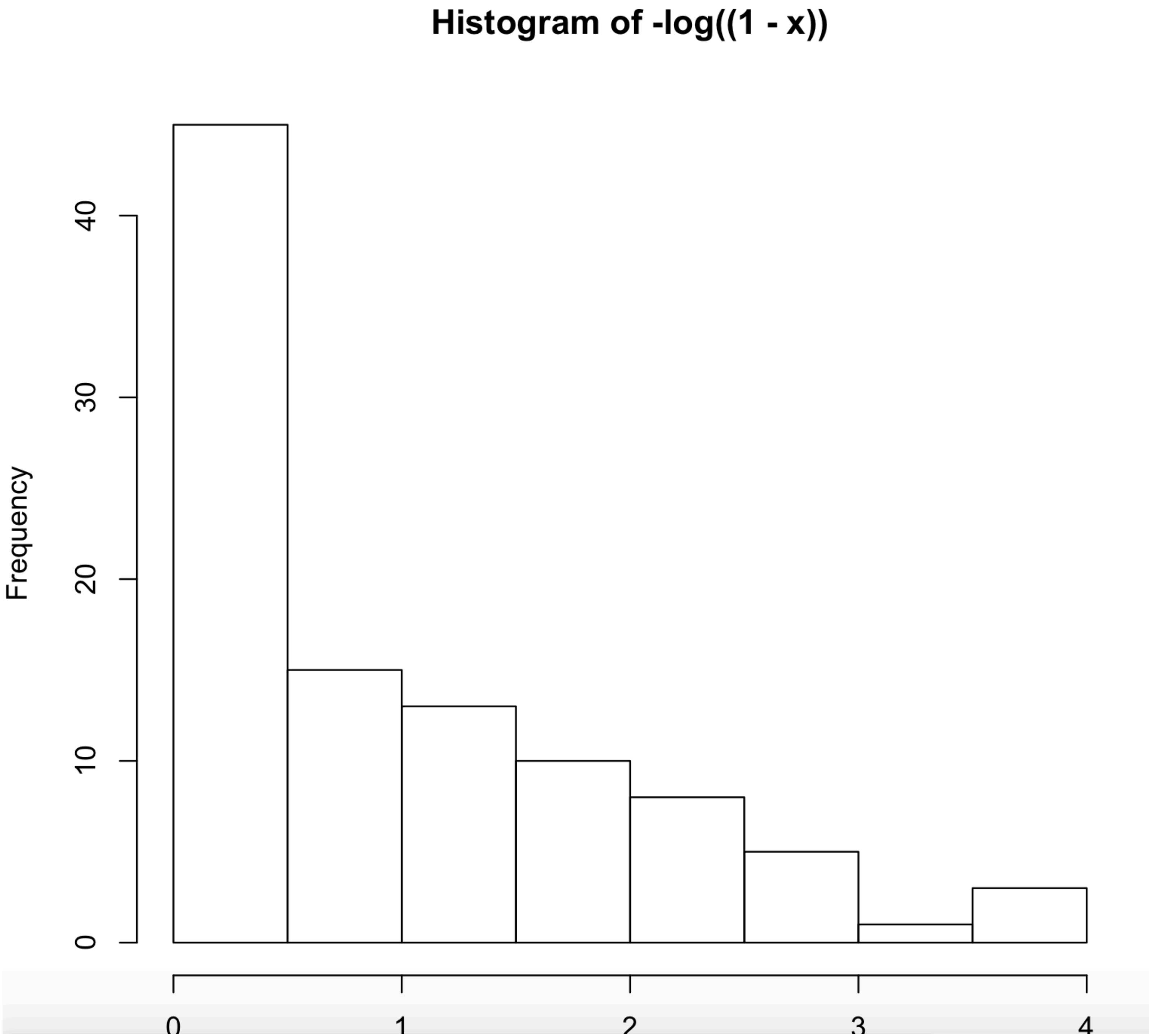
Proposed exercise

Extract N random numbers distributed as an exponential function with lifetime τ

Fill an histogram

Write the likelihood $L(t|\tau)$ in the binned and unbinned cases

```
N=100; x<-runif(N) ; x
  [1] 0.405059710 0.028044254 0.758571449 0.382914253 0.231949128 0.457176317
  [7] 0.736658152 0.038088207 0.104203774 0.513283288 0.742335360 0.368812945
 [13] 0.898926650 0.884993284 0.029905424 0.510855547 0.976764989 0.163296696
 [19] 0.312905139 0.172199152 0.789298260 0.518792378 0.076755612 0.187093519
 [25] 0.613189997 0.007589616 0.476067148 0.091391122 0.254679165 0.642145047
 [31] 0.068187724 0.213190998 0.284391620 0.652574104 0.375936000 0.938753973
 [37] 0.768648992 0.934079373 0.576549295 0.822300084 0.963397188 0.677318145
 [43] 0.804149516 0.278122875 0.918408046 0.161690666 0.816283114 0.219679127
 [49] 0.247514679 0.144359027 0.238819577 0.499138632 0.801599954 0.882881265
 [55] 0.817341159 0.484859340 0.865183191 0.866059658 0.375084123 0.287952191
 [61] 0.832247817 0.392507337 0.292606502 0.018239798 0.980023583 0.892270450
 [67] 0.843237637 0.927634800 0.204098272 0.763523759 0.545941953 0.600462520
 [73] 0.078878091 0.445519178 0.375912647 0.614324038 0.194723071 0.839467755
 [79] 0.265073122 0.870599505 0.696728359 0.085964346 0.004559065 0.710412472
 [85] 0.824518329 0.868817609 0.730170102 0.016328960 0.087571226 0.173662371
 [91] 0.367700928 0.491316323 0.085512807 0.738371863 0.977629644 0.378448315
 [97] 0.194459494 0.754219429 0.376693783 0.939928670
```


Histogram of -log((1 - x))
Frequency
0
10
20
30
40
0
1
2
3
4

Rate measurement

$$L(\underline{n}/\lambda) = \prod_{i=1}^{N} \frac{e^{-\lambda}\lambda^{n_i}}{n_i!}$$

$$\ln L = \sum_{i=1}^{N} (-\lambda + n_i \ln\lambda - \ln n_i!)$$

$$\frac{\partial \ln L}{\partial \lambda} = -N + \sum_{i=1}^{N} \frac{n_i}{\lambda}$$

$$-\frac{\partial^2 \ln L}{\partial \lambda^2}\bigg|_{\lambda=\hat{\lambda}} = \frac{\sum_{i=1}^{N} n_i}{\lambda^2}$$

$$\hat{\lambda} = \frac{\sum_{i=1}^{N} n_i}{N}$$

$$Var[\hat{\lambda}] = \frac{\hat{\lambda}^2}{\sum_{i=1}^{N} n_i} = \frac{\hat{\lambda}}{N}$$

$$\hat{r} = \frac{\hat{\lambda}}{\Delta t} \pm \frac{\sqrt{\hat{\lambda}}}{\sqrt{N}\Delta t}$$

Lifetime measurement

$$L(\underline{t}/\tau) = \prod_{i=1}^{N} \frac{1}{\tau} e^{-t_i/\tau}$$

$$\ln L = \sum_{i=1}^{N} \left(-\ln \tau - \frac{t_i}{\tau} \right)$$

$$\frac{\partial \ln L}{\partial \tau} = -\frac{N}{\tau} + \frac{1}{\tau^2} \sum_{i=1}^{N} t_i$$

$$-\frac{\partial^2 \ln L}{\partial \tau^2}\bigg|_{\tau=\hat{\tau}} = -\frac{1}{\hat{\tau}^2} \left(N - 2\frac{\sum_{i=1}^{N} t_i}{\hat{\tau}} \right) = \frac{N}{\hat{\tau}^2}$$

$$\hat{\tau} = \frac{\sum_{i=1}^{N} t_i}{N}$$

$$Var[\hat{\tau}] = -\hat{\tau}^2 \frac{1}{N - 2N} = \frac{\hat{\tau}^2}{N}$$

Gaussian measurement

$$L(\underline{x}/\mu, \sigma) = \prod_{i=1}^{N} \frac{1}{\sqrt{2\pi}\sigma_i} e^{-\frac{(x_i-\mu)^2}{2\sigma_i^2}}$$

Estimate μ, known σ

$$\frac{\partial \ln L}{\partial \mu} = \sum_i \frac{(x_i - \mu)}{\sigma_i^2}$$

$$-\frac{\partial^2 \ln L}{\partial \mu^2} = \sum_i \frac{1}{\sigma_i^2}$$

$$\hat{\mu} = \frac{\sum_i \frac{x_i}{\sigma_i^2}}{\sum_i \frac{1}{\sigma_i^2}}$$

$$Var[\hat{\mu}] = \frac{1}{\sum_i \frac{1}{\sigma_i^2}}$$

Gaussian measurement

$$L(\underline{x}/\mu,\sigma) = \prod_{i=1}^{N} \frac{1}{\sqrt{2\pi}\sigma_i} e^{-\frac{(x_i-\mu)^2}{2\sigma_i^2}}$$

Estimate σ, known μ

$$\frac{\partial \ln L}{\partial \sigma} = -\frac{N}{\sigma} + \frac{\sum_i (x_i-\mu)^2}{\sigma^3}$$

$$+\frac{\partial^2 \ln L}{\partial \sigma^2} = \frac{N}{\sigma^2} - 3\frac{\sum_i (x_i-\mu)^2}{\sigma^4}$$

$$\hat{\sigma^2} = \frac{\sum_i (x_i-\mu)^2}{N}$$

$$Var[\hat{\sigma}] = \frac{\hat{\sigma}^2}{2N}$$

Linear fit

- each measurement of y_i is characterized by a gaussian pdf with a known variance σ_i^2;
- the x_i values are assumed to be known with no or negligible uncertainty[26]; $(\sigma(x_i) << \sigma(y_i)/\hat{m})$
- the y_i measurements are not correlated;
- we make the hypothesis that the two physics quantities y and x are related by

$$y = mx + c$$

where m (the slope) and c (the intercept) are free parameters we want to measure from the data.

$$L(\underline{y}/m, c) = \prod_{i=1}^{N} \frac{1}{\sqrt{2\pi}\sigma_i} e^{-\frac{(y_i - mx_i - c)^2}{2\sigma_i^2}}$$

$$-2\ln(L) = \chi^2 = \sum_{i=1}^{N} \frac{(y_i - mx_i - c)^2}{\sigma_i^2}$$

that we have called χ^2 since, within the hypotheses done and discussed above, it is a test statistics with a χ^2 pdf with $N - 2$ degrees of freedom.

Linear fit

Minimizing X^2

$$\begin{aligned} \overline{x^2}m + \overline{x}c &= \overline{xy} \\ \overline{x}m + c &= \overline{y} \end{aligned}$$

$$\overline{z} = \frac{\sum_{i=1}^{N} \frac{z_i}{\sigma_i^2}}{\sum_{i=1}^{N} \frac{1}{\sigma_i^2}}$$

$$\begin{aligned} \hat{m} &= \frac{\overline{xy} - \overline{x} \cdot \overline{y}}{\overline{x^2} - \overline{x}^2} \\ \hat{c} &= \frac{\overline{x^2} \cdot \overline{y} - \overline{x} \cdot \overline{xy}}{\overline{x^2} - \overline{x}^2} \end{aligned}$$

Linear fit

The covariance matrix of the 2 parameters is determined evaluating first the Hessian matrix (see eq.125), and by inverting it with the usual methods of matrix inversions. The Fisher matrix is:

$$\begin{pmatrix} \sum_i \frac{x_i^2}{\sigma_i^2} & \sum_i \frac{x_i}{\sigma_i^2} \\ \sum_i \frac{x_i}{\sigma_i^2} & \sum_i \frac{1}{\sigma_i^2} \end{pmatrix}$$

and the covariance matrix is:

$$\begin{pmatrix} \frac{1}{\sum_i (1/\sigma_i^2) Var[x]} & \frac{-\overline{x}}{\sum_i (1/\sigma_i^2) Var[x]} \\ \frac{-\overline{x}}{\sum_i (1/\sigma_i^2) Var[x]} & \frac{\overline{x^2}}{\sum_i (1/\sigma_i^2) Var[x]} \end{pmatrix}$$

where the variance of x is not the uncertainty on x but the lever arm of the fit, namely the spread of the x values on the x axis.

$$\sigma(\hat{m}) = \frac{\sigma}{\sqrt{N}\sqrt{Var[x]}}$$

$$\sigma(\hat{c}) = \frac{\sqrt{\overline{x^2}}\sigma}{\sqrt{N}\sqrt{Var[x]}}$$

$$cov(\hat{m}, \hat{c}) = -\frac{\sqrt{\overline{x}}\sigma}{\sqrt{N}\sqrt{Var[x]}}$$

Generic linear fit

$$y = f(x/\underline{\theta}) = \sum_{k=1}^{M} \theta_k f_k(x)$$

$$\chi^2 = \sum_{i=1}^{N} \frac{(y_i - \sum_k \theta_k f_k(x_i))^2}{\sigma_i^2} = -2lnL$$

M equations

$$\frac{\partial \chi^2}{\partial \theta_j} = \sum_i \frac{-2 f_j(x_i)(y_i - \sum_k \theta_k f_k(x_i))}{\sigma_i^2} = 0$$

$$\sum_k \left[\sum_i \frac{f_j(x_i) f_k(x_i)}{\sigma_i^2} \right] \theta_k = \sum_i \frac{y_i f_j(x_i)}{\sigma_i^2}$$

$$\hat{\theta}_k = \sum_i \sum_j V_{kj} \frac{y_i f_j(x_i)}{\sigma_i^2}$$

Generic linear fit

i runs on events
j,k on coefficients

$$\hat{\theta}_k = \sum_i \sum_j V_{kj} \frac{y_i f_j(x_i)}{\sigma_i^2}$$

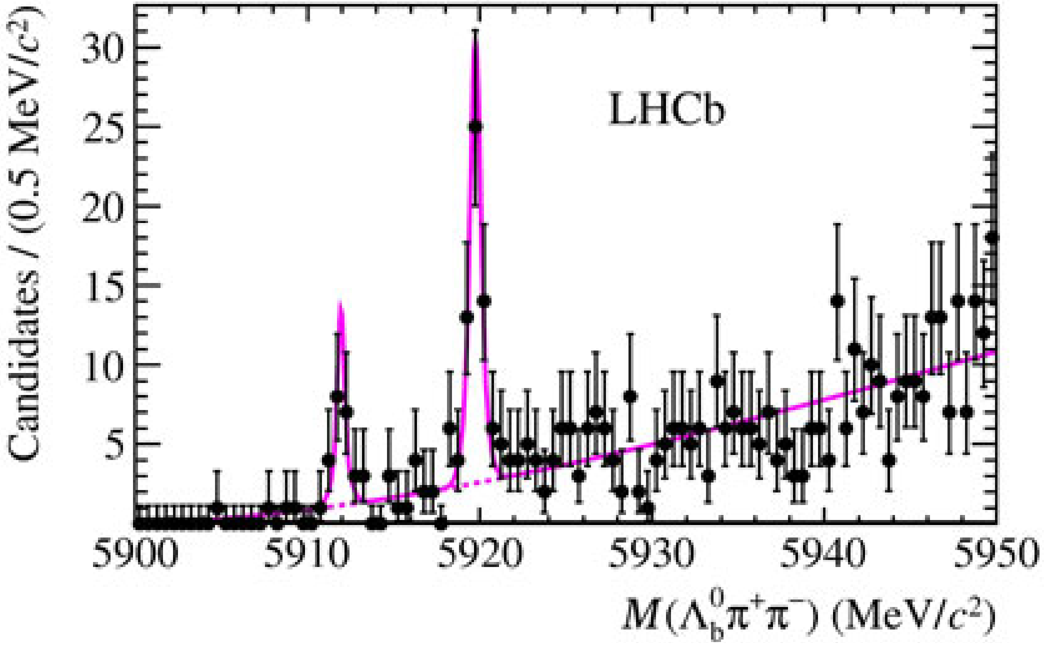
V_{kj} covariance matrix of the parameters

Parameter coefficient matrix

$$\left(V^{-1}\right)_{kj} = \sum_i \frac{f_k(x_i) f_j(x_i)}{\sigma_i^2}$$

Analytical solution

Nuisance parameters



We define[28] N_s and N_b the total number of signal and background events respectively, $f_s(x/M)$ and $f_b(x/\underline{\alpha})$ the two functions of the mass x describing the signal and background respectively. f_s is assumed to be gaussian with mean M and a width σ assumed to be known[29]:

(184) $$f_s(x/M) = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(x-M)^2}{2\sigma^2}}$$

f_b is assumed to be a polynomial function[30], $\underline{\alpha}$ being the vector of parameters describing the polynomial background (together with N_b). Both functions are normalized to 1. The parameters describing the background are free parameters and have to be evaluated by the fit or have to be known independently (e.g. from Montecarlo). However, since they have not a deep physical meaning they are called generically **nuisance parameters**. On the other hand N_s and M are the parameters we are interested in.

Nuisance parameters

Let's consider first the unbinned case. The test statistics can be written as an extended likelihood (N is the number of events entering the histogram):

$$L(\underline{x}/N_s, N_b, M, \underline{\alpha}) = \frac{e^{-(N_s+N_b)}(N_s+N_b)^N}{N!}\prod_{i=1}^{N}\left[N_s f_s(x_i/M) + N_b f_b(x_i/\underline{\alpha})\right]$$

For the histogram fit we have to define the signal and background contents s_i and b_i in each of the M bins of width δx:

$$\begin{aligned} s_i &= N_s \int_{\overline{x_i}-\delta x/2}^{\overline{x_i}+\delta x/2} f_s(x/M)dx \\ b_i &= N_b \int_{\overline{x_i}-\delta x/2}^{\overline{x_i}+\delta x/2} f_b(x/\underline{\alpha})dx \end{aligned}$$

$$L(\underline{n}/N_s, N_b, M, \underline{\alpha}) = \prod_{i=1}^{M}\frac{e^{-(s_i+b_i)}(s_i+b_i)^{n_i}}{n_i!}$$

where n_i is the experimental content in the bin i.

Nuisance parameters

In both cases the minimization and the evaluation of the hessian matrix of this likelihood will be done numerically. As a result we'll have estimates of the 2 relevant parameters N_s and M and of the nuisance parameters. Moreover the value of L at the minimum will be used for hypothesis test.

The possibility to move the nuisance parameters in the fit, allows to obtain a better agreement between data and theory at the expense of having larger uncertainties on the relevant parameters N_s and M. Any knowledge of the nuisance parameters can be added in the likelihood as additional constraint. For example if N_b is known to be $\overline{N_b} \pm \sigma(N_b)$ with a gaussian shape, an additional gaussian factor can be added to the likelihood forcing N_b to stay within its gaussian limits. The lower is $\sigma(N_b)$ the lower will be its impact on the final uncertainties on N_s and M. From this example we see that the method of the nuisance parameters can be used to include the evaluation of systematic uncertainties directly in the fit.