

## The Method of Maximum Likelihood

A very popular and powerful method of deriving estimators. Maximum Likelihood Estimators (MLEs) have many desirable properties.

Let's understand the basic concepts with our muon lifetime measurement example.

► What do we "know" before the experiment? <sup>(a priori)</sup>

⇒ Our data  $\{t_1, t_2, \dots, t_n\}$  is  $\sim \frac{1}{\tau} e^{-t/\tau}$

i.e.  $f_{\text{sample}}(\vec{t}; \tau) = \frac{1}{\tau^n} \prod_{i=1}^n e^{-t_i/\tau}$

► What we do not know:

⇒ value of  $\tau$

► What are we interested in knowing?

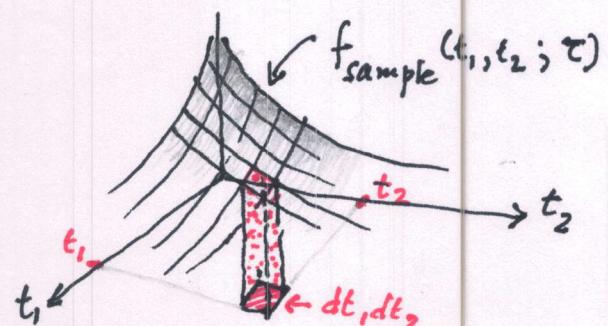
⇒ value of  $\tau$ , the parameter of interest (POI)

► What do we know after ('a posteriori') the experiment?

⇒ Measured values of  $t$ 's ( $t_1 = 1.2 \mu\text{s}, t_2 = 3.1 \mu\text{s}$  etc.)  
Those are now fixed

The probability of observing the dataset we have observed is (i.e. the probability that  $t_i$  lies between  $t_i$  and  $t_i + dt_i$ , etc.)

$$\prod_{i=1}^n f(t_i | \tau) dt_i \\ = f_{\text{sample}}(\vec{t}; \tau) \prod_{i=1}^n dt_i$$



quite clearly, the probability of observing a dataset in the neighbourhood (the little red box  $d\vec{t}_1 d\vec{t}_2$ ) of the data that we have observed depends on the unknown value of  $\tau$ .

It would be reasonable to expect:

If we calculate this probability ( $f_{\text{sample}}(\vec{t}; \tilde{\tau}) d^n t$ ) for various proposed (or hypothesized) values,  $\tilde{\tau}$  then, as  $\tilde{\tau} \rightarrow \tau^{\text{true}}$  the calculated value of the probability will maximize.

We will not get to the true value of  $\tau$  but we can hope to get a "best" estimate  $\hat{\tau}$  that is possible from this data

Let's try it out in our example.

**Important :** Remember that joint probability density is a function of  $\vec{x}$  for given value of parameter. But here our dataset is fixed  $\{\vec{t}_1, \dots, \vec{t}_n\}$  have all been measured. We want to see how this quantity changes as we vary the parameter  $\tau$ . To make it apparent let us define.

$$\underline{L}(\tau; \vec{t}) = f_{\text{sample}}(\vec{t}; \tau)$$

likelihoood function, a function of parameter  $\tau$ , equal in value with  $f_{\text{sample}}$

Note that :

$$\int_0^\infty L(\tau; \vec{t}) d\tau \neq 1 \quad \leftarrow \begin{array}{l} \text{likelihood function} \\ \text{can not be interpreted as} \\ \text{probability density function} \end{array}$$

while

$$\int_{t_1} \cdots \int_{t_n} f_{\text{sample}}(\vec{t}; \tau) dt_1 \cdots dt_n = 1$$

### Log likelihood

note that  $L(\tau)$  is a product of fractions. If we have many measurements, it becomes a very tiny fraction, hard to deal with numerically.

$l(\tau) = \log_e L(\tau)$  is often more convenient.

In our example

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

$$\Rightarrow l(\tau) = -n \ln(\tau) - \frac{1}{\tau} \left( \sum_{i=1}^n t_i \right)$$

Since  $\log L(\tau)$  is a monotonic function of  $L(\tau)$  maximizing  $l(\tau)$  w.r.t.  $\tau$  is equivalent of maximizing  $L(\tau)$ .

$$\text{So, let's try } \frac{\partial l}{\partial \tau} = 0$$

$$\frac{\partial L}{\partial \bar{t}} = -\frac{n}{2} + \frac{1}{\bar{t}^2} \sum_{i=1}^n t_i = 0$$

$$\Rightarrow \hat{\bar{t}} = \frac{1}{n} \sum_{i=1}^n t_i \quad \leftarrow \text{sample mean!}$$

In this case MLE gave consistent, unbiased estimator.

Note that if we multiply  $L(\bar{t})$  with a constant  $C$  result does not change

Definition: Likelihood function

Let  $f(\vec{x} | \vec{\theta})$  denote the joint pdf or pmf of the sample  $\vec{x} = \{x_i\}$ , then given that  $\vec{x}$  is observed, the function of  $\vec{\theta}$  defined by

$$L(\vec{\theta}; \vec{x}) = f(\vec{x}; \vec{\theta})$$

is called the likelihood function

Likelihood Principle:

If  $\vec{x}$  and  $\vec{y}$  are two sample points such that  $L(\vec{\theta}; \vec{x})$  is proportional to  $L(\vec{\theta}; \vec{y})$ , that is, there exists a constant  $C(\vec{x}, \vec{y})$  such that,

$$L(\vec{\theta}; \vec{x}) = C(\vec{x}, \vec{y}) L(\vec{\theta}; \vec{y}) \text{ for all } \vec{\theta}$$

then the conclusions drawn from  $\vec{x}$  and  $\vec{y}$  should be identical

[Casella & Berger]

## Definition (MLE)

For each sample point  $\vec{x}$  let  $\hat{\theta}(\vec{x})$  be a parameter value at which  $L(\vec{\theta}; \vec{x})$  attains its maximum as a function of  $\vec{\theta}$ , with  $\vec{x}$  held fixed. A maximal likelihood estimator (MLE) of the parameter (vector)  $\vec{\theta}$  based on sample  $\vec{x}$  is  $\hat{\theta}(\vec{x})$

[casella & Berger]

An useful and interesting property of MLE is invariance under transformation

Invariance property: If  $\hat{\theta}$  is the MLE of  $\theta$  then for any function  $a(\theta)$ , the MLE of  $a(\hat{\theta})$   $a(\theta)$  is  $a(\hat{\theta})$ , i.e.  $\hat{a} = a(\hat{\theta})$ .

If there is a one to one map between  $a$  and  $\theta$  this is quite obvious. Even if that is not the case (e.g.  $a(\theta) = \theta^2$ ) the invariance property holds.

example Using this property we can see that the MLE of decay constant  $\lambda = \frac{L}{T}$  is  $\hat{\lambda} = \frac{\bar{t}}{\bar{T}}$

However  $E[\hat{\lambda}] = \lambda \frac{n}{n-1}$ , so it is only asymptotically unbiased.

Normal MLE with both  $\mu, \sigma^2$  unknown.

$$\begin{aligned} L(\mu, \sigma^2; \vec{x}) &= \prod_{\substack{i=1 \\ \theta}}^n \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{1}{2} \left( \frac{x_i - \mu}{\sigma} \right)^2} \\ &= \frac{1}{(2\pi\sigma^2)^{n/2}} \cdot e^{-\frac{1}{2\sigma^2} \left( \sum_i (x_i - \mu)^2 \right)} \end{aligned}$$

Maximization w.r.t  $\mu, \sigma^2$  requires

$$\frac{\partial l(\mu, \sigma^2)}{\partial \mu} = \frac{1}{\sigma^2} \sum_{i=1}^n (x_i - \mu) = 0$$

$$\frac{\partial l(\mu, \sigma^2)}{\partial \sigma^2} = -\frac{n}{2\sigma^2} + \frac{1}{2\sigma^4} \sum_{i=1}^n (x_i - \underbrace{\mu}_{\theta_i})^2 = 0$$

$$\Rightarrow \hat{\mu} = \frac{1}{n} \sum_{i=1}^n x_i = \bar{x}_n, \quad \hat{\sigma}^2 = \frac{1}{n} \sum_{i=1}^n (x_i - \hat{\mu})^2$$

so we see that  $\hat{\sigma}^2$  is not unbiased.

$$E[\hat{\sigma}^2] = \frac{n-1}{n} \sigma^2 \text{ asymptotically unbiased.}$$

Question Setting  $\frac{\partial l}{\partial \sigma} = 0$  we could get  $\hat{\sigma}$  instead of  $\hat{\sigma}^2$ . What's the value of  $\hat{\sigma}$ ?

Recall, all these estimated parameters are random variables and will have an uncertainty, which can be quoted as error on the estimated parameter.

## Variance of ML estimators

### 1. Analytic Method

In the lifetime experiment example,

$$\begin{aligned} V[\hat{\tau}] &= E[\hat{\tau}^2] - (E[\hat{\tau}])^2 \\ &= \frac{\tau^2}{n} \quad \text{by CLT} \end{aligned}$$

- \* also one can explicitly write down the expectation values and work it out. (HOME WORK)

In practice we will calculate

$$\hat{V} = \widehat{\sigma^2(\hat{\tau})} = \frac{\hat{\tau}^2}{n} \quad \text{since } \tau \text{ is unknown}$$

Result of the experiment is reported as

$$\hat{\tau} = \underbrace{2.19}_{\text{MLE}} \pm \underbrace{0.18}_{\sqrt{\hat{V}}} \mu\text{s.}$$

However, this is not a standard interval if the distribution of  $\hat{\tau}$  is non-Gaussian.

### 2. Monte Carlo Method

- 1) Take  $\hat{\tau}$  as proxy for  $\tau$ .
- 2) Generate large no. of toy datasets  $\{\hat{\tau}_1, \dots, \hat{\tau}_m\}$
- 3) For each toy calculate  $\hat{\tau}$  by MLE
- 4) Find the  $s^2$  as the estimator of variance.

$$\frac{1}{m-1} \sum_{j=1}^m (\hat{\tau}_j - \bar{\hat{\tau}}_m)^2, \quad \bar{\hat{\tau}}_m = \frac{1}{m} \sum_{j=1}^m \hat{\tau}_j$$

↓ Computation intensive!

## Variance of MLE

Cramér-Rao Inequality (RCF bound)

Let  $\vec{x} = \{x_1, \dots, x_n\}$  be a sample with pdf  $f(x; \theta)$  and let  $\hat{\theta}(\vec{x})$  be any estimator satisfying

$$\frac{d}{d\theta} E_{\theta}[\hat{\theta}] = \int_{\text{sample space}} \frac{\partial}{\partial \theta} [\hat{\theta}(\vec{x}) f(\vec{x}; \theta)] d\vec{x}$$

and

$$V_{\theta}[\hat{\theta}] < \infty$$

Note: Subscript  $\theta$  in  $E_{\theta}, V_{\theta} \rightarrow$  Calculation done for given  $\theta$ . Will suppress.

Then

$$V_{\theta}[\hat{\theta}] \geq \frac{\left( \frac{d}{d\theta} E[\hat{\theta}] \right)^2}{E_{\theta}\left( \left( \frac{\partial}{\partial \theta} \ln f(\vec{x}; \theta) \right)^2 \right)}$$

[Casella & Berger 7.3.9  
P. 335]

Continuing with the lifetime example.

$$\hat{\theta}(\vec{x}) \Rightarrow \hat{\tau}(\vec{t}) = \bar{\tau}(t_1, \dots, t_n) = \frac{1}{n} \sum_{i=1}^n t_i$$

$$E[\hat{\theta}] = E[\hat{\tau}] = \tau \Rightarrow \frac{d}{d\theta} E[\hat{\theta}] = \frac{d}{d\tau} (\tau) = 1.$$

$$\therefore \text{Numerator} = (1)^2 = 1.$$

$$\ln f(\vec{x}; \theta) = \ln(t_1, \dots, t_n; \tau) = -n \ln \tau - \frac{1}{\tau} \left( \sum_{i=1}^n t_i \right) \equiv l$$

$$\frac{\partial l}{\partial \theta} = \frac{\partial l}{\partial \tau} = -\frac{n}{\tau} + \frac{1}{\tau^2} \left( \sum t_i \right) = -\frac{n}{\tau} + \frac{n \bar{t}_n}{\tau^2}$$

$$E\left[\left(\frac{\partial l}{\partial \tau}\right)^2\right] = \frac{n}{\tau^2} \quad \text{Homework}$$

$$\Rightarrow V[\hat{\tau}] \geq \frac{\tau^2}{n}$$

Solution to homework RCF bound for  $\hat{\tau}$

$$\begin{aligned} \left(\frac{\partial \ell}{\partial \tau}\right)^2 &= \left(-\frac{n}{\tau} + \frac{1}{\tau^2} \sum t_i\right)^2 \\ &= \frac{n^2}{\tau^2} + \left(\frac{n\bar{t}_n}{\tau^2}\right)^2 - 2 \cdot \frac{n}{\tau} \cdot \frac{1}{\tau^2} \cdot n\bar{t}_n \end{aligned}$$

$$\begin{aligned} E\left(\frac{\partial \ell}{\partial \tau}\right)^2 &= E\left(\frac{n^2}{\tau^2}\right) + \frac{n^2}{\tau^4} E(\bar{t}_n^2) - \frac{2n^2}{\tau^3} E[\bar{t}_n] \\ &= \frac{n^2}{\tau^2} + \frac{n^2}{\tau^4} E\left[\frac{1}{n^2} \sum_i \sum_j t_i t_j\right] - \frac{2n^2}{\tau^3} \cdot \tau \\ &= -\frac{n^2}{\tau^2} + \frac{1}{\tau^4} E\left[\sum_{i,j} t_i t_j\right] \end{aligned}$$

$$E\left[\sum_{i,j} t_i t_j\right] = E\left[\sum_{i=1}^n t_i^2 + \sum_{i \neq j} t_i t_j\right] \quad \text{--- (1)}$$

(n<sup>2</sup> terms)                                  (n terms)                                   $\uparrow 2 \sum_{i < j} t_i t_j$  (nC<sub>2</sub> terms)

$$\begin{aligned} E[t_i^2] &= \underbrace{\int_0^\infty t_i^2 \cdot \frac{1}{\tau} e^{-t_i/\tau} dt_i}_{\tau^2 \int_0^\infty x^2 e^{-x} dx} \underbrace{\int \dots \int_{i \neq j} \frac{1}{\tau} e^{-t_j/\tau} dt_j}_{1} \\ &= \Gamma(3) = 2 \\ &= 2\tau^2 \end{aligned}$$

$$\begin{aligned} E[t_i t_j] &= \underbrace{\int t_i \frac{1}{\tau} e^{-t_i/\tau} dt_i}_{\tau} \underbrace{\int t_j \cdot \frac{1}{\tau} e^{-t_j/\tau} dt_j}_{\tau} \underbrace{\int \dots \int_{k \neq i, j} \frac{1}{\tau} e^{-t_k/\tau} dt_k}_{k \neq i, j} d^{n-2} t_k \\ &= \tau^2 \end{aligned}$$

$$\therefore \textcircled{1} = 2n\tau^2 + 2 \cdot \frac{n(n-1)}{2} \cdot \tau^2 = (n^2+n)\tau^2$$

$$\therefore E\left[\left(\frac{\partial \ell}{\partial \tau}\right)^2\right] = -\frac{n^2}{\tau^2} + \frac{(n^2+n)\tau^2}{\tau^4} = \underline{\underline{\frac{n}{\tau^2}}}$$

We know that in the  $\hat{\tau}$  example  $V[\hat{\tau}] = \frac{\tau^2}{n}$

$\Rightarrow$  RCF bound is reached  $\Rightarrow$  efficient.

► If efficient estimator exists MLE will find it.

$E\left[\left(\frac{\partial l}{\partial \theta}\right)^2\right]$  is called information number or Fisher information  $\Rightarrow$  RCF is also called information inequality

Under certain (general enough) conditions

$$E\left[\left(\frac{\partial l}{\partial \theta}\right)^2\right] = -E\left[-\frac{\partial^2 l}{\partial \theta^2}\right]$$

and RCF ~~be~~ inequality can be put as

$$V[\hat{\theta}] \geq \left(1 + \frac{\partial b}{\partial \theta}\right)^2 / E\left[-\frac{\partial^2 l}{\partial \theta^2}\right]$$

Homework Verify that in the Gaussian example  $\hat{\sigma}^2$  does not reach RCF bound

$$\text{When } \vec{\theta} = \{\theta_1, \dots, \theta_n\} \quad \frac{\partial^2 l}{\partial \theta^2} \Rightarrow \frac{\partial^2 l}{\partial \theta_i \partial \theta_j}$$

For unbiased, efficient estimator, then

$$(V^{-1})_{ij} = E\left[-\frac{\partial^2 l}{\partial \theta_i \partial \theta_j}\right] \quad V_{ij} = \text{cov}(\hat{\theta}_i, \hat{\theta}_j)$$

In practice

$$(\widehat{V}^{-1})_{ij} = -\frac{\partial^2 l}{\partial \theta_i \partial \theta_j} \Big|_{\vec{\theta} = \hat{\vec{\theta}}} \quad \begin{array}{l} \leftarrow \text{Hessian} \\ \text{calculated numerically} \end{array}$$

(HESSE in MINUIT)

## Variance of MLE by graphical method

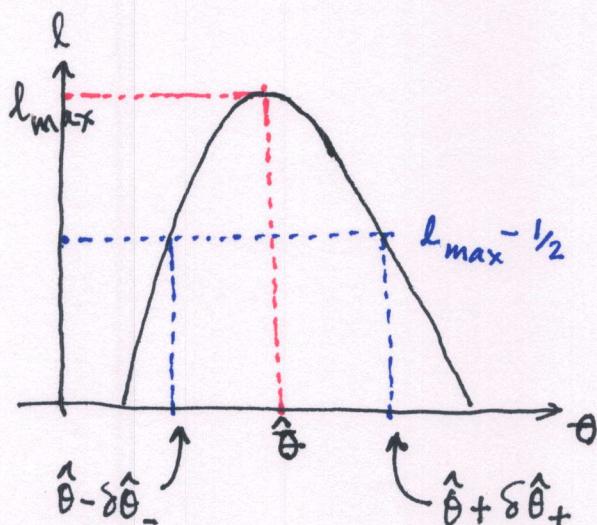
$l(\hat{\theta})$  and its derivatives at  $\theta = \hat{\theta}$  can be computed.

$$l(\theta) = l(\hat{\theta}) + \left. \frac{\partial l}{\partial \theta} \right|_{\theta=\hat{\theta}} (\theta - \hat{\theta}) + \frac{1}{2} \left. \frac{\partial^2 l}{\partial \theta^2} \right|_{\theta=\hat{\theta}} (\theta - \hat{\theta})^2 + \dots$$

In large sample limit  $L(\theta) \rightarrow \text{Gaussian}$

$\Rightarrow l(\theta)$  becomes parabola

$\Rightarrow$  Symmetric error



$$l(\hat{\theta} + \delta\hat{\theta}_+) = l_{\max} + \frac{1}{2} \left( -\frac{1}{\hat{\sigma}_{\hat{\theta}}} \right)^2 \cdot \delta\hat{\theta}_+^2$$

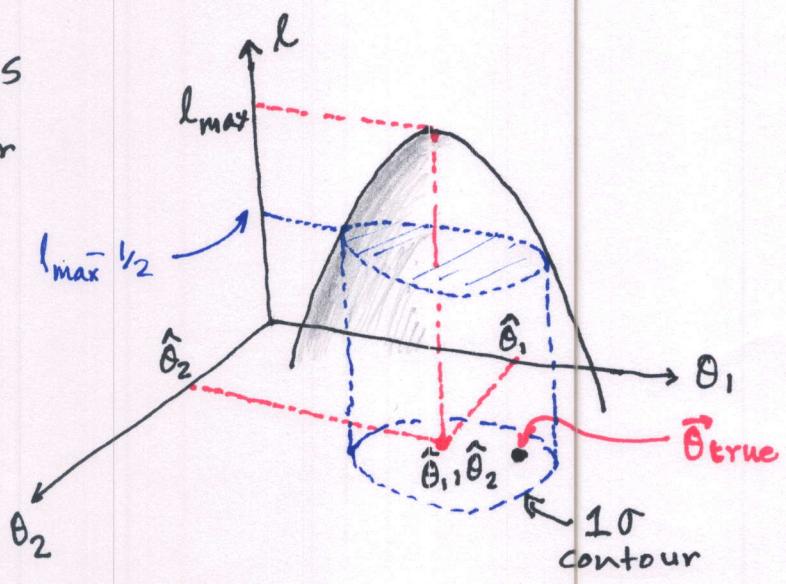
$$= l_{\max} - \frac{1}{2}$$

Note 1: In multidimension this  $l(\theta)$  becomes a (hyper)surface  $l(\vec{\theta}) = l(\theta_1, \dots, \theta_n)$

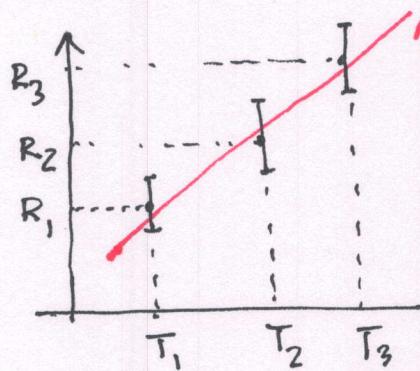
and  $l_{\max} - \frac{1}{2}$  points become a contour

Note 2: As data increases  
 $l(\theta)$  gets narrower

$$l_{\max}(n \rightarrow \infty) = -\frac{1}{2} \frac{1}{(1-\rho^2)} \left[ \frac{(\theta_1 - \hat{\theta}_1)^2}{\hat{\sigma}_{\hat{\theta}_1}^2} + \frac{(\theta_2 - \hat{\theta}_2)^2}{\hat{\sigma}_{\hat{\theta}_2}^2} - 2\rho \left( \frac{\theta_1 - \hat{\theta}_1}{\hat{\sigma}_{\hat{\theta}_1}} \right) \left( \frac{\theta_2 - \hat{\theta}_2}{\hat{\sigma}_{\hat{\theta}_2}} \right) \right]$$



## $\chi^2$ and likelihood



$$R_t(T) = \theta_1 + \theta_2 T \quad (\text{Theoretical model})$$

measured resistance

$$R = R_t + \delta R$$

Gaussian error

Probability of observing  $R$

$$= \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{1}{2} \frac{(R - R_t)^2}{\sigma^2}}$$

At three temperatures measured resistance

$R_1, R_2, R_3$  with measurement errors (s.d.)  $\sigma_1, \sigma_2, \sigma_3$

$$\begin{aligned} L(\theta_1, \theta_2; R_1, R_2, R_3) &= f(R_1; \theta_1, \theta_2) \cdot f(R_2; \theta_1, \theta_2) \cdot f(R_3; \theta_1, \theta_2) \\ &= \frac{1}{\sqrt{2\pi\sigma_1^2}} e^{-\frac{1}{2} \frac{(R_1 - R_{t1})^2}{\sigma_1^2}} \cdot \frac{1}{\sqrt{2\pi\sigma_2^2}} e^{-\frac{1}{2} \frac{(R_2 - R_{t2})^2}{\sigma_2^2}} \cdots \end{aligned}$$

$$\text{Here } R_{t1} = \theta_1 + \theta_2 T_1$$

$$R_{t2} = \theta_1 + \theta_2 T_2 \quad \text{etc.}$$

$$\Rightarrow l(\theta_1, \theta_2) = -\frac{1}{2} \left[ \frac{(R_1 - R_{t1})^2}{\sigma_1^2} + \frac{(R_2 - R_{t2})^2}{\sigma_2^2} + \frac{(R_3 - R_{t3})^2}{\sigma_3^2} \right]$$

↑ this is  $\chi^2$  of

3 measurements.

Whenever error is normal

$$\underline{l(\vec{\theta}; \vec{u}) = -\frac{1}{2} \chi^2}$$

So log likelihood maximization is the same as  $\chi^2$  minimization

## Extended Maximum Likelihood

When the size of data  $n \sim \text{Poisson}(\nu)$  for dataset  $\{x_1, \dots, x_n\}$ .

(e.g.  $x$  = invariant mass of final state particles in a particle search)

$$L(\nu, \vec{\theta}) = \frac{\nu^n}{n!} e^{-\nu} \prod_{i=1}^n f(x_i; \vec{\theta})$$

poisson prob of  $n$  observations.      usual likelihood

$$l(\nu, \vec{\theta}) = n \ln \nu(\vec{\theta}) - \nu(\vec{\theta}) + \sum_{i=1}^n \ln f(x_i; \vec{\theta}) + \text{const.}$$

[assume :  $\nu = \nu(\vec{\theta})$ ]

e.g. in a collision run  $\nu = \sigma L \epsilon$  ← efficiency.

$\sigma, x_i \rightarrow$  both depend on parameters like mass, coupling

In general will reduce stat error ✓

Sample with mixed signal and background

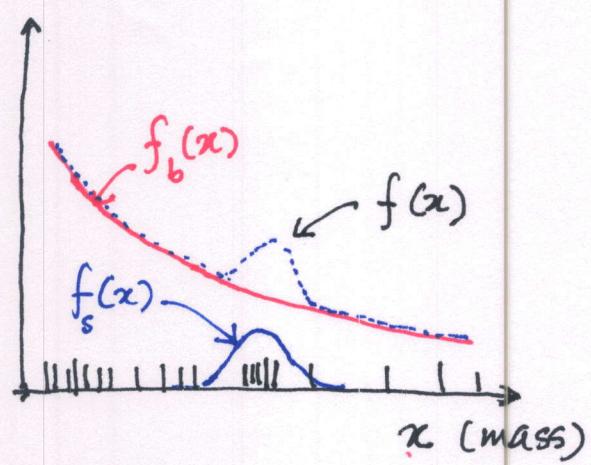
$$N = \underline{n_s} + \underline{n_b}$$

$$n_s \sim \frac{e^{-s} s^{n_s}}{n_s!}, \quad n_b \sim \frac{e^{-b} b^{n_b}}{n_b!}$$

$$f(x) = \frac{s}{s+b} f_s(x) + \frac{b}{s+b} f_b(x)$$

$f_s, f_b$  known

Interested in  $\underline{s}$



### Extended likelihood...

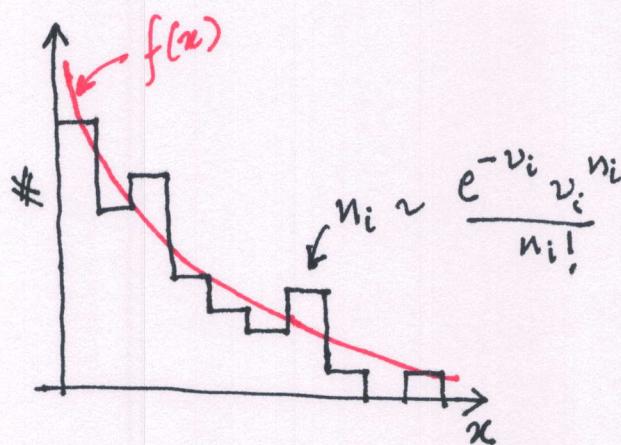
$$E(n) = E[n_s] + E[n_b]$$

$$\text{or } v = s+b$$

$$\begin{aligned}
 l(v, s, b, \vec{\theta}) &= -n \ln v - v + \sum_{i=1}^n \ln f(x_i; \vec{\theta}) \\
 &= -v + \sum_{i=1}^n \ln (v \cdot f(x_i)) \\
 &= -(s+b) + \sum_{i=1}^n \ln \left\{ (s+b) \left( \frac{s}{s+b} f_s + \frac{b}{s+b} f_b \right) \right\} \\
 &= -(s+b) + \sum_{i=1}^n \ln \left\{ s f_s + b f_b \right\}
 \end{aligned}$$

By setting  $\frac{\partial l}{\partial s} = 0$ ,  $\frac{\partial l}{\partial b} = 0$  one can estimate  $\hat{s}, \hat{b}$

## MLE of binned data



Histogram with  $N$  bins,  
 $n$  total events

or multinomial distributed.

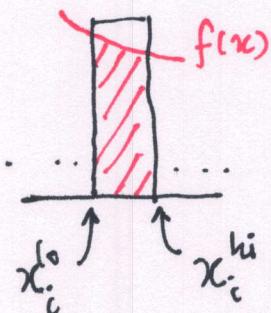
We have to fit  $f(x; \vec{\theta})$

$n_i$  = no. of events in bin  $i$

$$\sum_{i=1}^N n_i = n$$

$$E[n_i] = v_i \quad \begin{matrix} \text{fit parameters} \\ \text{in here.} \end{matrix}$$

$$v_i = \frac{n}{\int_{x_i^{lo}}^{x_i^{hi}} f(x) dx} \quad \begin{matrix} f(x) = f(x; \vec{\theta}) \\ \uparrow \end{matrix}$$



Joint probability of obtaining the histogram  $\vec{n} = \{n_1, \dots, n_N\}$

$$f_{\text{joint}} \{n_1, n_2, \dots, n_N; \vec{\theta}\} = \frac{n!}{n_1! \cdots n_N!} \prod_{i=1}^N \left(\frac{v_i}{n}\right)^{n_i}$$

$\frac{v_i}{n} = p_i$  : probability that entry is in  $i^{\text{th}}$  bin.

$$\Rightarrow \underline{l(\vec{\theta})} = \sum_{i=1}^N n_i \ln \left( \frac{v_i}{n} \right) + c = \sum_{i=1}^N n_i \ln (v_i(\vec{\theta})) + c'$$

If we take bin content Poisson distributed

$$f_{\text{joint}} \{ \vec{n}; \vec{\theta} \} = \prod_{i=1}^N \frac{e^{-v_i} v_i^{n_i}}{n_i!}$$

$$\Rightarrow l(\vec{\theta}) = \sum_{i=1}^N (-v_i + n_i \ln(v_i))$$

Note : Bin with 0 entry is not a problem

## Combining experiments with likelihood

Suppose two experiments measured  $\{x_1, \dots, x_n\}$  and  $\{y_1, \dots, y_m\}$  both aimed at measuring same parameters  $\vec{\theta}$ .

[e.g.  $Z_0$  mass and width from  $Z \rightarrow e^+ e^-$  and  $Z \rightarrow \mu^+ \mu^-$ ]

► Combined likelihood:

$$\begin{aligned} L(\vec{\theta}; \vec{x}, \vec{y}) &= L_x(\vec{\theta}; \vec{x}) L_y(\vec{\theta}; \vec{y}) \\ &= \prod_{i=1}^n f_x(x_i; \vec{\theta}) \prod_{j=1}^m f_y(y_j; \vec{\theta}) \end{aligned}$$

$$l(\vec{\theta}; \vec{x}, \vec{y}) = \sum_{i=1}^n \ln f_x(x_i; \vec{\theta}) + \sum_{j=1}^m \ln f_y(y_j; \vec{\theta})$$

► Suppose two experiments estimated some parameter  $\theta$   $\Rightarrow$  EXP 1 :  $\hat{\theta}_1 \pm \sigma_1$   
EXP 2 :  $\hat{\theta}_2 \pm \sigma_2$

For large sample the p.d.f.s of  $\hat{\theta}_1, \hat{\theta}_2$  become Gaussian, giving the joint probability

$$L(\theta; \hat{\theta}_1, \hat{\theta}_2) = \frac{1}{\sqrt{2\pi\sigma_1^2}} e^{-\frac{1}{2}(\hat{\theta}_1 - \theta)^2/\sigma_1^2} \cdot \frac{1}{\sqrt{2\pi\sigma_2^2}} e^{-\frac{1}{2}(\hat{\theta}_2 - \theta)^2/\sigma_2^2}$$

$$l(\theta) = -\frac{1}{2} \left[ \frac{(\hat{\theta}_1 - \theta)^2}{2\sigma_1^2} + \frac{(\hat{\theta}_2 - \theta)^2}{\sigma_2^2} \right] + C$$

$$\frac{dl}{d\theta} = 0 \Rightarrow \frac{(\hat{\theta}_1 - \theta)}{\sigma_1^2} + \frac{(\hat{\theta}_2 - \theta)}{\sigma_2^2} = 0$$

# Combining measurements...

Solving for  $\theta$ :

$$\frac{\hat{\theta}_1}{\sigma_1^2} + \frac{\hat{\theta}_2}{\sigma_2^2} = \theta \left( \frac{1}{\sigma_1^2} + \frac{1}{\sigma_2^2} \right)$$

$$\Rightarrow \hat{\theta} = \frac{\hat{\theta}_1/\sigma_1^2 + \hat{\theta}_2/\sigma_2^2}{1/\sigma_1^2 + 1/\sigma_2^2}$$

Error weighted average

Note that here  $\sigma_1, \sigma_2$  are shorthands for  $\sigma_{\hat{\theta}_1}, \sigma_{\hat{\theta}_2} \rightarrow$  errors on parameter estimations  $\hat{\theta}_1, \hat{\theta}_2$  from expt 1, 2 respectively.

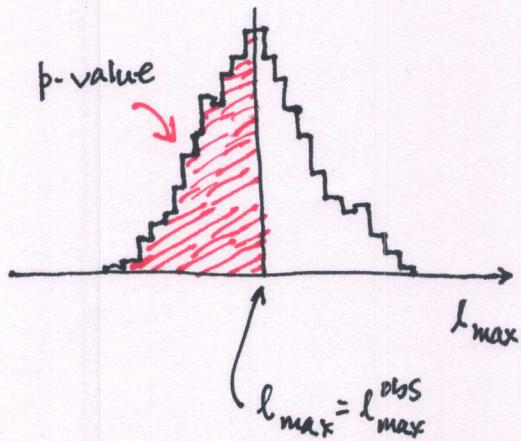
In practice we will use the estimators of the variances  $\hat{\sigma}_{\hat{\theta}_1}^2, \hat{\sigma}_{\hat{\theta}_2}^2$

estimated variance on the combined  $\hat{\theta}$

$$\hat{V}[\hat{\theta}] = \frac{1}{1/\hat{\sigma}_{\hat{\theta}_1}^2 + 1/\hat{\sigma}_{\hat{\theta}_2}^2}$$

## Goodness of fit of ML method

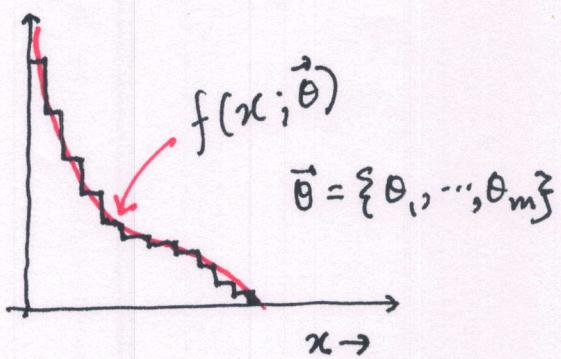
No direct way. Using Monte Carlo (MC) toy is one way



- take the estimated values of the parameters to construct p.d.f.
- generate toy dataset repeatedly.
- Estimate a p-value from distribution

One can also use methods like bootstrap

One way to visually inspect the quality of fit is to histogram the data (or do a kernel density) and compare with the fit.



For a quantitative comparison construct a statistic

$$\lambda = \frac{L(\vec{v}; \vec{n})}{L(\vec{u}; \vec{n})}$$

$$\vec{n} = \{n_1, n_2, \dots, n_N\}$$

$$\vec{v} = \{v_1, v_2, \dots, v_N\} = E[\vec{n}]$$

$$\chi^2_{\text{Mult}} = -2 \ln \lambda_M = 2 \sum n_i \ln \left( \frac{n_i}{\hat{v}_i} \right) \quad | \text{ [bin content multinomial]}$$

follows a  $\chi^2_{N-m-1}$  as  $n \rightarrow \infty$

$$\chi^2_{\text{Pois}} = -2 \ln \lambda_p = 2 \sum_{i=1}^N \left( n_i \ln \left( \frac{n_i}{\hat{v}_i} \right) + \hat{v}_i - n_i \right) \quad | \text{ [bin content Poisson]}$$

follows a  $\chi^2_{N-m}$  as  $n \rightarrow \infty$