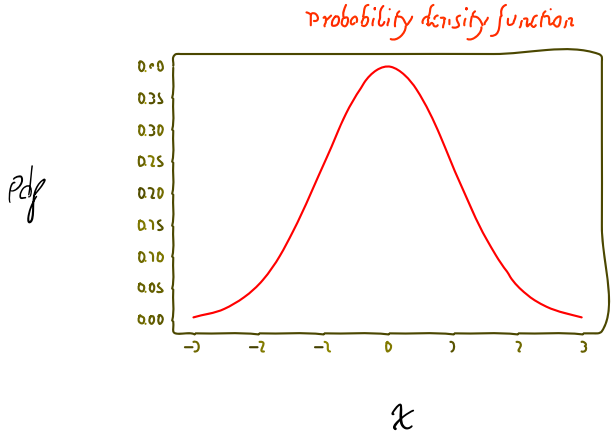
1. Basic Principles

Continuous Random Variable

Characterized by pdf $f_X(x;\theta)$



- Probability to be inside an interval

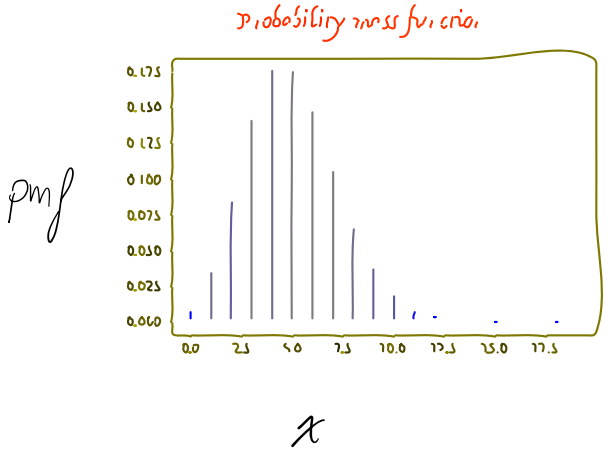
$$P(x \le X \le x+dx) = f_X(x;\theta)\,dx$$

- Expected Value

$$\mathbb{E}[X] = \int x f_X(x;\theta)\,dx$$

Discrete Random Variable

Characterized by pmf $P_X(x,\theta)$



- $P_X(x,\theta)$ = prob. that $X = x$
- $\mathbb{E}[X] = \sum_i x_i P_X(x_i;\theta)$

- $\text{var}[X] = \mathbb{E}\left[(X - \mathbb{E}[X])^2\right]$
- cdf: $F(t,\theta) = P(X \le t;\theta)$

Population Parameters

Expected value	$\mathbb{E}[X] = \mu$	• $\mathbb{E}[M] = \mathbb{E}[X]$
Variance	$\text{var}[X] = \mathbb{V}[X] = \sigma^2 = \mathbb{E}[(X-\mu)^2]$	⇝ • $\mathbb{V}[X] = \mathbb{E}[X^2] - (\mathbb{E}[X])^2$
Covariance	$\text{cov}(X,Y) = \sigma_{xy} = \mathbb{E}[(X-\mu_x)(Y-\mu_y)]$	• $\mathbb{V}[M] = \frac{\mathbb{V}[X]}{n}$
Moments	$\mathbb{E}[X^n] = m_n$	

Independence

X and Y are independent if $\boxed{f_{XY}(x,y) = f_X(x) f_Y(y)}$

Remark: Dependence ≠ Correlation

Correlation is measured by $\operatorname{cor}(X,Y) = \mathbb{E}[XY] - \mathbb{E}[X]\mathbb{E}[Y]$

$\rightsquigarrow \operatorname{cor}(X,Y) = 0$ if uncorrelated $\rightsquigarrow$ Independence $\Rightarrow$ Uncorrelation

BUT NOT CONTRARY

Empirical Quantities

Sample mean $\quad M = \frac{1}{n}\sum_{i=1}^{n} X_i$

Sample median: assuming $X_1 < X_2 < \ldots < X_n$

- median $= \frac{X_{n/2} + X_{1+n/2}}{2}$ (n even)
- median $= X_{(n+1)/2}$ (n odd)

Sample variance $\quad S_n^2 = \frac{1}{n}\sum_{i=1}^{n}(X_i - M)^2 = \frac{1}{n}\sum_{i=1}^{n} X_i^2 - M^2$ biased

$S_{n-1}^2 = \frac{1}{n-1}\sum_{i=1}^{n}(X_i - M)^2$ unbiased

Sample covariance matrix $\quad Q = \frac{1}{n-1}\sum_{i=1}^{n}(\vec{X}_i - \vec{M})(\vec{X}_i - \vec{M})^T$

Functions of Random Variables

Functions of random variables are also random variables

$Z = g(X_1, \ldots, X_n)$ — empirical quantity; g: function; $X_1, \ldots, X_n$: sample

The mean is

$$\mathbb{E}[Z] \simeq g(\mathbb{E}[X_1], \ldots, \mathbb{E}[X_n]) + \frac{1}{2}\sum_{i=1}^{n} \frac{\partial^2 g}{\partial X_i^2} \operatorname{var}[X_i] + \sum_{i,j} \frac{\partial^2 g}{\partial X_i \partial X_j} \operatorname{cov}(X_i, X_j)$$

Variance

$$\operatorname{var}[Z] \simeq \sum_{i=1}^{n}\sum_{j=1}^{n} \frac{\partial g}{\partial X_i}\frac{\partial g}{\partial X_j} \operatorname{cov}(X_i, X_j)$$

- If X_i are iid $\Rightarrow \frac{\partial g}{\partial X_i} = \frac{\partial g}{\partial X_j}$

$\Rightarrow \operatorname{cov}(X_i, X_j) = \operatorname{cov}(X_n, X_i) = \mathbb{V}(X_i)$

Binomial and Multinomial Distributions

Binomial

$$P(K; n, p) = \binom{n}{K} p^K (1-p)^{n-K}$$

$$= \frac{n!}{K!(n-K)!} p^K (1-p)^{n-K}$$

"Describe the probability of having K successes in n independent, identical trials where each trial has only 2 possible outcomes"

Properties:

- $\mathbb{E}[K] = np$
- $\text{var}[K] = np(1-p)$

if $n=1$ ⇒ Bernoulli

Multinomial

$$P(n_1, \dots, n_m; n, p_1, \dots, p_m) = \frac{n!}{n_1! n_2! \dots n_m!} p_1^{n_1} \dots p_m^{n_m}$$

↝ m: number of possible results in a trial
n_i: number of results of type i ($i \in [1, m]$), $\sum_i n_i = n$
p_i: probability that result in a trial is of type i

Properties:

- $\mathbb{E}[n_i] = np_i$
- $\text{var}[n_i] = np_i(1-p_i)$
- $\text{cov}(n_i, n_j) = -np_ip_j$

Example [edit]

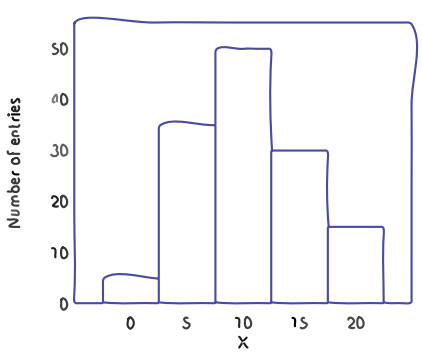
Suppose that in a three-way election for a large country, candidate A received 20% of the votes, candidate B received 30% of the votes, and candidate C received 50% of the votes. If six voters are selected randomly, what is the probability that there will be exactly one supporter for candidate A, two supporters for candidate B and three supporters for candidate C in the sample?

Note: Since we're assuming that the voting population is large, it is reasonable and permissible to think of the probabilities as unchanging once a voter is selected for the sample. Technically speaking this is sampling without replacement, so the correct distribution is the multivariate hypergeometric distribution, but the distributions converge as the population grows large in comparison to a fixed sample size[1].

$$\Pr(A = 1, B = 2, C = 3) = \frac{6!}{1!2!3!}(0.2^1)(0.3^2)(0.5^3) = 0.135$$

Filling an histogram from a FIXED sample is a MULTINOMIAL PROCESS — N FIXED

$$\Rightarrow \boxed{P(\text{histogram}) = \frac{n!}{\prod_i n_i!} \prod_i p_i^{n_i}}$$



N = # bins

↝ n_i: content of bin i
p_i: probability to be in bin i

$$p_i = \int_{x \in \text{bin}} f_x(x)dx$$

↝ Bins are not independent ⇒ What if n is not fixed?

The probability changes as follows:

$$P(n_1, \cdots, n_m; n, p_1, \cdots, p_m) \rightarrow P(n, n_1, \cdots, n_m; \nu, p_1, \cdots, p_m)$$

Parameters of the distribution of n

What is $P(n, n_1, \cdots, n_m)$?

$$\rightarrow P(n, n_1, \cdots, n_m) = \underbrace{P(n_1, \cdots, n_m; n)}_{\text{multinomial}} \times P(n)$$

Often n is a Poisson random variable → See next slide

Poisson Distribution

$$P(n; \nu) = \frac{\nu^n}{n!} e^{-\nu}$$

"Probability of a given number of events occurring in a fixed interval of time or space if those events occur with a known constant mean rate"

Properties

- $\mathbb{E}[n] = \nu$
- $\text{var}[n] = \nu$

Poisson = limit of binomial when $n \to \infty$ and $p \to 0$

Histogram Summary

n Fixed: histogram ~ multinomial

$n \sim$ Poisson: histogram ~ product of poisson in each bin

MULTIDIMENSIONAL SAMPLES

↝ Case where N measurements are performed of M different variables

$$\vec{x}_i = \{\vec{x}_1, \vec{x}_2, \dots, \vec{x}_N\} \quad \text{with} \quad \begin{cases} \vec{x}_1 : x_1^{(1)}, x_1^{(2)}, \dots, x_1^{(M)} \\ \vdots \\ \vec{x}_N : x_N^{(1)}, x_N^{(2)}, \dots, x_N^{(M)} \end{cases}$$

Def

$$\mathrm{cov}(x,y) = \frac{1}{N}\sum_{i=1}^{N}(x_i - \bar{x})(y_i - \bar{y}) = \overline{xy} - \bar{x}\,\bar{y}$$

Correlation Factor $\qquad \rho_{xy} = \dfrac{\mathrm{cov}(x,y)}{\sigma_x \sigma_y} \qquad$ with $\quad -1 \le \rho_{xy} \le 1$

- $\rho = 0 \Rightarrow$ x and y are uncorrelated ($\neq$ independent)
- $\rho_{xy} = \pm 1 \Rightarrow$ x and y are fully correlated

Covariance Matrix

$$C = \begin{pmatrix} \mathrm{cov}(x_1,x_1) & \cdots & \mathrm{cov}(x_1,x_N) \\ \vdots & \mathrm{cov}(x_i,x_j) & \vdots \\ \mathrm{cov}(x_N,x_1) & \cdots & \mathrm{cov}(x_N,x_N) \end{pmatrix} = \begin{pmatrix} \sigma_1^2 & \cdots & \rho_{1N}\sigma_1\sigma_N \\ \vdots & \rho_{ij}\sigma_i\sigma_j & \vdots \\ \rho_{N1}\sigma_N\sigma_1 & \cdots & \sigma_N^2 \end{pmatrix}$$

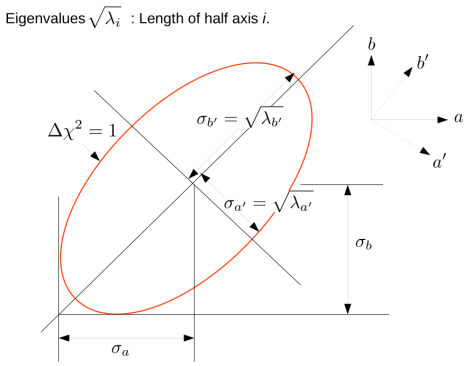
Correlation Matrix

$$\rho = \begin{pmatrix} 1 & \cdots & \rho_{1N} \\ \vdots & 1 & \vdots \\ \rho_{N1} & \cdots & 1 \end{pmatrix}$$

The error ellipse

From the covariance matrix we can find the eigenvalues λ_i

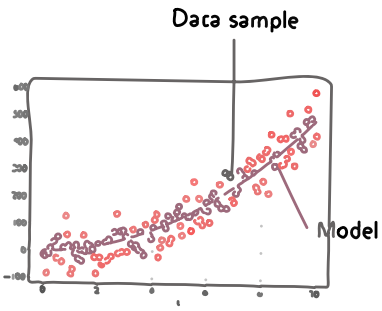
⇘ They are the variance of x_i $= \sigma_i^2$

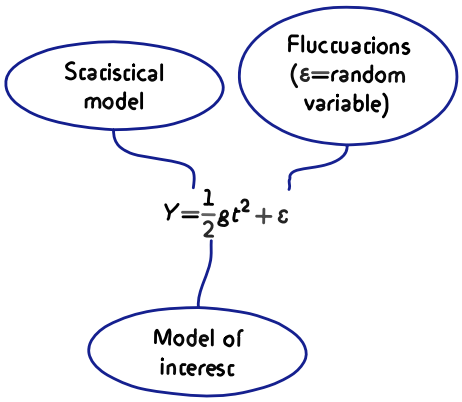


θ? → $\theta = \frac{1}{2}\tan^{-1}\left(\frac{2\rho\sigma_x\sigma_y}{\sigma_y^2 - \sigma_x^2}\right)$

2. Statistical Model

It's the mathematical expression describing your observation





Statistical model for iid samples

Suppose you have a sample $X=(X_1,\dots,X_n)$

The statistical model can be written as

$$f_{X_1\dots X_n}(x_1,\dots,x_n;\theta)=f_{X_1}(x_1|x_2\dots,x_n;\theta)\times f_{X_2}(x_2|x_3,\dots,x_n)\times\dots\times f_{X_n}(x_n;\theta)$$

For IID variables

$$f_{X_1\dots X_n}(x_1,\cdots,x_n;\theta)=f_X(x_1;\theta)\times f_X(x_2;\theta)\times\cdots\times f_X(x_n;\theta)$$

$$\Rightarrow \boxed{f_{X_1\dots X_n}(x_1,\cdots,x_n;\theta)=\prod_{i=1}^{n} f_X(x_i;\theta)}$$

ND If X_i's are normally distributed

$$\text{stat. mod} = \prod_{i=1}^{n}\frac{1}{\sqrt{2\pi}\sigma}\exp\left(\frac{-(x_i-\mu)^2}{2\sigma^2}\right)$$

$$=\frac{1}{(\sqrt{2\pi}\sigma)^n}\exp\left(-\sum_{i=1}^{n}\frac{(x_i-\mu)^2}{2\sigma^2}\right)$$

Likelihoods

$$\mathcal{L}(\theta,\vec{x})=\prod_{i=1}^{N}\mathcal{L}_i(\theta_i,\vec{x}) \qquad \text{where } \mathcal{L}_i=f_i$$

Extended Likelihood Function

If the size of the sample is not constant but follows Poisson Distribution

$$n \sim \text{Pois}(\nu)$$

In such cases the likelihood function has to be extended

$$\mathcal{L}_{ext}(\theta, \nu; x, n) = \frac{\nu^n e^{-\nu}}{n!} \times \mathcal{L}(\theta, \vec{x})$$

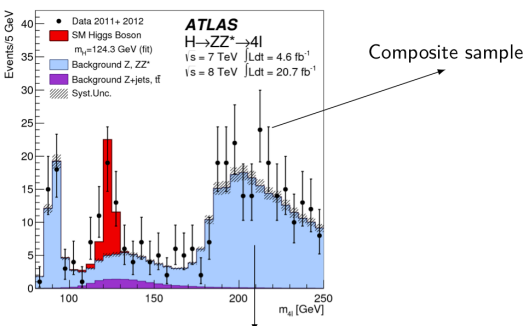
Composite Samples

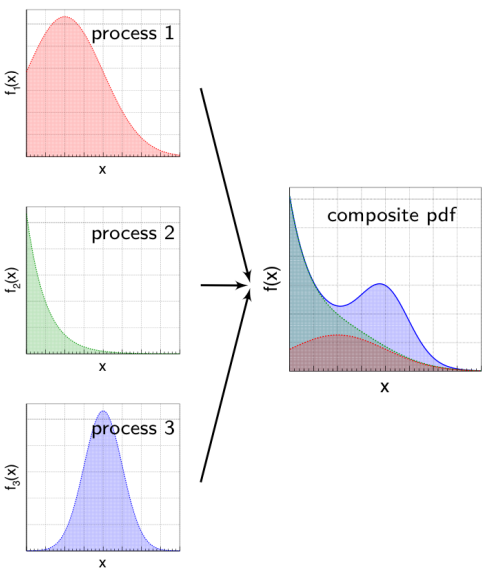
In realistic cases, samples are often composite

↝ Composite Sample = sample in which events can come from different origins

Example: sample = (E_1, E_2, .., E_{10})

E_1: source, E_2: bkg, E_{10}: source





- **Composite pdf**:

$$f(x; \{\mu_p\}, \theta) = \frac{\sum_{p=1}^{P} \mu_p f_p(x; \theta)}{\sum_{p=1}^{P} \mu_p}$$

where:

- p: process index
- μ_p: expected number of elements in sample from process p
- $f_p(x; \theta)$: pdf for process p

- **Remark:** the μ_p's are often unknown
  - → Determining them can be one of the objective

General extended unbinned $\mathcal{L}$

General extended unbinned likelihood:

$$\mathcal{L}_{\text{ext}}(\{\mu_p\}, \theta; \{x_i\}, n) = \underbrace{\frac{(\sum \mu_p)^n}{n!} e^{-\sum \mu_p}}_{\text{extended term}} \times \prod_{i=1}^{n} \underbrace{\frac{\sum_{p=1}^{P} \mu_p f_p(x_i; \theta)}{\sum_{p=1}^{P} \mu_p}}_{\text{composite pdf}}$$

Assuming the x_i's are iid

Treatment of nuisance parameters

Suppose your model has one parameter of interest and other parameter you do not care about

Two cases must be distinguished:

1. **Nuisance parameters perfectly known**: formalism presented previously works perfectly fine
2. **Nuisance parameters not perfectly known**: formalism presented previously should be extended

→ New term in likelihood

$$\mathcal{L}_{full} = \mathcal{L}(\theta_1, \theta_2, \dots, \theta_N) \times \underbrace{g(\theta_1, \theta_2, \dots, \theta_N)}_{\text{New const. term}}$$

Detailed Example

Lets consider a Poissonian process with background (Measurement of radioactivity)

$$\text{Statist. Model}: \quad \mathcal{L}(s,b) = \frac{(s+b)^n}{n!} e^{-(s+b)}$$

- → n: number of measured events (observable)
- → s: number of signal events (parameter of interest)
- → b: number of background events (nuisance parameter)

You can get a value for b ⟹ But + error : $b \pm \sigma_b$

Full likelihood:

- Incomplete Stat. Model: $\mathcal{L}(s,b) = \frac{(s+b)^n}{n!} e^{-(s+b)}$
- b determined in an auxiliar poissonian measurement $\mathcal{L}_{aux}(b) = \frac{(\alpha b)^{n_b}}{n_b!} e^{-(\alpha b)}$
- Full likelihood $\mathcal{L}_{FM} = \mathcal{L}(s,b) \times \mathcal{L}_{aux}(b) = \frac{(s+b)^n}{n!} e^{-(s+b)} \times \frac{(\alpha b)^{n_b}}{n_b!} e^{-(\alpha b)}$

3 Inference Generalities

Are what allow us to gain, from a SAMPLE and a STATISTICAL MODEL, some knowledge about the parameter of intrest

↝ Knowledge acquisition process is called inference

Two big schools of thought:

FREQUENTIST BAYESIAN

Different interpretation of probability

INTERPRETATION 1: Probability = Frequency of occurence

$$P(A) = \lim_{n \to \infty} \frac{n(A)}{n}$$ where $\begin{cases} A: \text{some event} \\ n(A): \text{number of times event } A \text{ occurs} \\ n: \text{total number of events} \end{cases}$

Example: dice roll

INTERPRETATION 2: Probability = Degree of belief on occurence

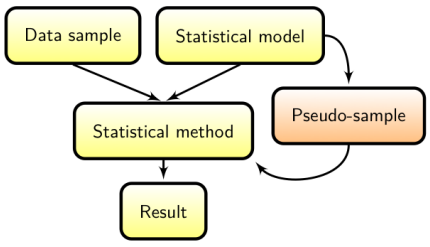
Often used for unique events (for which Frequentist calculation impossible)
↝ Probability tomorrow will rain

Frequentist

- ☐ Use frequentist interpretation of probability
- ☐ Model parameters are fixed, the only thing that can have a probability associated to it is the data

$$P(x;\theta)$$

- ☐ Inference based on notion of **pseudo-sample** or **pseudo-dataset**



Bayesian

- ☐ Use probabilities as degree of belief
- ☐ Model parameters have a probability

$$P(\theta;x)$$

→ They are not random variable as the x's (better qualified as **uncertain variables**)
→ Even if notation is the same, this probability doesn't have the same frequentist meaning

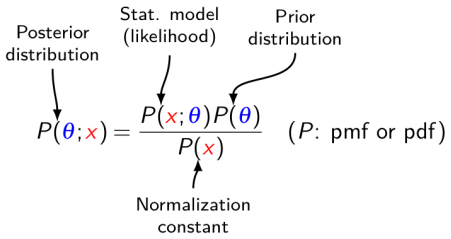
- ☐ Inference of parameter of interest based on this probability
- ☐ How is this probability computed ?

$$P(\theta;x) = \frac{P(x;\theta)P(\theta)}{P(x)} \text{ (Bayes thm)}$$

More on bayesian approach

In bayesian approach, inference proceeds as follows:

1. Write statistical model
2. Find out what is best prior
3. Compute postmor
4. Conclude from postenor



Bayes Theorem

For 2 arbitrary events A and B

$$P(A \cap B) = P(A|B)P(B)$$

$$P(B \cap A) = P(B|A)P(A)$$

As $P(A \cap B) = P(B \cap A)$, one has $P(A|B)P(B) = P(B|A)P(A)$

$$\Rightarrow \boxed{P(A|B) = \frac{P(B|A)P(A)}{P(B)}}$$

Simple explenation:

Scenario: You want to know the probability of having a rare disease.

- Initial belief (prior probability): 1% of the population has the disease (P(A) = 0.01).
- Medical test accuracy: The test is 95% accurate in detecting the disease (P(B|A) = 0.95) but has a 10% false positive rate (P(B|¬A) = 0.10).

Using Bayes' Theorem:

1. Calculate the overall probability of getting a positive test result (P(B)) using the law of total probability. $P(B) = P(B|A) \cdot P(A) + P(B|-A) \cdot P(-A)$
2. Apply Bayes' Theorem to find the updated probability of having the disease given a positive test result (P(A|B)).

Result: Even with a positive test result, the updated probability of having the disease is around 8.76%. This demonstrates how Bayes' Theorem adjusts your beliefs based on new evidence while considering the accuracy and potential uncertainties in the test.

4. Parameter Estimation

Frequentist Approach

↝ Let X be gaussian with $\sigma = 1$ and mean μ

Measurement providing a sample of 4 values: (x_1, x_2, x_3, x_4)

Problem: How we determine μ from these x_i?

Possible Solution: Take the sample mean $M = \frac{1}{4}\sum_{i=1}^{4} x_i$

Doing this we hope that we get a value that is close to the true value

$$\mu \simeq M$$

Questions

1. What does it mean when we say that an estimate is "close" to the true value $\hat{\theta} \simeq \theta$?

2. How are estimators determined in general?

Answer 1

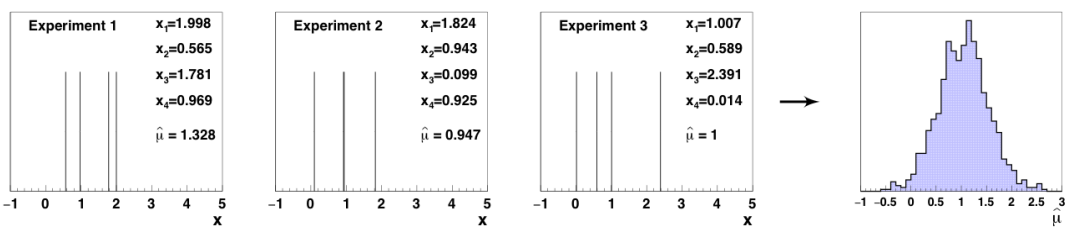
It means nothing → θ is not known

No sense random variable ≃ constant

In order to determine whether an estimator is good or not, need to get back to Fondamental frequentist interrogation:

What would one get if we repeat the same measurement many times?

→ We would get a distribution of $\hat{\theta}$ characterized by a MEAN & VARIANCE ..



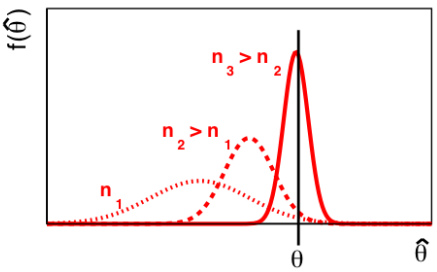
Frequentist Properties of estimators

Three main properties: CONSISTENCY, BIAS, EFFICIENCY

CONSISTENCY

An estimator is consistent when it converges in probability towards the true value as $n \to \infty$

$$P(|\hat{\theta} - \theta| > \varepsilon) \xrightarrow[n \to \infty]{} 0 \qquad \forall \varepsilon > 0$$

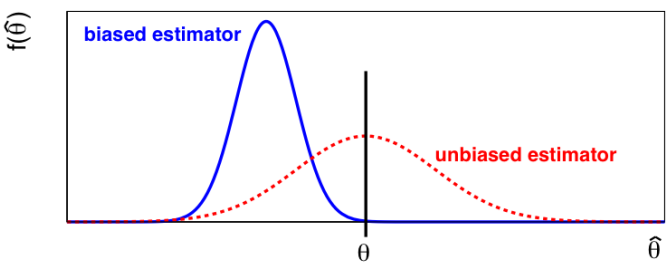


BIAS

An estimator is unbiased when its expectation value is equal to the true value (for all n)

$$\mathbb{E}[\hat{\theta}] = \theta$$

The bias of an estimator is defined as $b = \mathbb{E}[\hat{\theta}] - \theta$



EFFICIENCY

An estimator is efficient when its variance is equal to the RCF bound

RCF bound:

$$\mathrm{var}[\hat{\theta}] \geq \frac{\left(1 + \frac{\partial b}{\partial \theta}\right)^2}{\mathbb{E}\left[\left(\frac{\partial \ln \mathcal{L}}{\partial \theta}\right)^2\right]}$$

RFC bound

sometimes $\geq \frac{1 + \frac{\partial b}{\partial \theta}}{\mathbb{E}\left[\frac{\partial^2 \ln \mathcal{L}}{\partial \theta^2}\right]}$

Ideal estimators are consistent, unbiased and efficiency

↪ Not always possible ⇒ Should at least be consistent

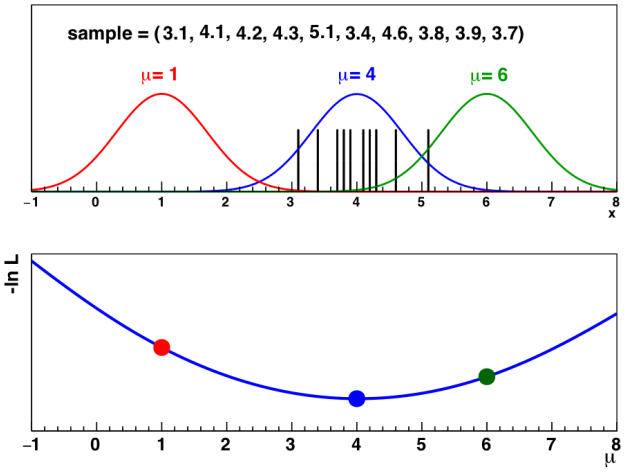
How estimators are determined in general?

1 Maximum Likelihood

↪ Let X be a gaussian r.v. with known σ and unknown mean μ

Suppose we make a measurement for the determination of μ providing 10 values $(x_1, \dots, x_{10})$

Of the 3 gaussian distribution presented the BLUE ONE has the minimum value of $-\ln\mathcal{L}$



The Maximum Likelihood estimator $\hat{\theta}_{ML}$ is the parameter that MAXIMIZES the Likelihood function

$$\left.\frac{\partial \mathcal{L}}{\partial \theta}\right|_{\theta=\theta_{ML}} = 0 \qquad \text{et} \qquad \left.\frac{\partial^2 \mathcal{L}}{\partial \theta^2}\right|_{\theta=\hat{\theta}_{ML}} < 0$$

equivalently

$$\left.\frac{\partial(-\ln\mathcal{L})}{\partial \theta}\right|_{\theta=\hat{\theta}_{ML}} = 0 \qquad \text{et} \qquad \left.\frac{\partial^2(-\ln\mathcal{L})}{\partial \theta^2}\right|_{\theta=\hat{\theta}_{ML}} > 0$$

Properties of estimators

The ML method provides estimators with nice properties

- Functional Invariance $\hat{\xi}_{ML} = \xi(\hat{\theta}_{ML})$
- ML estimators are consistent
- If an unbiased and efficient estimator exist $\Rightarrow$ it's variance is

$$\mathrm{var}[\hat{\xi}_{ML}] = \frac{-(\partial\xi/\partial\theta)^2}{\left.\frac{\partial^2 \ln \mathcal{L}}{\partial\theta^2}\right|_{\theta=\theta_{ML}}}$$

- ML method provides POINT ESTIMATION

$\rightsquigarrow$ It's possible to have an error bar in addition to a single value?

$\Rightarrow$ YES! GRAPHICAL METHOD

Graphical Method

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

but $\left.\frac{\partial(-\ln\mathcal{L})}{\partial\theta}\right|_{\theta=\hat{\theta}_{ML}} = 0$ thus

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

$\rightsquigarrow$ If estimator unbiased and efficient : $\mathrm{var}[\hat{\theta}_{ML}]^{-1} = \left.-\frac{\partial^2}{\partial\theta^2}\ln\mathcal{L}\right|_{\theta=\hat{\theta}_{ML}}$

Thus

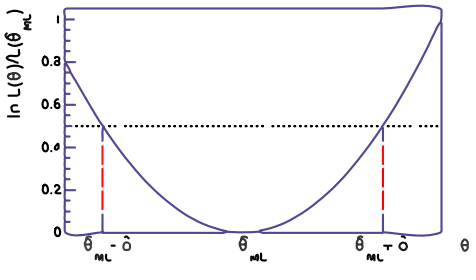
$$-\ln \mathcal{L}(\theta) \simeq -\ln \mathcal{L}(\hat{\theta}_{ML}) + \frac{(\theta-\hat{\theta}_{ML})^2}{2\,\mathrm{var}[\hat{\theta}_{ML}]}$$

Conclusion

$$\theta = \hat{\theta}_{ML} \pm \sqrt{\mathrm{var}[\hat{\theta}_{ML}]}$$

Then $-\ln \mathcal{L}(\theta) \simeq -\ln \mathcal{L}(\hat{\theta}_{ML}) + \frac{1}{2}$

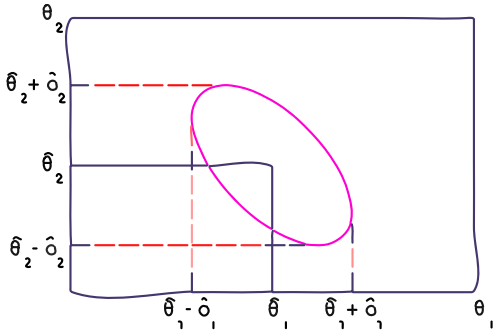
- This leads to the following graphical method:



Note: doing this we build a so-called "confidence interval" for the parameter of interest

→ We'll come back to this notion later

The graphical method also works with 2 parameters of interest



- Using this method we can:
  - → Estimate uncertainties on the 2 parameters
  - → Estimate correlation between the two estimates

2. Least Squares Method

LSM is useful when one is intrested in dependence of one variable with one or more other variables

Notations

Y_i : observation i

$m_i(\theta) = \mathbb{E}[Y_i]$ expected value of observation i

V: covariance matrix of the Y_i's

$$V = \begin{pmatrix} \sigma_1^2 & \cdots & \operatorname{cov}(Y_1, Y_n) \\ & \ddots & \\ \vdots & \operatorname{cov}(Y_i, Y_j) & \vdots \\ & & \ddots \\ \operatorname{cov}(Y_n, Y_1) & \cdots & \sigma_n^2 \end{pmatrix}$$

Method

Least Squares estimators are values that minimize $\chi^2(\theta) = (Y - m)^T V^{-1} (Y - m)$

Thus $\left.\frac{\partial^2 \chi}{\partial \theta}\right|_{\theta = \hat{\theta}_{LS}} = 0$ et $\left.\frac{\partial^2 \chi^2}{\partial \theta^2}\right|_{\theta = \hat{\theta}_{LS}} > 0$

⇝ We will consider independent observations

$$\chi^2(\theta) = \sum_{i=1}^{n} \frac{(Y_i - m_i(\theta))^2}{\sigma_i^2}$$

Least squared and Max likelihood method

For gaussian observations, the likelihood is

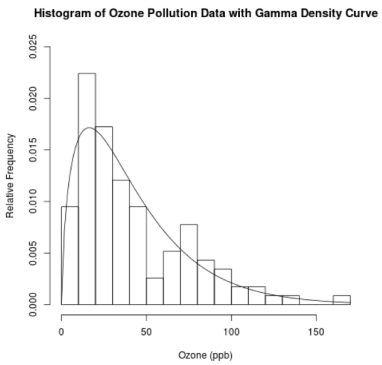
$$\mathcal{L} = e^{-\frac{1}{2}(y-m)^T V^{-1}(y-m)}$$

For independent observations

$$\mathcal{L} = \prod_{i=1}^{n} e^{-\frac{(y_i - m(x_i;\theta))^2}{2\sigma_i^2}}$$

We see that $-2\ln\mathcal{L} = \chi^2$ ⇒ Minimizing χ^2 is equivalent to maximizing the likelihood

Least Squared Method : binned sample



- **Hypothesis:** the Y_i's are independent and $Y_i \sim \text{Pois}(m_i)$

  $\Rightarrow \sigma_i = \sqrt{m_i}$

- **Remark:** m_i depends on θ

  $$m_i(\theta) = \nu p_i(\theta) = \nu \int_{C_i} f_X(x;\theta)\mathrm{d}x$$

  (ν=total expected number of events=$\mathbb{E}\left[\sum Y_i\right]$)

  $\Rightarrow \sigma_i(\theta) = \sqrt{m_i(\theta)}$

⇒

The χ^2 writes:

$$\boxed{\chi^2 = \sum_{i=1}^{n} \frac{(Y_i - m_i(\theta))^2}{m_i(\theta)}} \quad \text{(Pearson's } \chi^2\text{)}$$

which is often simplified to ("modified LS method")

$$\boxed{\chi^2 = \sum_{i=1}^{n} \frac{(Y_i - m_i(\theta))^2}{Y_i}} \quad \text{(Neyman's } \chi^2\text{)}$$

Remark: approximation $\sigma_i(\theta) = \sqrt{m_i(\theta)} \simeq \sqrt{Y_i}$ often done in various contexts

→ Equivalent to using for $\sigma(\theta)$ not the true value (unknown) but an estimate:

$$\hat{\sigma}_i(\theta) = \sqrt{Y_i}$$

Bayesian way of estimating parameters

Based on the posterior distribution:

Posterior distribution | Stat. model (likelihood) | Prior distribution

$$P(\theta;x) = \frac{P(x;\theta)P(\theta)}{P(x)} \quad (P\text{: pmf or pdf})$$

Normalization constant

$$\Rightarrow \quad P(\text{model} \mid \text{data}) = \frac{P(\text{data} \mid \text{model})\, P(\text{model})}{P(\text{data})}$$

Once the posterior $f(\theta, x)$ is found, parameters can be estimated using either the

Mean: $\hat{\theta} = \int \theta f(\theta|x) \mathrm{d}\theta$

Mode: $\hat{\theta} = \max_{\theta} f(\theta|x)$

Median: $F_\theta(\hat{\theta}) = \int_{-\infty}^{\hat{\theta}} f(\theta|x)\mathrm{d}\theta = 1/2$

Any other location measurement quantity

↝ Impact of prior on final result decreases as sample size increases.

2D Example

Poissonian measurement with signal and background

$$P(N|s,b) = \frac{(s+b)^N}{N!} e^{-(s+b)}$$

↝ b is measured and it's $b = b_0 \pm \sigma$ and you take the following normal prior

$$g(b|b_0,\sigma) = \frac{1}{\sqrt{2\pi}\,\sigma} e^{-\frac{(b-b_0)^2}{2\sigma^2}}$$

→ Suppose s is totally unknown prior to the measurement and you take a uniform prior: $g(s) \propto$ constant

2D posterior:

$$f(s,b|N) \propto P(N|s,b)\, \underbrace{g(b|b_0,\sigma)\, g(s)}_{\text{Priors}}$$

$$\propto \frac{(s+b)^N}{N!} e^{-(s+b)} \times \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(b-b_0)^2}{2\sigma^2}}$$

Also bayesian inference offers an easy way to describe how knowledge increases as measurements are performed

$$g(\theta) \xrightarrow{\times \mathcal{L}} f(\theta|x)$$

Suppose you make two measurements with likelihoods $\mathcal{L}_1$ and $\mathcal{L}_2$

→ After first measurement: $f_1(\theta|x_1) \propto \mathcal{L}_1 \times g(\theta)$

→ After second measurement: we could use

$$f_2(\theta|x_2) \propto \mathcal{L}_2 \times g(\theta)$$

but better to use knowledge acquired from first measurement:

$$f_2(\theta|x_2, x_1) \propto \mathcal{L}_2 \times f_1(\theta|x_1) \propto \mathcal{L}_2 \times \mathcal{L}_1 \times g(\theta)$$

→ Posterior of first measurement used as prior in second one

Generalization to n measurements:

$$f_n(\theta|x_n, \ldots, x_1) \propto \mathcal{L}_n \times \mathcal{L}_{n-1} \times \ldots \times \mathcal{L}_1 \times g(\theta)$$

→ As the number of measurement increases, the likelihood term "wins" and the prior becomes more and more subdominant in final result