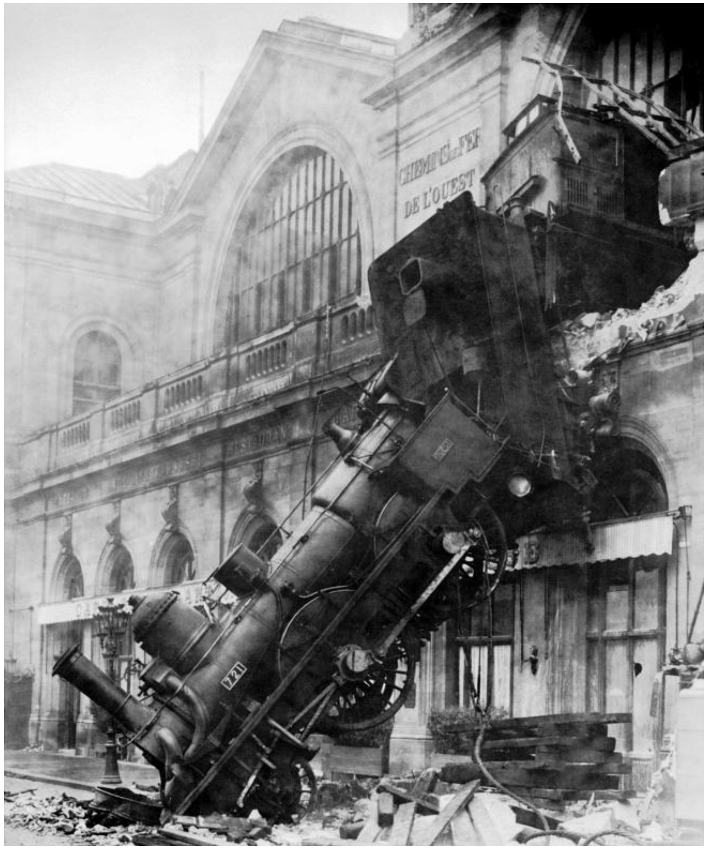
Unit 4



Model Selection

Luis Carlos Pardo

Escola d’Enginyeria de Barcelona Est

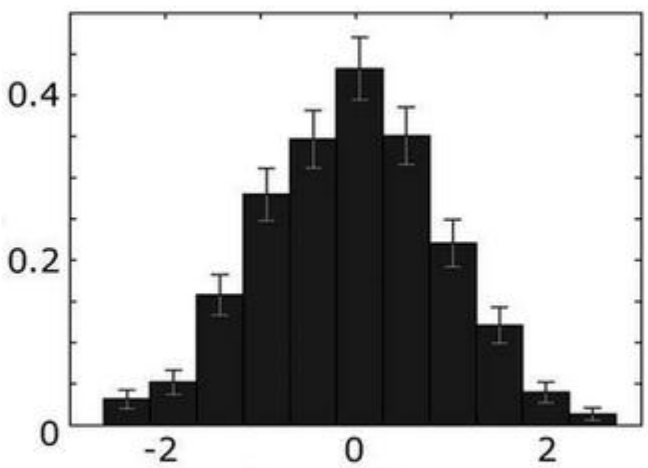
Summary

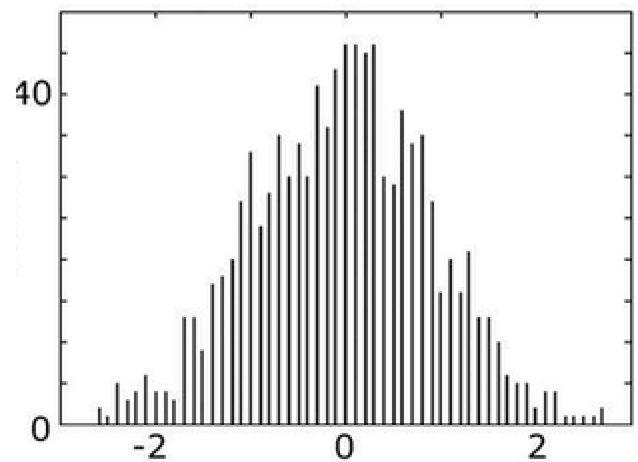
1.- Bayes theorem
2.- Maximum likelihood method
3.- Estimation of reliability parameters from tests
4.- Confidence limits of parameters
5.- Accelerated life testing
6.- Determination of distribution models
7.- Empirical determination of survival function
8.- Reliability growth
9.- Strength-stress models

So, we have up to now ASSUMED the distribution f(t), but...
<u>what if the distribution is not known???</u>

Let’s assume that there are n units each failing at a time t_i

First we have to distribute them inside boxes: binning
we must therefore decide the size of the binning!





- big bin size
- small error
- few points

- small bin size
- big error
- a lot of points

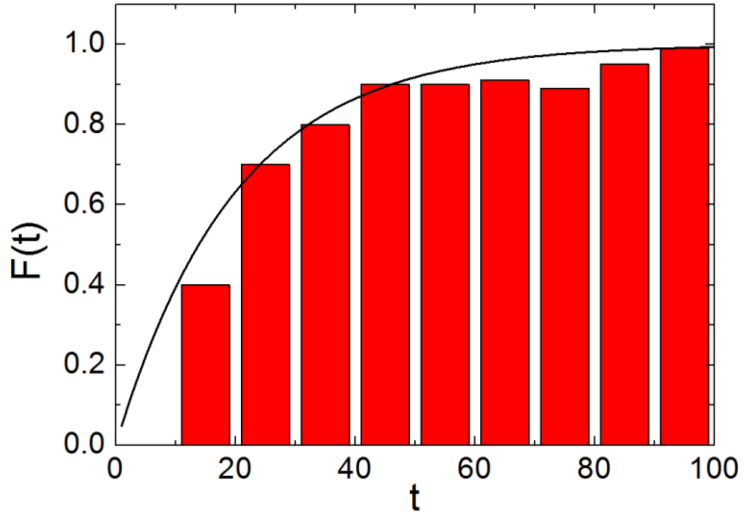
So, we have up to now ASSUMED the distribution f(t), but...
<u>what if the distribution is not known???</u>

Let's assume that there are n units each failing at a time t_i

Now we must see if the obtained curve is a gaussian, or exponential, or weibull...

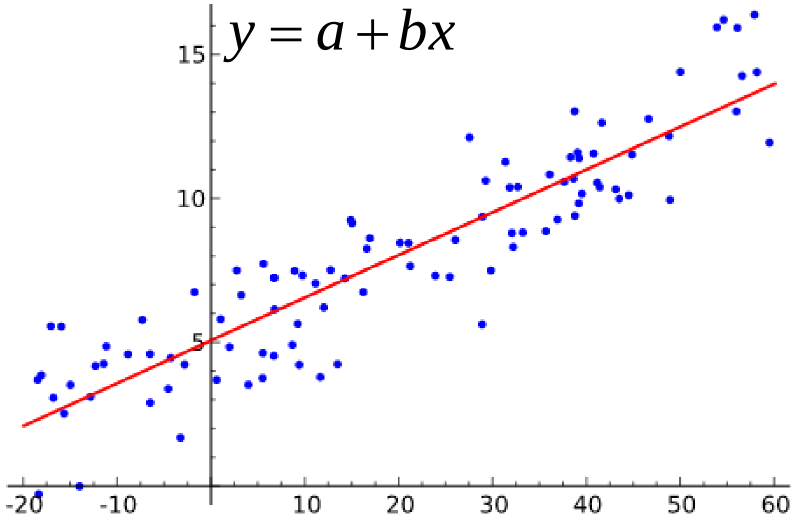
EXAMPLE: exponential. Normally this is done with the cumulative distribution

$$F_T(t) = 1 - e^{-\lambda t}$$



Do the points follow the simple exponential law? or maybe a Weibull $F_T(t) = 1 - e^{-(\lambda t)^{\beta}}$

Fitting data: least squares



Remember the goals:

1.- Parameter estimation
2.- Modelling (Hypothesis testing)

Goal: to minimize the "total distance" of n exp. points $d_1+d_2+\ldots d_n$ to the model

or more seriously.... to minimize the following variabel that follows a χ^2-distribution and therefore D **is normally distributed** around H

$$\chi^2 = \sum_{i=1}^{n} \frac{(D_i - H_i)^2}{H_i}$$

Where D_i are the data pints, and H_i are the model expected values (Hypothesis)

... since for n large a χ^2-distribution for point i tends to a gaussian with $\sigma^2=n=H_i$

$$\chi^2 = \sum_{i=1}^{n} \frac{(D_i - H_i)^2}{H_i} \rightarrow \sum_{i=1}^{n} \frac{(D_i - H_i)^2}{\sigma_i^2}$$

Fitting data: is the model right?

Let's do the question again (more precisse):

•Is the χ^2 arising after minimization
•When assuming that the data are normally distributerd around the model (hypothesis)
•Following a χ^2-distribution of (as it should?)

Rule of thumb (or the joys of the χ^2-distribution)

The χ^2 should follow a χ^2-distribution with n-m degrees of freedom arising for the data (with n points) and the model (with m=2 parameters)

For n data and 2 parameters μ=n-m and therefore the calculated χ^2 :

$$\chi^2 \approx (n\text{-}m)$$

For this reason it seems reasonable to define a reduced χ^2 that should be about one

$$\chi^2_{red} = \frac{\chi^2}{n-m} \approx 1$$

... for a good fit

"Hypothesis testing"

Kolmogorov-Smirnov test: WE DO NOT NEED TO DECIDE A BINNING!!!!!

1.- We calculate the cumulative distribution function CDF from the PDF and the model

$$CDF(\chi^2) = \int_0^{\chi^2} f(\chi^2) d\chi^2$$

2.- We look at the maximum distance between the model and the calculated CDF

3.- We look at the table for confience limits

confidence limit

number of data

Table 9.2: Critical Values of D_n^α in the Kolmogorov-Smirnov Test [9]

n \ α	0.20	0.10	0.05	0.01
5	0.45	0.51	0.56	0.67
10	0.32	0.37	0.41	0.49
15	0.27	0.30	0.34	0.40
20	0.23	0.26	0.29	0.36
25	0.21	0.24	0.27	0.32
30	0.19	0.22	0.24	0.29
35	0.18	0.20	0.23	0.27
40	0.17	0.19	0.21	0.25
45	0.16	0.18	0.20	0.24
50	0.15	0.17	0.19	0.23
>50	$\frac{1.07}{\sqrt{n}}$	$\frac{1.22}{\sqrt{n}}$	$\frac{1.36}{\sqrt{n}}$	$\frac{1.63}{\sqrt{n}}$