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 (a priori) the experiment?
  - ⇒ Our data $\{t_1, t_2 \cdots, t_n\}$ is $\sim \frac{1}{\tau} e^{-t_i/\tau}$

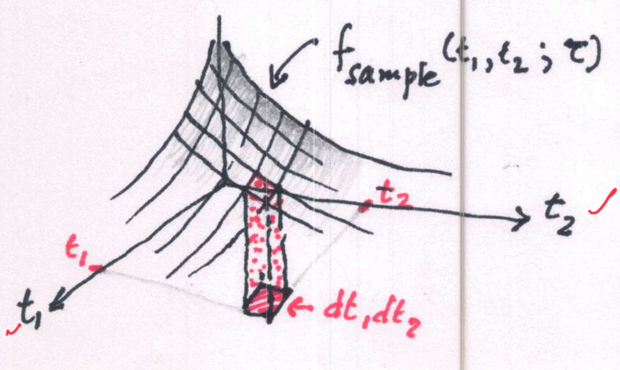
i.e. $f_{sample}(\underbrace{\{t_1, t_2, \cdots, t_n\}}_{\vec{t}}; \tau) = \frac{1}{\tau^n} \prod_{i=1}^{n} e^{-t_i/\tau}$

- What we do not know:
  - ⇒ value of τ
- What are we interested in knowing?
  - ⇒ value of τ, the parameter of interest (PoI)
- What do we know after ('a posteriori') the experiment?
  - ⇒ Measured values of t's ($t_1 = 1.2\,\mu s$, $t_2 = 3.1\,\mu s$ etc) Those are now fixed

The probability of observing the dataset we have observed is (i.e. the probability that t_1 lies between t_1 and $t_1 + dt_1$ etc.)

$$\prod_{i=1}^{n} f(t_i|\tau)\, dt_i$$

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



quite clearly, the probability of observing a dataset in the neighbourhood (the little red box $dt_1 dt_2$) of the data that we have observed depends on the unknown value of τ.

It would be reasonable to expect:

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

We will not get to the true value of τ 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.

$\boxed{\text{Important :}}$ Remember that joint probability density is a function of $\vec{x}$ for given value of parameter. But here our dataset is fixed $\{t_1, \cdots t_n\}$ have all been measured. We want to see how this quantity changes as we vary the parameter τ. To make it apparent let us define.

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

likelihood function, a function of parameter τ, equal in value with fsample

Note that :

$$\int_0^\infty L(\tau\,;\vec{t}\,)\,d\tau \neq 1 \quad \Leftarrow \quad \text{likilihood function can not be interpreted as probability density function}$$

while

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

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.

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

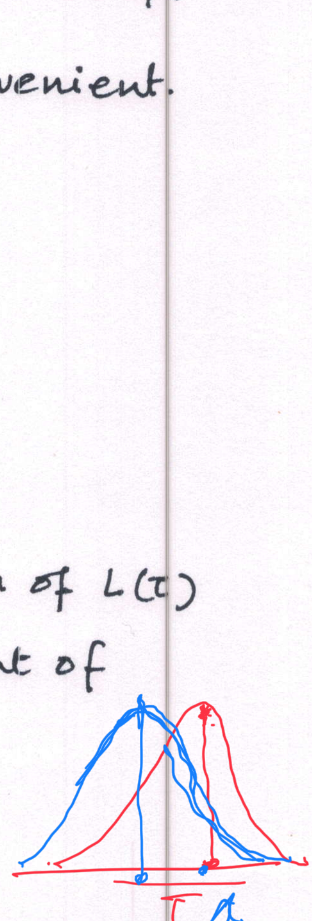
In our example

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

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

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

So, let's try $\dfrac{\partial \ell}{\partial \tau} = 0$



$$\underline{\underline{P(\tau\,|\,\underset{\vec{t}}{\vec{x}})}} \doteq \frac{P(\vec{x}\,|\,\tau)\,P(\tau)}{\text{norm.}}$$

$\uparrow \int \ldots d\tau$

$$\frac{\partial \ell}{\partial \tau} = -\frac{n}{\tau} + \frac{1}{\tau^2}\sum_{i=1}^{n} t_i = 0$$

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

In this case MLE gave consistent, unbiased estimator.

Note that if we multiply $L(\tau)$ 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,

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

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

[Casella & Berger]

Definition (MLE)

For each sample point $\vec{x}$ let $\hat{\vec{\theta}}(\vec{x})$ 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 $\hat{\vec{\theta}}(\vec{x})$

[Casella & Berger]

An useful and interesting property of MLE is invariance under transformation

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

If there is a one to one map between a and θ 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 $\lambda = \frac{1}{\tau}$ is $\hat{\lambda} = \frac{1}{\hat{\tau}}$

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

Normal MLE with both μ, σ unknown.

$$L(\underbrace{\mu, \sigma^2}_{\vec{\theta}}; \vec{x}) = \prod_{i=1}^{n} \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{1}{2}\left(\frac{x_i-\mu}{\sigma}\right)^2}$$

$$= \frac{1}{(2\pi\sigma^2)^{n/2}} \cdot e^{-\frac{1}{2\sigma^2}\left(\sum_i (x_i-\mu)^2\right)}$$

Maximization w.r.t μ, σ^2 requires

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

$$\frac{\partial l(\mu,\sigma^2)}{\partial \sigma^2} = -\frac{n}{2\sigma^2} + \frac{1}{2\sigma^4} \sum_{i=1}^{n} (x_i - \theta_1)^2$$

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

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

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

<u>Question</u> Setting $\frac{\partial l}{\partial \sigma} = 0$ we could get $\hat{\sigma}$ instead of $\widehat{\sigma^2}$. 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,

$$V[\hat{\tau}] = E[\hat{\tau}^2] - (E[\hat{\tau}])^2$$

$$= \frac{\tau^2}{n} \qquad \text{by CLT}$$

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

In practice we will calculate

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

Result of the experiment is reported as

$$\hat{\tau} = \underbrace{2.19}_{\tau_{MLE}} \pm \underbrace{0.18}_{\sqrt{\widehat{\sigma}^2}} \ \mu s.$$

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 τ.
2) Generate large no. of toy datasets $\{t_1' \ldots t_n'\}$
3) For each toy calculate $\hat{\tau}$ by MLE
4) Find the s^2 as the estimator of variance.

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

↓ computation intensive!

$\hat{\tau}_1 \leftarrow \boxed{t_1, t_2, t_3, t_4}\, n$

$\hat{\tau}_2 \leftarrow \boxed{t_2, t_3, t_2, t_4}\, n$

⋮

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

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

and

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

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

Then

$$V_\theta[\hat{\theta}] \geq \frac{\left(\frac{d}{d\theta} E[\hat{\theta}]\right)^2 = 1}{E_\theta\left(\left(\frac{\partial}{\partial\theta} \ln f(\vec{x};\theta)\right)^2\right) = \frac{n}{\tau^2}}$$

[Casella & Berger 7.3.9 P. 335

Continuing with the lifetime example.

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

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

$\therefore$ Numerator $= (1)^2 = 1$.

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

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

[$\bar{t}_n = \frac{1}{n}\sum t_i$

$$E\left[(\partial\ell/\partial\tau)^2\right] = \frac{n}{\tau^2}$$ ← Homework

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

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

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

$$= \frac{n^2}{\tau^2} + \left(\frac{n\bar{t}_n}{\tau^2}\right)^2 - 2\cdot\frac{n}{\tau}\cdot\frac{1}{\tau^2}\cdot n\bar{t}_n$$

$$E\left(\frac{\partial l}{\partial \tau}\right)^2 = E\left(\frac{n^2}{\tau^2}\right) + \frac{n^2}{\tau^4}E\left(\bar{t}_n^2\right) - \frac{2n^2}{\tau^3}E[\underbrace{\bar{t}_n}_{\tau}]$$

$$= \frac{n^2}{\tau^2} + \frac{n^2}{\tau^4}E\left[\frac{1}{n^2}\sum_i\sum_j t_i t_j\right] - \frac{2n^2}{\tau^3}\cdot\tau$$

$$= -\frac{n^2}{\tau^2} + \frac{1}{\tau^4}E\left[\sum_{i,j} t_i t_j\right]$$

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

$$2\sum_{i<j} t_i t_j \quad ({}^nC_2 \text{ terms})$$

$$E[t_i^2] = \underbrace{\int_0^\infty t_i^2\cdot\frac{1}{\tau}e^{-t_i/\tau}dt_i}_{\tau^2\int_0^\infty x^2 e^{-x}dx} \underbrace{\int\cdots\int\prod_{j\neq i}\frac{1}{\tau}e^{-t_j/\tau}d^{n-1}t_j}_{1}$$

$$\int_0^\infty x^2 e^{-x}dx = \Gamma(3) = 2$$

$$= 2\tau^2$$

$$E[t_i t_j] = \underbrace{\int t_i \frac{1}{\tau}e^{-t_i/\tau}dt_i}_{\tau} \underbrace{\int t_j\cdot\frac{1}{\tau}e^{-t_j/\tau}dt_j}_{\tau} \prod_{k\neq i,j}\int\cdots\int\frac{1}{\tau}e^{-t_k/\tau}d^{n-2}t_k$$

$$= \tau^2$$

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

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

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.

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

Under certain (general enough) conditions

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

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

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

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

When $\vec{\theta} = \{\theta_1, \ldots \theta_n\}$ $\quad \frac{\partial^2 \ell}{\partial \theta^2} \Rightarrow \frac{\partial^2 \ell}{\partial \theta_i \partial \theta_j}$

For unbiased, efficient estimator, ~~then~~

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

In practice

$$(\widehat{V^{-1}})_{ij} = -\frac{\partial^2 \ell}{\partial \theta_i \partial \theta_j}\Bigg|_{\vec{\theta} = \hat{\vec{\theta}}}$$

$\Leftarrow$ Hessian ~~calculat~~ computed numerically (HESSE in MINUIT)

Variance of MLE by graphical method

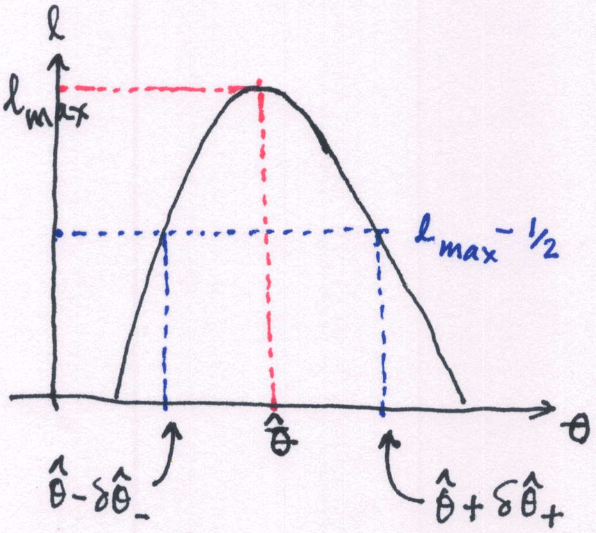
$\ell[\hat{\theta}]$ and its derivatives at $\theta = \hat{\theta}$ can be computed.

$$\ell(\theta) = \underbrace{\ell(\hat{\theta})}_{\ell_{max}} + \cancelto{0}{\frac{\partial \ell}{\partial \theta}\bigg|_{\theta=\hat{\theta}}} (\theta - \hat{\theta}) + \frac{1}{2}\frac{\partial^2 \ell}{\partial \theta^2}\bigg|_{\theta=\hat{\theta}} (\theta - \hat{\theta})^2 + \cdots$$

In large sample limit $L(\theta) \to$ Gaussian

$\Rightarrow \ell(\theta)$ becomes parabola

$\Rightarrow$ symmetric error



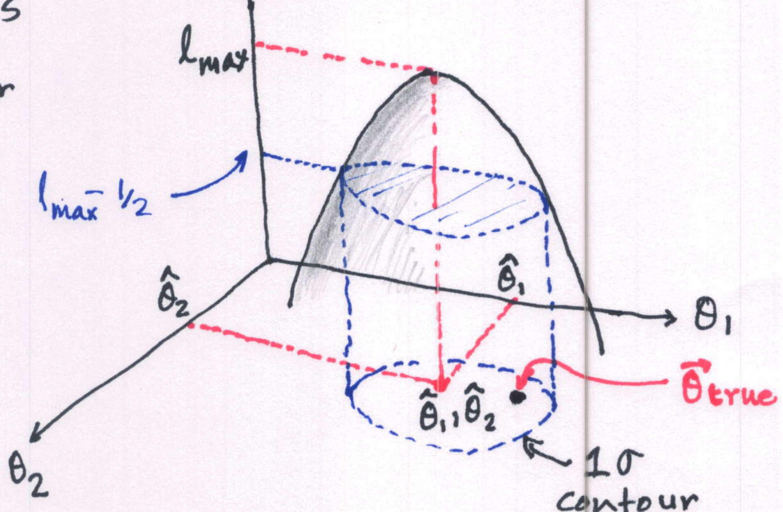
$$\ell(\hat{\theta} + \hat{\sigma}_{\hat{\theta}}) = \ell_{max} + \frac{1}{2}\left(-\frac{1}{\hat{\sigma}_{\hat{\theta}}}\right)^2 \cdot \hat{\sigma}_{\hat{\theta}}^2$$

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

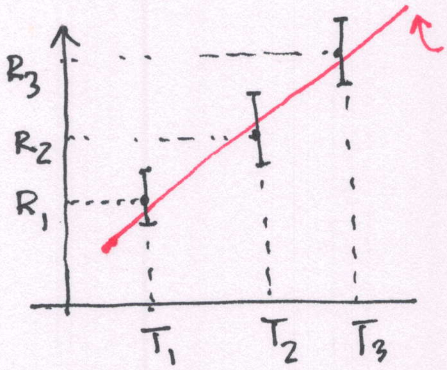
Note 1: In multidimension this $\ell(\theta)$ becomes a (hyper) surface $\ell(\vec{\theta}) = \ell(\theta_1, \cdots \theta_n)$ and $\ell_{max} - 1/2$ points become a contour

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

$$\ell_{max}(n \to \infty) = -\frac{1}{2}\frac{1}{(1-\rho^2)}\left[\frac{(\theta_1 - \hat{\theta}_1)^2}{\sigma_{\hat{\theta}_1}^2} + \frac{(\theta_2 - \hat{\theta}_2)^2}{\sigma_{\hat{\theta}_2}^2} - 2\rho\left(\frac{\theta_1 - \hat{\theta}_1}{\sigma_{\hat{\theta}_1}}\right)\left(\frac{\theta_2 - \hat{\theta}_2}{\sigma_{\hat{\theta}_2}}\right)\right]$$



χ^2 and likelihood



$R_t(T) = \theta_1 + \theta_2 T$ (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.) $\sigma_1, \sigma_2, \sigma_3$

$$L(\theta_1, \theta_2; R_1, R_2, R_3) = f(R_1; \theta_1, \theta_2) \cdot f(R_2; \theta_1, \theta_2) \cdot f(R_3; \theta_1, \theta_2)$$

$$= \frac{1}{\sqrt{2\pi\sigma_1^2}} e^{-\frac{1}{2}\frac{(R_1-R_{t1})^2}{\sigma_1^2}} \cdot \frac{1}{\sqrt{2\pi\sigma_2^2}} e^{-\frac{1}{2}\frac{(R_2-R_{t2})^2}{\sigma_2^2}} \cdots$$

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

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

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

this is χ^2 of 3 measurements.

Whenever error is normal

$$\ell(\vec{\theta}; \vec{x}) = -\frac{1}{2}\chi^2$$

So log likelihood maximization is the same as χ^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)

$$L(\nu, \vec{\theta}) = \underbrace{\frac{\nu^n}{n!} e^{-\nu}}_{\substack{\text{poisson} \\ \text{prob of } n \\ \text{observations.}}} \underbrace{\prod_{i=1}^{n} f(\vec{x}_i ; \vec{\theta})}_{\text{usual likelihood}}$$

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

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

e.g. in a collision run $\nu = \sigma L \varepsilon$

σ, x_i → both depend on parmeters like mass, coupling

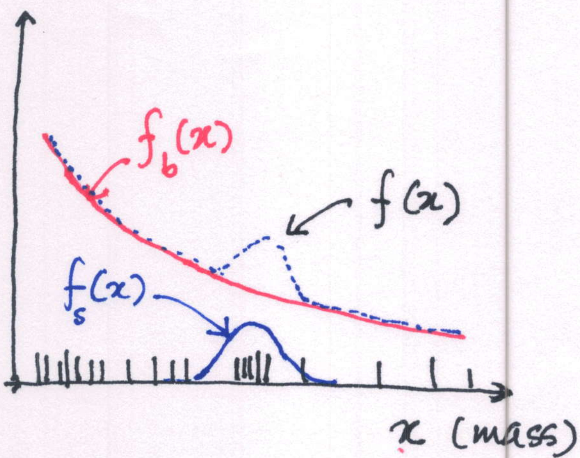
In general will reduce stat error ✓

Sample with mixed signal and background

$n = n_s + n_b$

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

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



f_s , f_b known

Interested in s

Extended likelihood...

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

or $\nu = s + b$

$$\ell(\nu, s, b, \vec{\theta}) = -n \ln \nu - \nu + \sum_{i=1}^{n} \ln f(x_i; \vec{\theta})$$

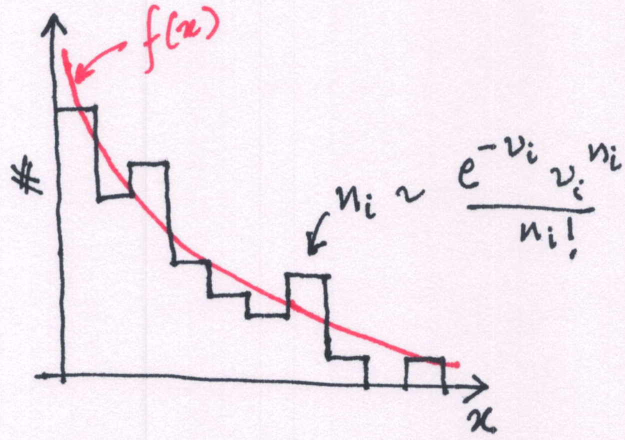
$$= -\nu + \sum_{i=1}^{n} \ln(\nu . f(x_i))$$

$$= -(s+b) + \sum_{i=1}^{n} \ln\left\{(s+b)\left(\frac{s}{s+b} f_s + \frac{b}{s+b} f_b\right)\right\}$$

$$= -(s+b) + \sum_{i=1}^{n} \ln\{s f_s + b f_b\}$$

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

MLE of binned data



Histogram with N bins,
n total events

or multinomial distributed.

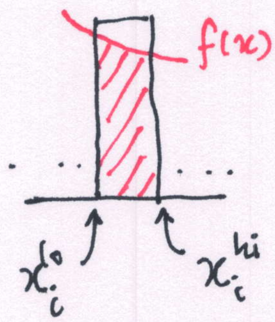
We have to fit $f(x;\vec{\theta})$

n_i = no. of events in bin i

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

$$E[n_i] = \nu_i$$

fit parameters in here.



$$\nu_i = n \int_{x_i^{lo}}^{x_i^{hi}} f(x)\, dx$$

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

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

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

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

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

If we take bin content Poisson distributed

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

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

Note : Bin with 0 entry is not a problem

Combining experiments with likelihood

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

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

▷ Combined likelihood:

$$L(\vec{\theta}; \vec{x}, \vec{y}) = L_x(\vec{\theta}; \vec{x})\, L_y(\vec{\theta}; \vec{y})$$

$$= \prod_{i=1}^{n} f_x(x_i; \vec{\theta}) \prod_{j=1}^{m} f_Y(y_j); \vec{\theta})$$

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

▷ Suppose two experiments estimated some parameter θ

$$\Rightarrow \quad \begin{array}{l} \text{EXP 1}: \ \hat{\theta}_1 \pm \sigma_1 \\ \text{EXP 2}: \ \hat{\theta}_2 \pm \sigma_2 \end{array}$$

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

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

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

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

Combining measurements...

Solving for θ!

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

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

Error weighted average

Note that here σ_1, σ_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.

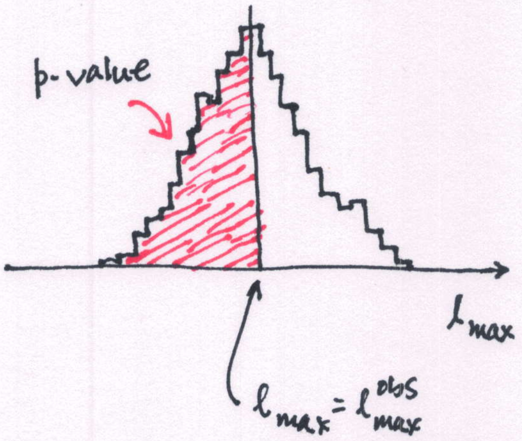
In practice we will use the estimators of the variances $\widehat{\sigma}^2_{\hat{\theta}_1}$, $\widehat{\sigma}^2_{\hat{\theta}_2}$

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

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

Goodness of fit of ML method

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



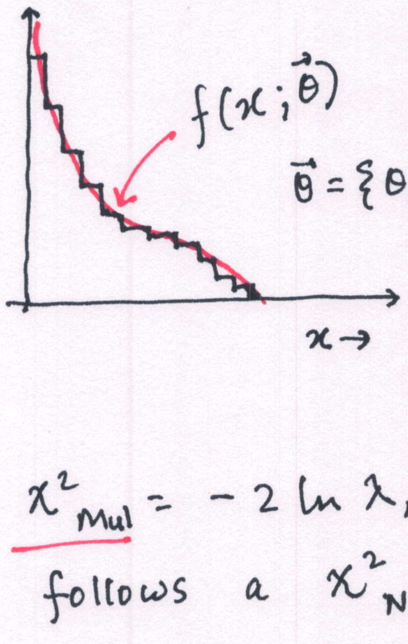
→ take the estimated values of the parameters to construct p.d.f.

→ 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

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

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

$$\vec{\nu} = \{\nu_1, \nu_2 \cdots \nu_N\} = E[\vec{n}]$$

$$\chi^2_{Mul} = -2\ln\lambda_M = 2\sum n_i \ln(n_i/\hat{\nu}_i)$$ [bin content multinomial]

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

$$\chi^2_{Pois} = -2\ln\lambda_P = 2\sum_{i=1}^{N}\left(n_i \ln\left(\frac{n_i}{\hat{\nu}_i}\right) + \hat{\nu}_i - n_i\right)$$ [bin content Poisson

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