Intensities and Poisson Models

In this chapter we return to a great extent to the notions of intensities. In the first section, the failure intensity is introduced; this gives the distribution of the waiting time to the first event. This intensity is of particular interest when lifetimes (of components or humans) are considered. Estimation procedures and statistical problems are discussed. Relative risks and risk exposure are the main topics of Section 7.2. In Section 7.3, models for Poisson counts are considered, leading to an introduction to the often-used Poisson regression models. This makes modelling possible of situations where the relation between an intensity and some explanatory, non-random variables, is given by a regression equation.

In Section 7.4, we introduce the notion of Poisson point process (PPP), an extension of Poisson streams discussed in Chapter 2. This enables modelling of events that can occur in spatial locations or at space and time locations, discussed in Section 7.5. Finally, we study superposition and decomposition of Poisson processes.

7.1 Time to the First Accident — Failure Intensity

7.1.1 Failure intensity

Before presenting new notions, let us revisit Example 4.1 (lifetimes of ball bearings) to analyse refined probabilistic modelling of lifetimes.

Example 7.1 (Lifetimes of ball bearings). In safety analysis, studies are often made of data of a type describing time to the first occurrence of an event. Time can sometimes be measured in rather strange units like the number of revolutions to failure, if lifetimes of ball bearings are studied (cf. Example 4.1, where an experiment with 22 observed lifetimes was presented).

An important issue is obviously to find a suitable distribution to describe the variability of lifetimes. In Example 4.1, the data were described using the empirical distribution \( F_n \), while in Example 9.1 a Weibull distribution will be used to model variability of lifetimes. In this chapter we introduce another
(equivalent) means to describe data: the so-called failure intensity $\lambda(s)$. The intensity measures risk for failure of a component of age $s$. For example, consider the risk that a ball bearing that has been used for 30 millions of revolutions will break in the next one. If the risk for failure increases with age, which is the case with ball bearings, then we say that lifetime of ball bearings has increasing failure rate (IFR).

Let $T$ denote a waiting time for the first failure (accident, death, etc.) Suppose the value of $T$ cannot be predicted and hence is modelled as a random variable. Let $F(t) = P(T \leq t)$ be the probability that the failure happens in the interval $[0, t]$. One sometimes speaks about the survival function

$$R(t) = P(T > t) = 1 - F(t)$$

which can equivalently be used to describe statistical properties of the lifetime.

The properties of the distribution $F(t)$ (or survival function $R(t)$) are often discussed in safety analysis, where failure times (life times) of components or structures are of interest. In such analysis, one is not only limited to failures that can be traced to accidents caused by environmental actions but also can be related to wear and other ageing processes. The distribution $F(t)$ may also reflect variability of quality (or strength) in some population of components: an element is chosen randomly from a population and then the lifetime of the chosen element is observed. Generally, any r.v. taking only positive values can be a model for the lifetime of members in some population.

Next we introduce a very important characterization of $T$ called the failure-intensity function, (for short, the failure intensity), alternatively, the hazard function.

**Definition 7.1 (Failure-intensity function).** For an r.v. $T \geq 0$ there is a function $\Lambda(t)$, called the cumulative failure-intensity function, such that

$$R(t) = e^{-\Lambda(t)}, \quad t \geq 0.$$  

If $T$ has a density, then

$$R(t) = \exp\left(-\int_0^t \lambda(s) \, ds\right)$$

where the function $\lambda(s) = \frac{d\Lambda(s)}{ds}$ is called the failure-intensity function.

The failure-intensity function defines the distribution of $T$ uniquely. If the distribution is of the continuous type, the failure intensity can also be calculated by

$$\lambda(s) = \frac{f(s)}{1 - F(s)}. \quad (7.1)$$
It can also be demonstrated that
\[ \lambda(s) = \lim_{t \to 0} \frac{P(T \leq s + t | T > s)}{t}, \]
which means that for small values of \( t \), \( \lambda(s) \cdot t \) is approximately the probability that an item of age \( s \) will break within the period of time \( t \).

Generally, failure-intensity functions are classified as **IFR** (increasing failure rate), where components wear with time, or **DFR** (decreasing failure rate) where the weak components fail first so the ones that rest are the strongest; consequently failures occur less frequently. Often both mechanisms are present simultaneously and we observe an increasing failure rate for the old components due to damaging processes. This is often experienced by owners of old cars. In the following somewhat artificial example the IFR and DFR failure intensities are given.

**Example 7.2 (Strength of a wire).** Suppose the strength \( R \) of a particular wire is modelled as a Weibull distribution, that is, with a distribution function
\[ F_R(r) = 1 - e^{-\left(r/a\right)^c}, \quad r \geq 0. \]

The wire is used under water and is exposed to a load increasing with time, due to growth of the organic material attached to its surface. The rate of growth is considered constant, \( \gamma \); hence, during a period of length \( t \), the load has increased by \( \gamma t \) (the initial weight is neglected).

At the lifetime \( T \), when the weight exceeds the strength, obviously \( R = \gamma T \) or equivalently, \( T = R/\gamma \). Hence the lifetime distribution is given by
\[ F_T(t) = P(T \leq t) = P(R/\gamma \leq t) = F_R(\gamma t) = 1 - e^{-\left(\gamma t/a\right)^c}. \]

Since \( R(t) = e^{-\left(\gamma t/a\right)^c} \), the cumulative failure-intensity function \( \Lambda(t) = (\gamma t/a)^c \) and hence \( \lambda(t) = c \gamma a (\gamma t/a)^{c-1} \). Suppose that in some units, \( a = 1 \) and \( \gamma = 0.1 \). For different choices of the shape parameter \( c \), the failure-intensity function is presented in Figure 7.1; from top to bottom \( c = 0.8 \), \( c = 1.0 \), and \( c = 1.2 \). Note that, depending on the choice of \( c \), the function might be classified as IFR, DFR, or have a constant failure intensity. \( \square \)

In the previous examples failure intensity described the properties of populations of some components. Principally, it was used to model the uncertainty of properties like quality or strength of a component. A different situation is met in the following example.

**Example 7.3 (Constant failure intensity).** Consider periods in days between serious earthquakes worldwide (presented in Example 1.1). This data set was investigated in many aspects in Chapter 4. Now assume that we at some fixed date \( s \) (say, today) start counting the time until the next earthquake. As in Chapter 2, we thus consider a stream of events with intensity \( \lambda_A \)
and the event \(A=\text{“Earthquake occurs”}\). Recall the properties I-III in Chapter 2 (page 40), which we assume to be satisfied in our situation; then Theorem 2.5 gives
\[
P(T > t) = P(N(s, s + t) = 0) = e^{-\lambda_A t}
\]
and hence \(A(t) = \lambda_A t\), giving by differentiation \(\lambda(t) = \lambda_A\). Thus, the failure-intensity function is constant and equal to the intensity of the stream.

If the intensity of events \(A\) is non-stationary, \(\lambda_A(s)\) say, then similar calculations lead to the failure-intensity function \(\lambda(t) = \lambda_A(s + t)\).

Often one is interested in whether the time to failure is longer than \(t\), if we know the age of the component to be \(t_0\), say, i.e. we wish to compute the probability \(P(T > t_0 + t | T > t_0)\). This conditional probability can be easily computed if the failure intensity is known:
\[
P(T > t_0 + t | T > t_0) = \frac{P(T > t_0 + t \text{ and } T > t_0)}{P(T > t_0)}
= \frac{P(T > t_0 + t)}{P(T > t_0)} = \frac{e^{-\int_{t_0}^{t_0+t} \lambda(s) \, ds}}{e^{-\int_{t_0}^{t} \lambda(s) \, ds}}
= e^{-\int_{t_0}^{t_0+t} \lambda(s) \, ds}.
\]
Note that for \(\lambda(s) = \lambda\) (being constant),
\[
P(T > t_0 + t | T > t_0) = e^{-\lambda t},
\]
that is, old components have the same distribution for their remaining life as new ones. This is sometimes stated as “memorylessness” of the exponential distribution.
We exemplify the use of Eq. (7.2) with an example from life insurance where the "components" are humans. The variability of lifetimes, and hence the failure intensity, depends on the choice of population. For example when considering the lifetimes of inhabitants in two countries, there can be different diseases common for each country, habits of smoking, traffic situation, frequency of catastrophes like earthquakes, etc. which leads to different functions $\lambda(s)$.

Example 7.4 (Life insurance). Let $T$ be a lifetime for a human. In life insurance, $P(T > t)$ is specified by standard tables, based on observed lifetimes of a huge number of people; one example is the Norwegian N-1963 standard. A popular choice of $\lambda(s)$ is the Gompertz-Makeham distribution (with roots to Makeham [53]), given by the failure-rate function

$$\lambda(s) = \alpha + \beta e^s,$$

$s$ measured in years. For example, for N-1963, the estimates are

$$\alpha^* = 9 \cdot 10^{-4}, \quad \beta^* = 4.4 \cdot 10^{-5}, \quad c^* = 10^{0.042}.$$

We want to solve the following problems:

(i) Calculate the probability that a person will reach the age of at least seventy.

(ii) A person is alive on the day he is thirty. Calculate the conditional probability that he will live to be seventy.

For problem (i), we obtain the solution as

$$P(T > 70) = \exp \left\{ - \int_0^{70} \lambda(s) \, ds \right\} = 0.63.$$

The solution to problem (ii) is given by Eq. (7.2) as

$$P(T > 70 | T > 30) = \exp \left\{ - \int_{30}^{70} \lambda(s) \, ds \right\} = 0.65.$$

Combining different risks for failure

In real life, there are often several different types of risks that may cause failures; one speaks of different failure modes. Each of these has an intensity $\lambda_i(s)$ and a lifetime $T_i$. We are interested in the distribution of $T$: the time instant when the first of the modes happen. If $T_i$ are independent then the event $T > t$ is equivalent to the statement that all lifetimes $T_i$ exceed $t$, i.e. $T_1 > t, T_2 > t, \ldots, T_n > t$ and hence

$$P(T > t) = P(T_1 > t) \cdot \ldots \cdot P(T_n > t) = e^{-\int_0^t \lambda_1(s) \, ds} \cdot \ldots \cdot e^{-\int_0^t \lambda_n(s) \, ds}$$

$$= e^{-\int_0^t \lambda_1(s) \, ds - \ldots - \int_0^t \lambda_n(s) \, ds} = e^{-\int_0^t \lambda_1(s) + \ldots + \lambda_n(s) \, ds}$$

(7.3)
which means that the failure intensity, including the \( n \) independent failure modes, is \( \lambda(s) = \sum \lambda_i(s) \).

**Remark 7.1.** In the special case when the failures can be related to external actions (accidents causing failures) constituting independent streams \( A_i \), each with constant intensity \( \lambda_i \), Eq. (7.3) was already derived in Chapter 2. There, a stream \( A = A_1 \cup \ldots \cup A_n \) was considered, with the interpretation that at least one of \( A_i \) happens. The intensity \( \lambda_A \) is equal to the sum of intensities \( \lambda_i \), (see Eq. (2.9)). If the streams are Poisson then the stream \( A \) is also Poisson (see Theorem 7.1, p. 188), and hence

\[
P(T \leq t) = 1 - e^{-\lambda_A t},
\]

i.e. \( T \) is exponentially distributed with intensity \( \lambda_A = \lambda_1 + \cdots + \lambda_n \). □

### 7.1.2 Estimation procedures

Earlier in this chapter, we have introduced the notions failure intensity and survival function when studying the distribution of the time \( T \) to failure for some items. In this section we discuss how these functions can be estimated from data. Obviously, a standard (parametric) method is to assume that \( F(t) \) belongs to a class of distributions \( F(t; \theta) \), estimate parameters, and finally calculate \( \lambda(s) \). We here instead present a non-parametric method, commonly used in applications with lifetime data.

In reliability studies as well as in clinical trials in the medical sciences, it is not always possible to wait for all units to reach their final “lifetime” (lifetime could mean time for failure, or death, or the appearance of a certain condition). An intricate issue is that censored data may occur; for example, an item may not have reached its lifetime until the study is finished or is lost during the time (e.g. people move). Efficient estimation procedures need to take censoring aspects into account.

In this section, we review some commonly used tools within statistical analysis of survival or reliability data: the Nelson–Aalen estimator for estimation of the cumulative failure-intensity function \( \Lambda(t) \), and the log-rank test for testing hypotheses about the failure-intensity functions of two samples. For further reading, we refer to Klein and Moeschberger [43] where a thorough presentation of methods in survival analysis is given.

**Nelson–Aalen estimator**

This estimator was first presented by Nelson [57] and later refined by Aalen [1]. It estimates the cumulative failure-intensity function

\[
\Lambda(t) = \int_0^t \lambda(s) \, ds.
\]
The Nelson–Aalen estimator is considered to have good small-sample performance, i.e. when \( n \) is small, when estimating the survival function.

Introduce the following notation:

- \( t_i \): Time points for failures
- \( d_i \): Number of failures at time \( t_i \)
- \( n_i \): Number of items at risk at time \( t_i \), i.e. number of items not yet failed prior to failure time \( t_i \).

The estimator is given by

\[
\Lambda^*(t) = \sum_{t_i \leq t} \frac{d_i}{n_i}
\]

and thus \( R^*(t) = \exp(-\Lambda^*(t)) \). (Note that \( R^*(t) \neq 1 - F_n(t) \).)

If censoring is present, the values of \( n_i \) will be affected, leading to a change in the value of the estimated survival function.

**Example 7.5 (Cycles to failure).** In an experiment, the number of cycles to failure for reinforced concrete beams was measured in seawater and air [37]. The observations (in thousands) were as follows:

- **Seawater:** 774, 633, 477, 268, 407, 576, 659, 963, 193
- **Air:** 734, 571, 520, 792, 773, 276, 411, 500, 672

**Parametric model.** A Weibull distribution is often used to model the strength of a material, and plots of the observations in Weibull probability paper indicate that Weibull might be a good choice. With

\[
F_T(t) = 1 - e^{-(t/a)^c} , \quad t \geq 0
\]

one finds by statistical software the ML estimates \( a^* = 620, \ c^* = 2.63 \) for seawater conditions. Based on these, an estimate of the cumulative failure-intensity function can easily be computed and is shown as the solid curve in Figure 7.2.

**Non-parametric model.** The following table gives the Nelson–Aalen estimate of the cumulative failure-intensity for seawater (creation of the corresponding scheme for air is left as an exercise):

<table>
<thead>
<tr>
<th>( i )</th>
<th>( t_i )</th>
<th>( n_i )</th>
<th>( d_i )</th>
<th>( \Lambda^*(t_i) )</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>193</td>
<td>9</td>
<td>1</td>
<td>0.1111</td>
</tr>
<tr>
<td>2</td>
<td>268</td>
<td>8</td>
<td>1</td>
<td>0.2361</td>
</tr>
<tr>
<td>3</td>
<td>407</td>
<td>7</td>
<td>1</td>
<td>0.3790</td>
</tr>
<tr>
<td>4</td>
<td>477</td>
<td>6</td>
<td>1</td>
<td>0.5456</td>
</tr>
<tr>
<td>5</td>
<td>576</td>
<td>5</td>
<td>1</td>
<td>0.7456</td>
</tr>
<tr>
<td>6</td>
<td>633</td>
<td>4</td>
<td>1</td>
<td>0.9956</td>
</tr>
<tr>
<td>7</td>
<td>659</td>
<td>3</td>
<td>1</td>
<td>1.3290</td>
</tr>
<tr>
<td>8</td>
<td>774</td>
<td>2</td>
<td>1</td>
<td>1.8290</td>
</tr>
<tr>
<td>9</td>
<td>963</td>
<td>1</td>
<td>1</td>
<td>2.8290</td>
</tr>
</tbody>
</table>

In Figure 7.2, the Nelson–Aalen estimate is shown (the stair-wise function). From the plot it can be judged that we have a case, which can be considered IFR. □
Log-rank test

Finally, we present a statistical test, called the log-rank test, for comparison of the intensities $\lambda_1(t)$ and $\lambda_2(t)$ in two groups (1 and 2). The aim is to test the hypothesis

$$H_0 : \lambda_1(t) = \lambda_2(t).$$

The test can be generalized to more than two groups, but we content ourselves in this exposition to the simplest case and refer to the literature for more specialized studies. (Note that two groups can have different number of elements.)

Consider the time points for failures $t_1, t_2, \ldots, t_D$, both groups considered. Introduce the following notation:

- $d_{i1}$: Number of failures in group 1 at times $t_i$
- $d_{i2}$: Number of failures in group 2 at times $t_i$
- $d_i$: $d_i = d_{i1} + d_{i2}$
- $n_{i1}$: Number of items in group 1 at risk at time $t_i$, i.e. number of items not yet failed prior to failure time $t_i$.
- $n_{i2}$: Number of items in group 2 at risk at time $t_i$, i.e. number of items not yet failed prior to failure time $t_i$.
- $n_i$: $n_i = n_{i1} + n_{i2}$

The test quantity is

$$Q = \frac{1}{s^2} \left( \sum_{i=1}^D d_{i1} - \sum_{i=1}^D d_i \frac{n_{i1}}{n_i} \right)^2$$

where

$$s^2 = \sum_{i=1}^D \frac{d_i}{n_i} \frac{n_i - d_i}{n_i} \frac{n_{i1}n_{i2}}{n_i(n_i - 1)}.$$
The test is similar to the $\chi^2$ test and is as follows: If $Q \geq \chi^2_\alpha(1)$, reject $H_0$. (Note that since for $X \in N(0, 1)$, $X^2 \in \chi^2_1$; hence, $\chi^2_\alpha(1) = \lambda^{2/2}_\alpha$.)

**Example 7.6 (Cycles to failure).** Consider again the experiment mentioned in Example 7.5. In Figure 7.3, the cumulative failure-intensity functions are shown for air (solid) and seawater (dashed). Does seawater seem to lessen the number of cycles to failure, or in other words, can we reject the hypothesis that $\lambda_1(s) = \lambda_2(s)$, where group 1 corresponds to seawater conditions, group 2 to air?

From the 18 observations of lifetimes, the quantities needed are computed and presented in the following table:

<table>
<thead>
<tr>
<th>$t_i$</th>
<th>$n_{i1}$</th>
<th>$d_{i1}$</th>
<th>$n_{i2}$</th>
<th>$d_{i2}$</th>
<th>$n_i$</th>
<th>$d_i$</th>
</tr>
</thead>
<tbody>
<tr>
<td>193</td>
<td>9</td>
<td>1</td>
<td>9</td>
<td>0</td>
<td>18</td>
<td>1</td>
</tr>
<tr>
<td>268</td>
<td>8</td>
<td>1</td>
<td>9</td>
<td>0</td>
<td>17</td>
<td>1</td>
</tr>
<tr>
<td>276</td>
<td>7</td>
<td>0</td>
<td>9</td>
<td>1</td>
<td>16</td>
<td>1</td>
</tr>
<tr>
<td>407</td>
<td>7</td>
<td>1</td>
<td>8</td>
<td>0</td>
<td>15</td>
<td>1</td>
</tr>
<tr>
<td>411</td>
<td>6</td>
<td>0</td>
<td>8</td>
<td>1</td>
<td>14</td>
<td>1</td>
</tr>
<tr>
<td>477</td>
<td>6</td>
<td>1</td>
<td>7</td>
<td>0</td>
<td>13</td>
<td>1</td>
</tr>
<tr>
<td>500</td>
<td>5</td>
<td>0</td>
<td>7</td>
<td>1</td>
<td>12</td>
<td>1</td>
</tr>
<tr>
<td>520</td>
<td>5</td>
<td>0</td>
<td>6</td>
<td>1</td>
<td>11</td>
<td>1</td>
</tr>
<tr>
<td>571</td>
<td>5</td>
<td>0</td>
<td>5</td>
<td>1</td>
<td>10</td>
<td>1</td>
</tr>
<tr>
<td>576</td>
<td>5</td>
<td>1</td>
<td>4</td>
<td>0</td>
<td>9</td>
<td>1</td>
</tr>
<tr>
<td>633</td>
<td>4</td>
<td>1</td>
<td>4</td>
<td>0</td>
<td>8</td>
<td>1</td>
</tr>
<tr>
<td>659</td>
<td>3</td>
<td>1</td>
<td>4</td>
<td>0</td>
<td>7</td>
<td>1</td>
</tr>
<tr>
<td>672</td>
<td>2</td>
<td>0</td>
<td>4</td>
<td>1</td>
<td>6</td>
<td>1</td>
</tr>
<tr>
<td>734</td>
<td>2</td>
<td>0</td>
<td>3</td>
<td>1</td>
<td>5</td>
<td>1</td>
</tr>
<tr>
<td>773</td>
<td>2</td>
<td>0</td>
<td>2</td>
<td>1</td>
<td>4</td>
<td>1</td>
</tr>
<tr>
<td>774</td>
<td>2</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>3</td>
<td>1</td>
</tr>
<tr>
<td>792</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>2</td>
<td>1</td>
</tr>
<tr>
<td>963</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>1</td>
</tr>
</tbody>
</table>
From this table, we find

\[ \sum_{i=1}^{18} d_{i1} = 9, \quad \sum_{i=1}^{18} d_{i1} \frac{n_{i1}}{n_i} = 8.02, \]

and \( s^2 = 4.16 \). It follows that \( Q = 0.23 \). Hence \( \chi^2_{0.05}(1) = 3.84 \) and we do not reject the hypothesis about equal failure intensity. \( \square \)

### 7.2 Absolute Risks

In the previous section we introduced the concept of failure intensity \( \lambda(s) \), which describes variability of lifelengths in a population of components, objects or human beings. Extensive statistical studies are needed to estimate \( \lambda(s) \). More often, observed information is not sufficient to determine the failure intensity. In this section we consider situations when information about failures is less detailed: instead of knowing the times for failures \( t_i \), access is available only to the total number; for example, failures during a specified period of time (or in a certain geographical region). Let us call failures “accidents”, and suppose that these cause serious hazards for humans. Absolute risk is meant as the chance for a person to be involved in a serious accident (fatal), or of developing a disease, over a time period. Chances for accidents due to different activities are often compared. A full treatment of such issues is outside of the scope of this book and hence we only mention some aspects of the problem.

**Poisson assumption**

Let \( N \) be the number of deaths due to an activity, in a specified population (a country), and period of time (often one year). The distribution of \( N \) may not be easy to choose. For example, if \( N \) is the number of accidents that occur independently with small probability then \( N \) may have a Poisson distribution, \( N \sim Po(\mu) \), where \( \mu = E[N] \). This is a consequence of the approximation of the binomial distribution by the Poisson distribution (the law of small numbers). For instance, it seems reasonable to model the number of commercial air-carrier crashes during one year by a Poisson variable. However, the number of people killed in those accidents is not Poisson distributed, since usually a large number of people are killed in a single accident. Since \( N \), for different activities, can have different types of distributions, risks are often compared by means of averages. However, as demonstrated next, such comparisons have to be made with care.

**Example 7.7 (Number of deaths in traffic).** In year 1998 it was reported about 41,500 died in traffic accidents in the United States while in Sweden the number was about 500 [7]. In order to compare these numbers, one needs to
compare the size of populations in both countries. A fraction of the numbers of deaths by the size of population, giving the frequencies of death, is often used to measure risk for death. In US the frequency was about 1 in 6000, circa $1.7 \cdot 10^{-4}$; while in Sweden, 1 in 17000, circa $0.6 \cdot 10^{-4}$, which is nearly three times lower. (Comparisons of chances to die in traffic accidents between countries can be difficult since statistics may use different definitions and have different accuracy.)

The last example turns our attention to a problem often discussed in the literature of reliability and risk analysis, namely when risks are acceptable (or tolerable).

**Tolerable risks**

Often a distinction is made between the so-called “voluntary risks” and the “background risks”. Clearly accidents due to an activity like mountaineering are obviously a voluntary risk, while the risk for death because of a collapse of a structure is an example of a background risk and is much smaller. (In United Kingdom one estimates that one hour of climbing has twice as high probability for a fatal accident than for a fatal accident in 100 years caused by structural failures, see table in [77].)

In the literature indicators of tolerable risks can be found, see e.g. Otway et al. [59]. The magnitudes of the risks specified in Table 7.1 are meant approximatively: the number of fatal accidents during a year divided by the size of the population exposed for the hazard. (Fatal accidents in traffic belongs to the second category of hazards.)

**Example 7.8 (Perished in traffic).** Continuation of Example 7.7. The estimated chances of dying in traffic in the U.S. was nearly three times as high as in Sweden. When looking for explanation for the difference, the first thing to be explored is the total exposure of the populations for the hazard, in other words if an average inhabitant of the U.S. spends more time in a car

<table>
<thead>
<tr>
<th>Risk of death per person per year</th>
<th>Characteristic response</th>
</tr>
</thead>
<tbody>
<tr>
<td>$10^{-3}$</td>
<td>Uncommon accidents; immediate action is taken to reduce the hazard</td>
</tr>
<tr>
<td>$10^{-4}$</td>
<td>People spend money, especially public money to control the hazard (e.g. traffic signs, police, laws)</td>
</tr>
<tr>
<td>$10^{-5}$</td>
<td>Parents warn their children of the hazard (e.g. fire, drowning, fire arms, poison)</td>
</tr>
<tr>
<td>$10^{-6}$</td>
<td>Not of great concern to average person; aware of hazard, but not of personal nature; act of God.</td>
</tr>
</tbody>
</table>
than a person in Sweden does. For traffic-related accidents, exposure is often measured by total vehicle kilometres.

Neglecting that the exposures are estimates and hence uncertain numbers, we found that in 1998 the intensity \( \lambda = \frac{E[N]}{t} \) in the U.S. was about 1 person per 100 million km driven, while in Sweden, \( \lambda \) is 1 per 125 million km. The conclusion is that a person who drives 0.01-million km during one year has a chance of the order \( 10^{-4} \), \( 0.8 \cdot 10^{-4} \), respectively, of dying as a result of traffic accidents in both countries. In other words, the chances are quite similar.

In our setup the absolute risk was derived for an average member of the population (a person chosen at random). However, the natural question is whether the same risk is valid for some subpopulations: geographical, stratified by age, income, etc. We return to this kind of question when Poisson regression is presented. Here we end with an example where we compare risk for fire in an average school in a country as a whole compared with a school in a smaller urban region.

**Example: Intensity of fire ignitions in schools in Sweden**

In published statistical tables ([74], [76]) one can find that in 2002 there were \( k = 13053 \) educational buildings in Sweden and \( n = 422 \) fires were recorded. (We ignore the fact that these two numbers are uncertain, taken from different statistical tables.) As is common practice in fire safety, the assumption is made that the stream of fire ignitions in a school is Poisson, and that ignitions in different schools happen independently (see examples in Chapter 2).

**Constant intensity**

The simplest approach is to assume that intensities in all schools are constant and equal to \( \lambda \) (per school). As derived in Example 4.8, the ML estimate is

\[
\lambda^* = \frac{n}{k} = \frac{422}{13053} = 0.032 \text{ [year}^{-1}]\text{].}
\]

Now the probability of at least one fire in a school in three years, \( P_1(A) \), \( t = 3 \) years, can be estimated as

\[
P_1(A) = 1 - e^{-\lambda^* t} = 1 - e^{-0.097} = 0.092.
\]

The expected number of fires in an average school during a three-year period is found to be \( \mu = 3 \cdot 0.032 = 0.096 \).

**Validation of the model: Schools in Stockholm**

Here we use a small data set presented in [69] and further analysed in [68]. Data contain the number of fires \( n_i \) for 20 schools in Stockholm, Sweden.
These have been chosen, at random, from Stockholm Fire Department files containing reports from actions in 2000-2002. The number of fires in each of the schools was $n_i$, $i = 1, \ldots, 20$

\[
1 \hspace{1em} 1 \hspace{1em} 3 \hspace{1em} 1 \hspace{1em} 1 \hspace{1em} 1 \hspace{1em} 3 \hspace{1em} 1 \hspace{1em} 2 \hspace{1em} 1 \hspace{1em} 1 \hspace{1em} 1 \hspace{1em} 1 \hspace{1em} 1 \hspace{1em} 2 \hspace{1em} 1 \hspace{1em} 1 \hspace{1em} 1 \hspace{1em} 1
\]

We now investigate if the risk for fire in schools in Stockholm differs from the average risk for the country, i.e. circa 0.1 fires on average during three years. (We neglect some uncertainties in the estimate $\mu^* = 0.032$ found above and assume stationarity of fire ignitions in years 2000-2002.) Our model is that the number of fires (during three years) in Stockholm schools is independent Poisson distributed variables with mean $\mu_S$, which has to be estimated. We suspect that $\mu_S > \mu$. However, there is a small difficulty here: namely the fire department files contain only addresses of schools where the fire started. Thus, schools with zero fires are not present in the data (see Remark 4.3 where the inspection paradox was discussed); hence the average of the data is an obviously biased estimate of $\mu_S$. In order to resolve the problem we need to work with conditional probabilities, conditionally that one knows that there was already a fire in a school.

We proceed as follows. Using data, we derive the ML estimate $\theta^*$ of the three-year average $\theta = \mu_S$. The asymptotic normality of the estimation error is used to construct a 0.95-confidence interval for $\mu_S$. If the country average (here consider as known constant) lies outside the interval we can reject the hypothesis that the intensity of Stockholm school fires is the same as the average in the country.

**ML estimate of $\mu_S$**

Let $N$ be the number of fires observed in schools that had at least one fire during the period. Clearly $N$ may take values $1, 2, \ldots$ with probabilities

$$P(N = n) = \frac{\theta^n}{n!} \frac{e^{-\theta}}{1 - e^{-\theta}},$$

where $\theta$ is the unknown average number of fires in a school in 3 years. Suppose $n_1, \ldots, n_k$ are independent observations from $k$ schools, then the likelihood-, log likelihood-, and the derivatives functions are given by

$$L(\theta) = \prod_{i=1}^{k} P(N = n_i) = \prod_{i=1}^{k} \frac{\theta^{n_i}}{n_i!} \frac{e^{-\theta}}{1 - e^{-\theta}},$$

$$l(\theta) = - \sum_{i=1}^{k} \ln(n_i!) + \ln(\theta) \sum_{i=1}^{k} n_i - k\theta - k \ln(1 - e^{-\theta}),$$

$$\dot{l}(\theta) = k \frac{\bar{n}}{\theta} - \frac{k}{1 - e^{-\theta}}, \quad \ddot{l}(\theta) = -k \left( \frac{\bar{n}}{\theta^2} - \frac{e^{-\theta}}{(1 - e^{-\theta})^2} \right),$$

(7.5)
where \( \bar{n} = \sum_{i=1}^{k} n_i/k \). The ML estimate \( \theta^* \) is the solution to the equation

\[ \theta^* = \bar{n}(1 - e^{-\theta^*}) \]

while \( \sigma^*_E = 1/\sqrt{-l(\theta^*)} \). The estimate \( \theta^* \) can be found by means of numerical procedures or a graphical method to solve the equation. For the data the solution is \( \theta^* = 0.5481 \) while \( \sigma^*_E = 0.2151 \). Since with approximately 0.95 confidence

\[ \mu_S \in [0.5481 - 1.96 \cdot 0.2151, 0.5481 + 1.96 \cdot 0.2151] = [0.13, 0.97] \]

we reject the hypothesis that \( \mu_S = 0.096 \), i.e. that the schools in Stockholm has the same yearly average number of fire as the country as whole.

### 7.3 Poisson Models for Counts

As we have seen the number of accidents \( N_i \) in different populations may vary; it also can change from year to year. Sometimes the differences can be explained as the result of random variability, i.e. when \( N_i \) are independent outcomes of the same random experiment. However, often the independence can be questionable or properties of the “experiment” changes with time, hence \( N_i \) are not iid.

In this section we study this type of problems closer. We do not treat it in full generality but assume that \( N_i \) are independent Poisson distributed variables, counting a number of failures (accidents) in different populations or time periods. Since the Poisson distribution has only one parameter, it means that our model is fully specified if \( \mu_i = \mathbb{E}[N_i] \) are estimated.

Let \( N \) be the number of people killed in traffic, for instance, the next year. We assume that \( N \in \text{Po}(\mu) \) where \( \mu = \mathbb{E}[N] \). In order to be able to make any statement of type \( P(N > 400) \), \( \mu \) needs to be estimated. This is usually done using historical data. Denote by \( N_i \) the number of people killed in year \( i \). We assume that \( N_i \) are independent, Poisson distributed with mean \( \mu_i = \mathbb{E}[N_i] \).

We have access to historical data and we know that \( N_i = n_i \).

Using the historical data, we wish to find a pattern of how \( \mu_i \) varies in order to extrapolate the variability to the future, i.e. the unknown value \( \mu \). Obviously, if there is no clear pattern in \( \mu_i \), the ML estimate \( \mu^* \) and the historical data can hardly be used to predict future. However, if the mechanism generating accidents can be assumed to be stationary, then \( \mu_i = \mu \) for all \( i \).

The average value \( \bar{n} = \sum n_i/k \) is the ML estimate \( \mu^* \) of \( \mu \).

In this section, we first briefly consider two data sets with regard to possible constant \( \mu \) over time. For a more thorough analysis, tests are then introduced: to test for Poisson distribution and \( \mu_i = \mu \) (constant mean). Finally, for the situation with \( \mu_i \) not constant, the expected value is modelled as a function of other, explanatory variables.

**Example 7.9 (Flight safety).** In Example 6.12 flight safety was studied. From “Statistical Abstract of the United States”, data for the number of crashes in the world during the years 1976-1985 are found:

\[
\begin{align*}
24 & \quad 25 & \quad 31 & \quad 31 & \quad 22 & \quad 21 & \quad 26 & \quad 20 & \quad 16 & \quad 22
\end{align*}
\]
Here a model for constant mean number of accidents for the period seems sensible.

Sometimes trends are observed in \( n_i \); these seem to increase (or decrease) over time. A possible model can be that the mean changes linearly viz.

\[ \mu_i = \mu + \beta \cdot i \]

where \( \beta \) is a constant and \( i \) the year. Historical data are used to find estimates \( \mu^* \) and \( \beta^* \). However, often more complicated models for the variable mean has to be used.

**Example 7.10 (Traffic accidents in Sweden).** Suppose we are interested in the number of deaths related to traffic accidents in Sweden. From official statistics [7], we find that during the years 1990-2004 the following number of people died due to accidents in Sweden:

\[
\begin{align*}
772 & \quad 745 & \quad 759 & \quad 632 & \quad 545 & \quad 531 & \quad 508 & \quad 507 & \quad 492 & \quad 536 & \quad 564 \\
551 & \quad 532 & \quad 529 & \quad 480
\end{align*}
\]

We can see that number of deaths is decreasing and obviously one cannot assume that the data are observations of independent Poisson variables with constant mean.

### 7.3.1 Test for Poisson distribution – constant mean

In the following subsection we test whether data contradict the iid Poisson model for \( N_i \), i.e. \( \mu_i = \mu \). However, first we present a useful approximation of the Poisson distribution valid for large populations, commonly assumed to be valid for \( \mu > 15 \).

For \( N \in \text{Po}(\mu) \) when \( \mu \) is large, a very effective tool is to approximate the Poisson distribution by a normal distribution (the so-called normal approximation).

**Normal approximation of Poisson distribution.** Let \( N \) be a Poisson distributed random variable with expectation \( \mu \),

\[ N \in \text{Po}(\mu) \]

If \( \mu \) is large (in practice, \( \mu > 15 \)), we have approximately that

\[ N \in \text{N}(\mu, \mu). \tag{7.6} \]

**Example 7.11 (Accidents in mines).** Consider Example 2.11, page 39. We there estimated the intensity of accidents in mines \( \lambda = 3 \text{ year}^{-1} \) and argued
that the stream of accidents is Poisson. Suppose we want to calculate the probability of at least 80 accidents during 25 years, that is, $P(N(25) \geq 80)$. Since the stream is Poisson, $N(25) \in \text{Po}(3 \cdot 25) = \text{Po}(75)$. For simplicity of notation, let $N = N(25)$; we compute

$$P(N \geq 80) = 1 - P(N \leq 79) = 1 - P(N = 0) - P(N = 1) - \ldots - P(N = 79),$$

which might be cumbersome\(^1\). An alternative solution is to employ the normal approximation instead to evaluate the probability:

$$P(K \geq 80) \approx 1 - \Phi((79.5 - 75)/\sqrt{75}) = 1 - \Phi(0.52) = 0.30.$$  

\(\square\)

Suppose we have $k$ observations $n_1, \ldots, n_k$ of Poisson distributed quantities $N_1, \ldots, N_k$. Our assumption is that all $N_i \in \text{Po}(\mu)$, i.e. stationarity (homogeneity) is present.

For small values of $\mu$ but large $k$ we can use the $\chi^2$ test presented in Section 4.2.2 to validate the model.

In the case when $\mu$ is large, to test whether data do not contradict the assumption of stationarity or constant mean, often the following property of a Poisson distribution is used: $\text{V}[N] = \text{E}[N] = \mu$. In the case of a Poisson distribution, the ratio $\text{V}[N]/\text{E}[N]$ is obviously equal to 1. The test to be presented below is based on this fact. If $\mu$ is large, by Eq. (7.6) $N \in \text{N}(\mu, \mu)$ and we can estimate $\text{E}[N]$ by $\bar{n}$ and $\text{V}[N]$ by $s^2_{k-1}$. As confidence interval for $\theta = \text{V}[N]/\text{E}[N]$ can be constructed, viz.

$$\frac{\bar{n} \chi^2_{1-\alpha/2}(k-1)}{k-1} \leq \frac{\text{V}[N]}{\text{E}[N]} \leq \frac{\bar{n} \chi^2_{\alpha/2}(k-1)}{k-1}$$  \hspace{8cm} (7.7)

with approximate confidence $1 - \alpha$. If $\theta = 1$ is not in that interval, the hypothesis that $N$ is Poisson distributed is rejected. For further reading about tests of this type, see Brown and Zhao [6].

**Remark 7.2.** The assumption of equal variance and mean is not always satisfied working with real data. If $\text{V}[X] > \text{E}[X]$, overdispersion is present. Additional statistical tests for this are found in the literature (cf. [18]).

In the following example we study how the hypothesis that $\mu_i$ are constant over time can be validated.

**Example 7.12 (Flight safety).** This is a continuation of Example 7.12 where number of crashes of commercial air carriers in the world during the years 1976-1985 were presented. Let us assume that the flight accidents form

\(^1\)Some of the probabilities can be hard to compute and Stirling’s formula $n! \approx \sqrt{2\pi n} n^{n+0.5} e^{-n}$ needs to be used.
a Poisson stream and hence \( n_i \) are independent observations of \( \text{Po}(\mu) \) distributed variables.

A point estimate is \( \mu^* = \bar{n} = 23.8 \). Since \( \mu^* > 15 \), the model implies that \( n_i \) can be considered independent observations of an \( \text{N}(\mu, \mu) \) distributed variable (cf. Eq. (7.6)). Consequently, we expect for the ML estimate of the variance \( s^2_{k-1} \approx \bar{n} \). For this data set, \( s^2_{k-1} = 22.2 \), which is close to 23.8.

Next, the confidence interval given in Eq. (7.7) is computed:

\[
\left[ \frac{23.8}{22.2} \cdot \frac{2.7}{9}, \frac{23.8}{22.2} \cdot \frac{19.02}{9} \right] = [0.32, 2.26]
\]

and hence the hypothesis of constant \( \mu \) is not rejected. \( \square \)

### 7.3.2 Test for constant mean – Poisson variables

Suppose it can be assumed that data are observations of independent Poisson distributed variables but we suspect that the mean is not constant. More precisely, we check if the data do not contradict the assumption that \( \text{E}[N_i] = \mu \). The test we wish to use is based on a quantity called deviance and is based on log-likelihood values. The specific of the test is that we do not need to assume that the mean \( \mu \) is high.

**Statistical test using deviance**

Let \( N_i \) be independent Poisson distributed variables and consider two models:

- a more general model, where no restriction are put on the means \( \mu_i = \text{E}[N_i] \),
- and a simpler where all means are equal, \( i.e. \mu_i = \mu \).

Let \( n_i \) be the observed values of \( N_i \). Using the ML method the optimal estimates \( \mu_i^* = n_i \) if the general model is assumed while the ML estimate is \( \mu^* = \frac{\sum n_i}{k} \) for the simpler, more restrictive model.

Since the more general model contains the simpler, the log-likelihood function \( l(\mu_1^*, \ldots, \mu_k^*) \) must be higher than \( l(\mu^*) \). Higher values of the log-likelihood function means that the observed data are more likely to occur under the model, hence the increase of the function is a measure of how much better the more complex model explains the data. It can be shown that the following test quantity, called deviance,

\[
\text{DEV} = 2 \cdot \left( l(\mu_1^*, \ldots, \mu_k^*) - l(\mu^*) \right),
\]

for large \( k \) is \( \chi^2(k - 1) \) distributed if the simpler model is true\(^2\). Thus if \( \text{DEV} > \chi^2(k - 1) \), the difference between log-likelihoods cannot be explained by the statistical variability and hence the simpler model should be rejected.

Straightforward calculations lead to the following formula

\[
\text{DEV} = 2 \sum_{i=1}^{k} n_i (\ln(\mu_i^*) - \ln(\mu^*)) = 2 \sum_{i=1}^{k} n_i (\ln(n_i) - \ln(\bar{n})),
\]

where for \( n_i = 0 \) we let \( n_i \ln(n_i) = 0 \).

\(^2\)The test can also be used for small \( k \) if \( \mu \) is large.
174 7 Intensities and Poisson Models

Example 7.13 (Daily rains). This is continuation of Example 2.14 where the data \( n_i, i = 1, \ldots, 12 \), are numbers of daily rains exceeding 50 mm observed in month \( i \), during years 1961-1999. We suspect that the simplest model of constant mean \( \mu_i = \mu \), estimated to be \( \mu^* = \bar{n} = 3.67 \), is not correct. Let us compute the deviance by Eq. (7.9)

\[
\text{DEV} = 2 \left\{ 4(\ln(4) - \ln(3.67)) + \cdots + 10(\ln(10) - \ln(3.67)) \right\} = 19.64.
\]

The value 19.64 should be compared with the 0.05 quantile found as \( \chi^2_{0.05}(11) = 19.68 \). Obviously this is a border case. Although DEV is slightly below the quantile we decide that with approximative confidence 0.95 the hypothesis of the means \( \mu_i = \mu \) can be rejected. □

Example 7.14 (Motorcycle data). Consider the data set from Problem 4.9 where the numbers of killed motorcycle riders in Sweden 1990-1999, are reported. We suspect that the simplest model that \( E[N_i] = \mu_i = \mu \) explains well the data and wish to test it against the more complex model that \( E[N_i] = \mu_i \).

\[
\text{DEV} = 2 \sum_{i=1}^{10} n_i (\ln(n_i) - \ln(\bar{n})) = 5.5,
\]

since \( \bar{n} = 33.1 \). The value 5.5 should be compared with the 0.05 quantile found as \( \chi^2_{0.05}(9) = 16.92 \). We conclude that the more complex model does not explain data better than the simpler one does. □

7.3.3 Formulation of Poisson regression model

As seen in the previous subsection, often the assumption of constant mean \( \mu \) for the number of accidents \( N_i \) has to be rejected. In such a situation it is desirable to find a model for the variability of the mean \( \mu_i \). A standard approach is to find (or select from available data) a collection of explanatory variables (quantities) that influence means. A method to find a functional relation between the explanatory variables and the means is the so-called Poisson regression.

Regression techniques are widely used in statistical applications found in most sciences, a standard reference is the book by Draper and Smith [22]. The random outcomes of an experiment \( Y_i \) (called responses or dependent variables) of the \( i \)th experiment have means related to a vector of \( p \), say, explanatory variables \( x_1, x_2, \ldots, x_p \).

A regression model

Consider a sequence of Poisson distributed counting variables \( N_i, \ i = 1, \ldots, k \), for example the number of accidents (failures) occurring in year \( i \). Let \( n_i \) be the observed values of \( N_i \). Suppose that for each \( i \) one observes

\(^3\)Several names exist in the literature: independent variables, regressor variables, predictor variables.
7.3 Poisson Models for Counts

$p$ different variables characterizing the population, or mechanisms generating accidents. Consequently, data consist of $n_i$ and a vector $x_{i1}, x_{i2}, \ldots, x_{ip}$, $i = 1, \ldots, k$. In addition in some models an extra quantity $t_i$, say, measuring the exposure for risk is selected and the model for $\mu_i = \mathbb{E}[N_i]$ is written down as follows:

$$\mu_i = t_i \exp(\beta_0 + \beta_1 x_{i1} + \ldots + \beta_p x_{ip}). \quad (7.10)$$

As before, one assumes that $N_i \in \text{Po}(\mu_i)$ are independent and hence the ML estimates of the parameters $\beta_i$ are readily available. The algorithm is given in Section 7.3.4.

**Example 7.15.** The simplest regression model is derived when $p = 0$, i.e. there are no explanatory variables $x_{ij}$ at all. Then with $\lambda = \exp(\beta_0)$ the model is $\mu_i = t_i \lambda$. The ML estimate of the unknown intensity $\lambda$ and standard deviation of the estimation error are given by

$$\lambda^* = \frac{\sum_{i=1}^k n_i}{\sum_{i=1}^k t_i}, \quad \sigma^*_E = \sqrt{\lambda^*/\sum t_i}. \quad (7.11)$$

Obviously if all exposures $t_i$ are equal, $t_i = 1$, then $\mu = \lambda$ giving the estimate $\mu^* = \bar{n}$.

The model in Eq. (7.10) is convenient for studying the influence of a variable $x_{ij}$ on the mean $\mu_i$. The rate ratio defined as

$$RR_j = \exp(\beta_j), \quad j = 1, \ldots, p \quad (7.12)$$

measures multiplicative increase of intensity of events when $x_{ij}$ increases by one unit. The rate ratio is estimated by $RR_j^* = \exp(\beta_j^*)$, where $\beta_j^*$ is the ML estimate of $\beta_j$. Using asymptotic normality of ML estimators, confidence intervals for $RR_j$ can easily be given.

**Example 7.16 (Traffic accidents in Sweden).** This is continuation of Example 7.10 where we presented the number of people killed in traffic in Sweden in years 1990-2004. Constant work on improving safety in traffic, new legislations, technical improvements in cars (ABS, airbags, etc.) as well as better standards of roads should result in a decrease of the death rate. However, the increase in traffic volume has contrary effects.

In the report [7] the following model was proposed, $\mu_i = a \cdot b^i \cdot x_i^c$, where $i = 1, 2, \ldots$ are the years and $a$, $b$, and $c$ unknown parameters. Further $x_i$ is the traffic index in year $i$. Since we do not have access to the traffic index we consider first a simplified model when $c = 0$, $\mu_i = a \cdot b^i$. However, we use the equivalent formulation from Eq. (7.10)

$$\mu_i = \exp(\beta_0 + x_{i1} \beta_1),$$

---

4The functional form in Eq. (7.10) follows the set-up of so-called generalized linear models [56].
where \( x_{i1} = i - 8.0 \). The parameters \( \beta_0, \beta_1 \) are estimated using the ML algorithm giving

\[
\beta_0^* = 6.35, \quad \beta_1^* = -0.0294.
\]

The estimated values \( \mu_i^* = \exp(6.37 - 0.0294 \cdot (i - 8.0)) \) are given in Figure 7.4 (left), solid line. These constitute a regression curve and are compared with observed values \( n_i \) shown as dots.

We can see that data \( n_i \) oscillate quite regularly around the regression curve \( \mu_i^* \), which contradicts the assumed independence of \( N_i \). However, the model can still be a useful, crude description of the data. The most important property of this model is that it indicates that the average number of deaths decreases with \( RR_1^* = \exp(\beta_1^*) = 0.97 \) by 3%. (This was one of the conclusions of the VTI report [7].) \( \square \)

By taking further explanatory variables in Eq. (7.10) more sophisticated models can be proposed. In Example 7.16 we had two parameters \( (p = 1 \text{ while } k = 15) \) and we concluded that a more complex model would be needed to adequately describe the traffic data. However, a higher number of parameters \( \beta_j \) will lead to higher uncertainty of the estimate \( \mu_i^* \). In the limiting case when \( p \geq k - 1 \) there are at least as many parameters to estimate as there are observations \( n_i \). Consequently, the estimates \( \mu_i^* = n_i \) can be used as well instead of \( \mu_i = \exp(\beta_0 + \sum \beta_j^* x_{ij}) \).

Clearly, more complex models better explain the observed variability in data; however, as the number of parameters increases the estimated values

\[\text{5The values of the explanatory variables are centred in order to obtain more well-conditioned covariance matrices, hence } x_{i1} = i - 8.0 \text{ since } (1/15) \sum_{i=1}^{15} i = 8.0.\]
often become more uncertain. When combining both types of uncertainty—
(1) the uncertainty of the future outcome of the experiment; (2) the uncer-
tainty of the parameter— the computed measures of risks can be more un-
certain for the complex model than for the simpler one. This leads us to
the next important issue, the model selection. We do not go deep into this
matter, but just indicate how the different models can be compared using the
already-introduced quantity, deviance (for the simplest case see Eq. (7.8-7.9)).

Model selection and use of deviance

The above-discussed Poisson regression is a very versatile approach to model
variability of counts. Applications are found in most sciences: technology,
medicine, etc. In this subsection we further discuss these models, more pre-
cisely, the number of explanatory variables to be used. Illustrating examples
will be given.

One way of comparing different models is to analyse the value of the log-
arithm of the ML function \( l(.) \) for different choices of explanatory variables.
Let us consider two models: a more general model, with \( p \) explanatory vari-
ables, and a simpler where only \( q < p \) of the variables \( x_i \) are used. (Here
\( q = 0 \) if no explanatory variables \( x \) are used.) Denote by \( \beta_p, \beta_q \), the \( \beta \)
parameters in the two models. Using the ML method optimal estimates \( \beta_p^* \) and
\( \beta_q^* \) are chosen. Since the more general model contains all the parameters of
the simpler one (and some additional) the log-likelihood function \( l(\beta_p^*) \) must
be higher than \( l(\beta_q^*) \). Since higher values of the log-likelihood function means
that the observed data are more likely to occur (if the model is true), the
increase of the function is a measure of how much better the more complex
model explains the data. It can be shown that the following test quantity,
called deviance,

\[
DEV = 2 \cdot (l(\beta_p^*) - l(\beta_q^*)),
\]

for large \( k \) is \( \chi^2(p - q) \) distributed if the less complex model is true. Thus if
\( DEV > \chi^2(p - q) \), the difference between log-likelihoods cannot be explained
by the statistical variability and hence the simpler model should be rejected.
In other words, the more complex model fits data significantly better. Further
discussion of this type of \( \chi^2 \) test can be found in [82], page 345 or [10],
Section 8.2.

Now the computation of the deviance \( DEV \) is relatively simple if the ML
estimates \( \beta_p^*, \beta_q^* \) are given. Using \( \beta_p^*, \beta_q^* \), the estimates of \( \mu_i = E[N_i] \) can
be readily computed

\[
\mu_i^* = t_i \cdot \exp(\beta_0^* + \beta_1^* x_{i1} + \cdots + \beta_l^* x_{il}),
\]

We return to this problem in Chapter 10 where 100-year values will be
estimated.
where \( l = p \) and \( l = q \), respectively. Denote by \( \mu_{iS}^* \) the estimates derived using \( \beta_q^* \) while \( \mu_{iC}^* \) the ones derived using \( \beta_p^* \). Then

\[
\text{DEV} = 2 \sum_{i=1}^{k} n_i (\ln(\mu_{iC}^*) - \ln(\mu_{iS}^*)).
\] (7.14)

**Example 7.17 (Traffic accidents in Sweden).** This is a continuation of Example 7.16 where we concluded that the proposed model for the expected number of perished in traffic in one year is too simple. We believe that the systematical variability (see Figure 7.4, left panel) of \( n_i \) around the estimated regression could be explained by changes in the amount of traffic. It is obvious that the years where the observations are below the average correspond to the years when traffic growth was slower.

In the report [7], estimates of the total vehicle kilometres during 1990-2004 in \( 10^9 \) kilometres, where \( i = 1 \) corresponds to year 1990, were also reported. The estimates \( y_i \), say, are as follows:

\[
\begin{align*}
64.3 & \quad 64.9 & \quad 65.5 & \quad 64.1 & \quad 64.9 & \quad 66.1 & \quad 66.5 & \quad 66.7 & \quad 67.4 & \quad 69.6 \\
70.6 & \quad 71.6 & \quad 74.0 & \quad 75.4 & \quad 76.1 
\end{align*}
\]

Now the new, more complex, model for \( \mu_i \) (with \( p = 2 \)) is

\[
\mu_i = \exp(\beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2}),
\]

where \( x_{i1} = i - 8.0 \) while \( x_{i2} = y_i - 68.5 \). The parameters \( \beta \) are estimated using the ML algorithm giving

\[
\beta_0^* = 6.35, \quad \beta_1^* = -0.082, \quad \beta_2^* = 0.063.
\]

In Figure 7.4 (right panel) we can see the estimated values of \( \mu_i \) as a solid line together with observations marked as dots. The two rate ratios \( RR_1 = \exp(\beta_1) \) and \( RR_2 = \exp(\beta_2) \) are estimated to be \( RR_1^* = 0.92 \) and \( RR_2^* = 1.065 \). The rough interpretation of the ratios is that the safety improvements led to a yearly decrease of about 8\% of the expected number of perished in the traffic but the increase in the traffic volume by \( 10^9 \) km increases the expectation by ca 6.5\%. Since on average the traffic volume increases by \( 0.84 \cdot 10^9 \) km, this leads to a yearly decrease of the expected number of perished by about 3\% (the same as given for the simpler model in Example 7.16). The more complete model seems to give more insight into the problem; however, we should also check whether the more complicated model explains the data significantly better than the simpler one does.

Consequently, let us compute the deviance. Again, let \( \mu_{iS}^* \) denote the estimated averages \( \mu_i^* \) presented in Figure 7.4 (left panel) for the simpler regression, \( p = 1 \), while \( \mu_{iC}^* \) be the corresponding estimates \( \mu_i^* \) presented in Figure 7.4 (right panel) for the more complex regression, \( p = 2 \). Then the deviance given by Eq. (7.14) is equal to
7.3 Poisson Models for Counts

\[
\text{DEV} = 2 \sum_{i=1}^{15} n_i (\ln(\mu_i^* C) - \ln(\mu_i^* S)) = 59.75
\]

which could be compared with the 0.001 quantile found as \(\chi^2_{0.001}(1) = 10.83\). Since \(\text{DEV} > 10.83\), we reject with high confidence the hypothesis that the more complex model explains the data equally well as the simpler one. \(\square\)

Example 7.18 (Derailments in Sweden). In [73], statistics for derailments in Sweden are given. Authorities are interested in the impact of usage of different track types. Data consist of derailments of passenger trains during 1 January 1985 – 1 May 1995, where \(n_i\) is the number of derailments on track type \(i\) and \(t_i\) is the corresponding exposure in \(10^6\) train kilometres. The following numbers are extracted from [73]. The observations \(n_i, t_i\) are given in columns two and three, respectively:

\[
\begin{array}{cccc}
i & n_i & t_i & \text{[Track Type]} \\
1 & 15 & 421 & \text{Welded track with concrete sleepers} \\
2 & 28 & 80 & \text{Welded track with wooden sleepers}
\end{array}
\]

A statistical test is needed to test for possible differences in safety; below, we use the deviance. The numbers of derailments that occur at tracks of type \(i\), denoted by \(N_i\), is assumed to be independent and Poisson distributed. Further, let \(\mu_i = \mathbb{E}[N_i] = \lambda_i t_i\), where \(t_i\) are exposures measured in \(10^6\) train km (tkm). The simpler model is that \(\lambda_1 = \lambda_2 = \lambda\) while the more complex is that \(\lambda_1\) and \(\lambda_2\) are different. We are interested in the rate ratio \(RR = \lambda_2/\lambda_1\).

Eq. (7.11) gives the estimate \(\lambda^* = (n_1 + n_2)/(t_1 + t_2) = 0.0858 \ [10^{-6}\text{tkm}^{-1}]\); consequently, \(\mu_{1S}^* = \lambda^* t_1 = 3.61\) and \(\mu_{2S}^* = \lambda^* t_2 = 6.9\). Next, for the complex model \(\mu_i^* = n_i\) and hence using Eq. (7.13)

\[
\text{DEV} = 2 \left( 15 (\ln(15) - \ln(3.61)) + 28 (\ln(28) - \ln(6.9)) \right) = 52.1.
\]

Since the more complex model has two parameters while the simpler has only one, one should compare the computed deviation with the quantile \(\chi^2_{0.001}(1) = 10.83\). Consequently, with very high confidence, we reject the simplest model. Hence in the following we consider only the more complex model.

The rate ratio \(RR\). The rate ratio measures how the increase of intensity of events changes between the two populations, here \(RR = \lambda_2/\lambda_1\) and is estimated by

\[
RR^* = \frac{\lambda^*_2}{\lambda^*_1} = \frac{28 \cdot 421}{15 \cdot 80} = 9.8,
\]

i.e. the risk for derailment is nearly ten times higher for the second type of track.

The Poisson-regression model. The estimations \(\mu_i\) can also be described as a Poisson-regression problem since we can write

\[
\mu_i = t_i \exp(\beta_0 + \beta_1 x_i).
\]

(7.15)
Here \(x_i\) is a dummy variable taking only two values: defined to be zero when \(i = 1\) and one when \(i = 2\). The parameter estimate \(\beta^*_1\) could be computed using the ML algorithm, however, here we take a shortcut and use that \(RR^*\) has already been estimated. Since \(RR^* = 9.8\), we find \(\beta^*_1 = \ln(9.8) = 2.28\).

Any statistical software would compute the estimates \(\beta^*_i\) and give the matrix with \([-\hat{l}(\beta^*)]^{-1}\) needed for computations of standard deviations of the estimation error, \(\sigma^*_E\). However, since in this simple example these can be easily derived analytically we present the complete solution for illustration of the methodology.

The main purpose of these computations is to derive an asymptotic confidence interval for \(RR\). (Asymptotic normality of ML estimators is utilized.) When the estimate \(\sigma^*_E\) associated with \(\beta^*_1\) is computed then, with approximately 0.95 confidence,

\[
\beta^*_1 - 1.96\sigma^*_E < \beta_1 < \beta^*_1 + 1.96\sigma^*_E
\]

and hence

\[
\exp(\beta^*_1 - 1.96\sigma^*_E) < RR < \exp(\beta^*_1 + 1.96\sigma^*_E).
\]

What remains is computation of the estimated variance \((\sigma^*_E)^2\). The variance is the second element of the diagonal of \(\Sigma = [-\hat{l}(\beta^*_0, \beta^*_1)]^{-1}\). Now, the matrix of second-order derivatives can be computed using Eq. (7.17) when the estimates \(\mu^*_i\) are known. From the definition of \(x_{ij}\), Eq. (7.17) gives

\[
[\hat{l}(\beta^*_0, \beta^*_1)] = - \begin{pmatrix} \mu^*_1 \mu^*_2 \\ \mu^*_2 \mu^*_2 \end{pmatrix} = - \begin{pmatrix} 43 & 28 \\ 28 & 28 \end{pmatrix}.
\]

Consequently

\[
\Sigma = \begin{pmatrix} 0.0667 & -0.0667 \\ -0.0667 & 0.1024 \end{pmatrix},
\]

and hence with approximately 0.95 confidence

\[
\exp(2.28 - 1.96\sqrt{0.1024}) < RR < \exp(2.28 + 1.96\sqrt{0.1024}),
\]

5.2 < \(RR\) < 18.3. Thus, rail type 1 is, with high confidence, at least five times safer to use than rail type 2 is.

\hspace{1cm} \Box

### 7.3.4 ML estimates of \(\beta_0, \ldots, \beta_p\)

For simplicity of derivations, let us introduce \(x_{i0} = 1\) and let

\[
\mathbb{E}[N_i] = \mu_i = t_i \exp \left( \sum_{j=0}^{p} \beta_j x_{ij} \right),
\]

where \(N_i \in \text{Po}(\mu_i), \ i = 1, \ldots, k\). Clearly \(N_i\) may take values 0, 1, 2, \ldots with probabilities
\[ P(N_i = n) = \frac{\mu^n_i}{n!} e^{-\mu_i}. \]

Denote by \( n_i \) the observed \( N_i \), i.e. the number of events that occurred in a period of time \( t_i \). The likelihood-, log likelihood-, and the derivative functions are given by

\[
L(\beta) = \prod_{i=1}^{k} P(N_i = n_i) = \prod_{i=1}^{k} \frac{\mu^n_i}{n_i!} e^{-\mu_i},
\]

\[
l(\beta) = -\sum_{i=1}^{k} \ln(n_i!) + \sum_{i=1}^{k} n_i \ln(\mu_i) - \sum_{i=1}^{k} \mu_i,
\]

\[
\dot{l}(\beta) = \sum_{i=1}^{k} \frac{d\mu_i}{d\beta} \left( \frac{n_i}{\mu_i} - 1 \right).
\]  \hfill (7.16)

Now Eq. (7.16), with \( \beta \) replaced by \( \beta_j \) can be used to compute the derivatives of the log-likelihood functions. Since \( \partial \mu_i / \partial \beta_j = x_{ij} \mu_i \) the derivatives and second-order derivatives of the log-likelihood are given by

\[
\frac{\partial l(\beta)}{\partial \beta_j} = \sum_{i=1}^{k} (n_i - \mu_i) x_{ij}, \quad \frac{\partial^2 l(\beta)}{\partial \beta_j \partial \beta_m} = -\sum_{i=1}^{k} \mu_i x_{ij} x_{im}.
\]  \hfill (7.17)

As before the ML estimate of \( \beta_p^* = (\beta_0^*, \ldots, \beta_p^*) \) are solutions to the system of \( (p + 1) \) non-linear equations in \( \beta \), viz. \( \sum_{i=1}^{k} (n_i - \mu_i) x_{ij} = 0 \). Often these cannot be solved analytically, but a numerical method, e.g. the recursive Newton–Raphson algorithm, can be used:

- The algorithm starts with a guess \( \beta^0 \), say, of the values of the vector \( \beta \), for example

  \[
  \beta^0_i = \ln(\sum n_i) - \ln(\sum t_i), \quad \beta^0_i = 0, \quad i > 0.
  \]

- If the values of the parameters after the \( m \)th iteration are denoted by \( \beta^m \) then the N–R algorithm renders the new estimates by the following formula

  \[
  \beta^{m+1} = \beta^m - [\hat{l}(\beta^m)]^{-1} \hat{l}(\beta^m),
  \]

  where \( [\hat{l}(\beta)] \) is a matrix with derivatives \( \frac{\partial^2 l(\beta)}{\partial \beta_j \partial \beta_m} \) while \( \hat{l}(\beta) \) is a column vector of \( \frac{\partial l(\beta)}{\partial \beta_j} \).

- The algorithm stops when all components in the vector \( \hat{l}(\beta^{m+1}) \) are small enough.
7.4 The Poisson Point process

The Poisson point process is an important tool, widely used not only in applications to risk and safety analysis, but also in telecommunication engineering, financial, and insurance mathematics. Applications to risk analysis and accidents were present already in the 1920s, cf. [30]. In Section 2.6.1, we introduced a Poisson stream of events, which is here renamed Poisson point process (PPP) in order to generalize the notion from a line (time) to higher-dimensional spaces.

We start with an alternative definition of a PPP on the line, i.e. in the case when the PPP is a Poisson stream of events $A$, say, and review some basic properties of a PPP. Of particular interest is the distribution of the time intervals $T_i$ between the occurrences of $A$.

**Definition 7.2 (Poisson Point process (PPP)).** If the time intervals $T_1, T_2, \ldots$ between occurrences of an event are independent, exponentially distributed variables with common failure intensity $\lambda$, then the times $0 < S_1 < S_2 < \ldots$ when the event $A$ occurs form a Poisson point process with intensity $\lambda$.

Let us recall the notation $N_A(s, t), N_A(t)$ from Definition 2.2. (In the following the subscript $A$ is omitted.) For fixed values $s, t$ the random variable $N(s, t)$ is the number of times an event $A$ occurred in the time interval $[s, s+t]$ while $N(t)$ is understood as $N(0, t)$. The variable $N(t)$ can also be seen as a function of time which (see Figure 7.5), is called a Poisson process.

![Fig. 7.5. Illustration of a Poisson process.](image-url)
We summarize the important properties of a PPP:

- Let $\lambda$ be the intensity of a PPP. Then
  - The time to the first event, $T$, is exponentially distributed:
    $$ P(T > t) = e^{-\lambda t}. $$
  - Times between events, $T_i$, are independent and exponentially distributed:
    $$ P(T_i > t) = e^{-\lambda t}. $$
  - The number of events $N(s, t) \in \text{Po}(m)$, i.e. is Poisson distributed with $m = \lambda t$.
  - The number of events in disjoint time intervals are independent and (obviously) Poisson distributed.

**Remark 7.3.** If we assume that real-world phenomenon can be modelled by means of a Poisson point process then the intensity $\lambda$ is the only parameter that is needed to compute probabilities of interest, since $N(t) \in \text{Po}(\lambda t)$. If the mean $E[N(t)] = \lambda t$ is small, then

$$ P(N(t) = 0) = e^{-\lambda t} \approx 1 - \lambda t, \quad P(N(t) = 1) \approx \lambda t = E[N(t)], $$

and the probability of more than one accident is of smaller order. □

Typical applications of a PPP often are to model variability of counting and book-keeping of times, for example between cars passing a checkpoint. The Poisson model implies that in any time period $t$, say, the number of cars that have been registered in the period $N(t)$, say, is Poisson distributed with mean equal to $\lambda t$.

In safety analysis of complex systems, e.g. an electrical power network in a country, transients that occur in the system need to be analysed as consequences of different types of failures (accidents). The failures are modelled using Poisson streams and the safety of the system is investigated by means of suitable (numerical) simulations of transients. The possibility of analytical computations is limited by the complexity of a system. One of the inputs is times of failures and hence Poisson streams with a given intensity $\lambda$ need to be simulated.

**Simulation of a Poisson point process**

Since the intensity of events $\lambda$ is constant, we expect that there are no specific patterns regarding the positions of the points in a PPP. This somewhat unprecise statement can be illuminated by the following method to simulate a PPP.
Step 1 First choose an interval of length $t$, for example $[0,t]$.

Step 2 Then, by some Monte Carlo method, generate the number of points in the interval $N(t)$, i.e. random numbers with distribution $\text{Po}(\lambda t)$ (see Chapter 3 for details). Denote the generated number by $n$ (for instance, if $n = 10$, then there are 10 points in $[0,t]$).

Step 3 What remains to find are the exact locations of the $n$ points. These should be totally random. In fact, the locations are independent and uniformly distributed variables. By this we mean that we need to simulate $n$ values $u_i$ of uniformly distributed (between zero and one) random numbers. Then the positions of the events are given by $t \cdot u_i$ (not ordered).

It is important to be able to motivate the correctness of the assumption that the sequence of events forms a Poisson stream. According to Section 2.6.1, conditions I-III, one needs to motivate that the mechanism generating accidents is stationary. One often limits oneself to check if the intensity of accidents is constant (see Examples 7.13-7.14). Next, one needs to argue that the number of accidents in disjoint intervals is independent and, finally, that two or more events cannot happen exactly at the same moment. Here the reason for the use of a PPP consists mainly of general arguments. This type of “validation” is often used when events occur rarely and hence use of statistical tests is limited.

Remark 7.4 (Barlow–Proschan test). Actually the property used in Step 3 in the simulation algorithm, that times when accidents occur are uniformly distributed, can be used to construct a test whether the ordered observed times $0 < S_1 < S_2 < \ldots < S_n$ do not contradict the assumption that those are the first $n$ times of the PPP. It can be shown that the statistic

$$Z = \frac{1}{S_n} \sum_{i=1}^{n-1} S_i$$

is approximately normally distributed. From Step 3, it can be seen that $Z$ has the distribution of the sum of $n-1$ uniformly distributed random variables $U_i$. Consequently, a table of means and variances gives that $E[Z] = (n-1)/2$ and $V[Z] = (n-1)/12$ and hence, with approximately probability $1 - \alpha$

$$\frac{1}{2}(n-1) - \lambda \alpha/2 \sqrt{\frac{n-1}{12}} < Z < \frac{1}{2}(n-1) + \lambda \alpha/2 \sqrt{\frac{n-1}{12}}. \quad (7.18)$$

Now having observed the times $s_i$, $i = 1, \ldots, n$, the value of $z = \sum_{i=1}^{n-1} s_i/s_n$ is computed. If $z$ is outside the interval given in Eq. (7.18) then the hypothesis that the times $s_i$ are outcomes of Poisson point process is rejected. This procedure is called Barlow–Proschan’s test.

Example 7.19 (Periods between earthquakes). Let us reconsider times between earthquakes $t_i$, first encountered in Example 1.1, later discussed e.g.
in Example 4.6, where a \(\chi^2\) test was used to test for exponentially distributed time intervals. Here we make use of Barlow–Proschan’s test outlined earlier.

Obviously \(s_k = \sum_{i=1}^{k} t_i, \ k = 1, \ldots, n\) and hence

\[
z = \sum_{k=1}^{n-1} \frac{\sum_{i=1}^{k} t_i}{\sum_{i=1}^{n} t_i}.
\]

(7.19)

For the data, \(n = 62\) and we find \(z = 31.06\). The interval \([30.5 - 1.96\sqrt{61/12}, 30.5 + 1.96\sqrt{61/12}] = [26.1, 34.9]\) contains \(z\) and hence the hypothesis that times for major earthquakes forms a PPP cannot be rejected. □

### 7.5 More General Poisson Processes

Earlier in this chapter, we have used the Poisson point process to describe when events occur in time, i.e. a Poisson stream. However, applications do not have to be restricted to events occurring in time. Consider for example cracks along an oil pipeline and think about how a PPP can be applied. The concept can be generalized even more.

**A general Poisson process**

Let \(N(B)\) denote the number of events (or accidents) occurring in a region \(B\). Consider the following list of assumptions (cf. Section 7.4):

(A) More than one event cannot happen simultaneously.
(B) \(N(B_1)\) is independent of \(N(B_2)\) if \(B_1\) and \(B_2\) are disjoint.
(C) Events happen in a stationary (in time) and homogeneous (in space) way, more precisely, the distribution of \(N(B)\) depends only on the size \(|B|\) of the region: for example \(N(B) \sim \text{Po}(\lambda |B|)\).

The process for which we can motivate that (A–B) are true is called a Poisson process. It is a stationary process with constant intensity \(\lambda\) if (A–C) holds.

An illustration of a Poisson process in the plane is given in Figure 7.6.

**Example 7.20 (Japanese black pines).** In Figure 7.7 are shown the locations of Japanese black pine samplings in a square sampling region in a natural region. The observations were originally collected by Numata [58] and the data are used as a standard example in the textbook by Diggle [20].

Having adequate biological information about the species region and other relevant information one could may be assume the validity of assumptions (A-C) leading to the Poisson model for the locations of the trees.

As statisticians we can also validate the model, i.e. check if some statistics do not contradict the assumed PPP. First, let us estimate the intensity \(\lambda\) of pines.
Fig. 7.6. Illustration of a Poisson process in the plane. Here $N(B) = 11$ while
$N(B_1) = 2$, $N(B_2) = 3$.

Fig. 7.7. Locations of Japanese black pines in a square sampling region.

The region studied was $5.7 \times 5.7$ m$^2$, which we refer to as one area unit
(au) in the following. There are 65 pines in a region of 1 au, and hence the
estimate of the intensity $\lambda^* = 65$ au$^{-1}$. We divide the region
in 25 smaller squares, each of size $0.2 \times 0.2 = 0.04$ au. Since we assumed
homogeneity of trees, we expect on average $0.04 \cdot 65 = 2.6$ trees in each of
such smaller regions. Obviously the true number differs from the average
and their variability is modelled as 25 independent Po$(2.6)$ distributed variables.

From Figure 7.7 are found 1, 5, 4, 11, 2, 1, 1 regions containing 0, 1, 2, 3, 4, 5, 6 pines. The probability-mass function for Po$(2.6)$ is $p_k = 2.6^k \exp(-2.6)/k!$ and hence one expects to have $25 \cdot p_i$ smaller regions to contain $k$ plants. The expectations are 1.9, 4.8, 6.3, 5.4, 3.5, 1.8, 0.8, respectively; how close is this to what the model predicts? We use a $\chi^2$ test:

$$Q = \frac{(1 - 25p_0)^2}{25p_0} + \frac{(5 - 25p_1)^2}{25p_1} + \frac{(4 - 25p_2)^2}{25p_2} + \frac{(11 - 25p_3)^2}{26p_3} + \frac{(2 - 25p_4)^2}{25p_4} + \frac{(1 - 25p_5)^2}{25p_5} + \frac{(1 - 25p_6)^2}{25p_6} = 8.1.$$
Since $\chi^2_{0.05}(7 - 1 - 1) = 11.07$, the hypothesis about Poisson distribution cannot be rejected (see Eq. (4.3)).

Example 7.21 (Bombing raids on London). During the bombing raids on London in World War II, one discussed whether the impacts tended to cluster or if the spatial distribution could be considered random. This was not merely a question of academic interest; one was interested in whether the bombs really targeted (as claimed by Germans) or fell at random. An area in the south of London was divided into 576 small areas of $1/4$ km$^2$ each; the Poisson distribution was found to be a good model. For further discussion, consult Chapter VI.7 in the classical book by Feller [25].

7.6 Decomposition and Superposition of Poisson Processes

The Poisson process is a mathematical tool in risk analysis to describe the occurrence of events of particular interest in some application. We now go one step further: to a given event, additional properties can be related.

Example 7.22. In Example 1.11, the event $A =$“Fire starts” was considered. The stream of $A$ is often modelled as a PPP. Now the fire was furthermore classified at the arrival by means of two scenarios: $B =$“Fire with flames” or (if $B$ was false) “Smoke without flames”. The date of the fire is written down and is marked with a star in case scenario $B$ followed fire ignition, i.e. fire with flames was recorded, was true. Otherwise, when “merely smoke” was recorded, a dot is marked.

As it was shown in Eq. (2.14), if the scenario $B$ were independent of stream of ignition then the point process of “stars” (dates of fires with flames) is a PPP too. Consider one type of fire, e.g. the one marked with stars. In this section we discuss generalizations of the presented splitting of a PPP into point processes of stars and dots.

Consider an event $A$ that is true at point $S_i$ and suppose that $S_i$ form a PPP with intensity $\lambda$. Consider for instance a Poisson process in the plane, as used

---

7This problem has even influenced literary texts, as the following excerpt from Pynchon’s *Gravity’s Rainbow*, [64], Part 1, Chapter 9:

Roger has tried to explain to her the V-bomb statistics; the difference between distribution, in angel’s-eye view, over the map of England, and their own chances, as seen from down there. She’s almost got it, nearly understands his Poisson equation...“...Couldn’t there be an equation for us too...”...“...There is no way, love, not as long as the mean density of the strikes is constant ...”

8The same is valid for the point process of dots, since if $B$ is independent of the stream $A$ then the complement $B^c$ is independent too.
in Example 7.20. Let $B$ be a scenario (a statement that can be true or false when $A$ occurs, i.e. at points $S_i$). Now at each point $S_i$ (when $A$ occurs) we put a mark “star” if $B$ is true. All remaining $S_i$ (when $B$ is false) are marked by dots (see Figure 7.8). If $B$ is independent of the PPP $A$, then the point processes of stars and dots are independent Poisson and have intensities $P(B)\lambda$, $(1 - P(B))\lambda$, respectively.

It is not surprising that the reverse operation of superposition of two (or more) independent Poisson processes gives a Poisson process.

**Theorem 7.1. Superposition Theorem:** Assume that we have two independent Poisson point processes $S^I_i$ and $S^{II}_i$ with intensities $\lambda^I$, $\lambda^{II}$, respectively. Consider a point process $S_i$, which is a union of the point processes $S^I_i$ and $S^{II}_i$. (If $S^I_i$, $S^{II}_i$ are marked by stars and dots, respectively, replace all symbols with a ring (◦) and let $S_i$ be positions of rings.) The point process of $S_i$ is a superposition of the two processes and is a PPP itself, with intensity $\lambda = \lambda^I + \lambda^{II}$.

For further reading about decomposition and superposition, including proofs, see the books by Gut [33] or Çinlar [11].

**Problems**

7.1. Assume that the lifetime process for humans has the death-rate function

$$\lambda(t) = a + b \cdot e^{t/c}, \quad t > 0,$$

where $a = 3 \cdot 10^{-3}$, $b = 6 \cdot 10^{-5}$, and $c = 10$. The unit of time is 1 year.

(a) Calculate the probability that a person will reach the age of at least fifty.
(b) A person is alive on the day he is thirty. Calculate the conditional probability that he will live to be fifty.

7.2. Consider the experiment presented in Example 7.6. Use the Nelson–Aalen estimator to estimate the cumulative failure-intensity function of the observed lifetimes for concrete beams in air.

7.3. At time $t = 0$, a satellite is put into orbit. Two transmitters have been installed. At $t = 0$, both of them are working, but they break down independently with constant failure rate $\lambda$ each. When both transmitters have failed to work, the satellite is out of order. Find the failure rate for the whole transmitter system.
7.4. The random variable $Z$ is Poisson distributed and has a coefficient of variation of 0.50. Calculate $P(Z = 0)$.

7.5. The number of cars passing a street corner is modelled by a Poisson process with intensity $\lambda = 20 \text{ h}^{-1}$. Calculate (approximately) the probability that more than 50 cars will pass during two hours (2 h).

7.6. Consider an oil pipeline. Suppose the number of imperfections $N(x)$ along a distance $x$ can be modelled by a Poisson process, that is, $N(x) \in \text{Po}(\lambda x)$, where $\lambda$ is the intensity (km$^{-1}$). Let $\lambda = 1.7 \text{ km}^{-1}$.

(a) Calculate the probability that there are more than 2 imperfections along a distance of 1 km.
(b) Calculate the probability that two consecutive imperfections are separated by a distance longer than 1200 m.

7.7. Consider again the data set with time intervals between failures given in Problem 6.4.

(a) Test if data do not contradict the assumption of a PPP.
(b) Modelling the occurrences of failures of the air-conditioning system as a PPP, use the observations to estimate the intensity $\lambda$ for plane 7914.

7.8. The number of defects (“specks”) in plates is described by a Poisson distribution. One has the following observations: 30 plates are of colour 1 and 45 plates of colour 2.

| Colour 1: | 1 3 1 0 0 0 2 1 1 0 2 0 0 2 0 1 0 2 0 0 1 1 0 0 1 0 0 |
|------:|:-----:|:-----:|:-----:|:-----:|:-----:|:-----:|:-----:|:-----:|:-----:|:-----:|:-----:|:-----:|:-----:|:-----:|:-----:|:-----:|:-----:|:-----:|:-----:|:-----:|:-----:|:-----:|:-----:|:-----:|:-----:|:-----:|
|       |       |       |       |       |       |       |       |       |       |       |       |       |       |       |       |       |       |       |       |       |       |       |       |       |       |
| Colour 2: | 0 0 0 0 0 2 0 0 1 1 1 0 0 0 0 0 1 0 0 1 0 0 0 0 0 0 2 1 0 1 1 0 0 0 |
|       |       |       |       |       |       |       |       |       |       |       |       |       |       |       |       |       |       |       |       |       |       |       |       |       |       |       |       |

(Observations from $\text{Po}(m_1)$)

(Observations from $\text{Po}(m_2)$)

Give an estimate of $m_1 - m_2$ and estimate the standard deviation of the proposed estimator of $m_1 - m_2$. Compute a 0.95-confidence interval for $m_1 - m_2$.

7.9. A group of parachutists is launched randomly over a region. Suppose the mean intensity of parachutists is $\lambda$ per unit area and assume a Poisson model; that is, the number of people in a region of area $A$ is Poisson distributed with mean $\lambda A$.

For a randomly selected person in this region, let $R$ denote the distance to the nearest neighbour.

(a) Find the distribution for $R$. *Hint:* Note that $P(R > r)$ is the same as the probability of seeing no people within a circle of radius $r$.
(b) Give the expected value, $E[R]$ (cf. Problem 3.7).
(c) Suppose that a group of 20 people are launched over a region of size 1 km$^2$.

An estimate of $\lambda$ is then $2 \cdot 10^{-5} \text{ m}^{-2}$. Use the previous results to compute the average distance between the parachutists.
7.10. Consider flying-bomb hits on London, discussed in Example 7.21. The total number of small areas was 576 and the total number of hits was 537. In [12], the following numbers are found (reprinted in [25]):

\[
\begin{array}{c|cccccc}
    k & 0 & 1 & 2 & 3 & 4 & \geq 5 \\
n_k & 229 & 211 & 93 & 35 & 7 & 1 \\
\end{array}
\]

where \( n_k \) is the number of areas with exactly \( k \) hits.

Test for a Poisson distribution using a \( \chi^2 \) test.

7.11. Consider the data set of hurricanes, given in Problem 4.12.

(a) Based on the given 55 yearly observations, estimate the intensity of hurricanes. Compute the probability of more than 10 hurricanes in a given year using normal approximation. Use this probability to compute the expected number of years with more than 10 hurricanes during a 55-year period.

(b) The question of a possible increase over time of the average number of hurricanes has been much discussed in media as well as in the specialized research literature on climatology. We here investigate this complex issue by a simple Poisson-regression model:

\[
E[N_i] = \exp(\beta_0 + \beta_1x_i), \quad i = 1, \ldots, 55
\]

where the explanatory variable \( x \) is time in years \( x = 0, \ldots, 54 \). A constant intensity over time means \( \beta_1 = 0 \). We want to test for a possible trend, i.e. the null hypothesis is \( \beta_1 = 0 \).

A software package returns the values of log-likelihood functions \( l(\beta_0^*, \beta_1^*) = -123.8366 \) (with \( \beta_0^* = 1.8000, \beta_1^* = 1.4 \cdot 10^{-4} \)) and \( l(\beta_0^*, 0) = -123.8374 \) (with \( \beta_0^* = 1.8038 \)). Calculate the deviance and draw conclusions.

7.12. In Figure 7.9 are shown the locations of 71 pines in a square sampling region in Sweden. Use the division into 25 small squares given in the figure and perform

![Fig. 7.9. Locations of pines in a square sampling region at a location in Sweden.](image)
calculations as in Example 7.20 to investigate whether the pines are distributed in the plane according to a Poisson process.

Hint. Some observations fall at the border between squares. In that situation, let the observation belong to the higher limit.

7.13. Consider lorries travelling over a bridge. Assume that the times \( S_i \) of arrivals form a Poisson process with intensity 2 000 per day. Consider a scenario \( B = \{A \text{ lorry transports hazardous material}\} \). Assume that the scenario is independent of the stream of lorries. (This would not be the case if a chemical company is usually sending a convoy of lorries with hazardous material to the same destination). From statistics, one has found that with probability \( p = 0.08 \), a lorry transports hazardous materials and with probability \( q = 0.92 \), other material is transported.

(a) In one week (Monday–Friday), on average 10 000 lorries will travel over the bridge. What is the probability that during a week a number of 300 more than the average will pass?

(b) What is the probability that during a week (Monday–Friday) there are more than 820 transports of hazardous materials?