independent, but the amounts S_1 and S_2 constituting the shares of the primary insurer are independent, and so are the claim amounts S_1 (to be paid by the primary insurer) and S_3 (to be paid by the reinsurer).

The situation with *proportional reinsurance* is different. Then the primary insurer covers the amount $p\int_{(a,b]\times(0,\infty)} x\, N(dt,dx)$ and the reinsurer $q\int_{(a,b]\times(0,\infty)} x\, N(dt,dx)$, where $p,q\in(0,1)$ and $p+q=1$. In this case, the two total claim amounts are strongly dependent. Indeed, they are linearly dependent and therefore their correlation is 1.

8.1.2 Decomposition by Year of Occurrence

As in the previous section, we assume that the points (T_i, X_i) in time-claim size space constitute a marked $\mathrm{PRM}(\mu\times F)$, denoted by N, on $(0,\infty)^2$. We also assume that the accounting of the total claim amounts is provided on an annual basis. This means that we decompose time into the mutually disjoint sets (years)

$$A_i = (i-1, i]\,,\quad i = 1,2,\ldots\,.$$

Then it is immediate from the PRM property that the claim numbers $N(A_i\times(0,\infty))$ through the different years A_i are mutually independent and $\mathrm{Pois}(\mu(A_i))$ distributed. In particular, if the points T_j constitute a homogeneous Poisson process, then the distribution of $N(A_i\times(0,\infty))$ does not depend on the year. Similarly, the annual total claim amounts $\int_{A_i\times(0,\infty)} x\, N(dt,dx)$ are independent and have compound Poisson representation $\mathrm{CP}(\mu(A_i),F)$. In particular, for a homogeneous Poisson arrival process, the total claim amounts through the years constitute an iid sequence. These are properties we have already derived in Section 3.3.2. In contrast to that part of the book, the results in this section are simple byproducts of the theory of general Poisson processes.

The top graphs in Figure 8.1.1 show both the annual claim numbers and total claim amounts of the Danish fire insurance data 1980–2002. Claim sizes are evaluated in prices of 2002[1] by using the Danish Consumer Price Index (CPI) which is available from the website of Danmarks Statistik:

www.dst.dk/Statistik/seneste/Indkomst/Priser/FPI_inflation.aspx

The increase of the claim numbers through time can be explained for different reasons. First, not all companies might have reported their claims to the

[1] In Part I of this book we used the Danish fire insurance data 1980–1990 expressed in prices of 1985.

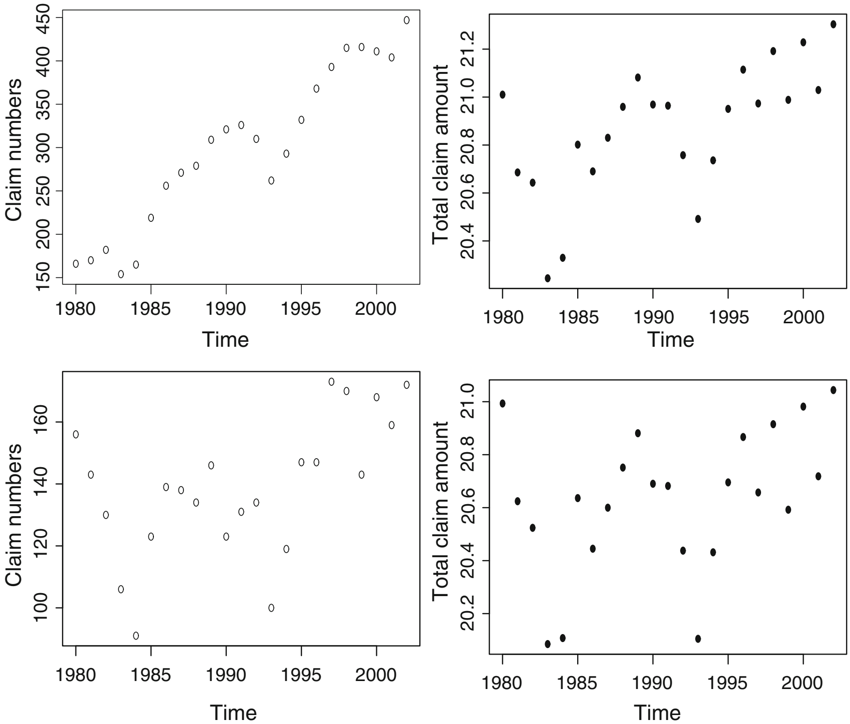

Figure 8.1.1 Top: *The Danish fire insurance data* 1980–2002 *in prices of* 2002*.* Bottom: *The data whose claim size exceeds* 2.244 *million Kroner in* 2002 *prices* (*corresponding to* 1 *million Kroner in* 1980 *prices*). Left column: *The annual claim numbers.* Right column: *The corresponding logarithmic annual total claim amounts. Notice that the bottom graphs are more in agreement with the hypothesis of iid annual claim numbers and total claim amounts than the top graphs.*

authorities in earlier years. Second, only those claims were reported which exceeded the value of 1 million Danish Kroner in the year of reporting. In prices of 2002, this threshold corresponds to 2.244 million Kroner in 1980. This means that many claims were not reported in 1980–2001 due to the use of different thresholds. For example, if the 1980 threshold of 2.244 million Kroner were applied in 2002, 172 out of the 447 reported claims (or 38%) would not be taken into account. Third, fire insurance and prices for buildings are rather closely linked. Therefore the CPI might not be the best indicator for evaluating fire insurance.

In order to show the influence of inflation, in the bottom graphs of Figure 8.1.1 we plot the annual claim numbers and total claim amounts of those claims exceeding 2.244 million Kroner in 2002 prices (1 million Kroner in 1980 prices). The new graphs give the impression that the distributions of the annual claim numbers and total claim amounts do not significantly change through the years, although a slight increase in both categories is plausible. The bottom graphs are more in agreement with the PRM assumption on the claim arrivals and claim sizes than the top graphs, resulting in iid annual claim numbers and total claim amounts.

8.1.3 Decomposition by Year of Reporting

In this section we assume that the ith claim occurs at the time point T_i of a homogeneous Poisson process on $(0,\infty)$ with intensity $\lambda > 0$. The corresponding claim size X_i is reported with delay D_i. In the language of point processes, every arrival T_i is marked with the pair (D_i, X_i) with values in $(0,\infty)^2$ and joint distribution $F_{D,X}$, possibly with dependent components. The sequence of marks (D_i, X_i), $i = 1, 2, \ldots$, constitutes an iid sequence, independent of (T_i). We write F_D for the distribution of D_i and F for the distribution of X_i.

In the remainder of this section we assume independence between D_i and X_i. We know from Example 7.3.9 that the points $(T_i + D_i, X_i)$ constitute $\mathrm{PRM}(\nu \times F)$, denoted by $N_{T+D,X}$, on $(0,\infty)^2$, where

$$\nu(0,t] = \lambda \int_0^t F_D(y)\, dy\,, \quad t \ge 0\,.$$

We split time $(0,\infty)$ into disjoint periods (years say) $A_i = (i-1, i]$, $i = 1, 2, \ldots$. The time component of $N_{T+D,X}$ counts the claims *reported* in A_i; they might have been *incurred* some periods ago. The corresponding pairs of claim numbers and total claim amounts

$$\left(N_{T+D,X}(A_i \times (0,\infty))\,, \int_{A_i \times (0,\infty)} x\, N_{T+D,X}(dt, dx) \right)\,, \quad i = 1, 2, \ldots\,,$$

are mutually independent. The claim number in the period A_i is $\mathrm{Pois}(\nu(A_i))$ distributed, the corresponding claim amount has a $\mathrm{CP}(\nu(A_i), F)$ distribution; see Example 7.3.10.

Write $N_{T,D}$ for the $\mathrm{PRM}(\lambda\, \mathrm{Leb} \times F_D)$ generated by the points (T_i, D_i). The number of claims that were incurred in the ith period but were reported d periods later is given by the quantity

$$\begin{aligned} N_{i,d} &= \#\{j \ge 1 : i-1 < T_j \le i\,, i+d-1 < T_j + D_j \le i+d\} \\ &= N_{T,D}(\{(t,y) : t \in A_i\,, t+y \in A_{i+d}\})\,, \\ &\quad i = 1, 2, \ldots\,, \quad d = 0, 1, \ldots\,. \end{aligned}$$

A straightforward calculation yields that $N_{i,d}$ is Poisson distributed with mean

$$\begin{aligned} EN_{i,d} &= (\lambda\,\mathrm{Leb} \times F_D)(\{(t,y) : t \in A_i\,, t+y \in A_{i+d}\}) \\ &= \lambda \int_{t\in(i-1,i]} \int_{t+y\in(i+d-1,i+d]} F_D(dy)\,dt \\ &= \lambda \int_d^{d+1} [F_D(z) - F_D(z-1)]\,dz\,. \end{aligned}$$

The distribution of $N_{i,d}$ is independent of i due to the homogeneity of the underlying Poisson process with points T_j. For different i, the quantities $N_{i,d}$ arise from disjoint subsets of the state space, hence $N_{i,d}$, $i = 1, 2, \ldots$, are iid. The Poisson property also ensures that the corresponding total claim amounts

$$\sum_{j:i-1<T_j\le i,\, i+d-1<T_j+D_j\le i+d} X_j\,, \quad i = 1, 2, \ldots\,, \tag{8.1.1}$$

are iid compound Poisson sums. It is left as Exercise 3(a) on p. 267 to calculate the parameters of their common distribution.

8.1.4 Effects of Dependence Between Delay in Reporting Time and Claim Size

We assume the conditions of Section 8.1.3, but we allow for possible dependence between the components D_i and X_i of the mutually independent pairs (D_i, X_i). Then the reporting time $T_i + D_i$ of the ith claim depends on the claim size X_i. This assumption can be realistic. For example, a large claim size is more likely to be reported as early as possible than a small claim size. For an illustration of this phenomenon, see Example 8.1.2 below.

The points (T_i, D_i, X_i) constitute a PRM$(\lambda\,\mathrm{Leb} \times F_{D,X})$, denoted by N, where $F_{D,X}$ denotes the joint distribution of (D_i, X_i) on $(0,\infty)^2$. This property implies, in particular, that for disjoint Borel sets $B_i \subset (0,\infty)^3$, the pairs $(N(B_i), \int_{B_i} x\,N(dt,dy,dx))$ of claim numbers and total claim amounts are mutually independent.

For any bounded Borel set $A \subset (0,\infty)^3$ the corresponding total claim amount $\int_A x\,N(dt,dy,dx)$ has CP$((\lambda\,\mathrm{Leb} \times F_{D,X})(A), F_Z)$ distribution given by the distribution function

$$\begin{aligned} F_Z(y) &= \frac{(\mathrm{Leb} \times F_{D,X})(A \cap \{(t,d,x) : x \le y\})}{(\mathrm{Leb} \times F_{D,X})(A)} \\ &= \frac{(\mathrm{Leb} \times F_{D,X})\left(A \cap \left((0,\infty)^2 \times [0,y]\right)\right)}{(\mathrm{Leb} \times F_{D,X})(A)}\,, \quad y > 0\,. \end{aligned}$$

Now specify the set A as follows:

$$A = (t_1, t_2] \times (d_1, d_2] \times (x_1, x_2]\,, \quad 0 < t_1 < t_2\,,\ 0 < d_1 < d_2\,,\ 0 < x_1 < x_2\,.$$

Then $\int_A x\, N(dt, dy, dx)$ has distribution

$$\mathrm{CP}(\lambda\,(t_2 - t_1)\, F_{D,X}((d_1, d_2] \times (x_1, x_2])\,,\ F_Z)$$

with corresponding distribution function

$$\begin{aligned} F_Z(y) &= \frac{F_{D,X}\left((d_1, d_2] \times (x_1, \min(x_2, y)]\right)}{F_{D,X}((d_1, d_2] \times (x_1, x_2])} \\ &= P(X_1 \in (x_1, \min(x_2, y)] \mid D_1 \in (d_1, d_2]\,, X_1 \in (x_1, x_2])\,, \quad y > 0\,. \end{aligned}$$

Example 8.1.2 (Large claims tend to be reported earlier than small ones) If one has more information about the dependence between D_i and X_i one can specify the distribution $F_{D,X}$ in a meaningful way. Norberg [114], p. 112, assumed that the conditional distribution of D_1 given $X_1 = x$, $x > 0$, is $\mathrm{Exp}(x\,\gamma)$ distributed for some positive $\gamma > 0$, and that X_1 has a $\Gamma(\alpha, \beta)$ distribution for some $\alpha, \beta > 0$.

The joint density $f_{D,X}$ of (D_i, X_i) can be calculated from the conditional density $f_D(y \mid X_1 = x)$ and the density f_X of X_1:

$$\begin{aligned} f_{D,X}(y, x) &= \frac{f_{D,X}(y, x)}{f_X(x)}\, f_X(x) = f_D(y \mid X_1 = x)\, f_X(x) \\ &= \left((x\,\gamma)\, \mathrm{e}^{-(x\,\gamma)\,y}\right) \left(\frac{\beta^\alpha}{\Gamma(\alpha)} x^{\alpha-1}\, \mathrm{e}^{-\beta\,x}\right) \\ &= \gamma\, \frac{\beta^\alpha}{\Gamma(\alpha)}\, x^\alpha\, \mathrm{e}^{-x\,(\gamma\,y + \beta)}\,, \qquad x, y \ge 0\,. \end{aligned} \tag{8.1.2}$$

The rationale for the choice of the density $f_{D,X}$ in (8.1.2) is that a large claim size $X_1 = x$ will increase the parameter of the exponential distribution $P(D_1 \le y \mid X_1 = x)$, hence large claims will tend to be reported faster than small claims. This fact is also immediate from the following comparison of the tails: for $0 < x_1 < x_2$,

$$P(D_1 > y \mid X_1 = x_1) = \mathrm{e}^{-(x_1\gamma)\,y} > \mathrm{e}^{-(x_2\gamma)\,y} = P(D_1 > y \mid X_1 = x_2)\,.$$

Integration with respect to x yields the density f_D of D_1:

$$\begin{aligned} f_D(y) &= \int_0^\infty f_{D,X}(y, x)\, dx \\ &= \gamma\, \frac{\beta^\alpha}{\Gamma(\alpha)}\, \frac{\Gamma(\alpha+1)}{(\gamma\, y + \beta)^{\alpha+1}} \int_0^\infty \frac{(\gamma\, y + \beta)^{\alpha+1}}{\Gamma(\alpha+1)} x^\alpha\, \mathrm{e}^{-x\,(\gamma\,y+\beta)}\, dx \\ &= \frac{\gamma\,\alpha\,\beta^\alpha}{(\gamma\,y + \beta)^{\alpha+1}}\,, \quad y \ge 0\,. \end{aligned}$$

This means that the distribution of D_1 is in the location-scale family of a Pareto distribution with tail parameter $\alpha > 0$. This result is surprising: from the forms of the conditional density $f_D(y \mid X_1 = x)$ and the density f_X, it is difficult to guess that a Pareto distributed delay time D_1 appears. Since F_D is heavy-tailed (see Section 3.2.5), it is not unlikely that some claims will be reported with a very long delay. □

8.1.5 Effects of Inflation and Interest

We again assume the conditions of Section 8.1.3. Then the reporting times $T_i + D_i$ of the claims constitute PRM(ν), denoted by N_{T+D}, with mean measure given by $\nu(0,t] = \lambda \int_0^t F_D(y)\,dy$, $t > 0$. We assume independence between (X_i) and (D_i). Therefore the points $(T_i + D_i, X_i)$ constitute PRM($\nu \times F$), denoted by $N_{T+D,X}$, on the state space $(0,\infty)^2$.

Let $f(y,x)$ be a non-negative measurable function on $\mathbb{R} \times (0,\infty)$ such that $f(y,x) = 0$ for $y < 0$. We consider the stochastic process

$$\begin{aligned} S(t) &= \int_E f(t-y,x)\,N_{T+D,X}(dy,dx) \\ &= \int_{(0,t]\times(0,\infty)} f(t-y,x)\,N_{T+D,X}(dy,dx) \\ &= \sum_{i=1}^{\infty} f(t-(T_i+D_i),X_i) \\ &= \sum_{i=1}^{N_{T+D}(0,t]} f(t-(T_i+D_i),X_i)\,, \quad t \ge 0\,. \end{aligned} \tag{8.1.3}$$

If we further specify

$$f(y,x) = \mathrm{e}^{-r\,y}\, I_{(0,\infty)}(y)\, x$$

for some $r \in \mathbb{R}$, we obtain

$$S(t) = \sum_{i=1}^{N_{T+D}(0,t]} \mathrm{e}^{-r\,(t-(T_i+D_i))}\, X_i\,, \quad t \ge 0\,. \tag{8.1.4}$$

If r is positive, then we can interpret r as the *inflation rate*. Assume that $t > 0$ is present time. Then the value of the claim size X_i which was reported at time $T_i + D_i$ in the past has the discounted value $\mathrm{e}^{-r\,(t-(T_i+D_i))}\, X_i$ in terms of present prices. If r is negative we can interpret r as *interest rate*. Then the present value of a payment X_i made at time $T_i + D_i$ in the past is given by the increased amount $\mathrm{e}^{-r\,(t-(T_i+D_i))}\, X_i$ due to compounded interest.

The stochastic process S considered in (8.1.4) is a modification of the total claim amount process in the Cramér-Lundberg model; the latter process is obtained by choosing $r = 0$ and $D_i = 0$ a.s. In contrast to the compound Poisson process in the original Cramér-Lundberg model, the process (8.1.4) has, in general, neither independent nor stationary increments even if one assumes no delay in reporting, i.e., $D_i = 0$ a.s. However, $S(t)$ has representation as a Poisson integral (8.1.3) and therefore, by Corollary 7.2.8, it has representation as a compound Poisson sum. We leave the verification of the details as Exercise 4 on p. 267.

Exercises

Sections 8.1.2

(1) Consider the situation in Section 8.1.2 from the point of view of a reinsurer who covers the amount $g_i(X_j)$ of any claim size X_j occurring in year i. Here g_i, $i = 1, 2, \ldots$, are non-negative measurable functions on $(0, \infty)$ with the property $0 \le g_i(x) \le x$.
  (a) Show that the reinsurer's annual total claim amounts $R_i = \sum_{j:T_j \in A_i} g_i(X_j)$, $i = 1, 2, \ldots$, are mutually independent.
  (b) Determine the distribution of R_i defined in (a).
  (c) Show that the amounts R_i covered by the reinsurer and $P_i = \sum_{j:T_j \in A_i} (X_j - g_i(X_j))$ covered by the primary insurer in year i are in general dependent. In which circumstances are R_i and P_i independent?

Section 8.1.3

(2) Consider the situation of Section 8.1.3 but assume that the arrival sequence (T_i) is PRM(μ) on $(0, \infty)$ with a positive intensity function $\lambda(t)$, $t \ge 0$. Derive the distributions of the claim number and total claim amount corresponding to the claims reported in the ith period $A_i = (i - 1, i]$.

(3) Assume the conditions of Section 8.1.3 and that D_i and X_i are independent for every i.
  (a) Determine the parameters of the compound Poisson representation of the total claim amounts (8.1.1).
  (b) Determine the joint distribution of the claim numbers

$$\#\{j \ge 1 :\ 0 < T_j \le 1 , \quad d - 1 < T_j + D_j \le d\} , \quad d = 1, 2, \ldots ,$$

  i.e., of those claims which occurred in the first period but were reported $d - 1$ years later. Determine the joint distribution of the corresponding total claim amounts.

Section 8.1.5

(4) Consider the process S in (8.1.4) with $r > 0$ and without delay in reporting, i.e., $D_i = 0$ a.s. Write N_T for the homogeneous Poisson process of the arrivals T_i with intensity $\lambda > 0$ and $N_T(t) = N_T(0, t]$, $t > 0$.

(a) Assume in addition that X_i has finite variance. Show that S neither has independent nor stationary increments on $(0,\infty)$.
(b) For fixed $t>0$ show that $S(t)$ has compound Poisson representation

$$S(t) \stackrel{d}{=} \sum_{i=1}^{N_T(t)} \mathrm{e}^{-r\,t\,U_i}\,X_i\,.$$

Here (U_i) is an iid uniform $\mathrm{U}(0,1)$ sequence and $N_T(t)$, (X_i), (U_i) are mutually independent.
(c) Show that for every $t>0$,

$$\mathrm{e}^{-rt}\sum_{i=1}^{N_T(t)} \mathrm{e}^{\,r\,T_i}\,X_i \stackrel{d}{=} \sum_{i=1}^{N_T(t)} \mathrm{e}^{-r\,T_i}\,X_i\,. \tag{8.1.5}$$

This identity in distribution has an interesting interpretation. We suppose that $r>0$ is the inflation rate in $[0,t]$. First assume that all claim sizes X_i are known at time 0. Then the quantity $\mathrm{e}^{\,r\,T_i}\,X_i$ stands for the value of X_i at time T_i. If we interpret T_i as the time when a payment to the insured is executed, $\mathrm{e}^{\,r\,T_i}\,X_i$ is the amount to be paid at time T_i. The quantity $\sum_{i=1}^{N_T(t)} \mathrm{e}^{\,r\,T_i}\,X_i$ is the total amount of all payments in $[0,t]$ in terms of inflated prices. The amount $\mathrm{e}^{-rt}\sum_{i=1}^{N_T(t)} \mathrm{e}^{\,r\,T_i}\,X_i$ is the corresponding deflated amount in terms of prices at time zero. The right-hand side of (8.1.5) has a different meaning. Here we assume that the claim size X_i occurs at time $T_i\le t$ and $\mathrm{e}^{-r\,T_i}\,X_i$ is its value in terms of prices at time 0. The quantity $\sum_{i=1}^{N_T(t)} \mathrm{e}^{-r\,T_i}\,X_i$ is then the total amount of those claims that were incurred in $[0,t]$ in terms of prices at time 0.

8.2 A General Model with Delay in Reporting and Settlement of Claim Payments

8.2.1 The Basic Model and the Basic Decomposition of Time-Claim Size Space

In this section we study an extension of the basic model used in Part I of the book. We again call it *the basic model*. It is given by the following conditions.

The Basic Model

- The ith claim is associated with the quadruple (T_i,D_i,S_i,X_i). The incident causing the claim arrives at time T_i with size X_i and is reported at time T_i+D_i. In the period $[T_i+D_i,T_i+D_i+S_i]$ the claim is settled, i.e., the amount X_i is paid to the insured.
- The *claim arrival sequence* (T_i) constitutes a homogeneous Poisson process on $(0,\infty)$ with intensity $\lambda>0$.
- The *claim size sequence* (X_i) is iid with common distribution F on $(0,\infty)$.

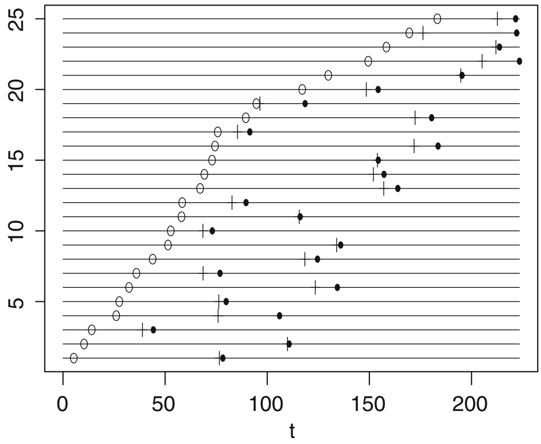

Figure 8.2.1 *Visualization of the time components in the basic model. Each line corresponds to one claim. On the* i*th line, the* claim arrival T_i (left dot)*, the* reporting time T_i+D_i (small vertical line) *and the* time of settlement $T_i+D_i+S_i$ (right bullet) *are shown.*

- The *delay sequence* (D_i) is iid with common distribution F_D on $(0,\infty)$.
- The *duration of settlement sequence* (S_i) is iid with common distribution F_S on $(0,\infty)$.
- The sequences (T_i), (X_i), (D_i) and (S_i) are *mutually independent.*

This is a simple model which takes into account some of the major ingredients of an insurance business. Of course, various of these assumptions deserve some criticism, for example, the homogeneity of the Poisson process, but also the independence of D_i, S_i and X_i. One also needs to specify in which way a claim is settled: one has to define a payment function on the settlement interval $[T_i + D_i, T_i + D_i + S_i]$ which yields the amount X_i at time $T_i + D_i + S_i$. In Section 8.2.3 we will discuss a simplistic payment function, and we will continue in Section 11.3 discussing a more realistic approach.

"More realistic" assumptions lead to a higher theoretical complexity. It is our aim to illustrate the problem of determining the distribution of the total claim amount of a portfolio under the "simple" but "still realistic" assumptions described in the basic model. This "simple model" will already turn out to be sufficiently complex.

Our first observation is that the points (T_i, D_i, S_i, X_i) constitute a

$$\text{marked PRM}(\lambda\,\text{Leb} \times F_D \times F_S \times F) \text{ on the state space } (0,\infty)^4.$$

Indeed, the sequence (T_i) constitutes PRM($\lambda\,$Leb), independent of the iid points (D_i, S_i, X_i) with common distribution $F_D \times F_S \times F$. Then the statement follows from Proposition 7.3.3.

Throughout the section, N denotes the basic process generated by the points (T_i, D_i, S_i, X_i). We will also work with other PRMs derived from N by transformations of its points; see Section 7.3.1 for the theoretical background. We have already introduced the PRM($\nu \times F$) of the points $(T_i + D_i, X_i)$ on $(0,\infty)^2$, denoted by $N_{T+D,X}$, with $\nu(0,t] = \lambda \int_0^t F_D(y)\,dy$, $t > 0$; see Example 7.3.9. We will also work with the PRM generated from the points

$$(T_i, T_i + D_i, T_i + D_i + S_i, X_i)\,.$$

In particular, we will use the fact that the points $(T_i + D_i + S_i, X_i) \in (0,\infty)^2$ constitute PRM($\gamma \times F$), denoted by $N_{T+D+S,X}$, where

$$\gamma(0,t] = \lambda \int_0^t F_{D+S}(y)\,dy = \lambda\, E(t - D_1 - S_1)_+\,, \quad t > 0\,, \qquad (8.2.6)$$

defines the mean measure of the PRM which consists of the points $T_i + D_i + S_i$. Here F_{D+S} is the distribution function of $D_i + S_i$ given by

$$F_{D+S}(y) = \int_0^y F_D(y-s)\,F_S(ds)\,, \quad y \ge 0\,.$$

We leave the verification of (8.2.6) as Exercise 1 on p. 286.

In the context of the basic model, we understand the state space $E = (0,\infty)^4$ of the point process N as the corresponding *time-claim size space*. At a given time $T > 0$ which we interpret as the present time we decompose this space into four disjoint subsets:

$$E = E_{\text{Settled}} \cup E_{\text{RBNS}} \cup E_{\text{IBNR}} \cup E_{\text{Not incurred}}\,.$$

They are characterized as follows.

The Basic Decomposition of Time-Claim Size Space

- The set

$$E_{\text{Settled}} = \{(t,d,s,x) : t + d + s \le T\}$$

  describes the claims which are settled by time T, i.e., the insurance company has paid the amount X_i to the insured by time T.
- The set

$$E_{\text{RBNS}} = \{(t,d,s,x) : t + d \le T < t + d + s\}$$

  describes the claims that have been incurred and are reported by time T, but they are not completely settled, i.e., the payment process for these claims is not finished yet. It is standard to call these claims **R***eported* **B***ut* **N***ot* **S***ettled* or simply *RBNS claims*.

- The set

$$E_{\mathrm{IBNR}} = \{(t, d, s, x) : t \le T < t + d\}$$

describes the claims that have been incurred but have not yet been reported at time T. It is standard to call these claims **I**_ncurred_ **B**_ut_ **N**_ot_ **R**_eported_ or simply _IBNR claims._
- The set

$$E_{\mathrm{Not\,incurred}} = \{(t, d, s, x) : T < t\} .$$

describes the claims that will be incurred after time T.

We notice that $E_{\mathrm{Not\,incurred}}$ contains infinitely many points of the point process N with probability 1. In order to avoid this situation one can consider the insurance business over a finite time horizon, $T \le T_0$, for some $T_0 < \infty$.

Although the sets of the basic decomposition depend on the time T we will often suppress this dependence in the notation.

8.2.2 The Basic Decomposition of the Claim Number Process

As a consequence of the disjointness of the sets in the basic decomposition of the time-claim space the claim number of concern for the insurance business at time T can be decomposed into three mutually independent Poisson numbers

$$\begin{aligned} N(E_{\mathrm{Settled}}) + N(E_{\mathrm{RBNS}}) + N(E_{\mathrm{IBNR}}) &= N((0,T] \times (0,\infty)^3) \\ &= \#\{i \ge 1 : T_i \le T\} . \end{aligned}$$

The points $T_i + D_i + S_i$ constitute PRM(γ) on $(0,\infty)$ with mean measure γ given in (8.2.6). Then the process

$$N(E_{\mathrm{Settled\,by\,time}\,T}) = \#\{i \ge 1 : T_i + D_i + S_i \le T\} , \quad T > 0 , \quad (8.2.7)$$

is inhomogeneous Poisson on $(0,\infty)$ with mean value function $\gamma(T) = \gamma(0,T]$, $T > 0$. In particular, $(N(E_{\mathrm{Settled\,by\,time}\,T}))_{T>0}$ has independent increments.

The processes $(N(E_{\mathrm{RBNS\,at\,time}\,T}))_{T>0}$ and $(N(E_{\mathrm{IBNR\,at\,time}\,T}))_{T>0}$ do not have the property of a Poisson process on $(0,\infty)$. For example, they do not have independent increments; see Exercise 2(a) on p. 286. However, at any fixed time $T > 0$, the random variables $N(E_{\mathrm{RBNS}})$ and $N(E_{\mathrm{IBNR}})$ are Poisson distributed.

Next we collect some characteristic properties of the claim numbers corresponding to the basic decomposition.

Lemma 8.2.2 (Characterization of the Poisson claim numbers of the basic decomposition)
For every $T > 0$ the claim numbers $N(E_{\mathrm{Settled}})$, $N(E_{\mathrm{RBNS}})$ and $N(E_{\mathrm{IBNR}})$ are independent Poisson random variables whose distribution has the following properties.

(1) *The process* $(N(E_{\text{Settled at time } T}))_{T>0}$ *is inhomogeneous Poisson on* $(0,\infty)$ *with mean value function*

$$\gamma(T) = \lambda \int_0^T F_{D+S}(y)\,dy$$
$$= \lambda\, E(T - D_1 - S_1)_+ \,, \quad T > 0\,. \tag{8.2.8}$$

(2) *For every* $T > 0$*, the Poisson random variable* $N(E_{\text{RBNS}})$ *has mean value*

$$\lambda\, E[(T - D_1)_+ - (T - D_1 - S_1)_+]$$
$$= \lambda\, E[S_1\, I_{\{D_1+S_1 \le T\}}] + \lambda\, E[(T - D_1) I_{\{D_1 \le T < D_1 + S_1\}}]\,.$$

(3) *For every* $T > 0$*, the Poisson random variable* $N(E_{\text{IBNR}})$ *has mean value*

$$\lambda\, E[T - (T - D_1)_+] = \lambda\, E[D_1\, I_{\{D_1 \le T\}}] + \lambda\, T\, P(D_1 > T)\,.$$

Proofs. (1) The Poisson process property is immediate from the representation (8.2.7). The mean value function γ was given in (8.2.6).

(2) Since the claim arrival process is homogeneous Poisson with intensity $\lambda > 0$, we observe that for $t > 0$,

$$\lambda\, t = E\#\{i \ge 1 : T_i \le t\} = E\left(\sum_{i=1}^{\infty} I_{(0,t]}(T_i)\right)$$
$$= \sum_{i=1}^{\infty} P(T_i \le t)\,. \tag{8.2.9}$$

A conditioning argument and an application of Fubini's theorem yield the following series of equations:

$$\begin{aligned} EN(E_{\text{RBNS}}) &= E\left(\sum_{i=1}^{\infty} I_{\{T_i + D_i \le T < T_i + D_i + S_i\}}\right) \\ &= \sum_{i=1}^{\infty} P(T_i + D_1 \le T < T_i + D_1 + S_1) \\ &= \sum_{i=1}^{\infty} E\left[\int_{((T-D_1-S_1)_+,(T-D_1)_+]} dP(T_i \le t)\right] \\ &= E\left(\int_{((T-D_1-S_1)_+,(T-D_1)_+]} d(\lambda\, t)\right)\,. \end{aligned}$$

In the last step we used Fubini's theorem and relation (8.2.9). Thus we arrive at the desired relation

$$EN(E_{\mathrm{RBNS}}) = \lambda\, [E(T - D_1)_+ - E(T - D_1 - S_1)_+]$$
$$= \lambda\, E[S_1\, I_{\{D_1+S_1 \le T\}}] + \lambda\, E[(T - D_1) I_{\{D_1 \le T < D_1+S_1\}}]\,.$$

(3) The calculations are similar to part (2). They are left as Exercise 2(b) on p. 286. □

8.2.3 The Basic Decomposition of the Total Claim Amount

In this section we study the total claim amounts at time T corresponding to the different parts in the basic decomposition of the time-claim size space. The total claim amount of the portfolio at time T is given by

$$\begin{aligned} S(T) &= \left(\int_{E_{\mathrm{Settled}}} + \int_{E_{\mathrm{RBNS}}} + \int_{E_{\mathrm{IBNR}}} \right) x\, N(dt, dr, ds, dx) \\ &= S_{\mathrm{Settled}} + S_{\mathrm{RBNS}} + S_{\mathrm{IBNR}} \\ &= \sum_{i: T_i \le T} X_i\,. \end{aligned}$$

As for the claim numbers, we will often suppress the dependence on T in the notation. Since the three Poisson integrals S_{Settled}, S_{RBNS} and S_{IBNR} are defined on disjoint sets of the state space, they are mutually independent. Each of them can be represented as a compound Poisson sum.

The Settled and IBNR Total Claim Amounts

The amount $S_{\mathrm{Settled\,at\,time}\,T}$ is that part of the total claim amount which corresponds to the claims arising from the set $E_{\mathrm{Settled\,at\,time}\,T}$. For the points of the latter set, the amounts X_i have been paid to the insured by time T. Hence the corresponding total claim amount process is given by

$$S_{\mathrm{Settled\ at\ time}\ T} = \sum_{i=1}^{N(E_{\mathrm{Settled\ at\ time\ T}})} X_i\,, \quad T > 0\,. \tag{8.2.10}$$

We know from Lemma 8.2.2(1) that the counting process $(N(E_{\mathrm{Settled\ at\ time\ T}}))$ constitutes an inhomogeneous Poisson process on $(0,\infty)$ with mean value function $\gamma(T) = \lambda\, E(T - D_1 - S_1)_+$, $T > 0$. The process (8.2.10) has independent but, in general, non-stationary increments. For every fixed $T > 0$, S_{Settled} has a $\mathrm{CP}(\gamma(T), F)$ representation. We leave the verification of the details as an exercise.

The IBNR part of the total claim amount by time T is dealt with in a similar way. Since the reporting times occur after time T, the following amount is outstanding:

$$S_{\mathrm{IBNR}} = \int_{E_{\mathrm{IBNR}}} x\, N(dt, dr, ds, dx) = \sum_{i:(T_i, D_i, S_i, X_i) \in E_{\mathrm{IBNR}}} X_i \,.$$

The condition defining E_{IBNR} only restricts the points (T_i, D_i). Therefore the claim number $N(E_{\mathrm{IBNR}})$ with a $\mathrm{Pois}(\lambda\, E(T - (T - D_1)_+))$ distribution is independent of (X_i), and S_{IBNR} has compound Poisson representation $\mathrm{CP}(\lambda\, E(T - (T - D_1)_+), F)$. The process $(S_{\mathrm{IBNR\ at\ time\ } T})_{T>0}$ does not have independent increments since the corresponding counting process does not constitute a Poisson process on $(0, \infty)$; see the discussion before Lemma 8.2.2.

The RBNS Total Claim Amount

For the RBNS part of the liability one has to make some assumptions about the cash flow from the insurer to the insured in the settlement period $[T_i + D_i, T_i + D_i + S_i]$. We assume that at each reporting time $T_i + D_i$ a stochastic (preferably non-decreasing càdlàg) *cash flow* or *payment process* starts which finishes at time $T_i + D_i + S_i$ with the settlement value X_i, i.e., with the actual claim size.

Although a stochastic payment process might be more realistic, we will restrict ourselves to a simplistic cash flow process which equals zero at $T_i + D_i$, is X_i at $T_i + D_i + S_i$ and increases linearly between these two instants of time. Then the settled part of the RBNS claims at time T amounts to

$$\begin{aligned} S_{\mathrm{Settled\,RBNS}} &= \int_{E_{\mathrm{RBNS}}} x\, \frac{T - t - r}{s}\, N(dt, dr, ds, dx) \\ &= \sum_{i:(T_i, D_i, S_i, X_i) \in E_{\mathrm{RBNS}}} X_i\, \frac{T - T_i - D_i}{S_i} \,. \end{aligned} \tag{8.2.11}$$

Since $S_{\mathrm{Settled\,RBNS}}$ is a Poisson integral, it has compound Poisson representation $\mathrm{CP}(EN(E_{\mathrm{RBNS}}), F_Z)$ according to Corollary 7.2.8. We know from Lemma 8.2.2(2) that

$$EN(E_{\mathrm{RBNS}}) = \lambda\, E[(T - D_1)_+ - (T - D_1 - S_1)_+] \,.$$

The integrand $f(t, r, s, x) = x\, s^{-1}\, (T - t - r)$ in the Poisson integral above is rather complex and therefore it seems difficult to evaluate F_Z. Writing

$$\pi = \pi(t, r, s) = s^{-1}\, (T - t - r) \,,$$

we obtain the following formula for $F_Z(y)$ from Corollary 7.2.8 for $y > 0$:

$$F_Z(y) = \frac{(\lambda\, \mathrm{Leb} \times F_D \times F_S \times F)\, (\{(t, d, s, x) : \pi \in [0, 1)\,, x\, \pi \le y\})}{EN(E_{\mathrm{RBNS}})} \,.$$

Here we have made use of the fact that $T_i + D_i \le T < T_i + D_i + S_i$ is equivalent to the fact that

$$\pi_i = S_i^{-1}(T - T_i - D_i) \in [0,1)\,.$$

The evaluation of $F_Z(y)$ simplifies by observing that the points π_i constitute a PRM on the state space $[0,1)$. Indeed, since $\pi\, I_{[0,1)}(\pi)$ is a measurable function of the points (t,r,s), we conclude from the results for the transformed points of a PRM that $\pi_i I_{[0,1)}(\pi_i)$ constitute the points of PRM(α) on $[0,1)$. The mean measure α is given by

$$\begin{aligned}
\alpha[0,z] &= (\lambda\,\mathrm{Leb} \times F_D \times F_S)\left(\{(t,d,s) : 0 \le s^{-1}(T-t-d) \le z\}\right)\\
&= (\lambda\,\mathrm{Leb} \times F_D \times F_S)\left(\{(t,d,s) : t+d \le T \le t+d+zs\}\right)\\
&= E\left(\#\{i \ge 1 : T_i + D_i \le T \le T_i + D_i + z\,S_i\}\right)\,, \quad z \in [0,1]\,.
\end{aligned}$$

The same calculations as in the proof of Lemma 8.2.2(2) yield that

$$\alpha[0,z] = \lambda\,E[(T-D_1)_+ - (T - D_1 - zS_1)_+]\,, \quad z \in [0,1]\,. \tag{8.2.12}$$

Also notice that $\alpha[0,1] = EN(E_{\mathrm{RBNS}})$, and

$$\widetilde{\alpha}(z) = \frac{\alpha[0,z]}{\alpha[0,1]}\,, \quad z \in [0,1]\,, \tag{8.2.13}$$

defines a distribution function on $[0,1]$.

Now we are in the position to rewrite $F_Z(y)$ in a much more accessible form:

$$\begin{aligned}
F_Z(y) &= (\widetilde{\alpha} \times F)(\{(\pi,x) : \pi \in [0,1)\,, x\,\pi \le y\})\\
&= \int_{\pi\in[0,1)} \int_{x\,\pi \le y} F(dx)\,\widetilde{\alpha}(d\pi)\\
&= \int_{\pi\in(0,1)} F(y/\pi)\,\widetilde{\alpha}(d\pi)\,.
\end{aligned} \tag{8.2.14}$$

This means that Z in the compound Poisson representation of the Poisson integral has representation as a product

$$Z \stackrel{d}{=} X\,\Pi\,,$$

where $X \stackrel{d}{=} X_1$, Π has distribution $\widetilde{\alpha}$, and X and Π are independent.

We summarize our findings.

Lemma 8.2.3 (Settled part of the RBNS claims by time T)
Assume the basic model of Section 8.2.1 and assume a linear cash flow function such that the insurer starts paying to the insured at time $T_i + D_i$ and finishes the payment X_i at time $T_i + D_i + S_i$. Then the amount $S_{\mathrm{Settled\,RBNS}}$ of the RBNS claims which is settled by time T has compound Poisson representation CP$(\alpha[0,1], F_Z)$, *where α is given by* (8.2.12) *and F_Z by* (8.2.14). *In particular,*

$$S_{\text{Settled RBNS}} \stackrel{d}{=} \sum_{i=1}^{M} X_i\, \Pi_i\,, \tag{8.2.15}$$

where M *is* $\mathrm{Pois}(\alpha[0,1])$ *distributed, independent of the mutually independent iid sequences* (X_i) *with common distribution* F *and* (Π_i) *with common distribution* $\widetilde{\alpha}$ *given in* (8.2.13)*.*

Notice the differences between the representations (8.2.11) and (8.2.15). In (8.2.11) the counting variable depends on the points $\pi_i = S_i^{-1}(T - T_i - D_i)$, whereas M in (8.2.15) is independent of (Π_i). Also notice that π_i and Π_i have different distributions. Whereas (Π_i) is an iid sequence, the points π_i are not independent and have different distributions.

We can deal with the *non-settled* or *outstanding part* of the RBNS claims in a similar way. This means we are interested in the remaining total claim amount

$$\begin{aligned} S_{\text{Outstanding RBNS}} &= \int_{E_{\text{RBNS}}} x \left[1 - \frac{T-t-r}{s}\right] N(dt,dr,ds,dx) \\ &= \sum_{i:(T_i,D_i,S_i,X_i)\in E_{\text{RBNS}}} X_i\,(1-\pi_i)\,. \end{aligned}$$

The latter sum is meaningful because $EN(E_{\text{RBNS}})$ is finite and hence, with probability 1, there are only finitely many points (T_i, D_i, S_i, X_i) in E_{RBNS}. Also notice that $1 - \pi_i \in (0,1]$ for any RBNS point (T_i, D_i, S_i, X_i).

Proceeding in the same way as above, one sees that the points $1 - \pi_i$ constitute $\mathrm{PRM}(\beta)$ on $(0,1]$ with mean measure β given by

$$\begin{aligned} \beta[0,z] &= (\lambda\,\mathrm{Leb}\times F_D \times F_S)\big(\{(t,d,s) : 0 \le 1 - s^{-1}(T-t-d) \le z\}\big) \\ &= (\lambda\,\mathrm{Leb}\times F_D \times F_S)\big(\{(t,d,s) : 1 - z \le s^{-1}(T-t-d) \le 1\}\big) \\ &= \alpha[1-z,1]\,, \quad z \in [0,1]\,. \end{aligned}$$

Notice that $\beta[0,1] = \alpha[0,1]$.

Now calculations similar to those which led to Lemma 8.2.3 yield an analogous result for the outstanding part of the RBNS claims.

Lemma 8.2.4 (Outstanding part of the RBNS claims at time T)
We assume the conditions of Lemma 8.2.3. Then the amount $S_{\text{Outstanding RBNS}}$ *of the RBNS claims which is outstanding at time* T *has compound Poisson representation* $\mathrm{CP}(\alpha[0,1], F_{Z'})$*, where* α *is given by* (8.2.12)*,* $Z' \stackrel{d}{=} X\,(1-\Pi)$*,* $X \stackrel{d}{=} X_1$ *and* Π *are independent and* Π *has distribution* $\widetilde{\alpha}$*. In particular,*

$$S_{\text{Outstanding RBNS}} \stackrel{d}{=} \sum_{i=1}^{M} X_i\,(1-\Pi_i)\,,$$

where M is $\mathrm{Pois}(\alpha[0,1])$ *distributed, independent of the mutually independent iid sequences* (X_i) *with common distribution* F *and* (Π_i) *with common distribution* $\widetilde{\alpha}$.

Finally, the distribution of

$$\begin{aligned} S_{\mathrm{RBNS}} &= S_{\mathrm{Settled\,RBNS}} + S_{\mathrm{Outstanding\,RBNS}} \\ &= \sum_{i:(T_i,D_i,S_i,X_i)\in E_{\mathrm{RBNS}}} X_i\,, \end{aligned}$$

has a $\mathrm{CP}(\alpha[0,1],F)$ distribution. Considered as a function of T, the process S_{RBNS} does not have independent increments.

Writing $N_{\pi,X}$ for the marked $\mathrm{PRM}(\alpha\times F)$ of the points (π_i,X_i), we have the following Poisson integral representations:

$$S_{\mathrm{Settled\ RBNS}} = \int_{\pi\in[0,1)} x\,\pi\,N_{\pi,X}(d\pi,dx)\,, \tag{8.2.16}$$

$$S_{\mathrm{Outstanding\ RBNS}} = \int_{\pi\in[0,1)} x\,(1-\pi)\,N_{\pi,X}(d\pi,dx)\,. \tag{8.2.17}$$

The integrands in these two integrals are non-negative and do not have disjoint support. Therefore the resulting Poisson integrals are dependent; see Corollary 7.2.13.

Using similar arguments, the theory above can be derived for any non-decreasing payment function f from $[0,1]$ to $[0,1]$ such that $f(0)=0$ and $f(1)=1$. The corresponding settled and outstanding amounts are then given by the Poisson integrals

$$S_{\mathrm{Settled\ RBNS}} = \int_{\pi\in[0,1)} x\,f(\pi)\,N_{\pi,X}(d\pi,dx)\,, \tag{8.2.18}$$

$$S_{\mathrm{Outstanding\ RBNS}} = \int_{\pi\in[0,1)} x\,(1-f(\pi))\,N_{\pi,X}(d\pi,dx)\,. \tag{8.2.19}$$

They have compound Poisson representations

$$S_{\mathrm{Settled\ RBNS}} \stackrel{d}{=} \sum_{i=1}^{M} X_i\,f(\Pi_i) \quad \text{and} \quad S_{\mathrm{Outstanding\ RBNS}} \stackrel{d}{=} \sum_{i=1}^{M} X_i\,(1-f(\Pi_i))\,,$$

where M, (X_i), (Π_i) have the same distributions and dependence structure as in Lemmas 8.2.3 and 8.2.4. We encourage the conscientious and non-passive reader to verify these formulae.

Even more generality is achieved if one chooses integrand functions $g(\pi,x)$ with $0\le g(\pi,x)\le x$ for $\pi\in[0,1]$, non-decreasing in the π-component and such that $g(1,x)=x$. Then, for example,

$$S_{\text{Settled RBNS}} = \int_{\pi \in [0,1]} g(\pi, x)\, N_{\pi,X}(d\pi, dx) \stackrel{d}{=} \sum_{i=1}^{M} g(\Pi_i, X_i)\,,$$

and $S_{\text{Outstanding RBNS}}$ is defined correspondingly. The choice of the function g allows one to determine the speed at which the insurer pays the insured in the settlement period of a claim. A simplistic (but not totally unrealistic) example is provided by the function

$$g(\pi, x) = \begin{cases} 0\,, & \pi \in [0,1)\,, \\ x\,, & \pi = 1\,. \end{cases}$$

Here the insurer pays nothing until the very end of the settlement period.

In real-life applications, the form of the payment function will depend on the circumstances surrounding each individual claim. In order to model such *random* phenomena, one would have to assume stochastic payment functions, i.e., at each reporting time $T_i + D_i$, a stochastic process starts which describes the individual settlement history of the claim. In Section 11.3 an attempt is made to take into account the stochastic nature of payments.

8.2.4 An Excursion to Teletraffic and Long Memory: The Stationary IBNR Claim Number Process

In this section we make an excursion into the active research area of large data or teletraffic networks such as the Internet or the Local Area Computer Network of a large company or a university. Such networks are highly complex and therefore the need for simple models arises which nevertheless describe some of the essential features of real-life networks. One of the properties of modern data networks is the presence of *long memory* or *long range dependence* in the counting process of packets registered by a sniffer or counter per time unit. We will give an explanation for this phenomenon in the model considered.

In what follows, we consider the counting process of the IBNR claims as a possible generating model for the activities in a large data network. Indeed, a modification of this process has been used for a long time as one of the standard models in the literature. Of course, the arrivals T_i and the delays D_i have a completely different meaning in this context. We will think of T_i as the *arrival of an activity* to the network. For example, this can be the arrival of a packet in your computer. The packet is processed, i.e., queued and routed to its destination. This activity creates an amount of work. The interval $[T_i, T_i + D_i]$ describes the period when the packet is processed. Assuming that the work is always processed at the same rate, the length D_i of this interval multiplied by the rate is then considered as a measure of the work initiated by the packet.

With these different meanings of T_i and D_i in mind, we modify the basic model of Section 8.2.1 insofar that we assume that the arrivals T_i come from a homogeneous Poisson process with intensity λ on the whole real line.[2] We consider an increasing enumeration of the points T_i such that

$$\cdots < T_{-2} < T_{-1} < 0 \le T_1 < T_2 < \cdots \quad \text{a.s.}$$

We mark each point T_i by a positive random variable D_i, which now stands for the amount of work brought into the system. The iid sequence $(D_i)_{i\in\mathbb{Z}_0}$ of positive random variables D_i is again independent of (T_i). Here we write $\mathbb{Z}_0 = \mathbb{Z}\backslash\{0\}$ for the set of the non-zero integers. Hence the points (T_i, D_i) constitute $\mathrm{PRM}(\lambda\,\mathrm{Leb}\times F_D)$, denoted by $N_{T,D}$, on the state space $\mathbb{R}\times(0,\infty)$.

We consider the following analog of the IBNR claim number process which in this context represents the number of active sources at time T:

$$\begin{aligned} &M(T) && (8.2.20)\\ &= N_{T,D}(\{(t,d) : t \le T < t+d\}) = \#\{i \in \mathbb{Z}_0 : T_i \le T < T_i + D_i\}\\ &= N_{T,D}(\{(t,d) : t < 0\,, T < t+d\}) + N_{T,D}(\{(t,d) : 0 \le t \le T < t+d\})\\ &= \#\{i \le -1 : T < T_i + D_i\,, T_i < 0\} + \#\{i \ge 1 : 0 \le T_i \le T < T_i + D_i\}\\ &= M_-(T) + M_+(T)\,, \quad T \ge 0\,. \end{aligned}$$

Notice that $M_-(T)$ counts the number of those arrivals which occurred before time 0 and whose activity period reaches into the future after time T. The claim number $M_+(T)$ coincides with the claim number $N(E_{\text{IBNR at time }T})$ considered in the previous sections. Since $M_+(T)$ and $M_-(T)$ arise from Poisson points (T_i, D_i) in disjoint subsets of the state space $\mathbb{R}\times(0,\infty)$ of $N_{T,D}$, they are independent and Poisson distributed.

In contrast to the quantities $\#\{i \in \mathbb{Z}_0 : T_i \le T\}$, which are infinite a.s. at any time T (see Exercise 6(a) on p. 287), the random variables $M(T)$ are finite a.s. for every $T \ge 0$, provided D_1 has finite expectation. This is easily seen since the mean values $EM_\pm(T)$ are finite for any $T \ge 0$. Indeed, from Lemma 8.2.2(3) we know that

$$EM_+(T) = EN(E_{\text{IBNR at time }T}) = \lambda\, E(T - (T - D_1)_+)\,.$$

For $M_-(T)$ we proceed in the same way as in the proof of Lemma 8.2.2(2). First observe that

$$(-T_i)_{i\le -1} \stackrel{d}{=} (T_i)_{i\ge 1}\,.$$

Then we have

[2] In an insurance context, this would mean that the business does not start at time 0 but it has always been running.

$$
\begin{aligned}
EM_-(T) &= E\left(\sum_{i=-\infty}^{-1} I_{\{T<T_i+D_i\}}\right) \\
&= \sum_{i=-\infty}^{-1} P(0 \le -T_i < (T-D_1)_-) \\
&= \sum_{i=1}^{\infty} P(0 \le T_i < (T-D_1)_-)\,.
\end{aligned}
$$

A Fubini argument and relation (8.2.9) yield:

$$
\begin{aligned}
EM_-(T) &= E\left(\int_{[0,(T-D_1)_-)} d\left[\sum_{i=1}^{\infty} P(T_i \le t)\right]\right) \\
&= \lambda\, E\left(\int_{[0,(T-D_1)_-)} dt\right) = \lambda\, E(T-D_1)_-\,.
\end{aligned}
$$

Thus we have proved that

$$
EM(T) = EM_+(T) + EM_-(T) = \lambda\, ED_1\,, \quad T \ge 0\,,
$$

and this quantity is finite for $ED_1 < \infty$.

Since the expectation $EM(T)$ does not depend on T this is a first indication of the fact that $(M(T))_{T\ge 0}$ constitutes a *strictly stationary process*, i.e.,

$$
(M(T))_{T\ge 0} \stackrel{d}{=} (M(T+h))_{T\ge 0} \quad \text{for } h \ge 0,
$$

where $\stackrel{d}{=}$ refers to equality of the finite-dimensional distributions. A proof of strict stationarity of the process M is left as Exercise 6(c) on p. 287.

We restrict ourselves to the problem of showing *second order stationarity*, in the sense that the covariance function $C_M(T,T+h)$ of the process M does not depend on T:

$$
C_M(T,T+h) = \operatorname{cov}(M(T), M(T+h))\,, \quad h \ge 0\,,\ T \ge 0\,.
$$

This property is easily verified since both $M(T)$ and $M(T+h)$ can be represented as Poisson integrals:

$$
M(s) = \int_{\mathbb{R}\times(0,\infty)} I_{\{t\le s<t+r\}}((t,r))\, N_{T,D}(dt,dr)\,, \quad s \in \mathbb{R}\,.
$$

Then we conclude from relation (7.2.22) that

$$
\begin{aligned}
C_M(T,T+h) &= \lambda \int_{\mathbb{R}\times(0,\infty)} I_{\{t\le T<t+r\}}((t,r))\, I_{\{t\le T+h<t+r\}}((t,r))\, F_D(dr)\,dt \\
&= \lambda \int_{t=-\infty}^{T} \int_{r=T-t+h}^{\infty} F_D(dr)\,dt \\
&= \lambda \int_{-\infty}^{T} \overline{F}_D(T-t+h)\,dt \\
&= \lambda \int_h^{\infty} \overline{F}_D(s)\,ds\,.
\end{aligned}
$$

We summarize these results.

Lemma 8.2.5 *The counting process* $(M(T))_{T\ge 0}$ *defined in* (8.2.20) *based on the points* (T_i, D_i) *of* $\mathrm{PRM}(\lambda\,\mathrm{Leb}\times F_D)$ *with* $ED_1 < \infty$ *is a strictly stationary process whose one-dimensional marginal distributions are Poisson. In particular,*

$$
\begin{aligned}
EM(T) &= \lambda\, ED_1\,, \\
\mathrm{cov}(M(T), M(T+h)) &= \lambda \int_h^{\infty} \overline{F}_D(s)\,ds \qquad (8.2.21)\\
&= \gamma_M(h)\,, \quad T\ge 0\,,\ h\ge 0\,.
\end{aligned}
$$

Since we have assumed $ED_1<\infty$, the covariance function $\gamma_M(h)$ is finite for any h and satisfies $\gamma_M(h)\downarrow 0$ as $h\uparrow\infty$. The decay rate of $\gamma_M(h)$ to zero as $h\to\infty$ is often interpreted as *range of memory* or *range of dependence* in the stationary process M. It is clear that the lighter the tail $\overline{F}_D$, the faster $\gamma_M(h)$ in (8.2.21) tends to zero. It is in general not possible to calculate γ_M more explicitly than (8.2.21). But then it often suffices to have results about the asymptotic behavior of $\gamma_M(h)$ as $h\to\infty$. One such case is described in the following example.

Example 8.2.6 (Regularly varying D_1)
Recall from Definition 3.2.20 on p. 99 that a positive random variable D_1 is said to be *regularly varying with index* $\alpha\ge 0$ if the right tail of F_D has form

$$
\overline{F}_D(x) = L(x)\,x^{-\alpha}\,, \quad x>0\,,
$$

for some $\alpha\ge 0$ and some slowly varying function L, i.e., a non-negative measurable function on $(0,\infty)$ satisfying $L(cx)/L(x)\to 1$ as $x\to\infty$ for every $c>0$. Regularly varying functions satisfy some asymptotic integration rule which runs under the name *Karamata's theorem*; see p. 181. An application of Karamata's theorem to (8.2.21) yields, for $\alpha>1$,

$$
\gamma_M(h) \sim (\alpha-1)^{-1}\,h\,\overline{F}_D(h)\,, \quad \text{as } h\to\infty. \qquad (8.2.22)
$$

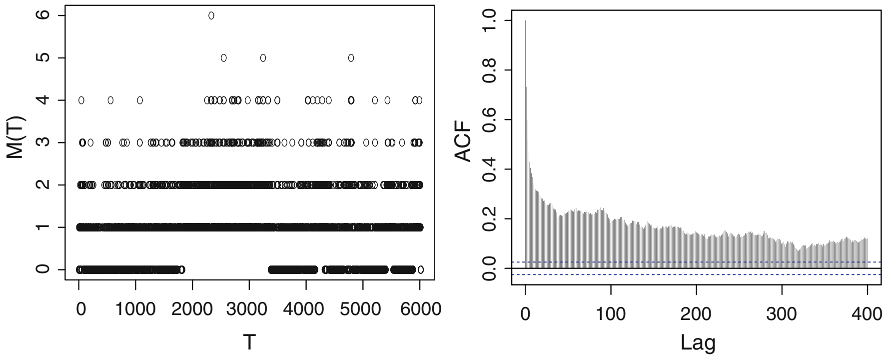

Figure 8.2.7 Left: *A realization of the strictly stationary process* $(M(n))_{n=1,\ldots,6000}$ *with Poisson intensity* $\lambda = 3$*,* D_1 *is regularly varying with index* 1.5. Right: *The sample autocorrelation function* $\widehat{\rho}_M$ *at the first* 400 *lags. It decays very slowly. The dashed lines indicate lag-wise* 95% *asymptotic confidence intervals for* $\widehat{\rho}_M(h)$ *under the hypothesis of iid Gaussian noise for the underlying sample.*

This means that the covariance function $\gamma_M(h)$ decays very slowly to zero — like a power law with exponent $1-\alpha < 0$. Another application of Karamata's theorem yields for $\alpha \in (1,2)$ that

$$\int_0^\infty \gamma_M(h)\, dh = \infty\,.$$

In Figure 8.2.7 we visualize the process M at the discrete instants of time $k = 1, 2, \ldots, 6000$. The D_i's are regularly varying with index $\alpha = 1.5$. Then the arguments above apply and the autocovariance function $\gamma_M(h)$ decays to zero as described in (8.2.22). Since the function γ_M is in general not explicitly known it is common practice in time series analysis to estimate its standardized version $\rho_M(h) = \gamma_M(h)/\gamma_M(0)$, the *autocorrelation function* of the stationary time series $(M(k))$, from the sample $(M(k))_{k=1,\ldots,n}$. The corresponding *sample autocovariances* $\widehat{\gamma}_M(h)$ and *sample autocorrelations* $\widehat{\rho}_M(h)$ are then given by

$$\widehat{\gamma}_M(h) = n^{-1} \sum_{k=1}^{n-h} (M(k) - \overline{M}_n)(M(k+h) - \overline{M}_n)\,,$$

$$\widehat{\rho}_M(h) = \frac{\widehat{\gamma}_M(h)}{\widehat{\gamma}_M(0)}\,, \quad h \geq 0\,,$$

where $\overline{M}_n$ denotes the sample mean. The sample autocovariances and sample autocorrelations are consistent estimators of their deterministic counterparts if the underlying process $(M(n))$ is strictly stationary and ergodic. This follows by an application of the ergodic theorem; see Krengel [89]. The process $(M(n))$ inherits ergodicity from the ergodicity of the underlying homogeneous Poisson process; see Daley and Vere-Jones [38].

Thus slow decay of the sample autocorrelation function $\widehat{\rho}_M(h)$ as a function of the lag h is an indication of slow decay of $\rho_M(h)$. $\square$

For any stationary process with covariance function γ the property

$$\int_0^\infty |\gamma(h)| = \infty$$

is often referred to as *long range dependence* or *long memory*. It describes extremely slow decay of the covariance function $\gamma(h)$ to zero as $h \to \infty$. This definition seems arbitrary and, indeed, there exist various other ones based on different arguments; see for example Samorodnitsky and Taqqu [131], Chapter 7, Doukhan et al. [41], and Samorodnitsky [130].

An alternative way of defining long memory is to require that $\gamma(h)$ is a regularly varying function with index in $(-1, 0)$, i.e., $\gamma(h) = L(h)h^{2(H-1)}$ for a slowly varying function L and a parameter $H \in (0.5, 1)$, called the *Hurst coefficient*. This assumption is satisfied for the process M provided D_1 is regularly varying with index $\alpha \in (1, 2)$. Then H assumes the value $(3 - \alpha)/2$; see (8.2.22). Such a definition is more reasonable from a statistical point of view since it allows one to estimate the Hurst parameter H, which characterizes the range of dependence in the stationary process.

The property of power law decay of the covariance function is also observed for certain fractional Gaussian noises, i.e., the increment process of fractional Brownian motion (see Exercise 9 on p. 289), and for certain fractional ARIMA processes. We refer to Samorodnitsky and Taqqu [131], Chapter 7, for fractional Brownian motion and Gaussian noise, and Brockwell and Davis [24] for fractional ARIMA processes.

The notion of long memory or long range dependence has attracted a lot of attention over the last 40 years. It is a phenomenon which is empirically observed in areas as diverse as physics, telecommunications, hydrology, climatology, and finance. We refer to Doukhan et al. [41] and Samorodnitsky [130] for recent surveys on the theory and applications of long memory processes.

The interest in the notion of long memory is explained by the fact that long memory stationary processes, in contrast to short memory processes, often exhibit asymptotic behavior different from standard central limit theory. For example, the workload process $\int_0^T M(t)\,dt$ of the strictly stationary teletraffic process M defined in (8.2.20) does not, in general, satisfy standard central limit theory in the sense of functional distributional convergence with limiting Brownian motion; cf. Billingsley [17]. On the contrary, the workload process with regularly varying D_1 with index $\alpha \in (1, 2)$ has a less familiar limiting process,[3] and the normalization in this limit result significantly differs from the common $\sqrt{T}$-scaling constants.

As a matter of fact, the process M defined in (8.2.20) with regularly varying D_1 with index $\alpha \in (1,2)$ has attracted a lot of attention in the teletraffic community; see for example Mikosch et al. [110], Faÿ et al. [50], Mikosch and Samorodnitsky [111], cf. Resnick [124] and the references given therein. The $\mathrm{PRM}(\lambda\,\mathrm{Leb} \times F_D)$ model generating the process M is often referred to as $M/G/\infty$ queue model or as *infinite source Poisson model* in the probability literature on queuing and telecommunications. It is a simple model for real-life teletraffic, in particular for the Internet. The model is simplistic but allows for the description of phenomena which are also observed in teletraffic data: long memory, heavy-tailed components (such as file sizes or transmission lengths) and self-similarity[4] of the limiting process of the workload process $\int_0^T M(t)\,dt$. In addition, the process M is easily simulated. In the teletraffic context, the quantities D_i are most relevant. Their size determines typical behavior of the whole system. The memory in the activity process $M(T)$ at a given time T is then determined by the range of the activities described by the length of the D_i's.

8.2.5 A Critique of the Basic Model

In the previous sections, we decomposed the time-claim size space E at the present time $T > 0$ into the disjoint sets E_{Settled}, E_{RBNS}, E_{IBNR} and $E_{\text{Not incurred}}$. The corresponding pairs of claim number and total claim amount

$$(N(E_{\text{Settled}}), S_{\text{Settled}})\,, \quad (N(E_{\text{RBNS}}), S_{\text{RBNS}})\,, \quad (N(E_{\text{IBNR}}), S_{\text{IBNR}})$$

are mutually independent. These quantities are functions of the PRM N with points (T_i, D_i, S_i, X_i). In our presentation we have assumed that the claim arrivals T_i come from a homogeneous Poisson process and that the three-dimensional iid marks (D_i, S_i, X_i) have independent components. These conditions can be weakened. For example, the arrivals may arise from an inhomogeneous Poisson process or the components D_i, S_i and X_i may be dependent; see Section 8.1.4 for an example of dependence between D_i and X_i. In this more general context, one can often follow the arguments given above without major difficulties. Of course, one has to pay a price for more generality: calculations become more tedious and the resulting formulae are more complex.

---

[3] The limiting process is spectrally positive α-stable Lévy motion; see Example 10.5.2 for its definition and properties. Alternatively, fractional Brownian motion may occur as limit if the intensity $\lambda = \lambda(T)$ of the underlying homogeneous Poisson process grows sufficiently fast with time T; see Mikosch et al. [110], cf. Resnick [124].

[4] See Exercise 9(d) on p. 289 for the definition and examples of some self-similar processes, including fractional Brownian motion.

The basic model is statistically tractable. Based on historical data one can estimate the underlying Poisson intensity and the distribution of the iid observations (D_i, S_i, X_i). If the components D_i, S_i and X_i are independent, one can use one-dimensional statistical methods to fit the distributions of F_D, F_S and F separately. The statistics become much more complicated if one aims at fitting a three-dimensional distribution with dependencies between D_i, S_i and X_i.

Based on historical information and on the fitted distributions one knows *in principle* the distributions of the settled and outstanding claim numbers and total claim amounts of an insurance business. In most cases of interest the distributions will not be tractable without Monte-Carlo or numerical methods. For example, we learned in Section 3.3.3 about *Panjer recursion* as a numerical technique for evaluating compound Poisson distributions.

An advantage of the presented theory is the consequent use of Poisson processes. The Poisson ideology allows one to decompose the total claim amount into its essential parts (IBNR, RBNS, settled and outstanding, say). These are independent due to the Poisson nature of the underlying counting process. One loses the elegance of the theory if one gives up the Poisson assumption. Nevertheless, even in the case of a non-Poissonian marked point process, several of the calculations given above can be provided by using general point process techniques: most of the moment and covariance calculations are still possible; see Daley and Vere-Jones [38, 39, 40].

As mentioned above, the basic model can be extended and generalized in different directions. A way of introducing a "more realistic" model is to assume a genuine stochastic process model Y_i which describes the payment of the ith claim size X_i in the period $[T_i + D_i, T_i + D_i + S_i]$. In Section 8.2.3 we have assumed a simple linear model $Y_i(T) = X_i\, S_i^{-1}(T - T_i - D_i)$ for $T \in [T_i + D_i, T_i + D_i + S_i]$. Unfortunately, every claim has its own characteristics and therefore it would be rather optimistic to believe that a linear function is in agreement with real-life data.

It is possible to assume very general pay-off functions Y_i and to develop some theory for the resulting total claim amounts; for some asymptotic results see Klüppelberg et al. [81, 82, 83] who worked with Poisson shot noise models. The latter class of models is closely related to the Poisson models considered above. In the language of marked Poisson processes, the claim arrivals T_i are then marked with an iid sequence of quadruples (D_i, S_i, X_i, Y_i), where the meaning of (D_i, S_i, X_i) is as above and Y_i is a stochastic process whose sample paths describe the payment process for the ith claim. In this context it is reasonable to let Y_i have non-decreasing sample paths on the interval of interest $[T_i + D_i, T_i + D_i + S_i]$: choose Y_i such that $Y_i(t) = 0$ a.s. for $t < 0$, the process $Y_i(T - T_i - D_i)$ gets activated at the reporting time $T = T_i + D_i$ (possibly with a positive initial payment), it does not decrease until time $T_i + D_i + S_i$, where it achieves its largest value X_i and becomes deactivated at times $T > T_i + D_i + S_i$, i.e., $Y_i(T - T_i - D_i) = 0$ a.s.

One faces a major problem: the choice of a reasonable model for the payment process Y_i. A practical solution would be to work with historical sample paths from a sufficiently large portfolio over a sufficiently long period of time. Then the distribution of the total claim amount could be approximated by Monte-Carlo simulations from the empirical distribution of the historical sample paths. This approach is close in spirit to the bootstrap; see Efron and Tibshirani [44] for an elementary introduction, cf. Section 3.3.5. However, this approach is ad hoc and requires a theoretical justification.

Motivated by Bayesian ideas, Norberg [114] suggested modeling the payment processes Y_i by suitable gamma and Dirichlet processes. He demonstrated that one can predict outstanding claims by calculating their expectation conditionally on information about past payments for the claim. While the required assumptions seem ad hoc, they are as realistic (or unrealistic) as assuming a non-decreasing smooth payment function, as we did on pp. 274–278.

In Chapter 11 we shall look at some models which we will call *cluster point processes*. There we will describe the payment processes for individual claims by stochastic processes. It will again be convenient to assume a simplifying Poisson structure of the points of these processes. In Section 11.3, this structure will allow us to get explicit expressions for predicted claim numbers and total claim amounts based on historical information.

Exercises

Section 8.2.1

(1) Let N be the PRM($\lambda\,\mathrm{Leb}\times F_D\times F_S\times F$) generated from the points (T_i, D_i, S_i, X_i) in the basic model; see p. 268.
   (a) Show that the point process $N_{T,T+D,T+D+S,X}$ of the points $(T_i, T_i+D_i, T_i+D_i+S_i, X_i)$ is PRM.
   (b) Determine the mean measure of $N_{T,T+D,T+D+S,X}$.
   (c) Show that the point process $N_{T+D+S,X}$ of the points $(T_i+D_i+S_i, X_i)$ is PRM($\gamma\times F$) on $(0,\infty)^2$, where γ is defined by (8.2.6).

Section 8.2.2

(2) Consider the basic decomposition of the time-claim size space; see p. 270.
   (a) Show that the processes $(N(E_{\mathrm{RBNS\,at\,time}\,T}))_{T>0}$, $(N(E_{\mathrm{IBNR\,at\,time}\,T}))_{T>0}$ do not have independent increments.
   Hint: It is advantageous to calculate the covariance of increments on disjoint intervals.
   (b) Prove Lemma 8.2.2(3).
   (c) Assume the conditions of the basic model (see p. 268) with one exception: the T_i's constitute a renewal sequence. This means that $T_n = Y_1+\cdots+Y_n$, $n\ge 1$, for iid positive random variables Y_i with finite mean value. Also assume that $ED_1<\infty$.

   Recall the notion of renewal function

$$m(t) = 1 + E\#\{i \ge 1 : T_i \le t\}\,, \quad t \ge 0\,;$$

see Section 2.2.2.

Show that the following relations hold:

$$EN(E_{\text{RBNS}}) = E\left[m((T - D_1)_+) - m((T - D_1 - S_1)_+)\right]\,,$$

$$EN(E_{\text{IBNR}}) = E\left[m(T) - m((T - D_1)_+)\right]\,.$$

Section 8.2.3

(3) Consider the process $(S_{\text{Settled at time }T})_{T>0}$ given in (8.2.10).
  (a) Prove that the process has independent increments.
  (b) Prove that S_{Settled} for fixed $T > 0$ has $\text{CP}(\gamma(T), F)$ representation, where the mean value function γ is given in (8.2.8). In particular, conclude that $(S_{\text{Settled at time }T})_{T>0}$ does in general not have stationary increments.

(4) Assume that $\text{var}(X_1) < \infty$. Calculate the covariance between $S_{\text{Settled RBNS}}$ and $S_{\text{Outsanding RBNS}}$.
  Hint: It is advantageous to use the Poisson integral representations (8.2.16) and (8.2.17).

(5) (a) Modify the calculations leading to Lemma 8.2.3 for $S_{\text{Settled RBNS}}$ such that you prove Lemma 8.2.4 for $S_{\text{Outstanding RBNS}}$.
  (b) Repeat the calculations in (a) for the claim amount

$$S_{\text{Settled RBNS}} = \int_{E_{\text{RBNS}}} g(s^{-1}(T - t - r), x)\, N(dt, dr, ds, dx)\,,$$

  where $g(\pi, x) \in [0, x]$ is continuous and non-decreasing in the π-component such that $g(1, x) = x$.

Section 8.2.4

(6) Consider a homogeneous Poisson process with points T_i, $i \in \mathbb{Z}_0$, on $\mathbb{R}$ such that $\cdots < T_{-2} < T_{-1} < 0 \le T_1 < T_2 < \cdots$, i.e., $T_i \ge 0$ for $i \ge 1$ and $T_i < 0$ for $i \le -1$. We mark the points T_i with positive random variables D_i such that the iid sequence (D_i) is independent of (T_i).
  (a) Show that for any $T \ge 0$, $\#\{i \in \mathbb{Z}_0 : T_i \le T\} = \infty$ a.s.
  (b) Show that

$$(-T_i)_{i \le -1} \stackrel{d}{=} (T_i)_{i \ge 1}\,.$$

  (c) Show strict stationarity of the process

$$M(T) = \#\{i \in \mathbb{Z}_0 : T_i \le T < T_i + D_i\}\,, \quad T \ge 0\,,$$

  in the sense that the finite-dimensional distributions of $(M(T + h))_{T \ge 0}$ do not depend on $h \ge 0$.
  Hint: Show that the PRMs with points (T_i, D_i) and $(T_i + h, D_i)$ have the same distribution for any $h \in \mathbb{R}$.
  (d) Calculate the covariance function γ_M of the strictly stationary process M given in (8.2.21) for (i) D_1 with an exponential $\text{Exp}(a)$, $a > 0$, distribution and (ii) a Pareto distribution with parameterization $\overline{F}_D(x) = x^{-\alpha}$, $x \ge 1$, for some $\alpha > 1$. Explain why the assumption $\alpha > 1$ is relevant.

(7) Let N be a homogeneous Poisson process on $\mathbb{R}$ with intensity $\lambda > 0$. For any $s > 0$, define the process

$$O(T) = \int_{-\infty}^{T} \mathrm{e}^{-s\,(T-t)}\, N(dt)\,, \quad T \in \mathbb{R}\,.$$

This process is an analog of the classical *Ornstein-Uhlenbeck process* where N is replaced by Brownian motion. The Ornstein-Uhlenbeck process is one of the most popular Gaussian processes. It has a multitude of applications in areas as diverse as physics and finance. For example, in finance a ramification of the Ornstein-Uhlenbeck process, known as the Vasicek model, is used as a model for interest rates; see Björk [20].

We refer to the process O as the *Ornstein-Uhlenbeck Poisson process.*

(a) Show that the Poisson integral $O(T)$ exists and is finite a.s. for every T.

(b) Calculate the mean value function $E[O(T)]$, $T \in \mathbb{R}$, and the covariance function $C_O(T, T+h) = \mathrm{cov}(O(T), O(T+h))$, $T, h \in \mathbb{R}$.

(c) Show that O is a strictly stationary process on $\mathbb{R}$.
Hint: It is convenient to use a Laplace-Stieltjes transform argument for the finite-dimensional distributions of the processes $(O(T))_{T\in\mathbb{R}}$ and $(O(T+h))_{T\in\mathbb{R}}$.

(d) Consider a discrete-time version of O given by $(O(n))_{n\in\mathbb{Z}}$. Conclude that this is a strictly stationary process, i.e., $(O(n))_{n\in\mathbb{Z}} \stackrel{d}{=} (O(n+k))_{n\in\mathbb{Z}}$ for any integer k. Calculate the mean value and covariance functions of this process.

(e) Prove that the discrete-time process $(O(n))_{n\in\mathbb{Z}}$ considered in (d) satisfies the difference equation

$$O(n) = \mathrm{e}^{-s}\, O(n-1) + Z_n\,, \quad n \in \mathbb{Z}\,, \tag{8.2.23}$$

where (Z_n) is an iid sequence. Determine the distribution of Z_n.

A discrete-time real-valued process with index set $\mathbb{Z}$ is often called a time series. For a general iid sequence (Z_n), a solution to the difference equation (8.2.23) defines an *autoregressive process of order* 1 or AR(1) process. The AR(1) processes are natural discrete-time analogs of an Ornstein-Uhlenbeck process. The AR(1) process is a prominent member of the class of stationary ARMA processes. The latter class consists of those time series models which are used most often in applications. We refer to the books by Brockwell and Davis [24, 25] for introductions to time series analysis and ARMA models.

Ornstein-Uhlenbeck processes can also be defined for classes of driving processes N much wider than Brownian motion or the Poisson process. For example, it can be defined for certain classes of Lévy processes, i.e., processes with independent stationary increments (see Section 10 for their definition and properties), or even for fractional Brownian motion (see Exercise 9 below), i.e., Gaussian Markov processes with stationary increments; see for example Mikosch and Norvaiša [109] for the definition and properties of such Ornstein-Uhlenbeck processes and Samorodnitsky and Taqqu [131], Chapter 7, for further reading on fractional Brownian motion.

(8) Let N be a homogeneous Poisson process on $\mathbb{R}$ with intensity $\lambda > 0$. Define the stochastic process

$$\eta(t) = \int_{-\infty}^{t} f(t-s)\, N(ds)\,, \quad t \in \mathbb{R}\,,$$

for a non-negative measurable function f on $\mathbb{R}$.

(a) Give a condition on f guaranteeing that $\eta(t) < \infty$ a.s. for every $t \in \mathbb{R}$.

(b) Assume that $\eta(t) < \infty$ a.s. for every $t \in \mathbb{R}$; see (a). Show that η is a strictly stationary process.

(c) Show that there exist processes $\eta^{(i)}(t) = \int_{-\infty}^{t} f_i(t-s)\, N(ds)$, $i = 1, 2$, with $f_1 \neq f_2$ a.e. with respect to Lebesgue measure but such that the autocovariance functions

$$\gamma_{\eta^{(i)}}(h) = \operatorname{cov}(\eta^{(i)}(s), \eta^{(i)}(t)), \quad s, t \in \mathbb{R}, \quad i = 1, 2,$$

of the processes $\eta^{(i)}$, $i = 1, 2$, coincide.

(9) *Fractional Brownian motion* $(B_t^{(H)})_{t \geq 0}$ is a mean zero Gaussian process given by its covariance function

$$C_B(t, s) = \operatorname{cov}(B_t^{(H)}, B_s^{(H)}) = \frac{\sigma^2}{2} \left(|t|^{2H} + |s|^{2H} - |t-s|^{2H} \right), \quad s, t \geq 0,$$

for $\sigma > 0$ and $H \in (0, 1]$. The increment process

$$\xi^{(H)}(n) = B_n^{(H)} - B_{n-1}^{(H)}, \quad n = 1, 2, \ldots,$$

is called *fractional Gaussian noise.*

(a) Show that $B^{(H)}$ has stationary increments and that $B_0^{(H)} = 0$ a.s.

(b) Show that $(\xi^{(H)}(n))_{n=1,2,\ldots}$ constitutes a strictly stationary process.

(c) Calculate the autocovariance function

$$\gamma_{\xi^{(H)}}(h) = \operatorname{cov}(\xi^{(H)}(1), \xi^{(H)}(1+h)), \quad h \geq 0,$$

and show that fractional Gaussian noise for $H \in (0, 0.5) \cup (0.5, 1)$ satisfies the relation

$$\gamma_{\xi^{(H)}}(h) \sim c\, h^{2(H-1)} \quad \text{as } h \to \infty \text{ for some constant } c > 0.$$

Conclude that $(\xi^{(H)}(n))$ for $H \in (0.5, 1)$ exhibits long range dependence in the sense that

$$\sum_{h=1}^{\infty} |\gamma_{\xi^{(H)}}(h)| = \infty.$$

(d) Verify that $B^{(H)}$, $0 < H < 1$ is a *self-similar process* in the sense that for any $c > 0$ (here $\stackrel{d}{=}$ refers to equality of the finite-dimensional distributions),

$$c^H\, (B_t^{(H)})_{t \geq 0} \stackrel{d}{=} (B_{ct}^{(H)})_{t \geq 0}.$$

Another self-similar process — symmetric α-stable Lévy motion — is discussed in Example 10.5.2 on p. 358.

(e) Show that the case $H = 0.5$ corresponds to Brownian motion, i.e., $B^{(0.5)}$ is mean zero Gaussian with independent stationary increments.

(f) Show that the distribution of any continuous-time mean zero Gaussian process $(\eta_t)_{t \geq 0}$ with stationary increments is determined by its variance function $\sigma_\eta^2(t) = \operatorname{var}(\eta_t)$, $t \geq 0$.