

Lecture 5. Time to failure - Failure intensity

Measures of Risk-Testing for Poisson cdf ¹

Igor Rychlik

Chalmers

Department of Mathematical Sciences

Probability, Statistics and Risk, MVE300 • Chalmers • April 2013. Click on **red text** for extra material.

¹Section 7.1.2 is not included in the course.

Survival function - Failure rate:

For a positive rv. T , for example life-time of a component, time to failure, accident etc.:

- ▶ $R(t)$ - **survival function** is defined by

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

- ▶ $\Lambda(t) = -\ln(R(t))$ - is called **cumulative failure-intensity function**

$$R(t) = e^{\ln R(t)} = e^{-\Lambda(t)}.$$

- ▶ $\lambda(t) = \dot{\Lambda}(t)$ - **failure intensity** function

$$R(t) = e^{\ln R(t)} = e^{-\int_0^t \lambda(s) ds}.$$

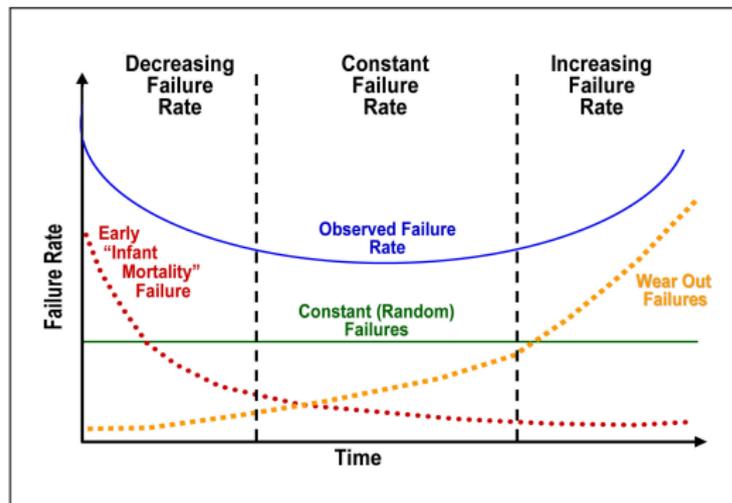
Problem 7.3

What is failure intensity measuring:

It can be demonstrated that

$$\lambda(s) = \lim_{t \rightarrow 0} \frac{P(T \leq s + t | T > s)}{t},$$

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 t time units.**



The lifetimes T is often classify as; IFR (increasing failure rate); DFR (decreasing failure rate);

bathtub.

Example 1

Constant failure rate

- ▶ For exponential $T \in \exp(a)$, $a = E[T]$

$$R(t) = e^{-t/a} \text{ hence } \Lambda(t) = t/a \text{ and } \lambda(t) = 1/a,$$

failure intensity is constant.²

- ▶ If it passed time s without failure, i.e. " $T > s$ " is true, then probability of no failure in the next t time units is

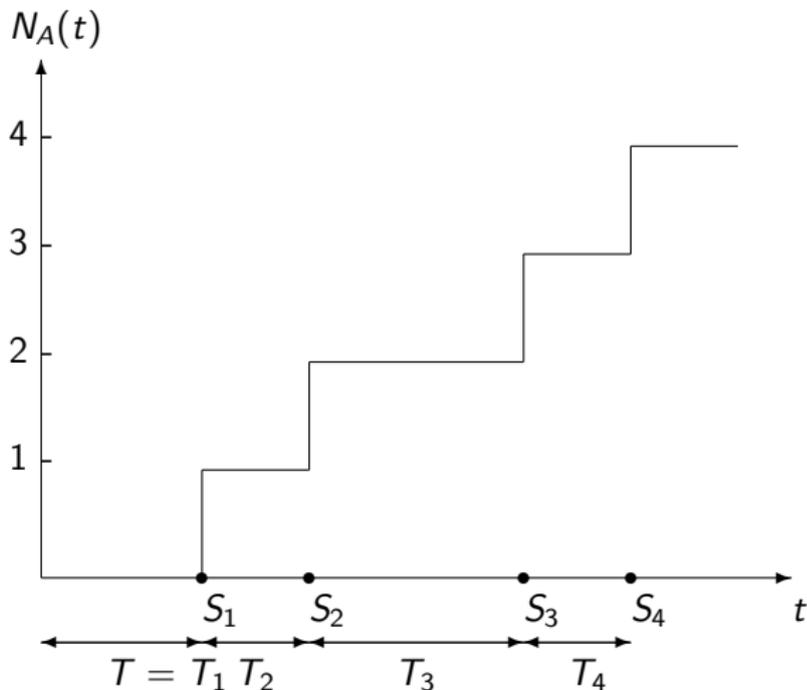
$$P(T > s + t | T > s) = \frac{P(T > s + t)}{P(T > s)} = \frac{e^{-(s+t)/a}}{e^{-s/a}} = P(T > t).$$

This is sometimes stated as "**memorylessness**" of exponential cdf.

- ▶ Consider components having exponential life-times. For example electrical fuse (Elsäkring) breaks a circuit when A ="overcurrent" occurred at time S_1 then it is immediately replaced by a new fuse. Again it breaks after exponentially distributed time (occurrence of the second overcurrent) at time S_2 etc. The sequence S_1, S_2, \dots is called a **point process**.

²Often failures are due to accidents occurring at random.

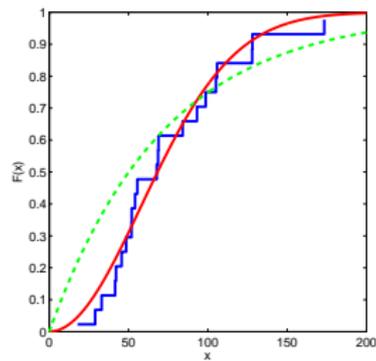
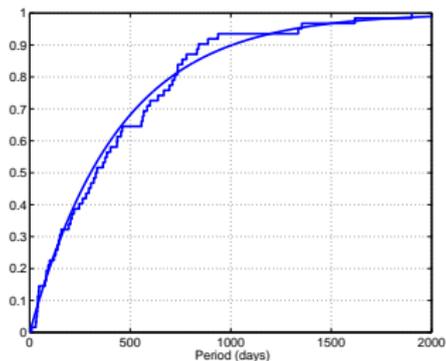
Constant failure rate - Poisson point process



S_i - times for accidents A (overcurrents), T_i - lifetimes of components,
 $N_A(t)$ - number of accidents in time interval $[0, t]$.

$N_A(t) \in Po(m)$, where $m = \lambda \cdot t$, thus point process S_i is named **Poisson**.

Examples:



Left figure - fitted exponential cdf to the earthquake data and ecdf.

Right figure - the fitted exponential cdf, with $a = 72.4$ (green dashed line) and Rayleigh cdf, with $a = 82$, (red line) compared with the ball-bearing life cdf.

Failure rate - examples:

- ▶ T - Rayleigh distributed life time of ball bearings then

$$R(t) = e^{-(t/a)^2} \text{ hence } \Lambda(t) = (t/a)^2 \text{ and } \lambda(t) = 2t/a^2.$$

This is **IFR** case, as expected. Failures are due to wear.

Example 2

Suppose that a structure contains four ball bearings of the type studied. The structure is working as long as all bearings are OK. *Compute the failure intensity of the system.*

Some useful formulas:

- ▶ For continuous positive rv. T

$$\lambda(t) = \frac{f_T(t)}{1 - F_T(t)}.$$

- ▶ The conditional probability that the s year old component will survive addition t years

$$P(T > s + t | T > s) = e^{-\int_s^{s+t} \lambda(x) dx}.$$

If T has constant failure rate than s time units old component is as good as a new one.

- ▶ Expected lifetime

$$E[T] = \int_0^{\infty} P(T > s) ds$$

Example 3

Consider a ball bearing that has been used for 60 millions revolutions. Compute its expected remaining life time. What is the probability that it will survive additional 60 millions of revolutions?

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.

The event $T > t$ is equivalent to the statement that all lifetimes T_i exceed t , i.e. $T_1 > t, T_2 > t, \dots, T_n > t$. If T_i are independent then

$$\begin{aligned} P(T > t) &= P(T_1 > t) \cdot \dots \cdot P(T_n > t) = e^{-\int_0^t \lambda_1(s) ds} \cdot \dots \cdot e^{-\int_0^t \lambda_n(s) ds} \\ &= e^{-\int_0^t \lambda_1(s) ds - \dots - \int_0^t \lambda_n(s) ds} = e^{-\int_0^t \lambda_1(s) + \dots + \lambda_n(s) ds} \end{aligned}$$

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

Absolute Risks

Failure intensity $\lambda(s)$ describes variability of life lengths 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.

Sometimes one has only access to the total number of failures; for example, number 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.

Often a distinction is made between so-called “**voluntary risks**” and the “**background risks**”. Clearly accidents due to an activity like mountaineering is 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 (about $10^6 : 1$ in Great Britain).

Tolerable risks

The magnitudes of the risks specified in the following table are meant approximately: 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.)

Risk of death per person per year	Characteristic response
10^{-3}	Uncommon accidents; immediate action is taken to reduce the hazard
10^{-4}	People spend money, especially public money to control the hazard (e.g. traffic signs, police, laws)
10^{-5}	Parents warn their children of the hazard (e.g. fire, (drowning, fire arms, poison)
10^{-6}	Not of great concern to average person; aware of hazard, but not of personal nature; act of God.

Example - Number of perished in traffic

Year 1998 it was reported about 41 500 perished in traffic accidents in the United States while in Sweden the number was about 500.

In order to compare these numbers, one needs to compensate for the size of populations in both countries. A fraction of the numbers of perished by the size of population, giving the "absolute risks" (frequencies of death).

In US the frequency was about 1 in 6 000, circa $1.7 \cdot 10^{-4}$, while in Sweden, 1 in 17 000, circa $0.6 \cdot 10^{-4}$, which is nearly three times lower.³

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 than a person in Sweden does.

We found that 1998 the risk in US was about 1 person per $100 \cdot 10^6$ km driven while in Sweden, 1 per $125 \cdot 10^6$ km. (Is this significant difference?)

³Comparisons of chances to die in traffic accidents between countries can be difficult since statistics may use different definitions and have different accuracy.

Comparative death risks,(average 1970-1973 in U.K.)

In the following list we will compare "activity/cause" with absolute risk for death measured per hour of exposure:

Mountaineering (international)	$2700 \cdot 10^{-8}$
Air travel (international)	$120 \cdot 10^{-8}$
Car travel	$56 \cdot 10^{-8}$
Accidents at home (all)	$2.1 \cdot 10^{-8}$
Accidents at home (able-bodied persons)	$0.7 \cdot 10^{-8}$
Fire at home	$0.1 \cdot 10^{-8}$

Let pretend that the risk for death in traffic in Sweden is of the same order as in U.K. and that the average person spend 15 minutes in a car per day and that there are 10^7 Swedes then the estimated average number of death in traffic would be

$$0.25 \cdot 365 \cdot 10^7 \cdot 50 \cdot 10^{-8} = 456.$$

Predicting N

Problem: Let N be the number of perished due to an activity, in a specified population (a country), and period of time (often one year). For example: N - number of perished in traffic next year in Sweden.

Uncertainty of N value can be model by means of probability distribution. Choice of model is often based on reasoning, experience (i.e. historical data), convenience. Prediction assumes, that N in close future, will vary in a similar way as in the past.

Model: If N is the number of accidents which occur independently with small probability then N may have a Poisson distribution, $N \in \text{Po}(m)$, where $m = E[N]$.⁴

Data: One needs data to estimate N 'th cdf or test the model. For example one may have observations of N_i during a number of years or have more detailed data, e.g. $N = N_1 + N_2 + \dots + N_k$ where N_i are the number perished in i th region. Are N_i independent Poisson rv. with the same mean m ?

⁴This is a consequence of the approximation of the binomial distribution by the Poisson distribution (the law of small numbers).

Testing the Poisson assumption for N

We consider two cases:

- ▶ $E[N] < 15$ then one could use the χ^2 test to check assumption that the data follows Poisson cdf. For example Horse-kick data considered in Chapter 4. Typically the data set has to be large.
- ▶ $E[N] > 15$ then one can use the following approximation. Here less data is needed to motivate the significance level of the test.

Normal approximation of Poisson distribution.

Let N be a Poisson distributed random variable with expectation m ,

$$N \in \text{Po}(m).$$

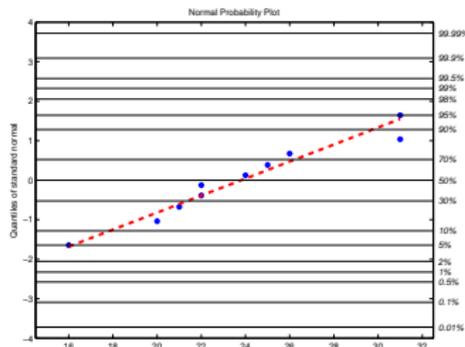
If m is large (in practice, $m > 15$), we have approximately that

$$N \in \text{N}(m, m).$$

Example:

From “Statistical Abstract of the United States”, data for the number of crashes in the world during the years 1976-1985 are found:

24 25 31 31 22 21 26 20 16 22



The estimated mean is 23.8 while
variance 22.2

$$P(N > 35) \approx 1 - \Phi((35 - 23.8)/\sqrt{23.8}) = 1 - \Phi(2.296) = 0.011.$$

$$\text{or } P(N > 35) \approx 1 - \Phi((35 + 0.5 - 23.8)/\sqrt{23.8}) = 0.008.$$

Test for overdispersion⁵

In the case when m is large, to test whether data do not contradict the assumption, often the following property of a Poisson distribution is used: $V[N] = E[N] = m$. In the case of a Poisson distribution, the ratio $V[N]/E[N]$ is obviously equal to 1. Let estimate $E[N]$ by \bar{n} and $V[N]$ by

$$s_{k-1}^2 = \frac{1}{k-1} \sum_{i=1}^k (n_i - \bar{n})^2.$$

Then an approximate confidence interval for $\theta = V[N]/E[N]$ can be constructed, viz.

$$\frac{\bar{n}}{s_{k-1}^2} \frac{\chi_{1-\alpha/2}^2(k-1)}{k-1} \leq \frac{V[N]}{E[N]} \leq \frac{\bar{n}}{s_{k-1}^2} \frac{\chi_{\alpha/2}^2(k-1)}{k-1}$$

with approximate confidence $1 - \alpha$. If $\theta = 1$ is not in that interval, the hypothesis about Poisson distribution is rejected.

⁵Overdispersion is the presence of greater variability in a data set than is expected. It is a very common feature in applied data analysis because in practice, populations are frequently heterogeneous, e.g. mean is not constant.

Example - Flight safety

Continuation of Example 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 forms a Poisson point process (exponential times between accidents) and hence n_i are independent observations of $P_0(m)$ distributed variables.

As we shown before $\bar{n} = 23.8$ while $s_{k-1}^2 = 22.2$. The approximate confidence interval for $V[N]/E[N]$ is

$$\frac{\bar{n}}{s_{k-1}^2} \frac{\chi_{1-\alpha/2}^2(k-1)}{k-1} \leq \frac{V[N]}{E[N]} \leq \frac{\bar{n}}{s_{k-1}^2} \frac{\chi_{\alpha/2}^2(k-1)}{k-1}$$

giving for the data

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

Since 1 is in the interval the hypothesis is not rejected.

Counting number of events N :

Data: Suppose we have observed values of N_1, \dots, N_k , which are equal to n_1, n_2, \dots, n_k , say. The first assumption is that N_i are independent Poisson with constant mean m (the same as N has). Suppose that the test for over-dispersion leads to rejection of the hypothesis that N_i are iid Poisson. **Over-dispersion can be caused by variable mean of N_i or that Poisson model is wrong.** What can we do?

The first step is to assume that N_i are Poisson but have different expectations m_i .

Little help for predicting future unless one can **model variability of m_i !**

We propose solution in the next lecture.

In this lecture we met following concepts:

- ▶ Failure intensity, IFR, DFR.
- ▶ Various risk measures.
- ▶ Poisson model for number of accidents.
- ▶ Over-dispersion test to control Poisson assumption.

Examples in this lecture "click"