PhD Training in Statistics

G.Wilquet

1. Recall on probability
2. Special distributions
3. Recall on statistics
4. Sampling distributions
5. Hypothesis tests
6. Estimation
7. Maximum likelihood
8. Least squares
9. Confidence levels in pathological cases
10. Monte-Carlo simulation

I– Recall of general notions of probability theory

Basic axioms of the theory of probability

1/ $0 \leq P(x) \leq 1$

if x certainly true: $P(x) = 1$

if x certainly false: $P(x) = 0$

2/ Addition and exclusion

$$P(x \cup y) = P(x) + P(y) - P(x \cap y)$$

if x and y are exclusif: $P(x \cap y) = 0$

$$P(x \cup y) = P(x) + P(y)$$

3/ Multiplication and independence

$$P(x \cap y) = P(x) \times P(y \mid x) = P(y) \times P(x \mid y)$$

if x and y are independent: $P(x \mid y) = P(x)$

$$P(y \mid x) = P(y)$$

$$P(x \cap y) = P(x) \times P(y)$$

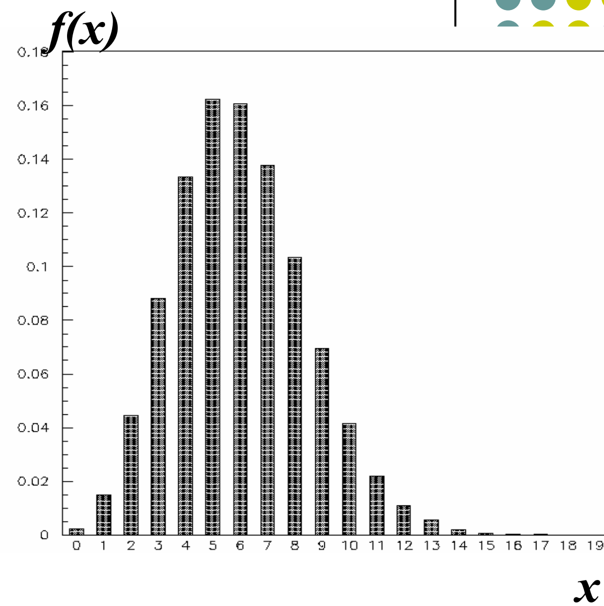
Density Probability Function (PDF)

Discrete variable

- k possible values $x_1, x_2, \ldots . x_k$ for x
- $p_i, i = 1, k$ the probability of occurence of x_i

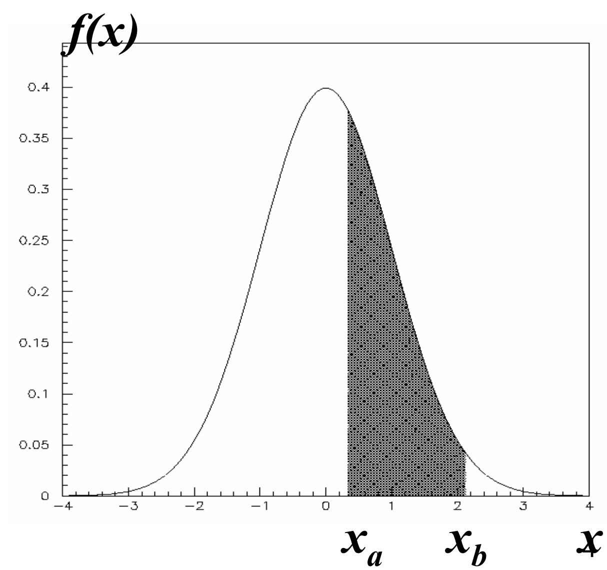
$$f(x_i) = p_i \qquad \sum_{i=1}^{k} p_i = 1$$

$$P(x_i \le x < x_j) = \sum_{k=1,}^{j} p_j$$



Continuous variable

- $f(x_i) = lim_{\Delta x \to 0} P(x_i \le x < x_i + \Delta x) / \Delta x$
- $f(x_i)\,dx = P([x_i, x_i + dx])$ $\qquad \int_{-\infty}^{\infty} f(x)\,dx = 1$
- $P(x_a \le x < x_b) = \int_{x_a}^{x_b} f(x)\,dx$



Change of variable – conservation of probability

- $f(x)$

  bijection $y = y(x)$

  $f'(y)$ **?**

- **Conservation of probability:**

  $f(x)\,dx = f'(y)\,dy$

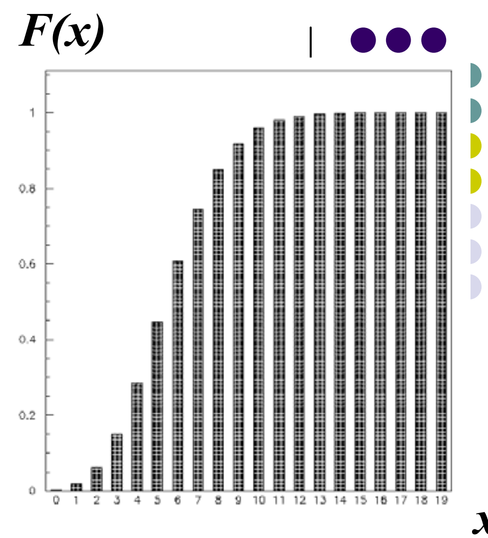
$$f'(y) = f(x)\left|\frac{dx}{dy}\right|$$

- ***n* variables**

$$f'(y_1, y_2, \ldots y_n) = |J|\, f(x_1, x_2, \ldots x_n)$$

$$J_{ij} = \frac{\partial x_i}{\partial y_j}$$

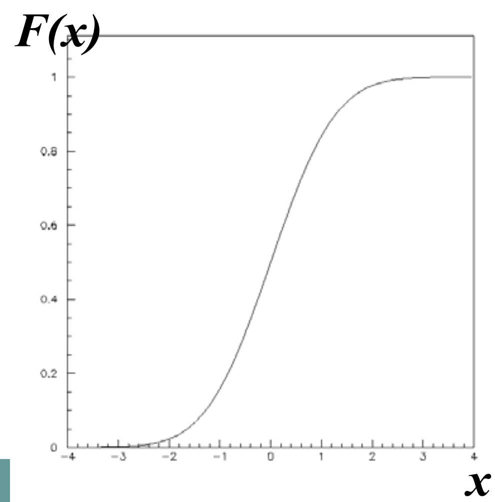
Distribution Function



Discrete variable

$$F(x_i) = \sum_{j=1}^{i} f(x_j) = \sum_{j=1}^{i} p_j \quad i = 1,n$$

Continuous variable



$$F(x) = \int_{-\infty}^{x} f(y)\,dy$$

$$dF(x) = f(x)\,dx$$

PDF of *F*(*x*) is uniform on [0,1]

$$F(x) = \int_{x}^{\infty} f(y)\,dy \qquad \text{or} \qquad dF(x) = f(x)\,dx$$

$$PDF(F(x)) = f(x)\left|\frac{dx}{dF(x)}\right| = \frac{f(x)\,dx}{f(x)\,dx} \equiv 1$$

Moments – Characteristic function

Non-centred moments

$$\mu_k = E\left[x^k\right] = \int_{-\infty}^{\infty} x^k \; f(x)\,dx$$

mean $\mu_1 = \mu = \int_{-\infty}^{\infty} x \; f(x)\,dx$

Centred moments

$$m_k = E\left[(x-\mu)^k\right] = \int_{-\infty}^{\infty} (x-\mu)^k \; f(x)\,dx$$

variance $m_2 = \sigma^2 = \int_{-\infty}^{\infty} (x-\mu)^2 \; f(x)\,dx$

standard deviation : $\sigma = \sqrt{\sigma^2}$

Characteristic Function

Fourier transform of the PDF
Expendable as a series of the moments

$$\phi(t) = \int_{-\infty}^{\infty} e^{itx} f(x)\,dx = \sum_{k=0}^{\infty} \frac{\mu_k (it)^k}{k!}$$

Equivalence between f(x), F(x), ϕ(t) and μ_k

Equivalent information are provided by

density probability function	$f(x)$
distribution function $F(x)$	$f(x) = d_x F(x)$
caractéristique function $\phi(t)$	$f(x) = \frac{1}{2\pi} \int_{-\infty}^{\infty} e^{-itx} \phi(t)\, dt$
non-null moment μ_k	$f(x) = \frac{1}{2\pi} \int_{-\infty}^{\infty} e^{-itx} \sum_{k=0}^{\infty} \frac{\mu_k (it)^k}{k\,!}\, dt$

Join PDF of several variables – dependence and correlation

$$f(x_1, x_2, \ldots x_n)\, dx_1 dx_2 \ldots dx_n = f(\underline{x})\, d\underline{x} = P([\underline{x}, \underline{x} + d\underline{x}])$$

normalisation $\int_\Omega f(\underline{x})\, d\underline{x} = 1$

means $\mu_i = E[x_i]$

variances $\sigma_i^2 = E\left[(x_i - \mu_i)^2\right]$

covariances $\sigma_{ij} = E\left[(x_i - \mu_i)(x_j - \mu_j)\right] = E\left[x_i \cdot x_j\right] - E[x_i]\; E[x_j]$

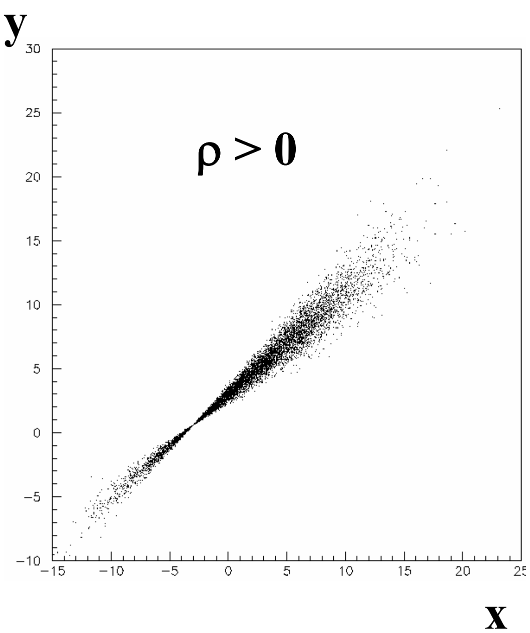
correlation coefficient $\rho_{ij} = \dfrac{\sigma_{ij}}{\sigma_i \sigma_j} = \dfrac{E\left[(x_i - \mu_i)(x_j - \mu_j)\right]}{\sqrt{E\left[(x_i - \mu_i)^2\right] E\left[(x_j - \mu_j)^2\right]}}$

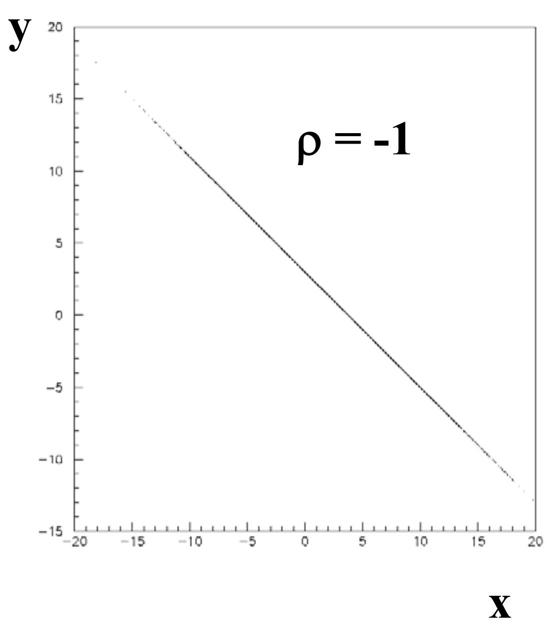
independence between variables x_i and x_j

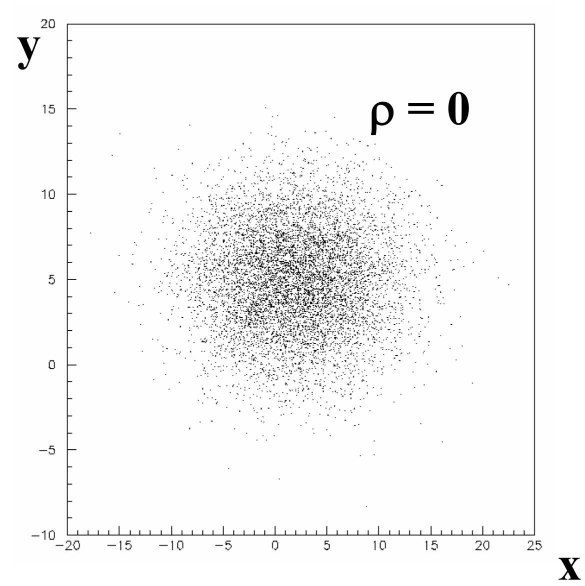
factorisation $f(\underline{x}) = f_i(x_i, \ldots x_{k \neq j}) \cdot f_j(x_j, \ldots x_{k \neq i})$

$$\Rightarrow \quad E\left[x_i \cdot x_j\right] = E[x_i]\; E[x_j] \Rightarrow \quad \sigma_{ij} = \rho = 0$$

Correlation coefficient







Marginal and conditional PDF of several variables

- **Marginal : projections of** $f(x_1, x_2)$ **on** x_1 **and** x_2

$$h_1(x_1) = \int_{-\infty}^{\infty} f(x_1, x_2)\,dx_2 \quad \text{and} \quad h_2(x_2) = \int_{-\infty}^{\infty} f(x_1, x_2)\,dx_1$$

if x_1 **and** x_2 **are independent:** $f(x_1, x_2) = f_1(x_1)\, f_2(x_2) \;\Rightarrow\; h_i(x_i) \equiv f(x_i)$

- **Conditional : PDF of** x_1 **for a given value** $x_2 = x_2^0$ **and conversaly**

$$g_1(x_1 \mid x_2 = x_2^0) = \frac{f(x_1, x_2^0)}{h_2(x_2^0)} \quad \text{and} \quad g_2(x_2 \mid x_1 = x_1^0) = \frac{f(x_1^0, x_2)}{h_1(x_1^0)}$$

- **if** x_1 **and** x_2 **are independent :** $f(x_1, x_2) = f_1(x_1)\; f_2(x_2)$ **et** $h_i(x_i) \equiv f_i(x_i)$

$$\Rightarrow g_1(x_1 \mid x_2 = x_2^0) \equiv f_1(x_1) \equiv h_1(x_1)$$

$$\Rightarrow g_2(x_2 \mid x_1 = x_1^0) \equiv f_2(x_2) \equiv h_2(x_2)$$

II – Special Distributions

Bernoulli process: binomial et multinomial

Poisson process: Poisson, exponential

Gauss process: gaussian or normal

Bernoulli process

Multinomial PDF

- *k* possible results
- Result *i* occurs with a probability p_i $\quad \sum_{i=1}^{k} p_i = 1$
- Probability to observe $\underline{r} = r_1, r_2, \ldots r_k$ résults of type $1, 2, \ldots k$

  on a total of $n = \sum_{i=1}^{k} r_i$ trials

$$f\left(\underline{r} \mid \underline{p}, n\right) = \frac{n!}{\prod_{i=1}^{k} r_i!} \prod_{i=1}^{k} p_i^{r_i}$$

- Mean $\mu_i = n\, p_i$
- Variance $\sigma_i^2 = n\, p_i \left(1 - p_i\right)$
- Covariance $\sigma_{ij} = -n\, p_i\, p_j$

Binomial PDF

- Probability to observe

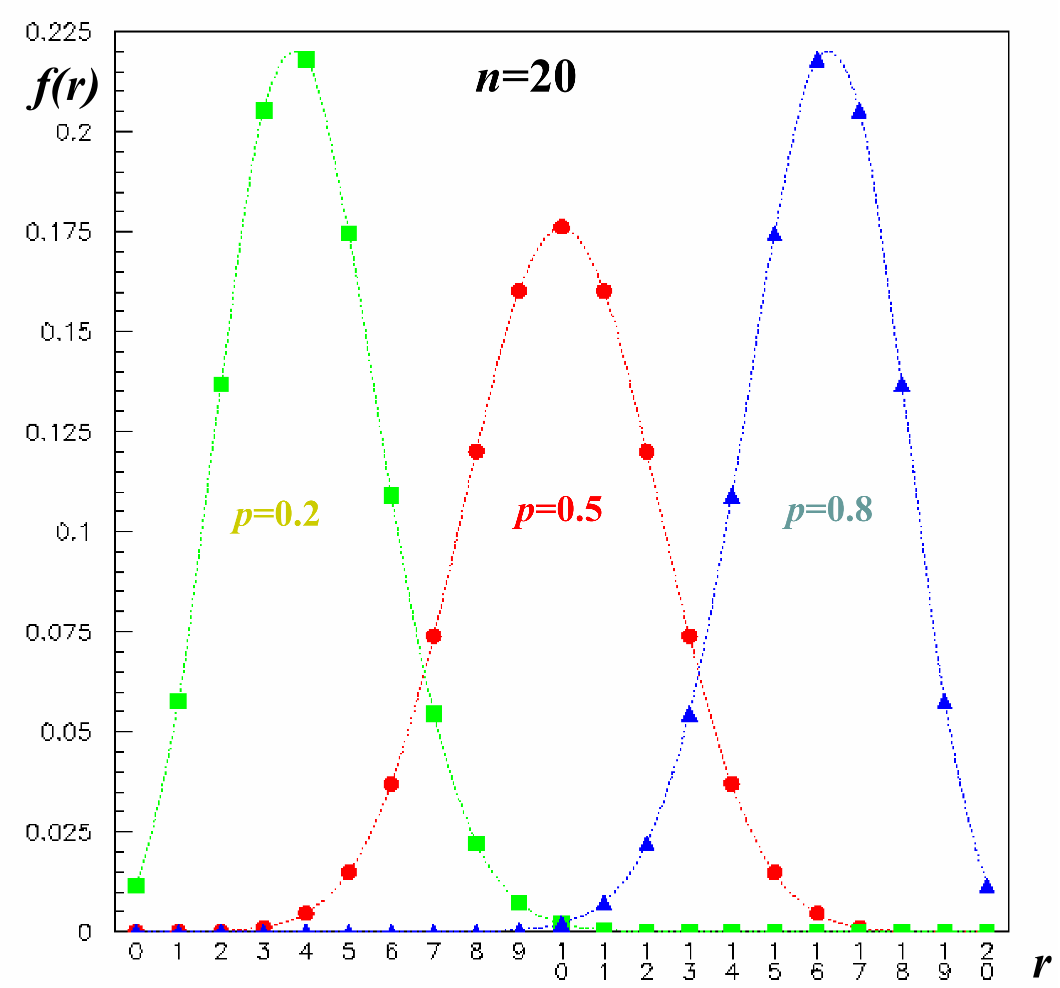
  $r_1 = r$ succeses that have probability $p_1 = p$

  $r_2 = n - r$ failurs that have probability $p_2 = 1 - p$

$$f\left(r \mid p, n\right) = \frac{n!}{r!\left(n-r\right)!} p^r \left(1-p\right)^{n-r}$$

- $\rho = -1$

Binomial PDF



Poisson process

Defining properties of the Poisson PDF

Process occurs or not on a small interval Δx : 0 or 1 success

$$P_1([x, x+\Delta x]) = \Delta x / \beta$$

$$P_0([x, x+\Delta x]) = 1 - \Delta x / \beta$$

- **do not depend on x**

β **mean interval between two succeses**

$1/\beta$ **density of succes per unitinterval**

Poisson PDF as the limit of the binomial PDF

$$n \to \infty$$

$$p \to 0$$

$np \to \mu$ **finate**

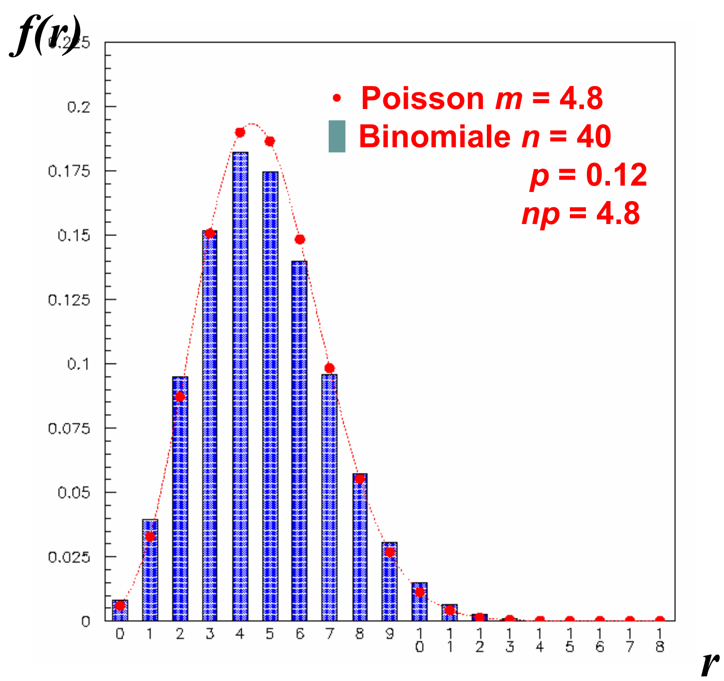
$$lim_{n\to\infty} \; n! = \sqrt{2\pi n} \; n^n \; e^{-n}$$

Poisson Distribution

- Probability of occurence of r (integer) successes on an interval x
  given a mean number of succes μ (real) $\quad \mu = \frac{x}{\beta}$

$$f(r \mid \mu) = \frac{1}{r!}\mu^r e^{-\mu}$$

- Mean μ
- Variance $\sigma^2 = lim_{p \to 0}\, np(1-p) = np = \mu$



Poisson Distribution : demonstration of the PDF

$$P_0(\Delta x) = 1 - P_1(\Delta x) = 1 - \frac{\Delta x}{\beta}$$

$$P_0(x+\Delta x) = P_0(x) \times P_0(\Delta x) = P_0(x) - \frac{\Delta x}{\beta} P_0(x)$$

$$\frac{P_0(x+\Delta x) - P_0(x)}{\Delta x} = -\frac{1}{\beta} P_0(x)$$

$$\left.\begin{array}{l} \dfrac{dP_0(x)}{dx} = -\dfrac{1}{\beta} P_0(x) \\ P_0(0) = 1 \end{array}\right\} \rightarrow P_0(x) = e^{-x/\beta}$$

$$P_r(x+\Delta x) = P_{r-1}(x) \times P_1(\Delta x) + P_r(x) \times P_0(\Delta x) = P_{r-1}(x)\frac{\Delta x}{\beta} + P_r(x)\left(1 - \frac{\Delta x}{\beta}\right)$$

$$\frac{P_r(x+\Delta x) - P_r(x)}{\Delta x} = -\frac{1}{\beta}\left(P_r(x) - P_{r-1}(x)\right)$$

$$\left.\begin{array}{l} \dfrac{dP_r(x)}{dx} = -\dfrac{1}{\beta}\left(P_r(x) - P_{r-1}(x)\right) \\ P_0(x) = e^{-x/\beta} \end{array}\right\} \rightarrow P_r(x) = \frac{1}{r!}\left(\frac{x}{\beta}\right)^r e^{-x/\beta}$$

$$\boxed{P_r(x) = f(r \mid \mu) = \frac{1}{r!}\mu^r e^{-\mu} \qquad avec \quad \mu = \frac{x}{\beta}}$$

Exponential Distribution

Poissonnian process

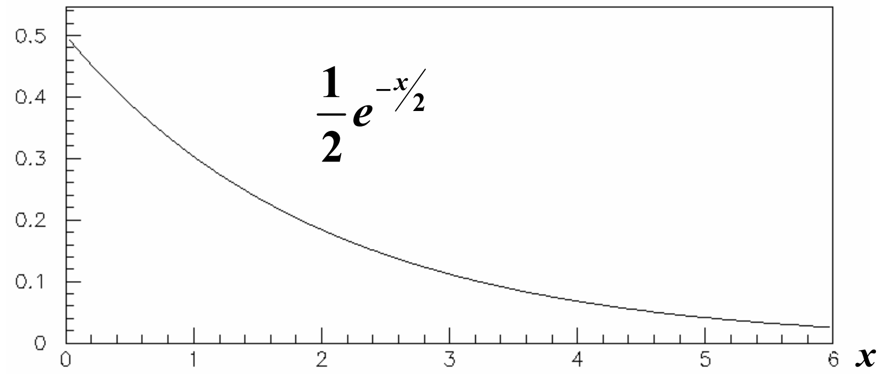
β = average interval between two successes

Probability of a first succès on interval $[x, x+dx]$.

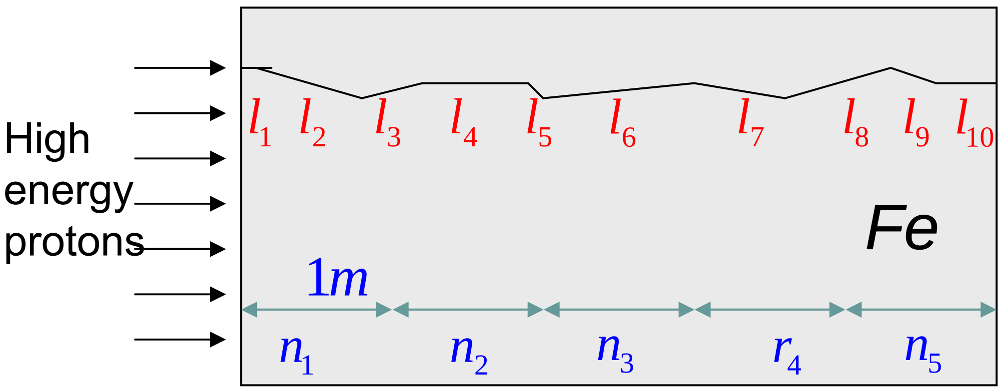
$$P_0([0,x]) \times P_1([x,x+dx]) = \frac{1}{0!}\left(\frac{x}{\beta}\right)^0 e^{-x/\beta} \times \frac{dx}{\beta} = \frac{1}{\beta} e^{-x/\beta} dx$$

$$f(x \mid \beta) = \frac{1}{\beta} e^{-x/\beta}$$

- **Mean $\mu = \beta$**
- **Variance $\sigma^2 = \beta^2$**



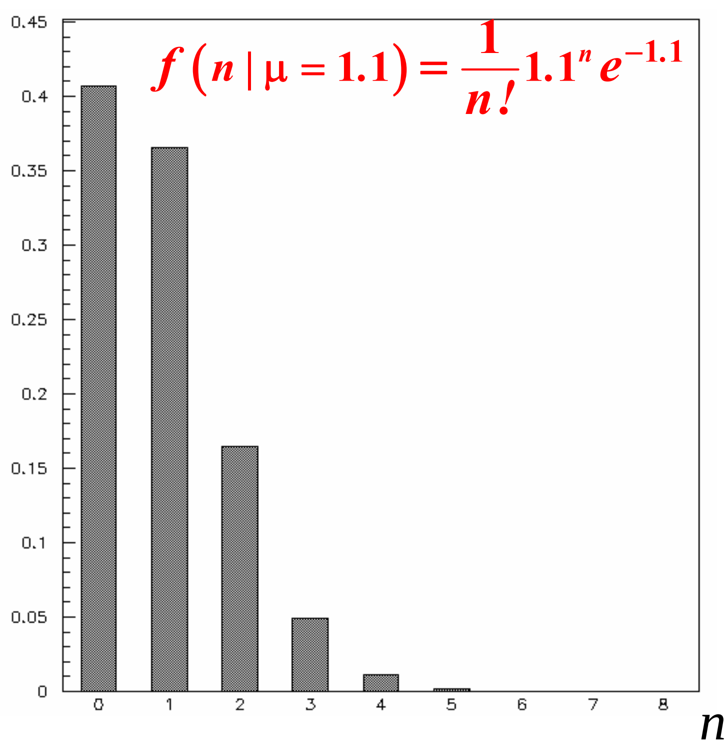
Example of relation between exponential and Poisson PDF

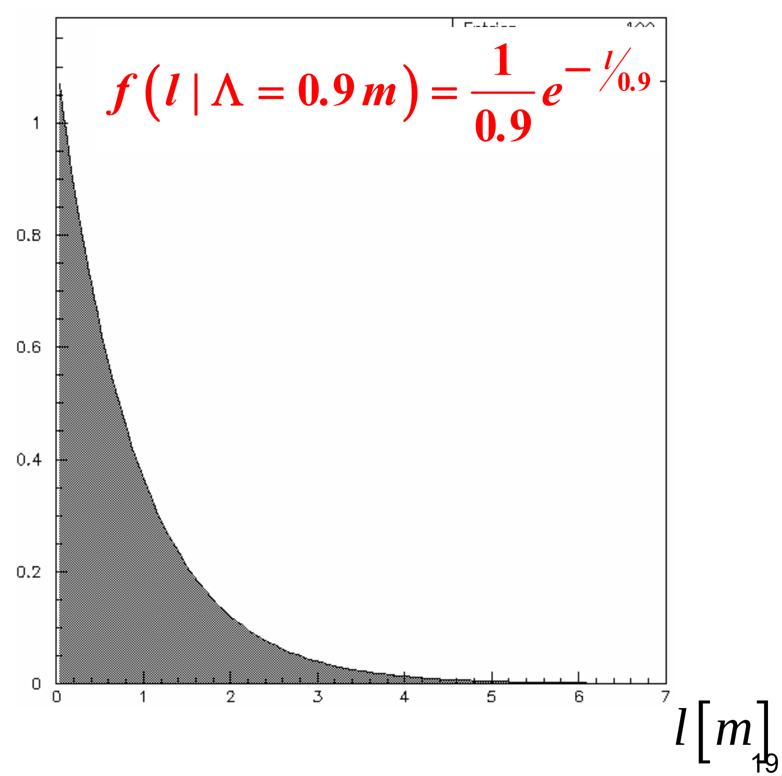


$\sigma_{pFe}^{elast} \approx 0.110 \cdot 10^{24}\, cm^2$ for $E \geq 50\, GeV$

$\Lambda = \dfrac{A_{Fe}}{\sigma_{pFe}^{elast} N_A \rho_{Fe}} \approx 0.9\, m$: mean of $\underline{l}$

$\mu = \dfrac{1\, m}{0.9\, m} \approx 1.1$:mean of $\underline{n}$





Convolutions of Poisson and Binomial PDF

r_1 and r_2 : independent from Poisson PDF of means μ_1 and $\mu_{2.} \Rightarrow$

$r_1 + r_2$: Poisson PDF of mean $\mu_1 + \mu_{2.}$

r : binomial PDF of success probability p and number of trials n.
n : Poisson PDF of mean μ. $\Rightarrow$

r : Poisson PDF of mean $p\mu$.

Convolution example

Plastic scintillator strip of thikness $0.1\ mm$ seen by a photosensor and crossed by a MIP

Mean photon emission : $\mu_\gamma = 200$

Distance to photocathode : $l = 1m$

Scintillator absorption length : $\lambda=0.5\ m$

Photocathode efficiency : $QE = 0.2$

PDF of n_γ **at emission: Poisson of mean** $\mu_\gamma = 200$

Fraction of photons reaching the photocathode $p_\lambda = \int_1^\infty \frac{1}{0.5} e^{-x/0.5} dx = 0.14$

PDF of n_γ **at photocathode input: Poisson of mean** $p_\lambda \times \mu_\gamma = 0.14 \times 200 = 28$

PDF of n_e **at photocathode output: Poisson of mean** $QE \times p_\lambda \times \mu_\gamma = 0.14 \times 200 \times 0.2 = 5.6$

Efficiency = probability to observe at least 1 e

$\varepsilon = 1 - P(0 \mid 5.6) = 1 - e^{-5.6} = 0.996$

Gauss Process

- Central place in statistics.

- No natural process is sricto sensu gaussian.

- Many PDF asymptotically tend towards a gaussian or normal PDF at the limit of large samples.

- Sums and means of large samples asymptotically follow a gaussian PDF (Central limit theorem).

Normal or gaussian PDF

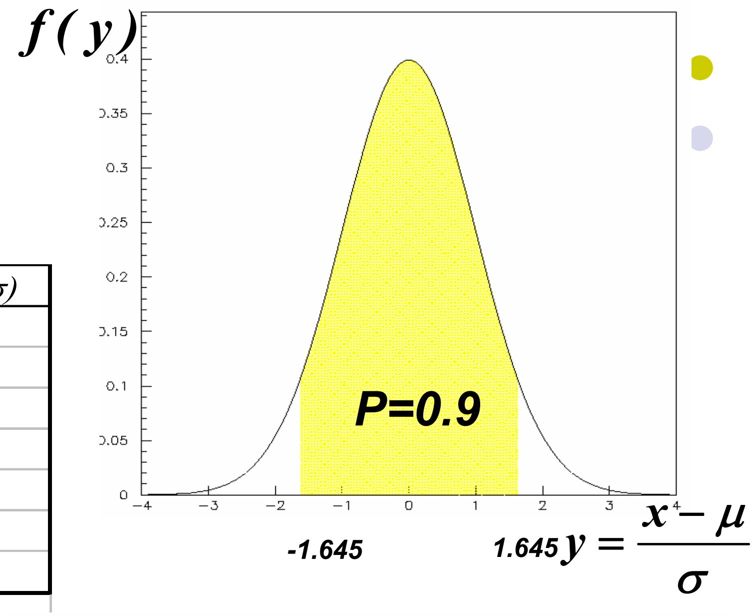
$$f(x\,|\,\mu,\sigma^2) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{1}{2}\frac{(x-\mu)^2}{\sigma^2}} = N(\mu,\sigma^2)$$

Standard normal PDF

$$y = \frac{x-\mu}{\sigma}$$

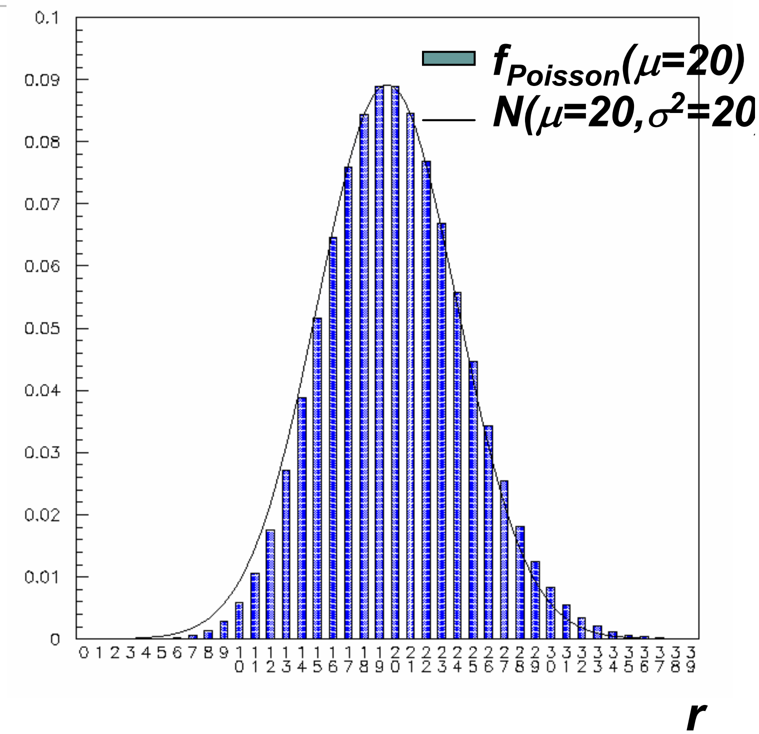
$$f(y\,|\,0,1) = \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}y^2} = N(0,1)$$

n	$P(\mu - n\,\sigma < x < \mu + n\,\sigma)$
1	0.683
1.645	0.900
1.960	0.950
2	0.955
2.576	0.990
3	0.997
3.29	0.999



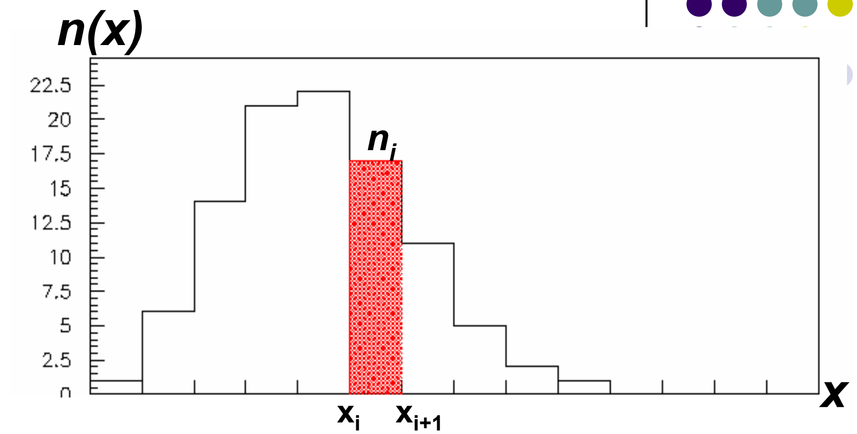
Normal PDF as a limit the Poisson PDF

$$lim_{\mu\to\infty} \frac{1}{r!}\mu^r e^{-\mu} = \frac{1}{\sqrt{2\pi\mu}} e^{-\frac{1}{2}\frac{(r-\mu)^2}{\mu}}$$



Contents of an histogram: from binomial to Poisson to Normal PDF

- n events into dans k classes
- n_i observed events class $i = 1,k$
- p_i probability in class i



Exact PDF : binomial PDF

n_i follows binomial PDF $f(n_i \mid n, p_i) \quad \forall i = 1,k$

$\mu_i = np_i \qquad \sigma_i^2 = np_i(1 - p_i)$

$n = \sum_{i=1}^{k} n_i$ is conserved

Many classes : Poisson PDF approximation

k large $\Rightarrow p_i \ll 1 \;\; \forall i = 1,k$

$\Rightarrow \sigma_i^2 = np_i(1 - p_i) \approx np_i \;\; \forall i = 1,k$

PDF of $n_i \Rightarrow$ Poisson PDF $f(n_i \mid \mu_i)$

$n \neq \sum_{i=1}^{k} n_i$ is not conserved

Many classes and events per class : Normal PDF approximation

$n_i >> 1 \;\; \Rightarrow \forall i = 1,k$

PDF of $n_i \Rightarrow N(n_i \mid n_i)$

$n \neq \sum_{i=1}^{k} n_i$ is not conserved

In practice : $n_i > 20 \;\; \Rightarrow \forall i = 1,k$

Binormal and Multi-Normal Distributions

Independent variables

$$f(x_1, x_2) = \frac{1}{2\pi\sqrt{\sigma_1^2\sigma_2^2}} e^{-\frac{1}{2}\left(\frac{(x_1-\mu_1)^2}{\sigma_1^2}+\frac{(x_2-\mu_2)^2}{\sigma_2^2}\right)}$$

$$f(\underline{x}) = \frac{1}{(2\pi)^{\frac{n}{2}}\prod_{i=1}^{n}\sigma_i} e^{-\frac{1}{2}\sum_{i=1}^{n}\frac{(x_i-\mu_i)^2}{\sigma_i^2}} \qquad \underline{x} = (x_1, \ldots, x_n)$$

Correlated variables

$$f(x_1, x_2) = \frac{1}{\sqrt{1-\rho^2}} \frac{1}{2\pi\sqrt{\sigma_1^2\sigma_2^2}} e^{-\frac{1}{2(1-\rho^2)}\left(\frac{(x_1-\mu_1)^2}{\sigma_1^2}+\frac{(x_2-\mu_2)^2}{\sigma_2^2}-2\rho\frac{(x_1-\mu_1)}{\sigma_1}\frac{(x_2-\mu_2)}{\sigma_2}\right)}$$

III – Recall of general notions de statistics

Sampling

PDF *f(x)* defines the probability for each value of a population to occur.

Unbiased experiment = random sample of *n* observations ($x_1, x_2, \ldots x_n$) differing from the population only by statistical fluctuations due to its limited size

For the experiment to make sense, if the sample is biased, either the bias can be corrected for or it is small enough to be neglected.

Concepts of statistic and estimator

A statistic : a random variables that depends only on the sample of observations and known parameters

An estimator : a statistic the value of which provides an estimation $\hat{\theta}$ of a parametre θ of unknown true value θ_0 .

A non-bias estimator : $E\left[\hat{\theta}\right] = \theta_0$

A coherent estimator : $lim_{n\to\infty}\, \hat{\theta} = \theta_0$

Invariance of the solution ; non propagation of the non-biasness :

If $\tau = \tau(\theta) \quad \Rightarrow \quad \hat{\tau} = \tau\left(\hat{\theta}\right)$

If $E\left[\hat{\theta}\right] = \theta_0 \quad \Rightarrow \quad E[\hat{\tau}] = E\left[\tau\left(\hat{\theta}\right)\right] \neq \tau\left(E\left[\hat{\theta}\right]\right) = \tau(\theta_0) = \tau_0$

$\Rightarrow E[\hat{\tau}] \neq \tau_0$ **in general**

Estimation of the mean and the variance of a sample

- **Mean** $\hat{\mu} = \bar{x} = \frac{1}{n}\sum_{i=1}^{n} x_i$

Non-biased estimator : $E[\bar{x}] = E\left[\frac{1}{n}\sum_{i=1}^{n} x_i\right] = \frac{1}{n}\sum_{i=1}^{n} E[x_i] = \mu$

Coherent estimator : $lim_{n\to\infty}\ \bar{x} = \mu$

intuitivley $\sigma_{\bar{x}}^2 = \frac{\sigma^2}{n}$

- **Variance - μ known**

$$\hat{\sigma}^2 = S^2 = \frac{1}{n}\sum_{i=1}^{n}(x_i - \mu)^2$$

Non baised estimator : $E[S^2] = \sigma^2$

Variance - μ unknown

$$\hat{\sigma}^2 = s^2 = \frac{1}{n-1}\sum_{i=1}^{n}(x_i - \bar{x})^2$$

Non baised estimator : $E[s^2] = \sigma^2$

Intuitively : 1 degree of freedom used to compute $\bar{x}$

Coherent estimators : $lim_{n\to\infty}\ S^2, s^2 = \sigma^2$

Intuitively : $\sigma_{\sigma^2}^2 = \frac{(m_4 - \sigma^4)}{n} + \frac{2\sigma^4}{(n-1)n}$

Note : $(\bar{x}, S^2)$ **and** $(\bar{x}, s^2)$ **are pairs of independent variables**

Size of the statistical fluctuations

Inequality of Bienaymé - Chebyshev

For any PDF $$P\left(x \notin [\mu\text{-}\lambda\sigma, \mu+\lambda\sigma]\right) \leq \frac{1}{\lambda^2}$$

The Law of Large Numbers

Application of Bienaymé-Chebyshev to the estimated mean of $\bar{x}$ of a sample

$$\sigma_{\bar{x}}^2 = \frac{\sigma^2}{n}$$

$$P\left(|\bar{x}-\mu| \geq \lambda\sigma\right) = P\left(|\bar{x}-\mu| \geq \lambda\sqrt{n}\sigma_{\bar{x}}\right) =\leq \frac{1}{n\lambda^2} \quad \Rightarrow \quad P\left(|\bar{x}-\mu| \geq \varepsilon\right) = \frac{\sigma^2}{n\varepsilon^2}$$

IV – Sampling Distributions

The central role of the Normal Distribution: The Central Limit Theorem

$\underline{x} = (x_1, x_2, \ldots, x_n)$ a set of independant random variables from any PDF provided

means $(\mu_1, \mu_2, \ldots, \mu_n)$ and variances $(\sigma_1^2, \sigma_2^2, \ldots, \sigma_n^2)$ are defined

$$X = \sum_{i=1}^{n} x_i \qquad \Rightarrow \qquad \mu_X = \sum_{i=1}^{n} \mu_i \quad \text{et} \quad \sigma_X^2 = \sum_{i=1}^{n} \sigma_i^2$$

If $n \to \infty$, $X = \sum_{i=1}^{n} x_i$ distributed following $N\left(\sum_{i=1}^{n} \mu_i, \sum_{i=1}^{n} \sigma_i^2\right)$

$\underline{x} = (x_1, x_2, \ldots, x_n)$ a set of independant random variables from the same PDF provided

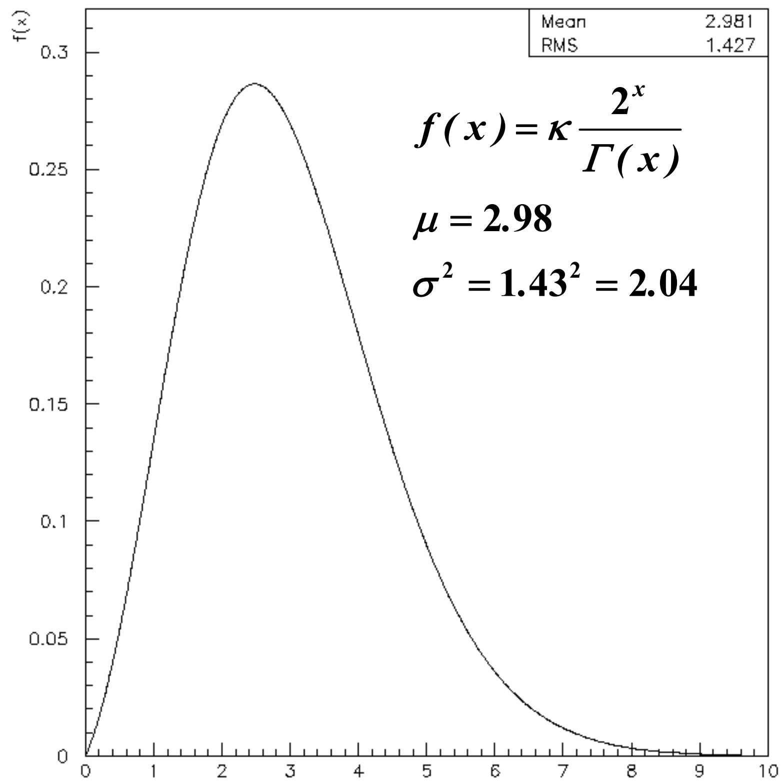
mean μ and variance σ^2 are defined

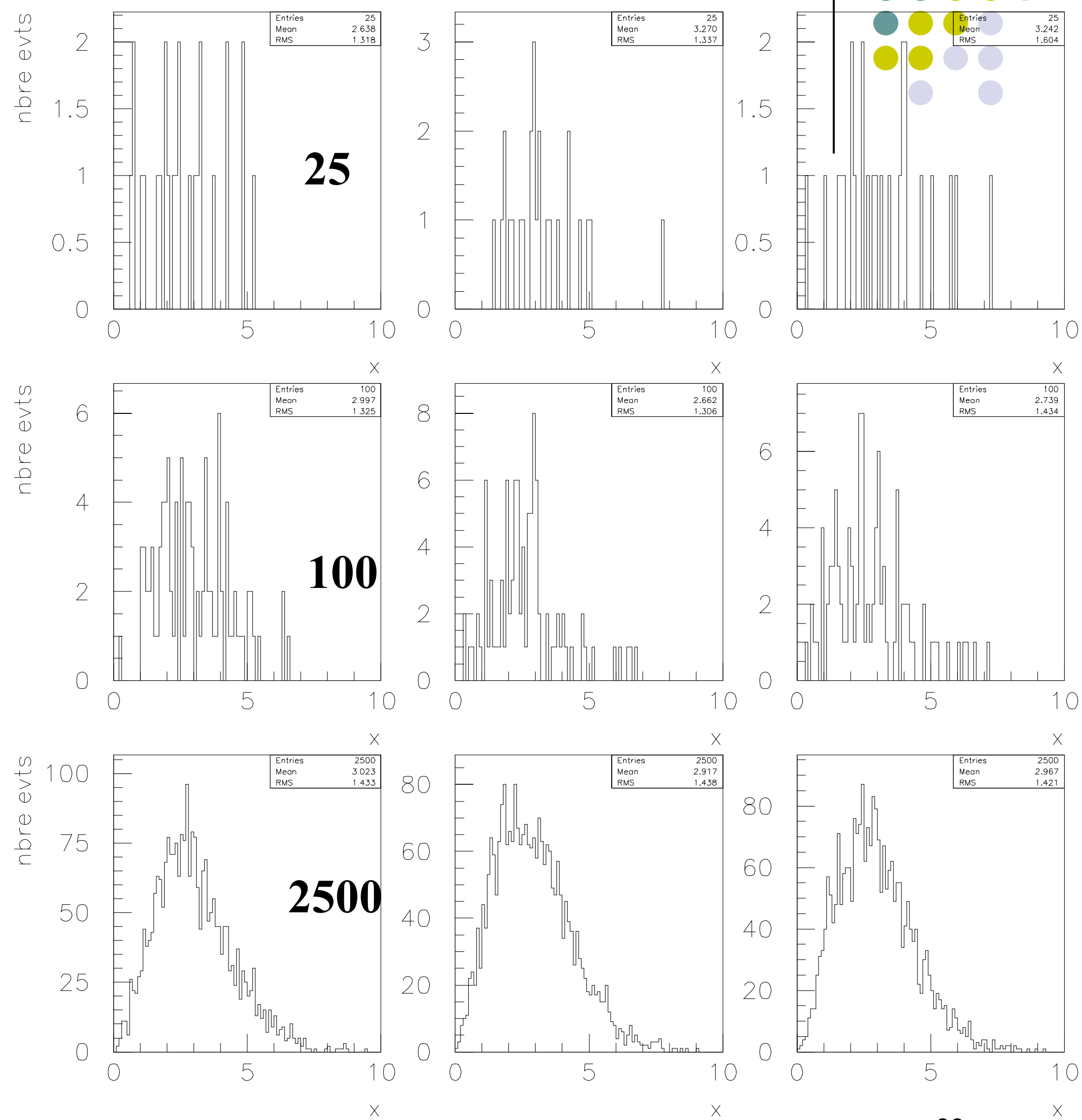
If $n \to \infty$, $\bar{x} = \frac{1}{n}\sum_{i=1}^{n} x_i$ distributed following $N\left(\mu, \frac{\sigma^2}{n}\right)$

If $n \to \infty$, $\frac{\bar{x} - \mu}{\sigma/\sqrt{n}}$ distributed following $N(0,1)$

Central Limit Theorem: example (1)

3 samples of each size





Extraction (par simulation) of
10000 samples de 25 events
10000 samples de 100 events
10000 samples de 2500 events

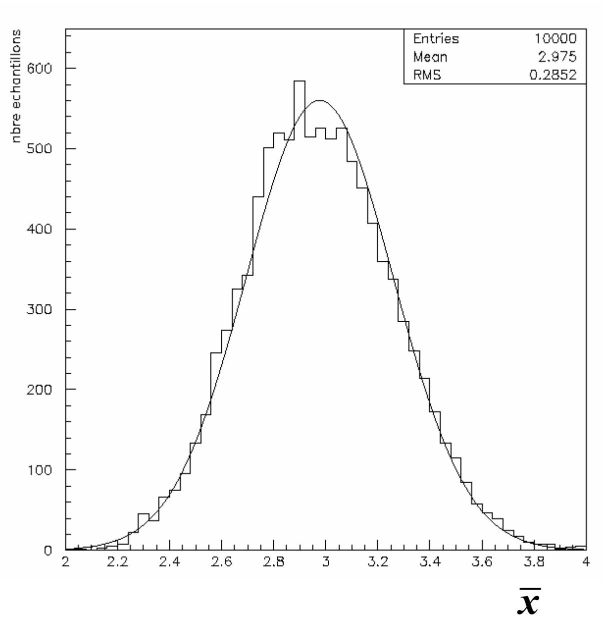
Central Limit Theorem: example (2)

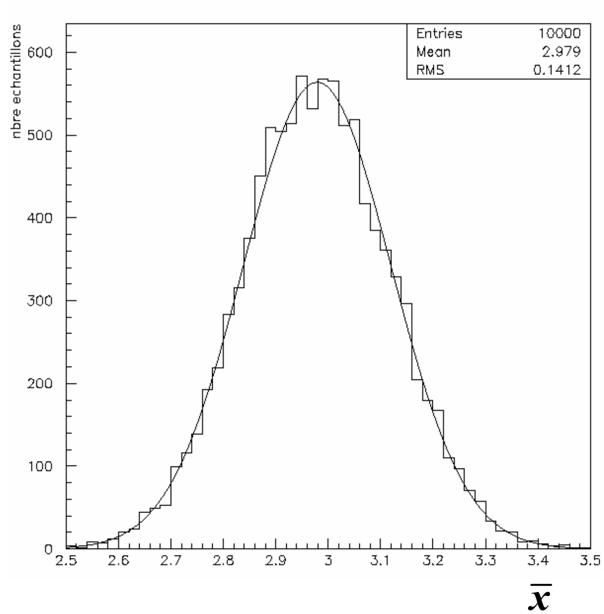
Distributions of the means of the 10000 samples of each of the 3 sample sizes

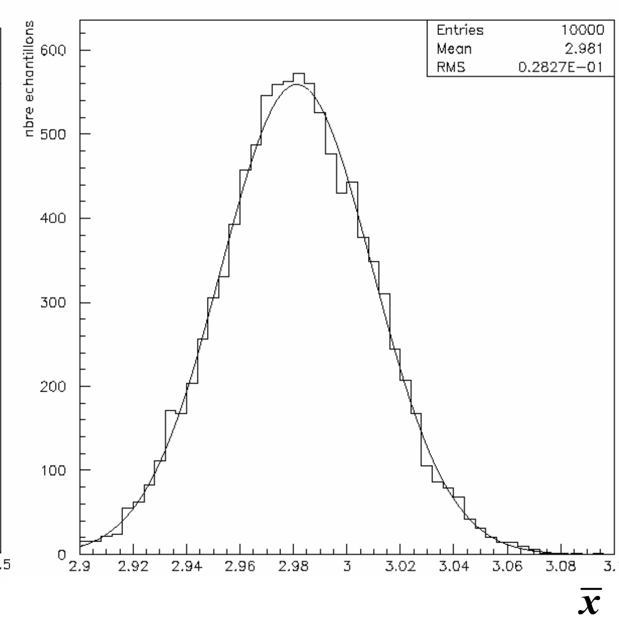
$$n = 25 \qquad \frac{\sigma = 2.04}{\sqrt{25}} = 0.28$$

$$n = 100 \qquad \frac{\sigma = 2.04}{\sqrt{100}} = 0.14$$

$$n = 2500 \qquad \frac{\sigma = 2.04}{\sqrt{2500}} = 0.028$$

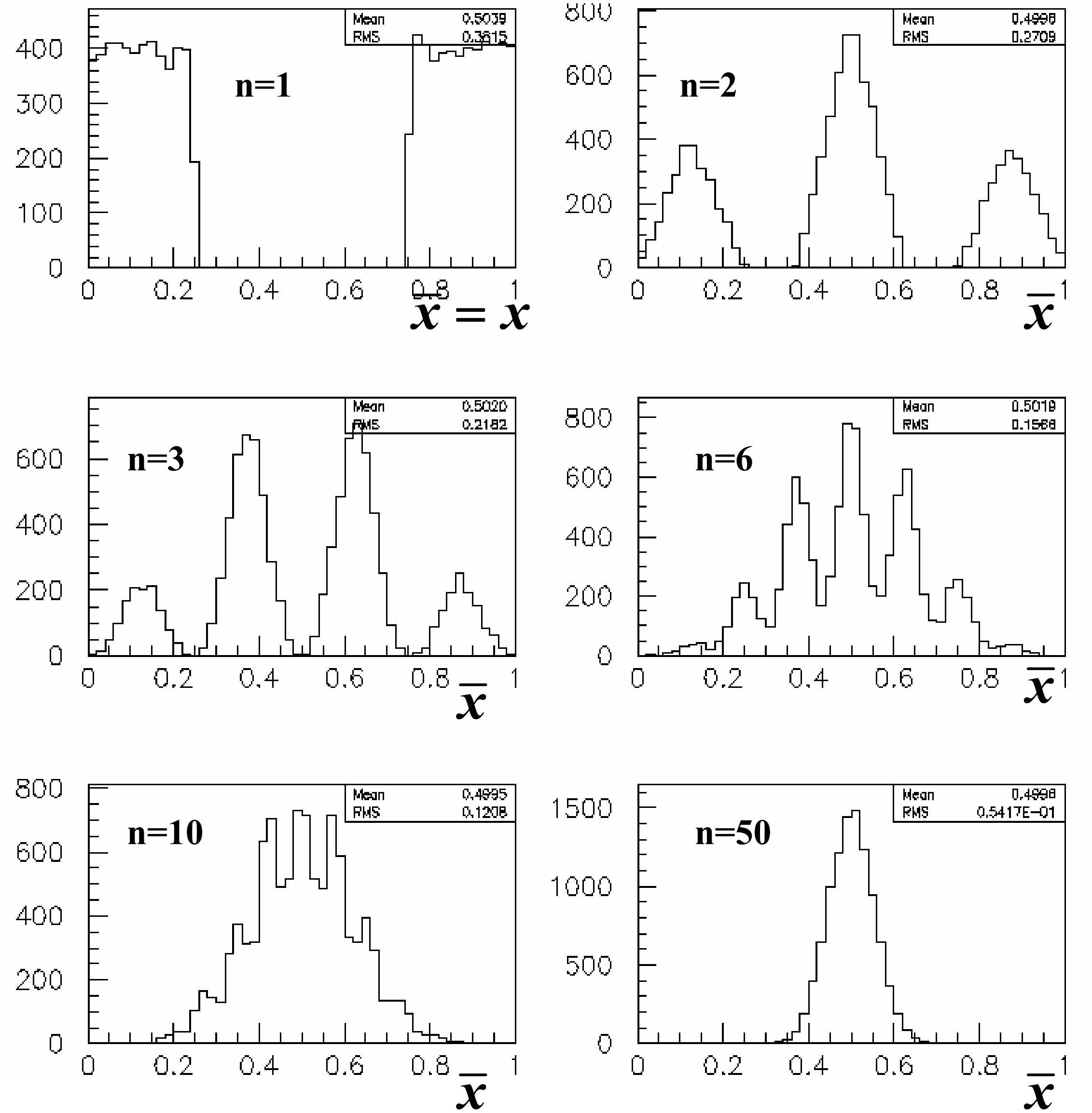






Central Limit Theorem: example (3)

f(x): uniform on [0,¼] and [¾,1]



Central Limit theorem and the Standard Error

The measurements of a distance of 3 km obtained by reporting 10 000 times a 30 cm ruler are distributed approximately mormaly with a standard deviation of $\sim\sqrt{10\,000}\times 1mm = 1\ m$

The Standard Error on a measurement is the standard deviation of the approximate normal distribution along which an hypothetical large number of measurements would distribute around the true value.

The concept of Standard Error applies only if the final measurement is the convolution of a rather large number of rather independant measurements.

Errors Propagation

One variable

Knowing the measured values $\hat{\underline{x}}$ of $\underline{x}$ and the standard errors $\underline{\sigma}$

what is the error on $\hat{y} = y(\hat{\underline{x}})$?

Unknown true values : $\underline{x}_0$ and $y_0 = y(\underline{x}_0)$

First order Taylor series development around y_0

$$y(\underline{x}) = y_0 + \sum_{i=1}^{n} (x_i - x_{i,0}) \left.\frac{\partial y}{\partial x_i}\right|_{\underline{x}=\underline{x}_0} + \text{ ....} = \sum_{i=1}^{n} x_i \left.\frac{\partial y}{\partial x_i}\right|_{\underline{x}=\underline{x}_0} + \text{Cste}$$

$\Rightarrow \quad \sigma_y^2 = \sum_{i=1}^{n} \left(\left.\frac{\partial y}{\partial x_i}\right|_{\underline{x}=\underline{x}_0} \right)^2 \sigma_i^2$ **and as $\underline{x}_0$ is unknown, approximation by $\hat{\underline{x}}$** $\Rightarrow$

$$\sigma_y^2 = \sum_{i=1}^{n} \left(\left.\frac{\partial y}{\partial x_i}\right|_{\underline{x}=\hat{\underline{x}}} \right)^2 \sigma_i^2$$

Several variables

n variables directly accessible to measurement $\underline{x} = (x_1, x_2, \dots, x_n)$

m variables $\underline{y} = (y_1, y_2, \dots, y_m)$ measured through relations $\underline{y} = \underline{y}(\underline{x})$

$$\sigma_{y_{kl}} = \sum_{i=1}^{n} \sum_{j=1}^{n} \left.\frac{\partial y_k}{\partial x_i}\right|_{\underline{x}=\hat{\underline{x}}} \left.\frac{\partial y_l}{\partial x_j}\right|_{\underline{x}=\hat{\underline{x}}} \sigma_{x_{ij}}$$

$k, l = 1, m$

where $\sigma_{ii} = \sigma_i^2$

χ^2 Distribution

n independent variables $\underline{x} = \left(x_1, \ldots, x_n\right)$ distributed following $N\left(\mu_i, \sigma_i^2\right)$

$X^2 = \sum_{i=1}^{n} \frac{\left(x_i - \mu_i\right)^2}{\sigma_i^2}$ follows a χ_n^2 distribution with n degrees of freedom

X^2 measures the sum of the square distance between point $\underline{x}$

and its expectation value $\underline{\mu}$

in a n-dimension space, using $\underline{\sigma}$ as unit length.

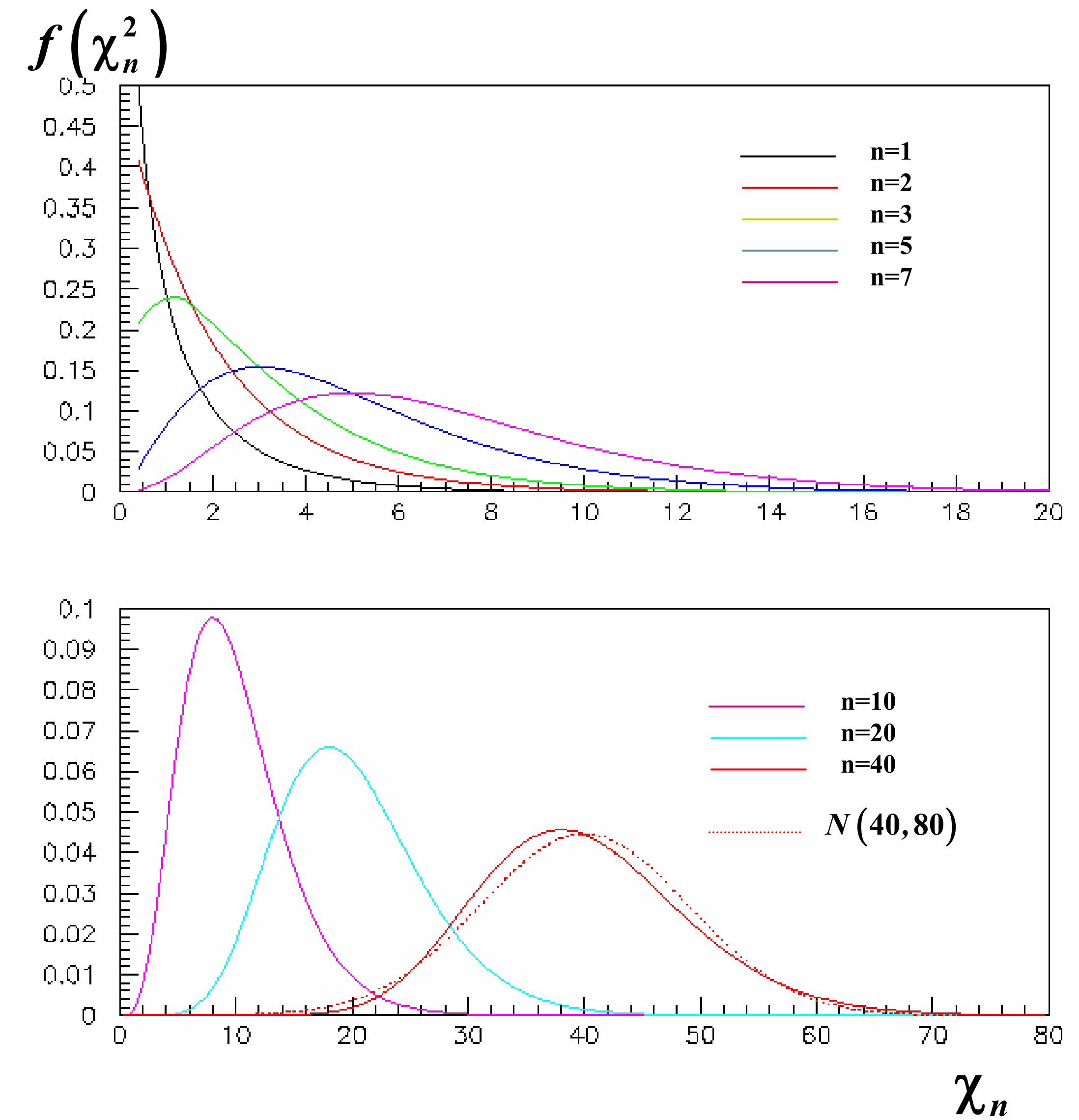
PDF:

$$f\left(X^2 \mid n\right) = \frac{1}{2^{n/2}\Gamma\left(n/2\right)}\left(X^2\right)^{n/2-1} e^{-X^2/2}$$

$$\mu = n$$

$$\sigma^2 = 2n$$

χ^2 Distribution: shape of PDF and normal asymptotic convergence



$$lim_{n\to\infty} f\left(\chi_n^2\right) = N(n, 2n)$$

Statistics following a χ^2 distribution

$\underline{x} = (x_1, \ldots, x_n)$ independent and follow (μ, σ^2)

$\sum_{i=1}^{n}\left(\frac{x_i - \mu}{\sigma}\right)^2$ follows a χ^2_n

$S^2 = \frac{1}{n}\sum_{i=1}^{n}(x_i - \mu)^2$

$\Rightarrow$ $n\frac{S^2}{\sigma^2}$ follows a χ^2_n

$\sum_{i=1}^{n}\left(\frac{x_i - \bar{x}}{\sigma}\right)^2$ follows a χ^2_{n-1}

$s^2 = \frac{1}{n-1}\sum_{i=1}^{n}(x_i - \bar{x})^2$

$\Rightarrow$ $(n-1)\frac{s^2}{\sigma^2}$ follows a χ^2_{n-1}

Student t Distribution : small samples

- Given - x distributed following a $N(0,1)$
  - u distributed following a χ^2_n
  - x and u independent

$$t_n = \frac{x}{\sqrt{u/n}}$$

follows a Student t distribution with n degrees of freedom

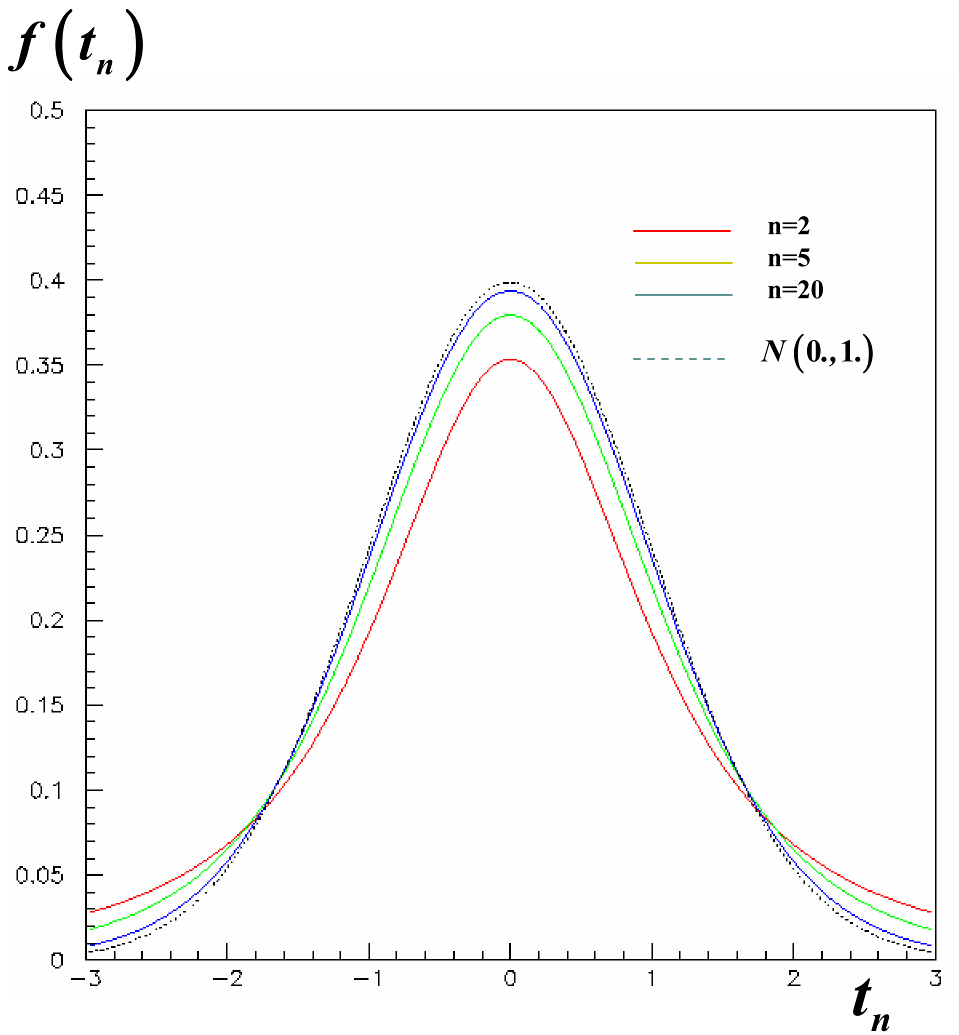
PDF :

$$f(t_n \mid n) = \frac{\Gamma\left(\frac{n+1}{2}\right)}{\sqrt{\pi n}\,\Gamma\left(\frac{n}{2}\right)} \frac{1}{\left(1 + t_n^2\right)^{\frac{n+1}{2}}}$$

$$\mu = 0$$

$$\sigma^2 = \frac{n}{n-2} \quad si \;\; n > 2$$

t Distribution: shape of PDF and normal asymptotic convergence



limit for $n \to \infty$

$$lim_{n\to\infty}(t_n) = N(0,1)$$

Statistics following a t distribution

$\underline{x} = (x_1, \dots, x_n)$ independent and follow $N(\mu, \sigma^2)$

$\bar{x} = \frac{1}{n}\sum_{i=1}^{n} x_i$ and $\frac{\bar{x}-\mu}{\sigma/\sqrt{n}}$ distributed following $N(0,1)$

$S^2 = \frac{1}{n}\sum_{i=1}^{n}(x_i - \mu)^2$ and $n\frac{S^2}{\sigma^2}$ distributed following χ^2_n

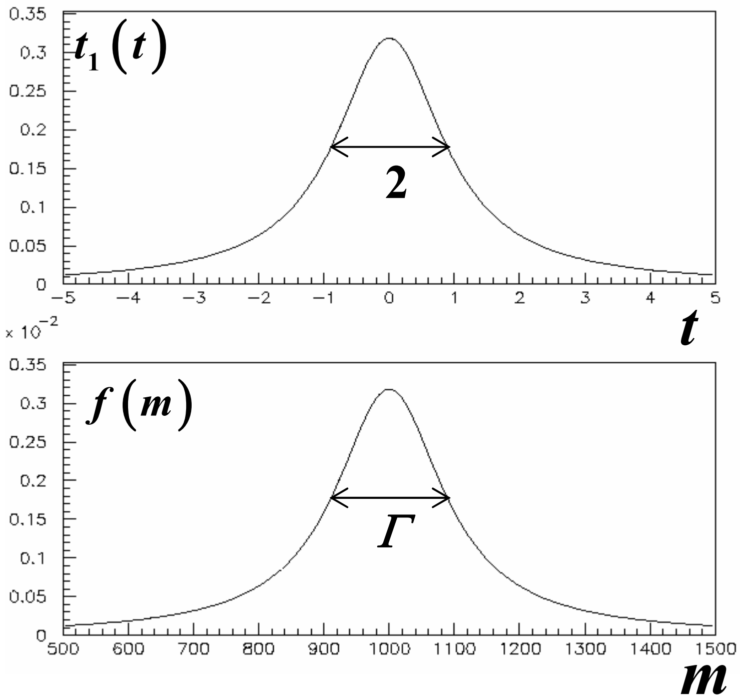
$$\frac{\frac{\bar{x}-\mu}{\sigma/\sqrt{n}}}{\sqrt{\frac{n\frac{S^2}{\sigma^2}}{n}}} = \frac{\bar{x}-\mu}{S/\sqrt{n}} \text{ follows } t_n \qquad \frac{\bar{x}-\mu}{\sigma/\sqrt{n}} \text{ follows } N(0,1)$$

$s^2 = \frac{1}{n-1}\sum_{i=1}^{n}(x_i - \bar{x})^2$ and $(n-1)\frac{s^2}{\sigma^2}$ distributed following χ^2_{n-1}

$$\frac{\frac{\bar{x}-\mu}{\sigma/\sqrt{n}}}{\sqrt{\frac{(n-1)\frac{s^2}{\sigma^2}}{(n-1)}}} = \frac{\bar{x}-\mu}{s/\sqrt{n}} \text{ follows } t_{n-1} \qquad \frac{\bar{x}-\mu}{\sigma/\sqrt{n}} \text{ follows } N(0,1)$$

$$lim_{n\to\infty} \frac{\bar{x}-\mu}{s/\sqrt{n}}, \frac{\bar{x}-\mu}{S/\sqrt{n}} = \frac{\bar{x}-\mu}{\sigma/\sqrt{n}} \quad \Leftrightarrow \quad lim_{n\to\infty} f(t_n) = N(0,1)$$

Cauchy or t_1 Student and Breit-Wigner distributions



Cauchy or Student t_1 Distribution

$$f(t_1) = \frac{1}{\pi\left(1 + t_1^2\right)}$$

$$\int_{-\infty}^{\infty} t_1^2 (t_1) dt = \infty$$

$$\Rightarrow \begin{cases} \sigma^2 \text{ undefined} \\ \text{central limit theorem not applicable} \\ \text{full width at half maximum FWHM = 2} \end{cases}$$

Breit-Wigner Distribution

$\Gamma = \frac{\hbar}{\tau}$ = intrinsic particle mass width.

Strong interaction : $\tau \approx 10^{-23} s \Rightarrow \Gamma \approx 100\, MeV$

Change of variable $m = m_0 + t_1 \cdot \Gamma/2$

$$f(m) = \frac{\Gamma/2}{\pi} \frac{1}{(m - m_0)^2 + (\Gamma/2)^2} \qquad \Gamma = \text{FWHM}$$

Mass distribution of a spin 0 particle

Fisher-Snedecor F Distribution

- u_1, u_2 distributed following $\chi^2_{n_1}, \chi^2_{n_2}$
- u_1 and u_2 independent

$$F_{n1,n2} = \frac{u_1/n_1}{u_2/n_2}$$ follows a Fisher distribution with n_1, n_2 degrees of freedom

PDF:

$$f\left(F_{n_1 n_2} \mid n_1 . n_2\right) = \frac{\Gamma\left(\frac{n_1+n_2}{2}\right)}{\Gamma\left(\frac{n_1}{2}\right)\Gamma\left(\frac{n_2}{2}\right)}\left(\frac{n_1}{n_2}\right)^{\frac{n_1}{2}} \times \frac{F_{n_1 n_2}^{\frac{n_1}{2}-1}}{\left(1+\frac{1}{n_2}F_{n_1 n_2}\right)^{\frac{n_1+n_2}{2}}}$$

$$\mu = \frac{n_2}{n_2-2} \text{ si } n_2 > 2$$

$$\sigma^2 = \frac{2n_2^{\,2}\left(n_1+n_2-2\right)}{n_1\left(n_2-2\right)^2\left(n_2-4\right)} \text{ si } n_2 > 4$$

$$lim_{n_2\to\infty} f\left(n_1 F_{n_1 n_2}\right) = f\left(\chi^2_{n_1}\right)$$

$$lim_{n_1, n_2\to\infty} f\left(F_{n_1 n_2}\right) = N(0,1)$$

Statistics following an F distribution

$\underline{x} = (x_1, \ldots, x_n)$ independent and follow $N(\mu_x, \sigma_x^2)$

$\underline{y} = (y_1, \ldots, y_m)$ independent and follow $N(\mu_y, \sigma_y^2)$

$\dfrac{S_x^2 / \sigma_x^2}{S_y^2 / \sigma_y^2}$ follows a F_{nm}

$\dfrac{s_x^2 / \sigma_x^2}{s_y^2 / \sigma_y^2}$ follows a $F_{(n-1)(m-1)}$

Central role of the Normal Distribution

Binomiale (n, p)

$\mu = np$

$\sigma^2 = np(1-p)$

$n \to \infty$, $p \to 0$, $np \to \mu$ → Poisson

Poisson (μ)

μ

$\sigma^2 = \mu$

$\mu \to \infty$ → Normal

Chi-square χ_n^2

$\mu = n$

$\sigma^2 = 2n$

$n \to \infty$ → Normal

Student t_n

$\mu = 0$

$$\sigma^2 = \frac{n}{n-2}$$

$n \to \infty$ → Normal

Normal

Fisher F_{n_1, n_2}

$$\mu = \frac{n_2}{n_2 - 2}$$

$$\sigma^2 = \frac{2n_2^2(n_1 + n_2 - 2)}{n_1(n_2-2)^2(n_2-4)}$$

$n_2 \to \infty$ → Chi-square

$n_1, n_2 \to \infty$ → Normal

Central Limit Theorem ($n \to \infty$ → Normal)

– $\underline{x}$ independent following $f_i(\mu_i, \sigma_i^2)$

$$X = \sum_{i=1}^{n} x_i \quad \Rightarrow \mu_X = \sum_{i=1}^{n} \mu_i \quad \sigma_X^2 = \sum_{i=1}^{n} \sigma_i^2$$

– $\underline{x}$ independent following $f(\mu, \sigma^2)$

$$\bar{x} = \frac{1}{n}\sum_{i=1}^{n} x_i \quad \Rightarrow \mu_{\bar{x}} = \mu \qquad \sigma_{\bar{x}}^2 = \frac{\sigma^2}{n}$$

V – Hypothesis Tests

Principle

Decide if the hypothesis **H_0**, the null or tested hypothesis, that an observation (value of a parameter, distribution of a variable) is compatible with a reference (expectation from a model, existing observation) is true while accepting to reject the hypothesis though it is true (commit a type I error) with an a priori probability α, the significance or level of significance of the test.

The test only makes sense if there exists an alternative hypothesis **H_1** with a non null probability to occur. The most trivial form of H_1 is that H_0 is false

If H_1 is fully specified, the probability β to accept H_0 though H_1 is true (commit a **type II error**) can be computed. The best test maximises the power **1-β** of the test.

If H_1 is not fully specified, β cannot be computed but it is often possible to define the test that maximises 1-β.

Definitions – Best critical zone

Parametric test: test the value of a parameter

Non-parametric test: test the shape of a distribution

Simple hypothesis: fully specified

Composite hypothesis: partly specified or unspecified

Best critical region R_α : domain of rejection of values of the parameter that maximise the power of the test.
If PDF f_0 defines H_0 :

$$\alpha = \int_{R_\alpha} f_0(x)\,dx$$

$$1 - \beta = \int_{R_\alpha} f_1(x)\,dx$$ **is maximal**

Acceptance region $A_\alpha = W - R_\alpha$: complement of the critical region

Best critical region R_α : Neyman-Pearson Lemma

One measurement :

$$\alpha = \int_{R_\alpha} f_0(x)\,dx$$

$$1-\beta = \int_{R_\alpha} f_1(x)\,dx = \int_{R_\alpha} \frac{f_1(x)}{f_0(x)} f_0(x)\,dx = \left\langle \frac{f_1(x)}{f_0(x)} \right\rangle_{H_0\ true}$$

$$R_\alpha \subset \begin{cases} f_0(x) = 0 \\ \dfrac{f_1(x)}{f_0(x)} > k_\alpha \end{cases}$$

n measurements : critical region difficult to get

$$\alpha = \int_{R_\alpha} \prod_{i=1}^{n} f_0(x_i)\,dx \quad \Rightarrow$$ **n-dimensional integral**

$$R_\alpha \subset \begin{cases} f_0(x) = 0 \\ \dfrac{\prod_{i=1}^{n} f_1(x_i)}{\prod_{i=1}^{n} f_0(x_i)} > k_\alpha \end{cases}$$

n large : use $\bar{x} = \dfrac{\sum_{i=1,}^{n} x_i}{n}$ instead of $\underline{x}$ and the Central Limit Theorem

Test of the mean of a normal distribution

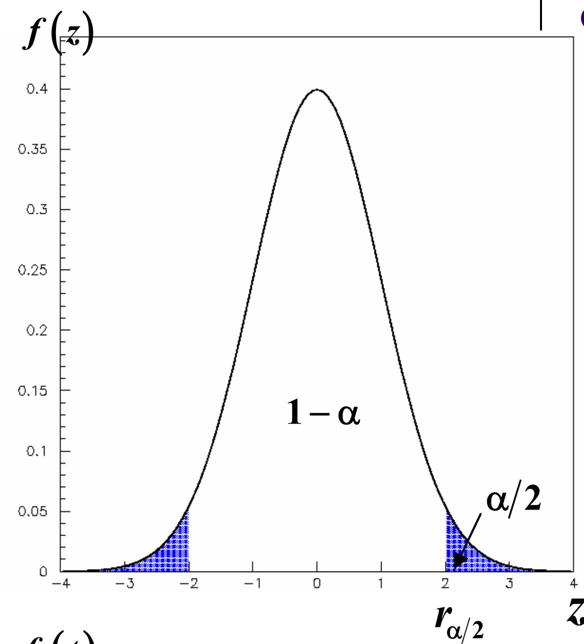
H_0: sample $N\left(\mu = \mu_0, \sigma^2 = \sigma_0^2 \text{ known}\right)$
H_1: $\mu \neq \mu_0$: composite hypothesis

If H_0 true:

$$z = \frac{\bar{x} - \mu_0}{\sigma_0 / \sqrt{n}}$$ **follows a $N(0,1)$**

R_α: **domain of large values of** $|z| > r_{a/2}$

$$\rightarrow \alpha/2 = \int_{-\infty}^{-r_{a/2}} N(z \mid 0,1)\, dz = \int_{r_{a/2}}^{\infty} N(z \mid 0,1)\, dz$$



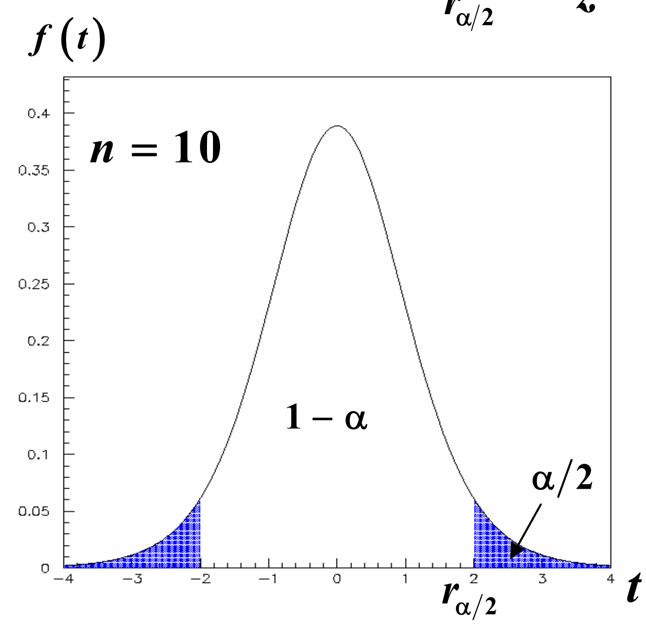
H_0: sample $N\left(\mu = \mu_0, \sigma^2 \text{ unknown}\right)$
H_1: $\mu \neq \mu_0$: composite hypothesis

$$t = \frac{\bar{x} - \mu_0}{s / \sqrt{n}}$$ **follows a Student t_{n-1}**

$$s^2 = \frac{1}{n-1} \sum_{i=1}^{n} \left(x_i - \bar{x}\right)^2$$

R_α: **domain of large values of** $|t| > r_{a/2}$

$$\rightarrow \alpha/2 = \int_{-\infty}^{-r_{a/2}} f(t_{n-1})\, dt = \int_{r_{a/2}}^{\infty} f(t_{n-1})\, dt$$



Test of the means of two normal distributions of known variances

$\bar{x} = \frac{1}{n}\sum_{i=1}^{n} x_i$: sample of size n from $N(\mu_x, \sigma_x^2)$

$\bar{y} = \frac{1}{m}\sum_{i=1}^{n} y_i$ sample of size m from $N(\mu_y, \sigma_y^2)$

H_0: $\mu_x = \mu_y$ - σ_x^2 and σ_y^2 known
H_1: $\mu_x \neq \mu_y$ $\rightarrow$ composite hypothesis

If H_0: $(\bar{x} - \bar{y})$ follows $N\left(\mu_x - \mu_y = 0, \sigma_x^2/n + \sigma_y^2/m\right)$

$$z = \frac{\bar{x} - \bar{y}}{\sqrt{\sigma_x^2/n + \sigma_y^2/m}}$$ follows $N(0,1)$

R_α: domain of large values of $|z| > r_{a/2}$

$$\rightarrow \alpha/2 = \int_{-\infty}^{-r_{a/2}} N(z \mid 0,1)\,dz = \int_{r_{a/2}}^{\infty} N(z \mid 0,1)\,dz$$

Test of the means of two normal distributions of unknown equal variances

$\bar{x} = \frac{1}{n}\sum_{i=1}^{n} x_i$: sample of size n from $N\left(\mu_x, \sigma_x^2\right)$

$\bar{y} = \frac{1}{m}\sum_{i=1}^{n} y_i$ sample of size m from $N\left(\mu_y, \sigma_y^2\right)$

H$_0$: $\mu_x = \mu_y$ - $\sigma_x^2 = \sigma_y^2 = \sigma_0^2$ unknown

sensible if similar experimental procedures

H$_1$: $\mu_x \neq \mu_y$ $\rightarrow$ composite hypothesis

If H$_0$: $z = \frac{\bar{x} - \bar{y}}{\sqrt{\sigma_0^2/n + \sigma_0^2/m}}$ follows $N(0,1)$

$(n-1)s_x^2/\sigma_0^2$ et $(m-1)s_y^2/\sigma_0^2$ follows χ^2_{n-1} et χ^2_{m-1}

$\rightarrow (n-1)s_x^2/\sigma_0^2 + (m-1)s_y^2/\sigma_0^2$ follows χ^2_{n+m-2}

$$t = \frac{\dfrac{\bar{x}-\bar{y}}{\sqrt{\sigma_0^2/n + \sigma_0^2/m}}}{\sqrt{\dfrac{(n-1)s_x^2/\sigma_0^2 + (m-1)s_y^2/\sigma_0^2}{n+m-2}}} = \frac{\bar{x}-\bar{y}}{\sqrt{\dfrac{\left((n-1)s_x^2 + (m-1)s_y^2\right)(n+m)}{nm(n+m-2)}}}$$ follows t_{n+m-2}

Test of the means of two normal distributions of unknown variance

H$_0$: $\mu_x = \mu_y$ - σ_x^2 , σ_y^2 unknown
H$_1$: $\mu_x \neq \mu_y$ $\rightarrow$ composite hypothesis

$$t = \frac{\dfrac{\bar{x}-\bar{y}}{\sqrt{\sigma_x^2/n+\sigma_y^2/m}}}{\sqrt{\dfrac{(n-1)s_x^2/\sigma_x^2 + (m-1)s_y^2/\sigma_y^2}{n+m-2}}} \quad \text{follows } t_{n+m-2}$$

t depends on unknown σ_x^2 et σ_y^2 and is not a statistic : the PDF of t is unknown

Approximation $s_x^2 \approx \sigma_x^2$ et $s_y^2 \approx \sigma_y^2$

$$z = \frac{\bar{x}-\bar{y}}{\sqrt{s_x^2/n+s_y^2/m}} \quad \text{follows } N(0,1)$$

The larger n, m the better the approximation

Non-parametric tests

Theoretical model for the PDF of x : $f_0(x)$

Set of n observations $\underline{x} = (x_1, x_2, \ldots, x_n)$

H_0: $\underline{x}$ is a sample extracted from population of PDF $f_0(x)$

H_1: H_0 false

Variant : the model also predicts the size of the sample

Set of n observations $\underline{x} = (x_1, x_2, \ldots, x_n)$

Set of m observations $\underline{y} = (y_1, y_2, \ldots, y_n)$

H_0: sample $\underline{x}$ and $\underline{y}$ extracted from the same population

H_1: H_0 false

Non-parametric Pearson's χ^2 test – partition in exclusive classes

Partition of $\underline{x}$ into $N < n$ classes of contents $n_i, i = 1, N$ with normalisation $n = \sum_{i=1}^{N} n_i$

If x numerical : class i defined by $X_i \le x < X_{i+1}$

Remember :

n_i follows a binomial of probability p_i

$lim_{\substack{p_i \to 0 \;\forall i \\ n \to \infty}}$ binomial $\to$ Poisson

$lim_{n_i \to \infty \;\forall i}$ Poisson $\to$ normal

n_i follows $N(np_i, np_i)$

If H_0 true: $p_i = p_{0i} = \int_{X_i}^{X_{i+1}} f_0(x)\,dx \quad \forall i = 1, N$

Test statistic $X^2 = \sum_{i=1}^{N} \frac{(n_i - np_{0i})^2}{np_i}$ follows χ^2_{N-1}

Contents of the last class $n_N = n - \sum_{i}^{N-1} n_i$

Number of degrees of freedom $\nu = N - 1$

If H_0 true: $n_i = n_{0i}$

Test statistic $X^2 = \sum_{i=1}^{N} \frac{(n_i - n_{0i})^2}{np_i}$ follows χ^2_N

Contents of the last class $\sum_{i}^{N} n_i \ne \sum_{i}^{N} n_{0i}$

Number of degrees of freedom $\nu = N$

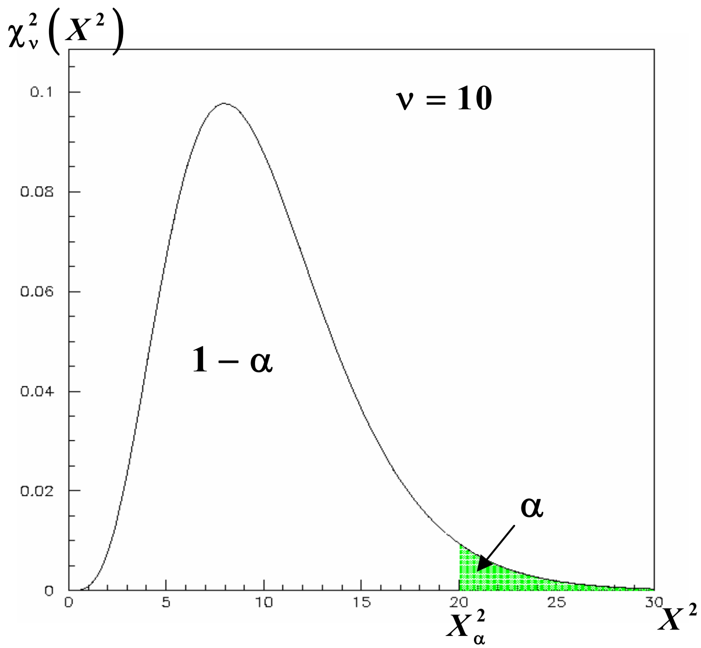
Pearson's χ^2 test – critical zone

Critical zone R_α: $X^2 > X^2_\alpha$

$$X^2_\alpha \Rightarrow \int_{X^2_\alpha}^{\infty} \chi^2_\nu\left(X^2\right) dX^2 = \alpha$$

If H_0 true: $E\left[X^2_\alpha\right] = \nu \quad \Rightarrow$ **small values of X^2 are improbable**

small values of X^2 are even less probable if H_1 true



Pearson's χ^2 test – choice of the classes

Contradictory requirements unless the sample is very large:

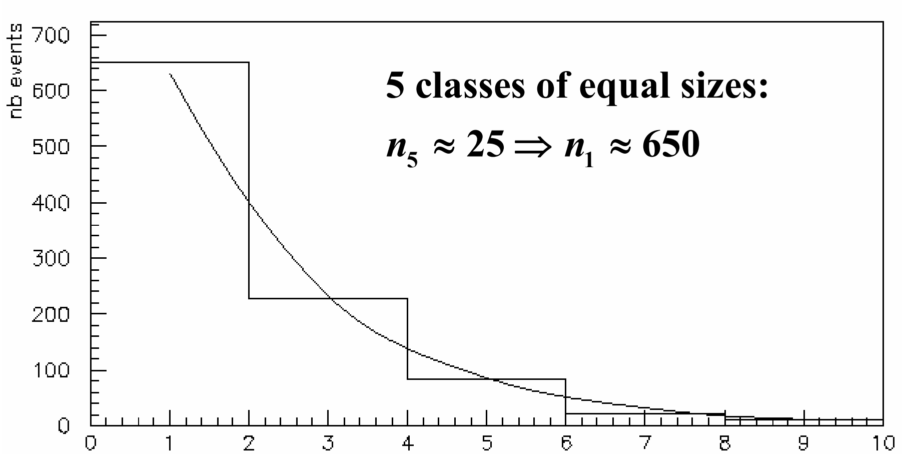
- **binomial → Poisson ⇒ small p_i ⇒ many classes**
- **Poisson → normal ⇒ large n_i ⇒ many entries per class**

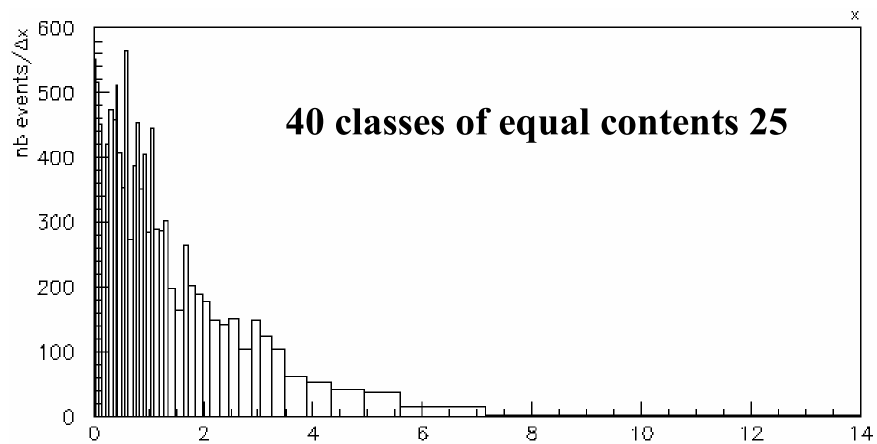
Loss of information:

- **many entries per class ⇒ large classes**

Two methods:

- **classes of equal size: simpler**
- **classes of large (≈25) equal content : minimises the loss of information**





Non-parametric Kolmogorov-Smirnov test

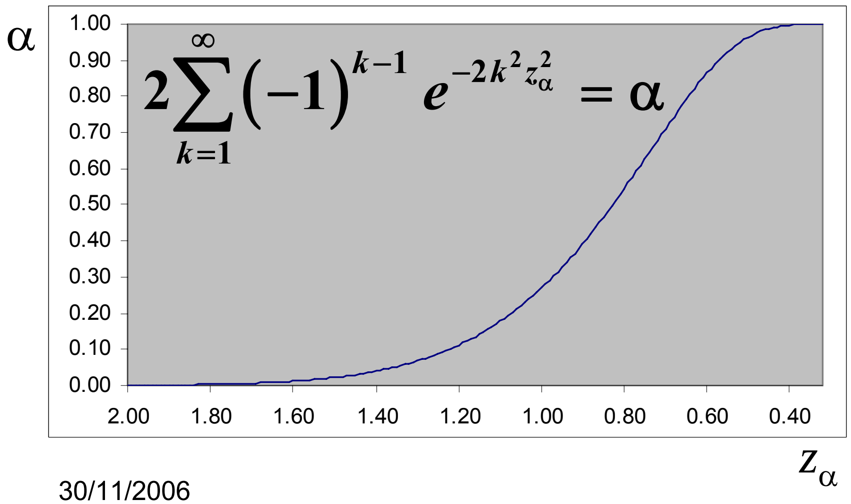
Ordering of $\underline{x}$ **:** $x_i \leq x_{i+1} \quad \forall i = 1, n-1$

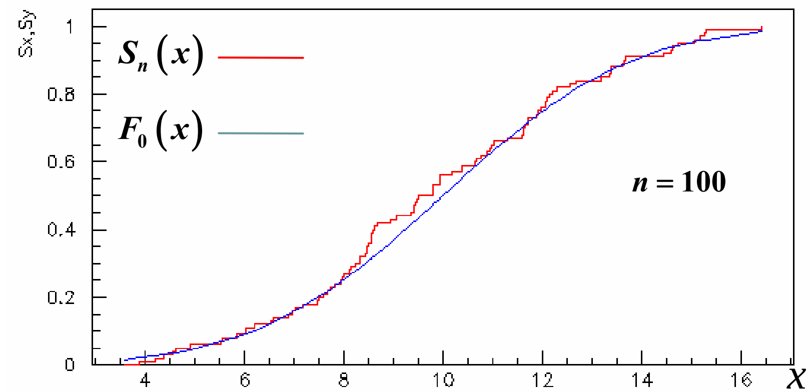
Observed distribution function $S_n(x)\begin{cases} 0 & \text{si} \quad x < x_1 \\ \dfrac{i}{n} & \text{si} \quad x_i \leq x < x_{i+1} \\ 1 & \text{si} \quad x > x_n \end{cases}$ $\quad 0 \leq S_n(x) \leq 1$

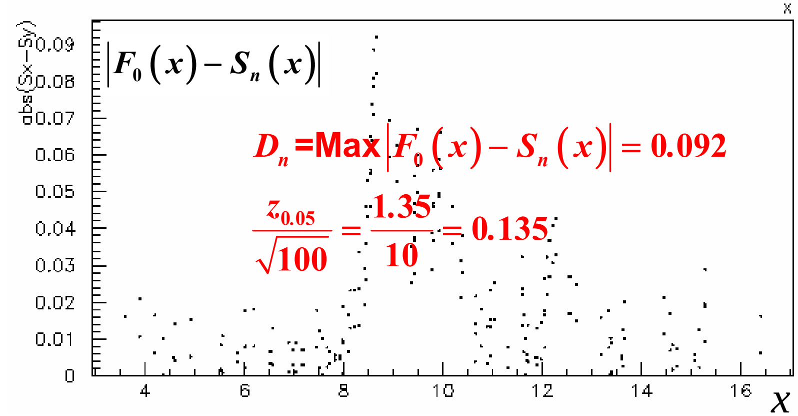
If H_0 **true: distribution function is** $F_0(x)$

Test Statistic $D_n = Max\left(\left|F_0(x) - S_n(x)\right|\right)$

Critical zone : $D_n > d_{n,\alpha} \approx \dfrac{z_a}{\sqrt{n}}$ **for** $n \geq 10$







Kolmogorov-Smirnov test between two samples

Two measurements samples $\underline{x} = \left(x_1, x_2, \ldots, x_n \right)$ **and** $\underline{y} = \left(y_1, y_2, \ldots, y_m \right)$

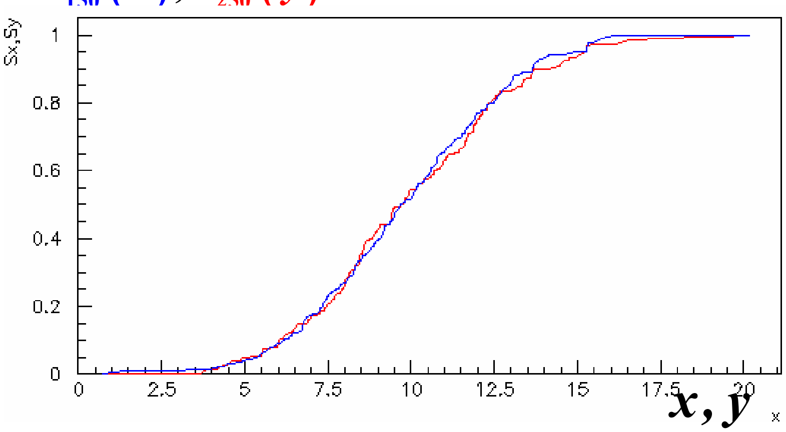
H$_0$: $\underline{x}$ **and** $\underline{y}$ **are samples of the same population - have the same PDF**

H$_1$: H$_0$ is false

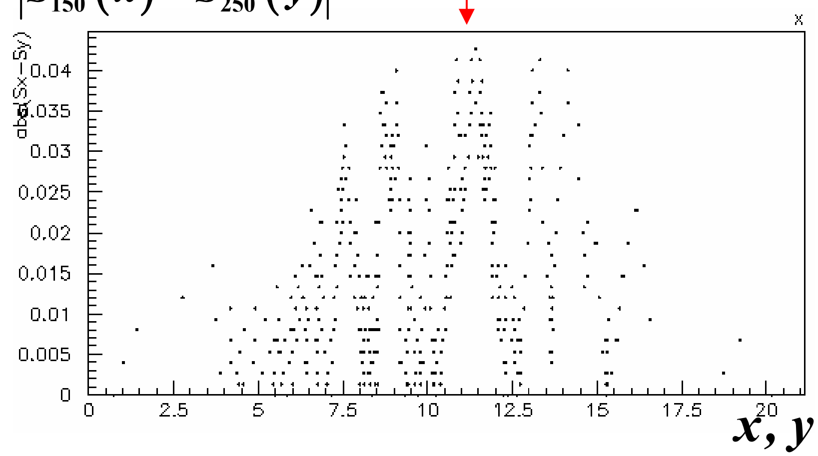
Test statistics $D_{m,n} = Max\left(\left| S_n\left(x\right) - S_m\left(y\right) \right| \right)$

Critical zone $D_{n,m} \geq d_{\alpha,n,m} = d_{\alpha,n}\sqrt{1+\frac{n}{m}} = d_{\alpha,m}\sqrt{1+\frac{m}{n}} \simeq z_{\alpha}\sqrt{\frac{1}{n}+\frac{1}{m}}$ **for** $n \geq 10$

$S_{150}\left(x\right), S_{250}\left(y\right)$



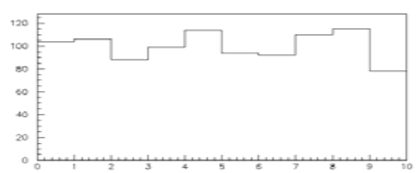
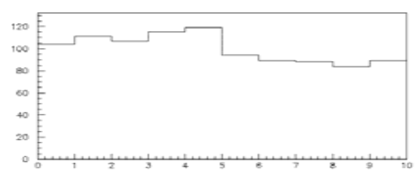
$\left| S_{150}\left(x\right) - S_{250}\left(y\right) \right|$



$$D_{150,250} = 0.043$$

$$d_{0.05,150,250} = z_{0.05}\sqrt{\frac{1}{m}+\frac{1}{n}} = 1.35\sqrt{\frac{1}{150}+\frac{1}{120}} = 0.162 > 0.043$$

Comparison between the two non-parametric tests

Pearson	Kolmogorov - Smirnov
Strictly exact pour a sample of infinite size	**Strictly exact pour a sample of any size.**
Partition in class → loss of information	**No loss of information**
Test not sensitive to the sign of differences  	**Test sensitive to the sign differences**
Test correct if *K* parameters of model f_0 are estimated by the least square method from the sample under test. **The PDF of statistic X^2 is know:** $\chi_n^2 \rightarrow \chi^2_{n-K}$	Test not correct if *K* parameters of model F_0 are estimated from the sample under test. The PDF of statistic D_n is not know.

Level of significance α and confidence level *C.L.*

Pearson

$$\alpha = \int_{X_\alpha^2}^{\infty} \chi_\nu^2\left(x^2\right)dx^2$$

$$C.L. = \int_{X^2}^{\infty} \chi_\nu^2\left(x^2\right)dx^2$$

$$P\left(X^2 \leq X_\alpha^2\right) \equiv P\left(C.L. \geq \alpha\right)$$

Kolmogorov-Smirnov

$$\alpha = 2\sum_{k=1}^{\infty}(-1)^{k-1} e^{-2k^2 z_\alpha^2}$$

$$C.L. = 2\sum_{k=1}^{\infty}(-1)^{k-1} e^{-2k^2 n D_n^2}$$

$$P\left(D_n \leq \frac{z_\alpha}{\sqrt{n}}\right) \equiv P\left(C.L. \geq \alpha\right)$$

$C.L.$ provides a straightforward probabilistic information on how well the hypothesis under test is verified

$$C.L. = 1. - \int_{-\infty}^{x} f\left(x\right)dx = 1. - F\left(x\right)$$

The PDF of $C.L.$ is uniform on [0,1] if H_0 is true

VI- Estimation

Principle of the estimation methods

Random set of measurements $\underline{x} = (x_1, x_2, \ldots, x_n)$ extracted from a population defined by $f(x \mid \underline{\theta}_0)$ with k unknown true parameters $\underline{\theta}_0 = (\theta_{0,1} \cdots \theta_{0,k})$ to be estimated from the sample.

- Point estimations: best estimation set $\hat{\underline{\theta}}$ of $\underline{\theta}_0$ given sample $\underline{x}$
- Variance-covariance matrix estimation

  Confidence level and confidence interval

Définitions et notations

True unknown value paramètre of paramter θ : θ_0

Statistic: $t = t\left(\underline{x}, \not{\theta}\right)$

Estimator: statistic the value of which is an estimation of paramètre θ

Estimation: value taken by the estimator for the observed sample $\underline{x}$

$$\hat{\theta} = t\left(\underline{x}\right)$$

Likelihood: joint ptobability to observe sample $\underline{x}$ for a given value of θ

$$\mathcal{L}\left(\underline{x} \mid \theta\right) = \prod_{i=1}^{n} f\left(x_i \mid \theta\right)$$

Likelihood fonction : $\mathcal{L}\left(\theta \mid \underline{x}\right)$ **as a fonction of θ given the observed sample $\underline{x}$**

Confidence interval in frequentist view: Neyman belts

Given :

- **a measurements sample** $\underline{x} = \left(x_1\ , x_2, \ldots, x_n\right)$ **from population** $f\left(x \mid \theta_0\right)$
- **estimator** $t\left(\underline{x}\right)$ **of** θ_0
- **estimation** $\hat{\theta} = t\left(\underline{x}\right)$ **of** θ_0
- $g\left(t \mid \theta\right)$ **the PDF to observe** t **given** θ

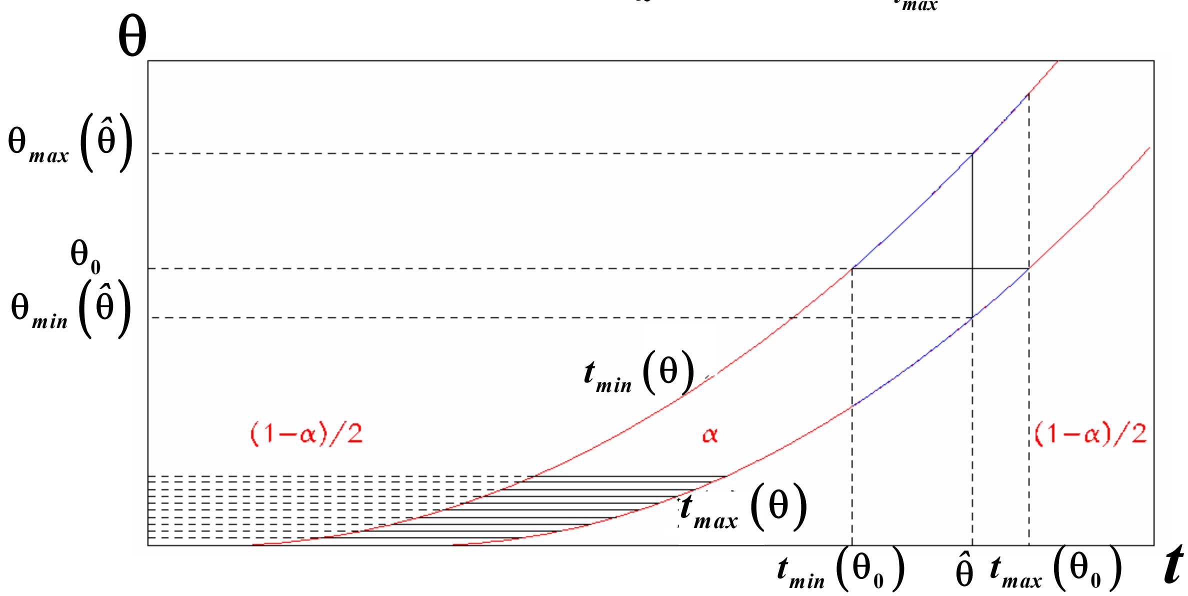
Knowing $g\left(t \mid \theta\right)$ **for all sensible values of** θ**,or at least for the restricted domain of values where the experiment claims sensitivity, is mandatory for the experiment to make sens.**

Confidence interval in frequentist view: Neyman centred belts

Neyman belts associated to a given $C.L.\ \alpha$

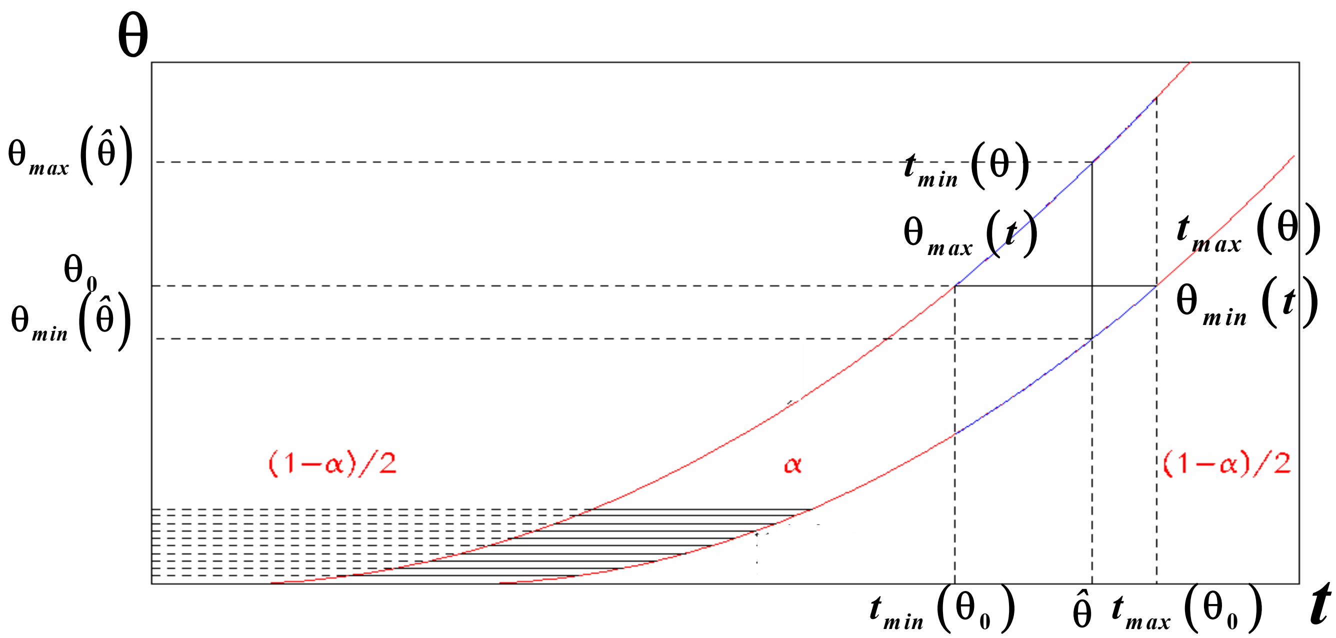
Compute for a finite values of θ in the sensible domain $[\theta_{min}, \theta_{max}]$ contours

$$t_{min}(\theta) \text{ and } t_{max}(\theta) \text{ such that } \int_{-\infty}^{t_{min}} g(t|\theta)\,dt = \int_{t_{max}}^{\infty} g(t|\theta)\,dt = \frac{1-\alpha}{2} \Rightarrow \int_{t_{min}}^{t_{max}} g(t|\theta)\,dt = \alpha$$



$$P\left(\hat{\theta} \in \left[t_{min}(\theta_0), t_{max}(\theta_0)\right]\right) = \alpha$$

Confidence interval in frequentist view: correct interpretation



$t_{b,a}(\theta) \Rightarrow \theta_{a,b}(t)$

$P\left(\hat\theta \in \left[t_{min}(\theta_0), t_{max}(\theta_0)\right]\right) = \alpha \Rightarrow P\left(\theta_0 \in \left[\theta_{min}(\hat\theta), \theta_{max}(\hat\theta)\right]\right) = \alpha$

$\hat\theta$ **is a random variable** $\Rightarrow$ θ_0 **is a constant**

$t_{min}(\theta_0), t_{min}(\theta_0)$ **are unknown constants** $\Rightarrow$ $\theta_{min}(\hat\theta), \theta_{max}(\hat\theta)$ **are known random variables**

Confidence interval in frequentist view: correct interpretation

The experiment determines a particular interval $[\theta_{min},\theta_{max}]$ of values of θ belonging to a large set of intervals that would be obtained by an ensemble of similar experiments such that a fraction α of these intervals contain (covers) the true value θ_0.

$\left[\theta_{min},\theta_{max}\right]$ = **confidence interval at** $C.L. = \alpha$

Statement $\theta_0 \in \left[\theta_{min},\theta_{max}\right]$ **is randomly true** α**% of the times.**

Statement $\theta_0 \in \left[\theta_{min},\theta_{max}\right]$ **is randomly true** $(1\text{-}\alpha)$**% of the times.**

Confidence interval in frequentist view: Neyman upper/lower belts

The Neyman centred belts are constructed with the particular prescription:

$$\int_{-\infty}^{t_{min}} g(t \mid \theta)\,dt = \int_{t_{max}}^{\infty} g(t \mid \theta)\,dt = \frac{1-\alpha}{2} \Rightarrow \int_{t_{min}}^{t_{max}} g(t \mid \theta)\,dt = \alpha$$

There is an infinite number of prescriptions to construct belts with correct coverage corresponding to *C.L.* α that lead to different confidence intervals but are all equally correct from the statistical point of view.

The Neyman upper and lower belts are constructed with the particular prescription:

$$\int_{t_{max}}^{+\infty} g(t \mid \theta)\,dt = \alpha \qquad \int_{-\infty}^{t_{min}} g(t \mid \theta)\,dt = \alpha$$

Probabilistic or Bayes confidence intervals

The probability to observe a value for an observable x depends on the value of parameter θ of true unknown values θ_0 with known PDF $P(x \mid \theta)$.

Bayes' theorem states $P(A \mid B)P(B) = P(B \mid A)P(A)$

Application to the particular observed value $\hat{x}$:

$P(\theta \mid \hat{x}) = \dfrac{P(\hat{x} \mid \theta)P(\theta)}{P(\hat{x})}$ the posteriory PDF that observation $\hat{x}$ result from a value of θ

Bayesian credible interval $[\theta_1, \theta_2]$ at $C.L. = \alpha \Rightarrow \int_{\theta_1}^{\theta_2} P(\theta \mid \hat{x})\, d\theta = \alpha$

$P(\hat{x} \mid \theta)$ the know likelihood to observe $\hat{x}$ given θ

$P(\hat{x})$ a normalisation factor $\Rightarrow \int_{-\infty}^{+\infty} P(\theta \mid \hat{x})\, d\theta = 1$

$P(\theta)$ the priory PDF or prior that θ is the true value.

What to use for $P(\theta)$? Degree of believe in θ based on ignorance, on knowledge from previous experiments, on subjectivity. The main difficulty with the Bayrsian approch is to define an objective informative prior.

Consistent and unbiased estimators

Consistency

$$lim_{n\to\infty}\hat{\theta}=\theta_0 \qquad \bar{x}\to\mu, \quad s^2, S^2\to\sigma^2$$

Unbiasedness

$$E\left[\hat{\theta}\right]=\theta_0 \qquad E\left[\bar{x}\right]=\mu, \quad E\left[s^2, S^2\right]=\sigma^2$$

$\tau=\tau(\theta)$ univocal and reciprocal : $\hat{\tau}=\tau\left(\hat{\theta}\right)$

$$E\left[\tau\left(\hat{\theta}\right)\right]\neq\tau\left(E\left[\hat{\theta}\right]\right)$$

$\Rightarrow \hat{\theta}$ **and** $\hat{\tau}$ **are not simultaneously unbiased**

Minimal variance – Efficient estimator

Given the PDF – the narrowest, the best - and the size of the sample – the largest, the best - the minimal variance on the estimation is given by the Cramer-Rao inequality:

$$V(t) = \sigma^2_{\hat{\theta}} = E\left[(t-\theta)^2\right] \geq \frac{1}{A(\underline{x},\theta)}$$

$$A(\underline{x},\theta) = E\left[\left(\frac{\partial\, log\, \mathcal{L}}{\partial\theta}\right)^2\right] = E\left[\left(\frac{1}{\mathcal{L}}\frac{\partial \mathcal{L}}{\partial\theta}\right)^2\right] = -E\left[\frac{\partial^2 log\, \mathcal{L}}{\partial\theta^2}\right]$$

Efficient estimator: variance = minimal variance

$$\boxed{\frac{\partial\, log\, \mathcal{L}}{\partial\theta} = A(\theta)\times(t-\theta)}$$

↑ **indépendent of $\underline{x}$**

$$-\frac{\partial^2 log\, \mathcal{L}}{\partial\theta^2} = -\frac{\partial A}{\partial\theta}(t-\theta) + A(\theta)$$

$$E\left[-\frac{\partial^2 log\, \mathcal{L}}{\partial\theta^2}\right] = -\frac{\partial A}{\partial\theta}E[t-\theta] + A(\theta) = A(\theta)$$

↑ **=0 if unbiased estimator**

$$\boxed{V(t) = \sigma^2_{\hat{\theta}} = \frac{1}{A(\theta)}}$$

Sufficiency

The whole information about θ is contained in the estimator.

Condition : $\mathcal{L}(\underline{x}\,|\,\theta) = \prod_{i=1}^{n} f(x_i\,|\,\theta) = g(\underline{x}) \cdot h(t(\underline{x})\,|\,\theta)$

no information on θ ↑ ↑ depends on $\underline{x}$ through estimator t

Sufficiency condition $\dfrac{\partial \log \mathcal{L}}{\partial\theta} = \dfrac{\partial h(t(\underline{x})\,|\,\theta)}{\partial\theta}$ **contained in**

Efficiency condition $\dfrac{\partial \log \mathcal{L}}{\partial\theta} = A(\theta) \times (t - \theta)$

VII - Maximum Likelihood

Principle of the maximum likelihood method

Random set of n measurements $\underline{x} = (x_1, x_2, \ldots, x_n)$ extracted from a population defined by $f(x \mid \underline{\theta}_0)$ with k unknown true parameters $\underline{\theta}_0 = (\theta_{0,1} \cdots \theta_{0,k})$ to be estimated from the sample.

Likelihood function $\mathcal{L}(\underline{x} \mid \underline{\theta}) = \prod_{i=1}^{n} f(x_i \mid \underline{\theta})$ calculable for any set of values $\underline{\theta}$

Estimation $\hat{\underline{\theta}}$ of $\underline{\theta}$ maximises $\mathcal{L}(\underline{x} \mid \underline{\theta})$ and thus $log\,\mathcal{L}(\underline{x} \mid \underline{\theta})$, given $\underline{x}$

$$\left\{ \begin{aligned} \left.\frac{\partial\, log\, \mathcal{L}(\underline{\theta})}{\partial \theta_j}\right|_{\underline{\theta}=\hat{\underline{\theta}}} &= \sum_{i=1}^{n} \left.\frac{\partial\, log\, f(x_i \mid \underline{\theta})}{\partial \theta_j}\right|_{\underline{\theta}=\hat{\underline{\theta}}} = 0 \\ \left.\frac{\partial^2 log\, \mathcal{L}(\underline{\theta})}{\partial \theta_j^{\,2}}\right|_{\underline{\theta}=\hat{\underline{\theta}}} &= \sum_{i=1}^{n} \left.\frac{\partial^2 log\, f(x_i \mid \underline{\theta})}{\partial \theta_j^{\,2}}\right|_{\underline{\theta}=\hat{\underline{\theta}}} < 0 \end{aligned} \right\} j = 1, k$$

Invariance of the solution

Conservation of probability $\mathcal{L}(\underline{x} \mid \theta) = \mathcal{L}(\underline{x} \mid \tau(\theta))$

Call $\tau^* = \tau(\hat{\theta})$

$$\mathcal{L}(\tau^*) = \mathcal{L}(\tau(\hat{\theta})) = \mathcal{L}(\hat{\theta}) \geq \mathcal{L}(\theta) = \mathcal{L}(\tau(\theta))$$

$$\mathcal{L}(\tau^*) \geq \mathcal{L}(\tau) \quad \forall \ \tau \qquad \Rightarrow \tau^* = \hat{\tau}$$

If $\tau = \tau(\theta)$ **univocal and reciprocal** $\Rightarrow \hat{\tau} = \tau(\hat{\theta})$

Example of analytic solution: mean and variance of a gaussian

$\underline{x} = (x_1, x_2, \dots, x_n)$ sample from $f(x \mid \mu, \sigma^2) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{1}{2}\frac{(x-\mu)^2}{\sigma^2}}$

$$\mathcal{L}(\underline{x} \mid \mu, \sigma^2) = \left(\frac{1}{\sqrt{2\pi\sigma^2}}\right)^n \prod_{i=1}^{n} e^{-\frac{1}{2}\frac{(x_i-\mu)^2}{\sigma^2}}$$

$$\log \mathcal{L}(\mu, \sigma^2) = -\frac{n}{2}\left(\log 2\pi + \log \sigma^2 + \frac{1}{n}\sum_{i=1}^{n}\frac{(x_i-\mu)^2}{\sigma^2}\right)$$

$$\frac{\partial \log \mathcal{L}}{\partial \mu} = \frac{\sum_{i=1}^{n} x_i - n\mu}{\sigma^2} = 0 \quad \Rightarrow \quad \hat{\mu} = \bar{x} = \frac{1}{n}\sum_{i=1}^{n} x_i$$

$$\frac{\partial \log \mathcal{L}}{\partial \sigma^2} = -\frac{n}{2\sigma^2} + \frac{\sum_{i=1}^{n}(x_i-\mu)^2}{2\sigma^4} = 0$$

replace μ by $\bar{x}$: $-\frac{n}{2\sigma^2} + \frac{\sum_{i=1}^{n}(x_i-\bar{x})^2}{2\sigma^4} = 0 \quad \Rightarrow \quad \hat{\sigma}^2 = s'^2 = \frac{1}{n}\sum_{i=1}^{n}(x_i-\bar{x})^2$

s'^2 biased: $E\left[s'^2\right] = \frac{n-1}{n}\sigma^2$

Example of analytic solution: weighted mean and standard error

$\hat{\theta}_1,\dots,\hat{\theta}_n$ n estimations of θ with standard errors $\sigma_1,\dots,\sigma_n$

How to combine to measurements into $\hat{\theta}\pm\sigma_{\hat{\theta}}$?

Each $\hat{\theta}_i$ is extracted from PDF $N\left(\theta_0,\sigma_i^2\right)$

$$\mathcal{L}\left(\underline{\hat{\theta}}\,|\,\theta\right)=\prod_{i=1}^{n}\frac{1}{\sqrt{2\pi\sigma_i^2}}e^{-\frac{1}{2}\frac{\left(\hat{\theta}_i-\theta\right)^2}{\sigma_i^2}}$$

$$\log\mathcal{L}=-\frac{1}{2}\sum_{i=1}^{n}\frac{\left(\hat{\theta}_i-\theta\right)^2}{\sigma_i^2}+Cste$$

$$\frac{\partial\log\mathcal{L}}{\partial\theta}=\sum_{i=1}^{n}\frac{\hat{\theta}_i-\theta}{\sigma_i^2}=0$$

$$\hat{\theta}=\frac{\sum_{i=1}^{n}\frac{\hat{\theta}_i}{\sigma_i^2}}{\sum_{i=1}^{n}\frac{1}{\sigma_i^2}}$$

weighted means with weights $\frac{1}{\sigma_i^2}$

$$\frac{\partial\log\mathcal{L}}{\partial\theta}=\sum_{i=1}^{n}\frac{1}{\sigma_i^2}\left(\hat{\theta}-\theta\right)\;\Rightarrow\;A(\theta)=\sum_{i=1}^{n}\frac{1}{\sigma_i^2}\;\Rightarrow\;\sigma_{\hat{\theta}}^2=\frac{1}{\sum_{i=1}^{n}\frac{1}{\sigma_i^2}}\;\Rightarrow\;\sigma_{\hat{\theta}}=\sqrt{\frac{1}{\sum_{i=1}^{n}\frac{1}{\sigma_i^2}}}$$

Asymptotic Consistency, Efficiency, Sufficiency, Normality of $\mathcal{L}$

For $n \to \infty$:

consistency: $lim_{n\to\infty} \hat{\theta} = \theta_0$

efficiency : $\dfrac{\partial \log \mathcal{L}}{\partial \theta} = A(\theta) \times (t - \theta)$ and $V(t) = \sigma_\theta^2 = \dfrac{1}{A(\theta)}$

sufficiency : from efficiency

normality : $\mathcal{L}(\theta \mid \underline{x}) = \prod_{i=1.}^{n} f(\theta \mid x_i)$ takes the shape $\mathcal{L}(\theta) = N(\hat{\theta}, \sigma_\theta^2) = \dfrac{1}{\sqrt{2\pi\sigma_\theta^2}} e^{-\frac{1}{2}\frac{(\hat{\theta}-\theta)^2}{\sigma_\theta^2}}$

$$\log \mathcal{L}(\theta) = -\frac{1}{2}\frac{(\hat{\theta}-\theta)^2}{\sigma_\theta^2} - \kappa$$

$$\frac{\partial \log \mathcal{L}(\theta)}{\partial \theta} = \frac{\hat{\theta}-\theta}{\sigma_\theta^2}$$

$$\boxed{\sigma_\theta^2 = -\left(\frac{\partial^2 \log \mathcal{L}(\theta)}{\partial \theta^2}\right)^{-1}}$$

Analytical resolution possible only if $\dfrac{\partial \log \mathcal{L}(\theta)}{\partial \theta}$ and $\dfrac{\partial^2 \log \mathcal{L}(\theta)}{\partial \theta^2}$ are analytical

Asymptotic Consistency, Efficiency, Sufficiency, Normality of $\mathcal{L}$

If k parameters $\underline{\theta} = (\theta_1, \ldots, \theta_k)$

$$\left.\boxed{\begin{aligned} &\frac{\partial \log \mathcal{L}(\underline{\theta})}{\partial \theta_i} = \frac{\hat{\theta}_i - \theta_i}{\sigma^2_{\theta_i}} \\ &\sigma^2_{\theta_j} = -\left(\frac{\partial^2 \log \mathcal{L}(\underline{\theta})}{\partial {\theta_j}^2}\right)^{-1} \\ &\sigma_{\theta_i \theta_j} = -\left(\frac{\partial^2 \log \mathcal{L}(\underline{\theta})}{\partial \theta_i \partial \theta_j}\right)^{-1} \end{aligned}}\right\} i, j = 1, k$$

Asymptotic normality : numerical estimation of the confidence interval

If $\partial_\theta \log \mathcal{L}(\theta)$ and $\partial^2_{\theta^2} \log \mathcal{L}(\theta)$ cannot be calculated analitically

$$\log \mathcal{L}^*(\theta) = \log \mathcal{L}(\theta) + \kappa = -\frac{1}{2}\frac{\left(\hat{\theta}-\theta\right)^2}{\sigma_\theta^2}$$ **is a parabol**

Point estimation:

$$\hat{\theta} \Leftrightarrow \log \mathcal{L}^*\left(\hat{\theta}\right) = 0$$

Standard error = Confidence interval at $C.L. = 0.683$

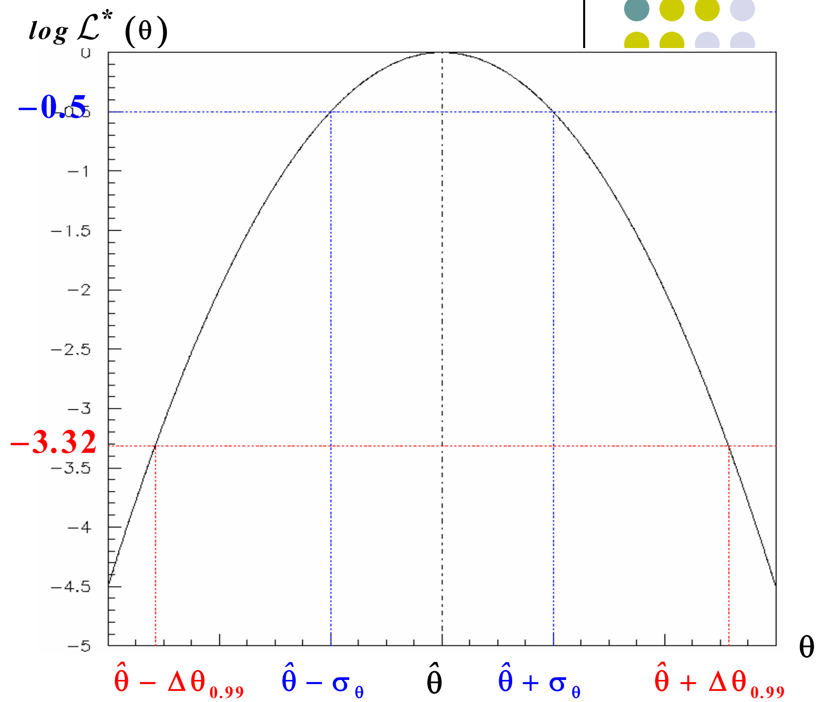
$$\log \mathcal{L}^*\left(\theta^\pm\right) = -\frac{1}{2} \Rightarrow \sigma_\theta = \left|\theta^\pm - \hat{\theta}\right|$$

Confidence interval at $C.L. = \alpha$

$$r_\alpha \rightarrow \alpha = \int_{-r_a}^{r_a} N(0,1)\,dx \text{ or equivalently } \int_0^{r_\alpha^2} f\left(\chi_1^2\right) d\chi_1^2$$

$$\log \mathcal{L}^*\left(\theta^\pm\right) = -\frac{r_\alpha^2}{2} \Rightarrow \Delta\theta_\alpha = \left|\theta^\pm - \hat{\theta}\right| \text{ at } C.L. = \alpha$$

Scaling relation : $\Delta\theta_\alpha = r_\alpha \sigma_\theta = \frac{r_\alpha}{r_{\alpha'}} \Delta\theta_{\alpha'}$



α	r_α	$r_\alpha^2/2$
0.683	1	0.5
0.900	1.65	1.35
0.950	1.96	1.92
0.990	2.58	3.32
0.999	3.29	5.41

Small samples : numerical estimation of the confidence interval

n small $\Rightarrow \log \mathcal{L}^*(\theta) \neq$ parabol

Assume $\tau = \tau(\theta)$ univocal and reciprocal such that $\log \mathcal{L}^*(\tau) = -\frac{1}{2}\frac{(\hat{\tau}-\tau)^2}{\sigma_\tau^2}$

$$\hat{\tau} \Leftrightarrow \log \mathcal{L}^*(\hat{\tau}) = 0$$

$$\log \mathcal{L}^*(\tau^\pm) = -\frac{1}{2} \Rightarrow \sigma_\tau = |\tau^\pm - \hat{\tau}|\sigma_\tau$$

Standard error results from probability conservation:

$$\log \mathcal{L}^*(\tau^+) = \log \mathcal{L}^*\left(\tau^{-1}(\tau^+) = \theta^+\right) = -\frac{1}{2}, \quad \sigma_\theta^+ = \theta^+ - \hat{\theta}$$

$$\log \mathcal{L}^*(\tau^-) = \log \mathcal{L}^*\left(\tau^{-1}(\tau^-) = \theta^+\right) = -\frac{1}{2}, \quad \sigma_\theta^- = \theta^- - \hat{\theta}$$

$$P\left(\theta_0 \in [\theta^-, \theta^+]\right) = 0.683$$

$$\theta = \hat{\theta}^{+\sigma_\theta^+}_{-\sigma_\theta^-}$$

Confidence interval at $C.L. = \alpha$: get $r_\alpha \rightarrow \alpha = \int_{-r_a}^{r_a} N(0,1)\,dx$ or equivalently $\int_0^{r_\alpha^2} f(\chi_1^2)\,d\chi_1^2$

$$\left.\begin{array}{l} \log \mathcal{L}^*(\theta_\alpha^+) = -\frac{r_\alpha^2}{2} \Rightarrow \Delta\theta_\alpha^+ = |\theta_\alpha^+ - \hat{\theta}| \\ \log \mathcal{L}^*(\theta_\alpha^-) = -\frac{r_\alpha^2}{2} \Rightarrow \Delta\theta_\alpha^- = |\theta_\alpha^- - \hat{\theta}| \end{array}\right\} \text{at } C.L. = \alpha \Rightarrow P\left(\theta_0 \in [\theta_\alpha^-, \theta_\alpha^+]\right) = \alpha$$

Example of small sample : confidence interval on the variance of a normal PDF

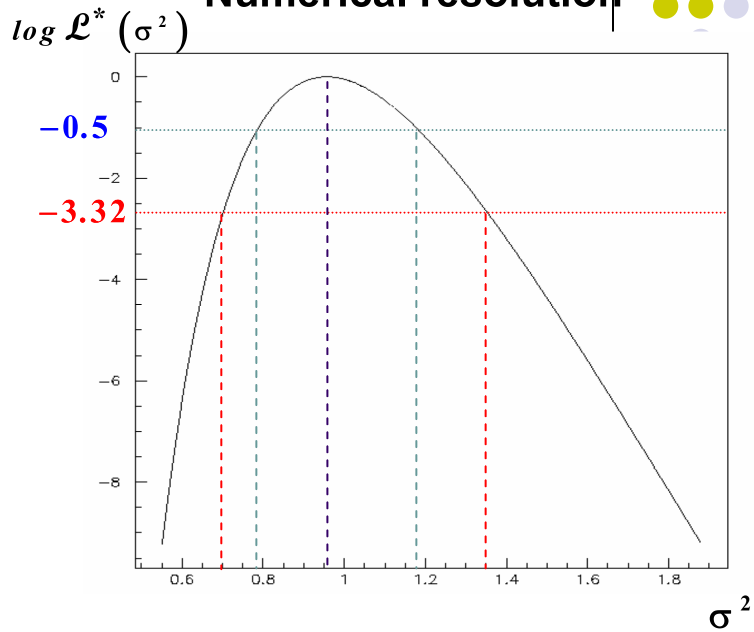
Random sample of $n = 100$ extracted from $N(0,1)$

Analytical resolution

$\sigma^2 = 0.96 \pm 0.14$

but $P\left(\sigma_0^2 \in [0.96 - 0.14, 0.96 + 0.14]\right) \neq 0.683$

Numerical resolution



but $\Delta\theta^+_{0.99} = 0.46 > \sigma^+ \times r_{0.99} = 0.15 \times 2.58 = 0.39$

$\Delta\theta^-_{0.99} = 0.28 < \sigma^- \times r_{0.99} = 0.12 \times 2.58 = 0.31$

Confidence interval at $C.L. = 0.683$: $0.96^{+0.15}_{-012}$

Confidence interval at $C.L. = 0.99$: $0.96^{+0.46}_{-028}$

No scaling relation between the size of the confidence interval and the $C.L.$

$$\Delta\theta^+_\alpha \neq r_\alpha \sigma^+_\theta \qquad \Delta\theta^-_\alpha \neq r_\alpha \sigma^-_\theta \qquad \frac{\Delta\theta^+_\alpha}{\sigma^+_\theta} \neq \frac{\Delta\theta^-_\alpha}{\sigma^-_\theta}$$

Asymptotic normality : extension to two independent variables

Asymptotic likelihood function for large samples and two independent variables $\theta=(\theta_1,\theta_2)$

$$\mathcal{L}(\theta_1,\theta_2 \mid \hat\theta_1,\hat\theta_2)=\frac{1}{2\pi\sqrt{\sigma_1^2\sigma_2^2}}e^{-\frac{1}{2}\left(\frac{(\theta_1-\hat\theta_1)^2}{\sigma_1^2}+\frac{(\theta_2-\hat\theta_2)^2}{\sigma_2^2}\right)}$$

$$log\, L^*(\theta_1,\theta_2)=log\, L(\theta_1,\theta_2)+\kappa=-\frac{1}{2}\sum_{i=1,}^{2}\frac{(\theta_i-\hat\theta_i)^2}{\sigma_i^2}$$ **is a paraboloid**

Point estimation: $\underline{\hat\theta} \Rightarrow log\, \mathcal{L}^*(\underline{\hat\theta})=0$

Confidence interval at $C.L.=\alpha$

$$-2\, log\, \mathcal{L}^*(\theta_1,\theta_2)=\sum_{i=1,}^{2}\frac{(\theta_i-\hat\theta_i)^2}{\sigma_i^2}$$ **follows a** χ_2^2 **PDF**

$$r_\alpha^2 \Rightarrow \int_0^{r_\alpha^2} f(\chi_2^2)\,d\chi_2^2=\alpha$$

α	r_α	$r_\alpha * r_\alpha/2$
0.393	1.00	0.50
0.632	1.41	1.00
0.683	1.51	1.14
0.865	2.00	2.00
0.900	2.14	2.30
0.950	2.45	3.00
0.990	3.03	4.60

confidence interval: area contained in the ellipse defined by

intersection of
- **paraboloid** $log\, \mathcal{L}^*(\theta_1,\theta_2)$
- **plane** $log\, \mathcal{L}^*(\theta_1,\theta_2)=-\frac{r_\alpha^2}{2}$

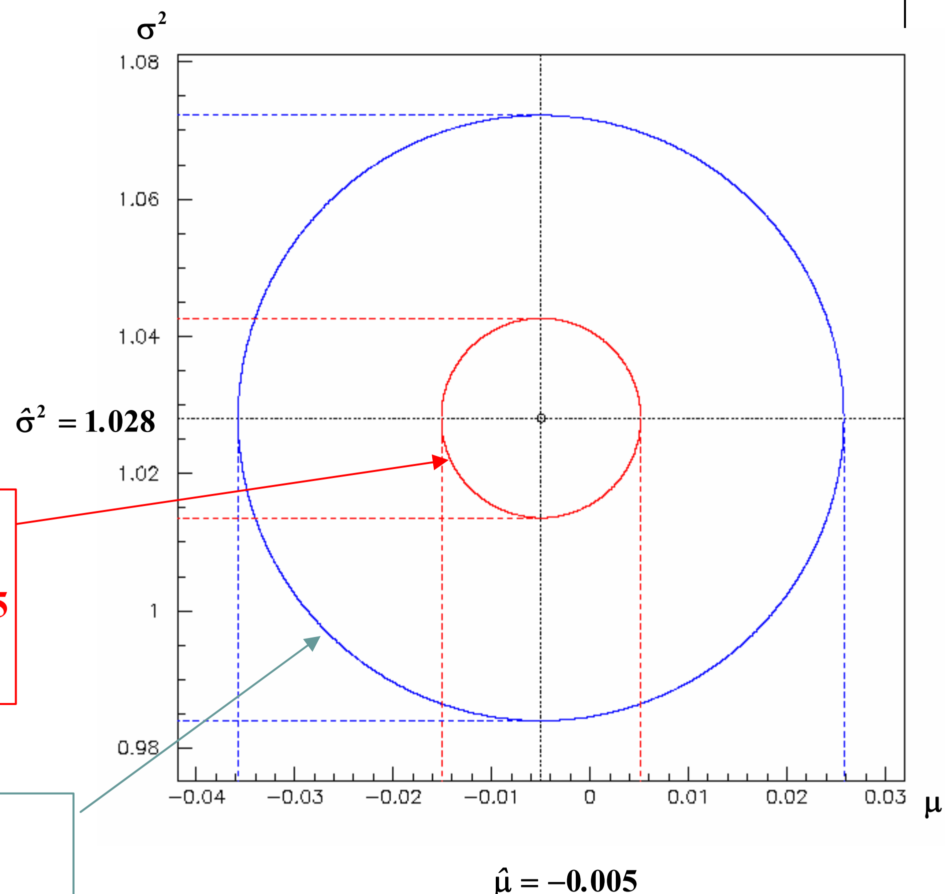
Example : confidence interval for mean and variance of a *N*(0,1)

Analytic method

$$\hat{\mu} = \bar{x} = 0.005 \pm 0.010$$

$$\hat{\sigma}^2 = s'^2 = 1.028 \pm 0.015$$

Numeric method



$r_\alpha = 1$

$$\log \mathcal{L}^*(\theta) = -\frac{r_\alpha^2}{2} = -0.5$$

$\alpha = 0.393$

$r_\alpha = 3.03$

$$\log \mathcal{L}^*(\theta) = -\frac{r_\alpha^2}{2} = -4.60$$

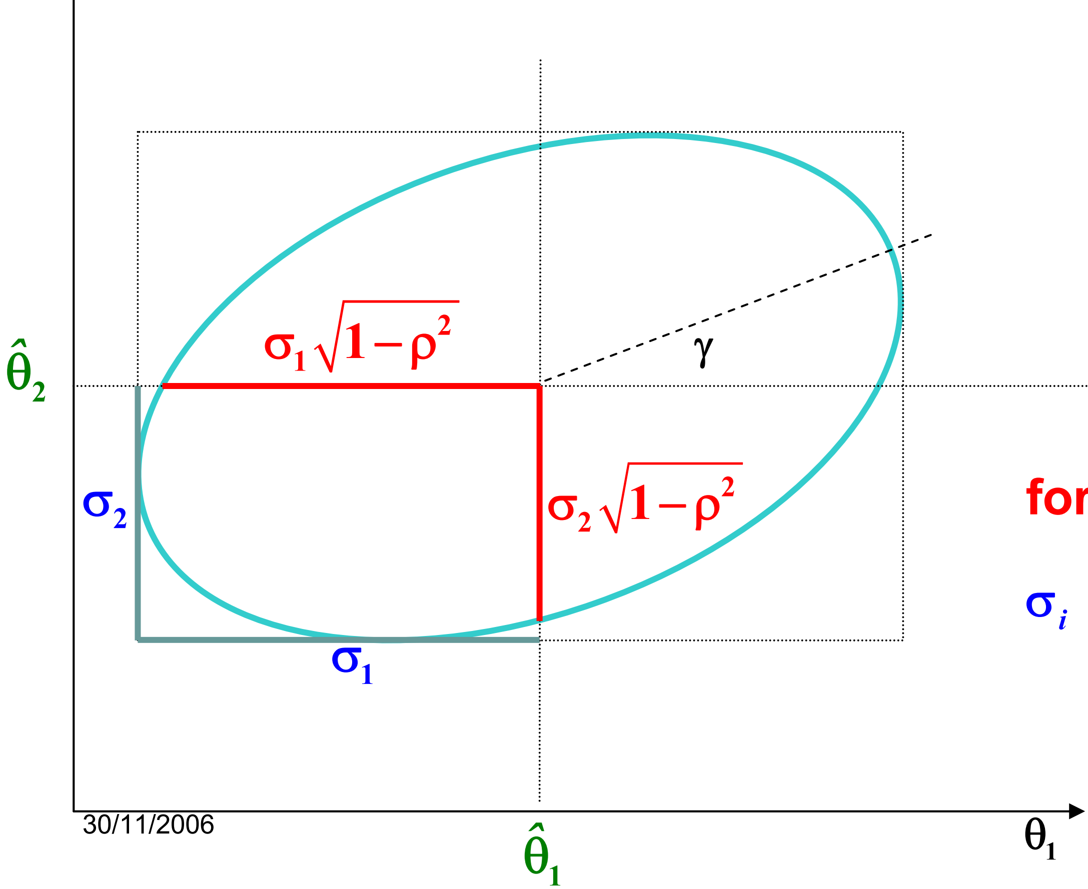
$\alpha = 0.99$

Asymptotic normality : extension to two correlated variables

Contour $log\ \mathcal{L}^*(\theta_1,\theta_2) = -\frac{1}{2(1-\rho^2)}\left(\frac{(\theta_1-\hat\theta_1)^2}{\sigma_1^2}+\frac{(\theta_2-\hat\theta_2)^2}{\sigma_2^2}-2\rho\frac{(\theta_1-\hat\theta_1)}{\sigma_1}\frac{(\theta_2-\hat\theta_2)}{\sigma_2}\right) = -\frac{1}{2}$

ellipse centred on $(\hat\theta_1,\hat\theta_2)$ **with** $tan(2\gamma) = \frac{2\rho\sigma_1\sigma_2}{\sigma_1^2-\sigma_2^2}$

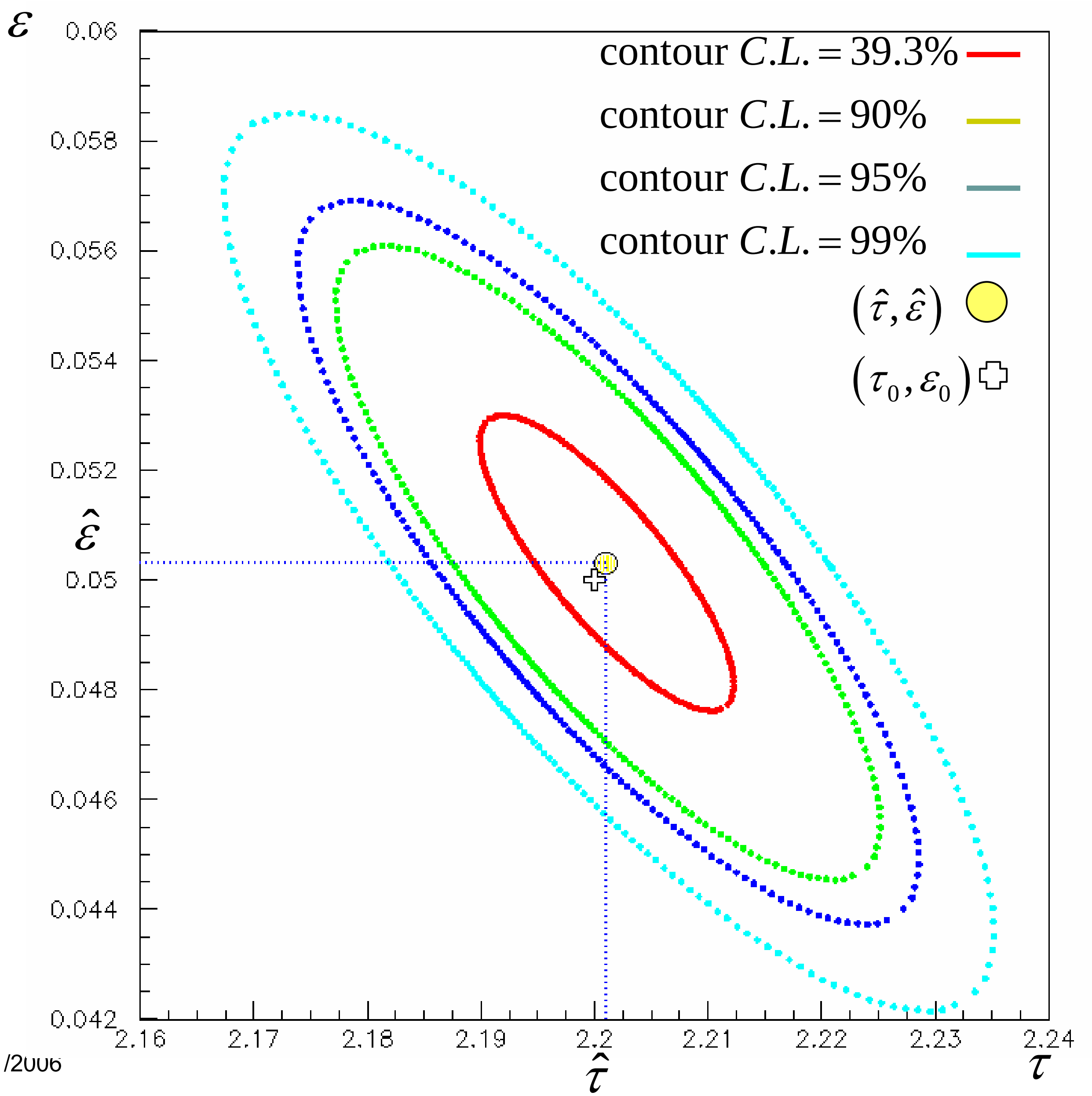
inscribed in rectangle $\hat\theta_1 \pm \sigma_1, \hat\theta_2 \pm \sigma_2$



for $\theta_i = \hat\theta_i\ :\ \left|\theta_j - \hat\theta_j\right| = \sigma_j\sqrt{1-\rho^2}$

$\sigma_i = \theta_i^{max} - \hat\theta_i = \hat\theta_i - \theta_i^{min}$

Asymptotic normality : extension to two correlated variables



Extension to small samples and *N* correlated variables

Method stays formally correct : $\chi^2_2 \rightarrow \chi^2_N$

$$r_\alpha \;\Rightarrow\; \int_0^{r_\alpha^2} f\left(\chi^2_N\right) d\chi^2_N = \alpha$$

Difficult in practice if n small, N large et correlations

VIII – Least Squares Method

Principle of the least squares method

- $y = f\left(x \mid \underline{\theta}_0\right)$ functional relation between variables y and x
- k unknown parameters $\underline{\theta} = \left(\theta_1, \theta_2, \ldots, \theta_k\right)$ of true values $\underline{\theta}_0$
- $\underline{y} = \left(y_1, \ldots, y_n\right)$ the $n > k$ measured values of $f\left(x \mid \underline{\theta}\right)$ at points $\underline{x} = \left(x_1, \ldots, x_n\right)$ with standard errors $\underline{\sigma} = \left(\sigma_1, \ldots, \sigma_n\right)$
- PDF of y_i : $N\left(f\left(x_i \mid \underline{\theta}_0\right), \sigma_i^2\right)$

Estimations $\hat{\underline{\theta}}$ of $\underline{\theta}_0$ minimise

$$X^2\left(\underline{\theta}\right) = \sum_{i=1}^{n} \frac{\left(y_i - f\left(x_i \mid \underline{\theta}\right)\right)^2}{\sigma_i^2} \quad \text{given } \underline{y} \pm \underline{\sigma}$$

Example : histogram

$$X^2\left(\underline{\theta}\right) = \sum_{i=1}^{N} \frac{\left(n_i - n p_{0i}\left(\underline{\theta}\right)\right)^2}{n p_i\left(\underline{\theta}\right)}$$

$n_i =$ number of events in class $\left[X_i \le x \le X_{i+1}\right]$ $\qquad n = \sum_{i=1}^{N} n_i$

$$p_{0i}\left(\underline{\theta}\right) = \int_{X_i}^{X_{i+1}} f\left(x \mid \underline{\theta}\right) dx$$

Equivalence between least squares and maximum likelihood for large samples

Large samples $\Rightarrow$ gaussian approximation of $\mathcal{L}$

$$\mathcal{L}(\underline{\theta}) = \prod_{i=1}^{n} \frac{1}{\sqrt{2\pi\sigma_i^2}} e^{-\frac{1}{2}\frac{\left(y_i - f(x_i|\underline{\theta})\right)^2}{\sigma_i^2}}$$

$$log\,\mathcal{L}^*(\underline{\theta}) = -\frac{1}{2}\sum_{i=1}^{n} \frac{\left(y_i - f(x_i|\underline{\theta})\right)^2}{\sigma_i^2}$$

$$X^2(\underline{\theta}) = \sum_{i=1}^{n} \frac{\left(y_i - f(x_i|\underline{\theta})\right)^2}{\sigma_i^2}$$

$$-2\,log\,\mathcal{L}^*(\underline{\theta}) = X^2(\underline{\theta})$$

The least squares method is asymptotically coherent, efficient and sufficient

Maximum likelihood

intersection of $log\,\mathcal{L}^*(\underline{\theta})$ **with hyperplan parallel to** $(\underline{\theta})$ **at** $log\,\mathcal{L}^*(\underline{\theta}) = -r_\alpha^2/2$

Least squares

intersection of $X^2(\underline{\theta})$ **with hyperplan parallel to** $(\underline{\theta})$ **at** $X^2(\underline{\theta}) = Min\left(X^2(\underline{\theta})\right) + r_\alpha^2$

$$r_\alpha^2 \Rightarrow \int_0^{r_\alpha^2} f(\chi_k^2)\, d\chi_k^2 = \alpha$$

Analytical resolution of the linear model

- $y = f(x) = \sum_{l=1}^{L} a_l(x)\,\theta_l$

$$X^2(\underline{\theta}) = \sum_{n=1}^{N} \frac{\left(y_n - \sum_{l=1}^{L} a_{nl}\,\theta_l\right)^2}{\sigma_n^2}$$ with $a_{nl} = a_l(x_n)$

$$= \left(\underline{y} - A\underline{\theta}\right)^T V^{-1} \left(\underline{y} - A\underline{\theta}\right)$$ **Matrix notation also valid for correlated parameters**

$$A = \begin{pmatrix} a_{11} & \cdots & a_{1L} \\ \vdots & \ddots & \vdots \\ a_{N1} & \cdots & a_{NL} \end{pmatrix}$$ and $$V = \begin{pmatrix} \sigma_1^2 & \cdots & \sigma_{1N} \\ \vdots & \ddots & \vdots \\ \sigma_{N1} & \cdots & \sigma_N^2 \end{pmatrix}$$

Point estimations

$$\frac{\partial X^2(\underline{\theta})}{\partial \underline{\theta}} = -2A^T V^{-1}\underline{y} + 2A^T V^{-1} A\underline{\theta} = 0 \quad \Rightarrow \quad \boxed{\hat{\underline{\theta}} = \left(A^T V^{-1} A\right)^{-1} A^T V^{-1} \underline{y}}$$

Variances

$$\hat{\underline{\theta}} = f(\underline{y})$$

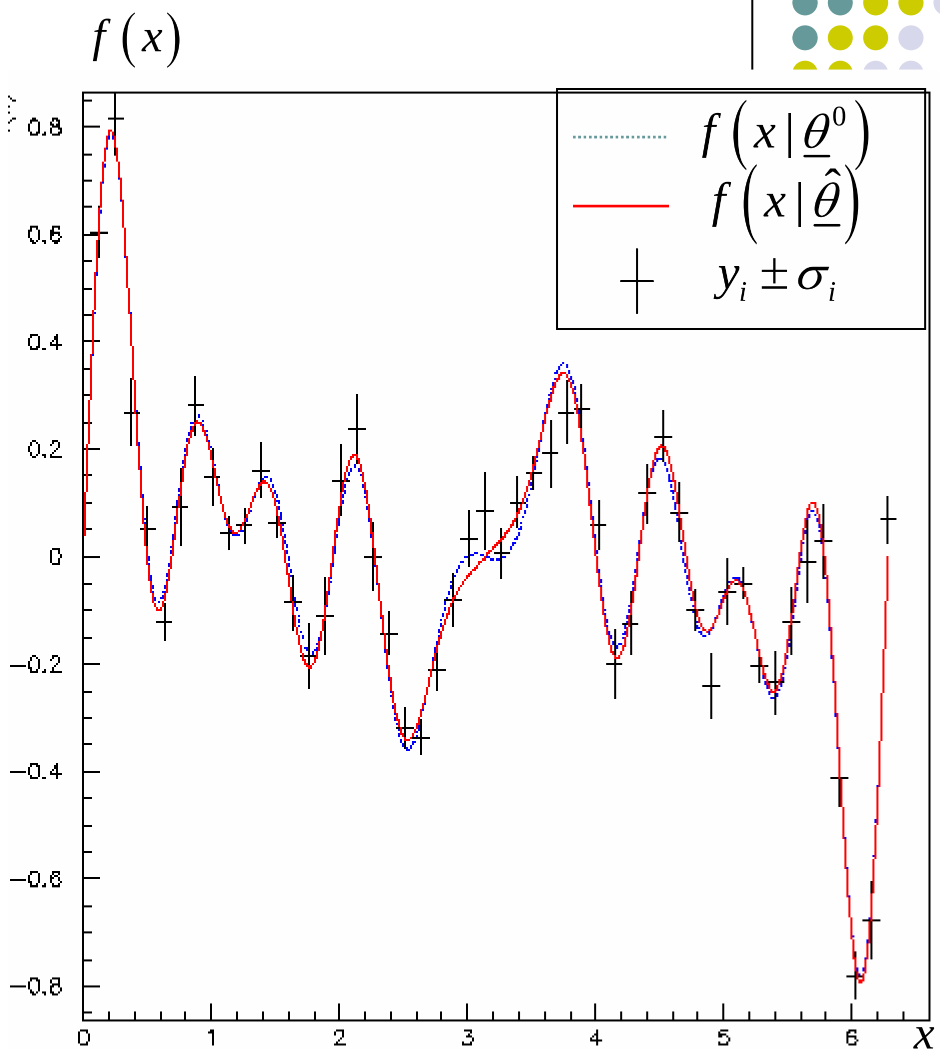
$$V(\hat{\underline{\theta}}) = \left(\frac{\partial \hat{\underline{\theta}}}{\partial \underline{y}}\right) V \left(\frac{\partial \hat{\underline{\theta}}}{\partial \underline{y}}\right)^T = \left(\left(A^T V^{-1} A\right)^{-1} A^T V^{-1}\right) V \left(\left(A^T V^{-1} A\right)^{-1} A^T V^{-1}\right)^T$$

$$\boxed{V(\hat{\underline{\theta}}) = \left(A^T V^{-1} A\right)^{-1}}$$

Example: superposition de 10 sinusoids of known frequencies

- $f\left(x \mid \underline{\theta}^0\right) = \sum_{l=1}^{10} sin\left(\omega_l x\right)\, \theta_l^0$ $\qquad L = 10$
- $y_i \pm \sigma_i \,, i = 1, N = 50$

$\underline{\omega}$	$\underline{\theta}^0$	$\underline{\hat{\theta}}$	$\underline{\sigma}\left(\hat{\theta}\right)$
1.0	0.040	0.034	0.010
2.0	0.182	0.175	0.010
3.0	0.026	0.028	0.010
4.0	0.123	0.126	0.009
5.0	0.041	0.028	0.010
6.0	0.116	0.130	0.010
7.0	0.174	0.178	0.009
8.0	0.032	0.037	0.010
9.0	0.158	0.149	0.009
10.0	0.107	0.116	0.010



The non-linear model with constraints

System of $M = N + L$ paremeters

- N measurable paremeters $\underline{\eta} = (\eta_1, \eta_2, \ldots, \eta_N)$
  - estimations $\underline{\hat{\eta}} = (\hat{\eta}_1, \hat{\eta}_2, \ldots \hat{\eta}_N)$
  - covariance matrix $V_{\hat{\eta}}$
- L non-measurable paremeters $\underline{\theta} = (\theta_1, \theta_2, \ldots, \theta_L)$
- $K > L$ constraints between the M paremeters $\underline{f}(\underline{\eta}, \underline{\theta}) = 0$

Purpose :

- improved estimation of the N measurable paremeters $\underline{\hat{\hat{\eta}}}$
- estimation of the L non-measurable paremeters $\underline{\hat{\theta}}$
- calculation of the covariance matrix between the $M = N + L$ paremeters

The non-linear model with constraints : Example

Kinematical analysis of 2-body $\rightarrow$ F-body

$$M = 2 + F \text{ bodies}$$

Number of variables : $4 \times M$

$$\left(p_m, \theta_m, \phi_m, E_m\right), m = 1, M$$

$K = $ **4 constraints : energy-momentum conservation**

$$\sum_{i=1}^{2} p_i \sin\theta_i \cos\phi_i - \sum_{f=1}^{F} p_f \sin\theta_f \cos\phi_f = 0$$

$$\sum_{i=1}^{2} p_i \sin\theta_i \sin\phi_i - \sum_{f=1}^{F} p_f \sin\theta_f \sin\phi_f = 0$$

$$\sum_{i=1}^{2} p_i \cos\theta_i - \sum_{f=1}^{F} p_f \cos\theta_f = 0$$

$$\sum_{i=1}^{2} E_i - \sum_{f=1}^{F} E_f = 0$$

L **unmeasurable variables**

$L > 4$ **: indefined system**

$L = 4$ **: solvable system**

$L < 4$ **: ajustable system with least squares method**

If the particles are identified, there masses are known and the number of variables is reduced to 3 $\times M$ **or equivalently, there** M **additionnal constraints** $E_m^2 = p_m^2 + m_m^2, m = 1, M$.

Lagrange parameters method

$$\left.\begin{array}{l} K \text{ constraints } \underline{f}\left(\underline{\eta},\underline{\theta}\right)=0 \\ \text{minimisation of } X^2=\left(\underline{\hat{\eta}}-\underline{\eta}\right)^T V_{\hat{\eta}}^{-1}\left(\underline{\hat{\eta}}-\underline{\eta}\right) \end{array}\right\}\Rightarrow$$

minimisation of $X'^2=\left(\underline{\hat{\eta}}-\eta\right)^T V_{\hat{\eta}}^{-1}\left(\underline{\hat{\eta}}-\eta\right)+2\underline{\lambda}^T \underline{f}\left(\underline{\eta},\underline{\theta}\right)$

K additionnal unknown parametres of Lagrange $\underline{\lambda}$

Minimisation of X'^2:

$$\left.\begin{array}{l} \dfrac{\partial X'^2}{\partial \underline{\eta}}=-2V_{\hat{\eta}}^{-1}\left(\underline{\hat{\eta}}-\eta\right)+2f_{\eta}^T\underline{\lambda}=0 \\ \dfrac{\partial X'^2}{\partial \underline{\theta}}=2f_{\theta}^T\underline{\lambda}=0 \\ \dfrac{\partial X'^2}{\partial \underline{\lambda}}=2\underline{f}=0 \end{array}\right\}\begin{array}{l} N+K+L \text{ normal equations} \\ N+K+L \text{ unknowns} \\ \underline{\eta},\underline{\theta},\underline{\lambda} \end{array}$$

$$\text{with}\left\{\begin{array}{ll} \left(f_{\eta}\right)_{kn}=\dfrac{\partial f_k}{\partial \eta_n} & K\times N \text{ matrix} \\ \left(f_{\theta}\right)_{kl}=\dfrac{\partial f_k}{\partial \theta_l} & K\times L \text{ matrix} \end{array}\right.$$

Lagrange parameters method : iterative procedure

$\underline{f}^{(\nu)} \equiv \underline{f}\left(\underline{\eta}^{(\nu)}, \underline{\theta}^{(\nu)}\right)$ at itération ν

$$\left.\begin{aligned} & V_{\hat{\eta}}^{-1}\left(\underline{\eta}^{(\nu+1)} - \hat{\underline{\eta}}\right) + f_{\eta}^{(\nu)T}\underline{\lambda}^{(\nu+1)} = 0 \\ & f_{\theta}^{(\nu)T}\underline{\lambda}^{(\nu+1)} = 0 \\ & \underline{f}^{(\nu+1)} = \underline{f}^{(\nu)} + f_{\eta}^{(\nu)}\left(\underline{\eta}^{(\nu+1)} - \underline{\eta}^{(\nu)}\right) + f_{\theta}^{(\nu)}\left(\underline{\theta}^{(\nu+1)} - \underline{\theta}^{(\nu)}\right) \end{aligned}\right\}$$

$N+K+L$ normal equations

$N+K+L$ unknowns $\underline{\eta}^{(\nu+1)}, \underline{\theta}^{(\nu+1)}, \underline{\lambda}^{(\nu+1)}$

Iteration 0

$\underline{\eta}^{(0)} = \hat{\underline{\eta}}$

$\underline{\theta}^{(0)}$: solution of a subset of L among K constraints

$\underline{\lambda}^{(0)} = \underline{I}_K$ arbitrarily

Conditions to stop the iterative process at iteration ν+1

$$\left\{\begin{aligned} & \left|\theta_l^{(\nu+1)} - \theta_l^{(\nu)}\right| \text{ or } \frac{\left|\theta_l^{(\nu+1)} - \theta_l^{(\nu)}\right|}{\theta_l^{(\nu)}} \le \varepsilon_l \quad \forall l = 1, L \\ & \left|\eta_n^{(\nu+1)} - \eta_n^{(\nu)}\right| \text{ or } \frac{\left|\eta_n^{(\nu+1)} - \eta_n^{(\nu)}\right|}{\eta_n^{(\nu)}} \le \varepsilon'_n \quad \forall n = 1, N \\ & f_k\left(\underline{\eta}^{(\nu+1)}, \theta^{(\nu+1)}\right) \le \varepsilon''_k \quad \forall k = 1, K \end{aligned}\right\} \Rightarrow \boxed{\begin{aligned} \hat{\underline{\theta}} &= \theta^{(\nu+1)} \\ \hat{\underline{\eta}} &= \underline{\eta}^{(\nu+1)} \end{aligned}}$$

Lagrange parameters method : solution

Point solutions

$$\left.\begin{aligned}\underline{\theta}^{(\nu+1)} &= \underline{\theta}^{(\nu)} - \boldsymbol{H}^{-1} \boldsymbol{f}_{\theta}^{(\nu)T} \boldsymbol{S}^{-1} \underline{\boldsymbol{r}} \\ \underline{\lambda}^{(\nu+1)} &= \boldsymbol{S}^{-1}\left(\underline{\boldsymbol{r}} - \boldsymbol{f}_{\theta}^{(\nu)}\left(\underline{\theta}^{(\nu+1)} - \underline{\theta}^{(\nu)}\right)\right) \\ \underline{\eta}^{(\nu+1)} &= \underline{\hat{\eta}} - \boldsymbol{V}_{\hat{\eta}}^{-1} \boldsymbol{f}_{\eta}^{(\nu)T} \underline{\lambda}^{(\nu+1)}\end{aligned}\right\} \quad \textbf{with} \quad \begin{aligned}\boldsymbol{S} &= \left(\boldsymbol{f}_{\eta}^{(\nu)} \boldsymbol{V}_{\hat{\eta}}^{-1} \boldsymbol{f}_{\eta}^{(\nu)T}\right) \\ \boldsymbol{H} &= \left(\boldsymbol{f}_{\theta}^{(\nu)T} \boldsymbol{S}^{-1} \boldsymbol{f}_{\theta}^{(\nu)}\right) \\ \underline{\boldsymbol{r}} &= \underline{\boldsymbol{f}}^{(\nu)} + \boldsymbol{f}_{\eta}^{(\nu)}\left(\underline{\hat{\eta}} - \underline{\eta}^{(\nu)}\right)\end{aligned}$$

Covariance matrix

$$\begin{aligned}\underline{\hat{\hat{\eta}}} &= \underline{\boldsymbol{g}}\left(\underline{\hat{\eta}}\right) \\ \hat{\theta} &= \boldsymbol{h}\left(\underline{\hat{\eta}}\right)\end{aligned} \Rightarrow \left.\begin{aligned}\boldsymbol{V}_{\hat{\hat{\eta}}} &= \left(\frac{\partial \underline{\boldsymbol{g}}}{\partial \underline{\hat{\eta}}}\right) \boldsymbol{V}_{\hat{\eta}} \left(\frac{\partial \underline{\boldsymbol{g}}}{\partial \underline{\hat{\eta}}}\right)^{T} = \boldsymbol{V}_{\hat{\eta}}\left(\underline{\boldsymbol{I}}_{N} - \left(\boldsymbol{G} - \boldsymbol{F}\boldsymbol{H}^{-1}\boldsymbol{F}^{T}\right)\boldsymbol{V}_{\hat{\eta}}\right) \\ \boldsymbol{V}_{\hat{\theta}} &= \left(\frac{\partial \underline{\boldsymbol{h}}}{\partial \underline{\hat{\theta}}}\right) \boldsymbol{V}_{\hat{\eta}} \left(\frac{\partial \underline{\boldsymbol{h}}}{\partial \underline{\hat{\theta}}}\right)^{T} = \boldsymbol{H}^{-1} \\ \boldsymbol{Cov}_{\hat{\eta}\hat{\theta}} &= \left(\frac{\partial \underline{\boldsymbol{g}}}{\partial \underline{\hat{\eta}}}\right) \boldsymbol{V}_{\hat{\eta}} \left(\frac{\partial \underline{\boldsymbol{h}}}{\partial \underline{\hat{\theta}}}\right)^{T} = \boldsymbol{V}_{\hat{\eta}} \boldsymbol{F}\boldsymbol{H}^{-1}\end{aligned}\right\} \quad \textbf{with} \quad \left\{\begin{aligned}\boldsymbol{G} &= \boldsymbol{f}_{\eta}^{T} \boldsymbol{S}^{-1} \boldsymbol{f}_{\eta} \\ \boldsymbol{H} &= \boldsymbol{f}_{\theta}^{T} \boldsymbol{S}^{-1} \boldsymbol{f}_{\theta} \\ \boldsymbol{F} &= \boldsymbol{f}_{\eta}^{T} \boldsymbol{S}^{-1} \boldsymbol{f}_{\theta}\end{aligned}\right.$$

Quality of the fit of the data to the model

Estimation methods provide values for model parameters that fit best to the experimental data. The best fit, however, may be a very poor fit. The quality of the fit requires an hypothesis test.

The only combination of an estimation method followed by an hypothesis test that is formally, but asymptotically, correct for large sample is:

1. **The least square method to estimate the *L* parameters of the model.**
2. **The Pearson χ^2 test to estimate the quality of the fit.**

   The PDF of statistic X^2 follows a χ^2 with number of degrees of freedom = ν-*L*.

If the maximum likelihood is used instead, the number of degrees of freedom is undefined in the range [ν, ν-*L*].

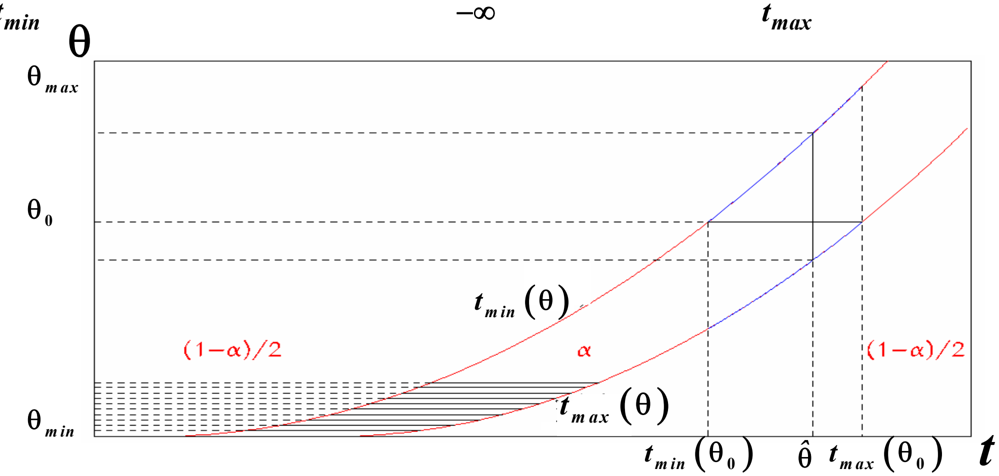
IX – Confidence intervals for pathological cases

- Small signals above small background.
- Measurements with standard errors extending over a physical limit.

Neyman prescription : belts

The Neyman centred belts are constructed with a particular prescription:

$$\int_{t_{min}}^{t_{max}} g(t\,|\,\theta)\,dt = \alpha \text{ and } \int_{-\infty}^{t_{min}} g(t\,|\,\theta)\,dt = \int_{t_{max}}^{\infty} g(t\,|\,\theta)\,dt = \frac{1-\alpha}{2}$$



There is an infinite number of prescriptions to construct belts with correct coverage corresponding to *C.L.* α that lead to different confidence intervals but are all equally correct from the statistical point of view.!

The Neyman upper and lower belts are constructed with the particular prescriptions:

$$\int_{t_{max}}^{+\infty} g(t\,|\,\theta)\,dt = \alpha \qquad \int_{-\infty}^{t_{min}} g(t\,|\,\theta)\,dt = \alpha$$

Gaussian error near physical limit:

Exemples :

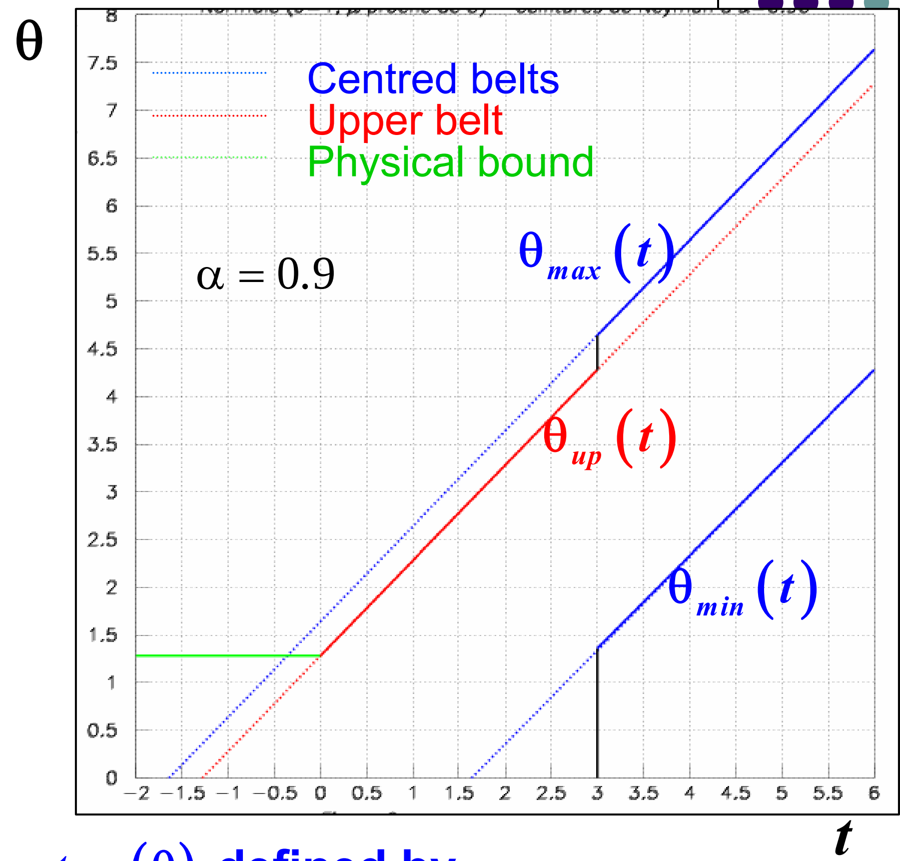
- sine of an angle compatible with being >1 within error,
- mass of a particle compatible with being <0 within error.

Appropriate change of variable:
Positive variable near physical bound 0 with standard error $\sigma = 1$

Two sensible prescriptions to build confidence intervals with exact coverage for a *C.L.* α :

- The centred Neyman belts defining lower and upper limits.
- Upper limit Neyman belt, the lower limit being the physical bound 0.

Gaussian error near physical limit: Neyman belts



Lower limit $\theta_{up}(t) \equiv t_{low}(\theta)$ defined by

$$\int_{t_{up}(\theta)}^{+\infty} N(t \mid \theta, 1)\, dt = \alpha$$

Centred limits $\theta_{max}(t) \equiv t_{min}(\theta)$ et $\theta_{min}(t) \equiv t_{max}(\theta)$ defined by

$$\int_{t_{min}(\theta)}^{t_{max}(\theta)} N(t \mid \theta, 1)\, dt = \alpha \text{ et } \int_{-\infty}^{t_{min}(\theta)} N(t \mid \theta, 1)\, dt = \int_{t_{max}(\theta)}^{\infty} N(t \mid \theta, 1)\, dt = \frac{1-\alpha}{2}$$

Belts are straight lines of slope $\pi/4$

Gaussian error near physical limit: Neyman hybrid belts

- If $\hat{\theta} > 3\sigma = 3$:

  $\theta > 0$ is sensible $\Rightarrow$

  use centred belts $\theta_a(t), \theta_b(t)$

  $P\left(\theta_0 \in \left[\theta_a\left(\hat{\theta}\right), \theta_b\left(\hat{\theta}\right)\right]\right) = \alpha = 0.9$

  if $\hat{\theta} = 4$: $P\left(\theta_0 \in [2.3, 5.6]\right) = \alpha = 0.9$

- If $\hat{\theta} < 3\sigma$:

  θ compatible with 0 is sensible $\Rightarrow$

  use upper belt $\theta_l(t)$

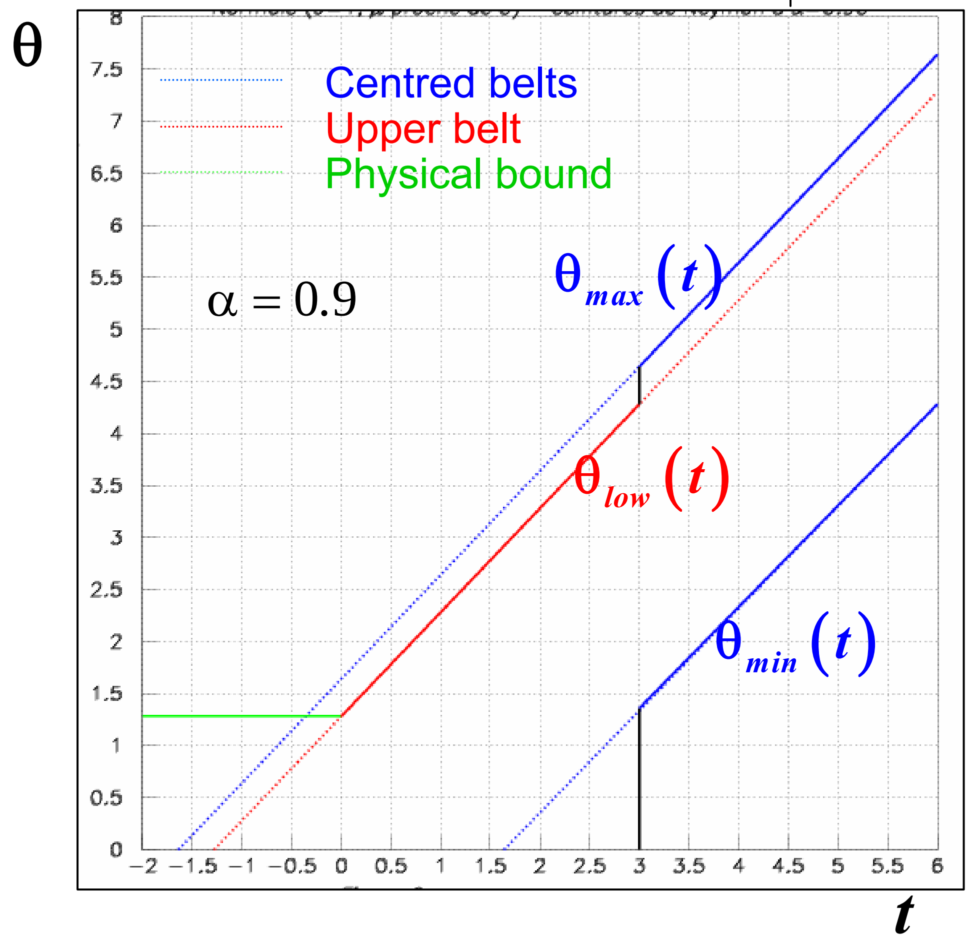
  $P\left(\theta_0 \le \theta_l\left(\hat{\theta}\right)\right) = \alpha$

  if $\hat{\theta} = 1$: $P\left(\theta_0 < 2.2\right) = 0.9$

- If $\hat{\theta} < 0$

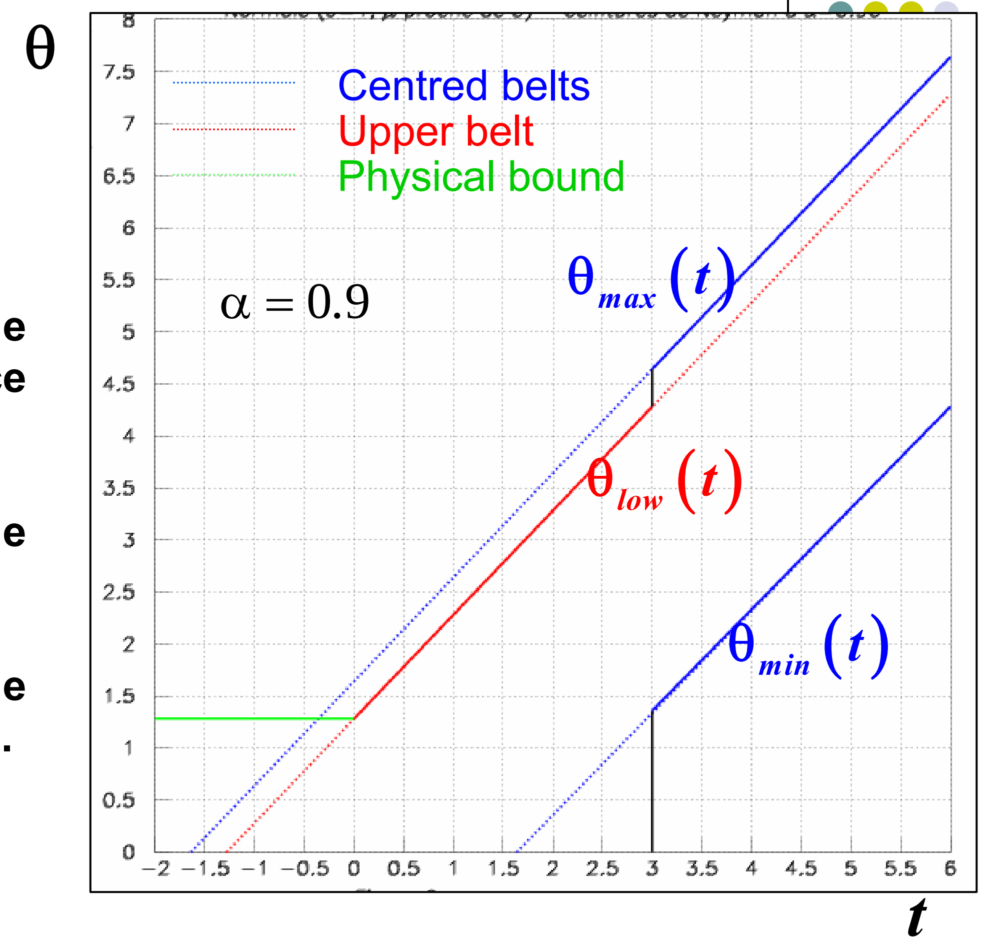
  $\theta = 0$

  $P\left(\theta_0 \le \theta_S(0) = 1.3\right) = 0.9$



Gaussian error near physical limit: what is wrong with Neyman hybrid belts?

- If $\hat{\theta} > 3$: $\theta_0 \in \left[\theta_{min}\left(\hat{\theta}\right), \theta_{max}\left(\hat{\theta}\right)\right]$
- If $0 < \hat{\theta} < 3$: $\theta_0 \leq \theta_{low}\left(\hat{\theta}\right)$
- If $\hat{\theta} < 0$: $\theta_0 < \theta_s\left(0\right)$



Choosing belts a priori is correct from the statistical point of view but the choice may be meaningless:

- **compute an upper limit when the measurement is clearly positive**
- **compute a pair of limits when the measurement is clearly compatible with 0.**

To make the choice a posteriori is statistically incorrect.

The coverage does not correspond $\alpha = 0.9$

Small event count with background

r_S : unknwon number of signal events - confidence interval to be calculated

r_B : expected number of background events

n : estimator of the total number of events (signal + background)

$\hat{n}$: estimation, particular value taken by n

$$P(n \mid r_S) = \frac{(r_S + r_B)^n}{n!} e^{-(r_S + r_B)}$$

Small event count with background : Neyman prescription

Centred belts:

Compute contours $n_{min}(r_S)$ et $n_{max}(r_S)$ for a discrete set of values of r_S in a sensible range:

$$\sum_{k=0}^{n_{min}(r_S)} P(k \mid r_S) > \frac{1-\alpha}{2} \text{ and } \sum_{k=0}^{n_{min}(r_S)-1} P(k \mid r_S) \le \frac{1-\alpha}{2}$$

$$\sum_{k=n_{max}(r_S)}^{\infty} P(k \mid r_S) > \frac{1-\alpha}{2} \text{ and } \sum_{k=n_{max}(r_S)+1}^{\infty} P(k \mid r_S) \le \frac{1-\alpha}{2}$$

$$\Rightarrow \sum_{k=n_{min}(r_S)}^{n_{max}(r_S)} P(k \mid r_S) > \alpha$$

n_{min}, n_{max} **are discrete $\Rightarrow$ exact coverage for $C.L. = \alpha$ not possible**

$P\left(r_{S0} \in \left[r_{S\,min}(\hat{n}), r_{S\,max}(\hat{n})\right]\right) \ge \alpha$ **with** $r_{S\,min}(n) = n_{max}^{-1}\left(n_{max}(r_S)\right)$, $r_{S\,max}(n) = n_{min}^{-1}\left(n_{min}(r_S)\right)$

Upper belt:

Compute contour $n_{up}(r_S)$ for a discrete set of values of r_S in a sensible range:

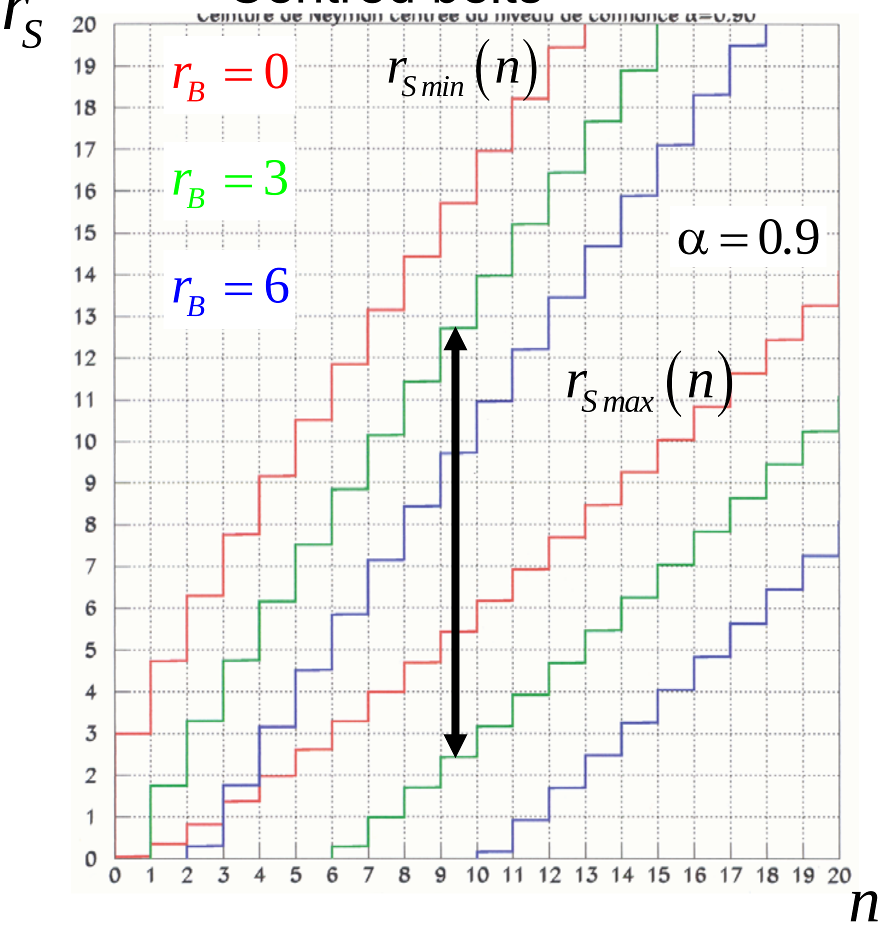
$$\sum_{k=n_{up}(r_S)}^{\infty} P(k \mid r_S) > 1-\alpha \text{ and } \sum_{k=n_{up}(r_S)+1}^{\infty} P(k \mid r_S) \le 1-\alpha \quad \Rightarrow \sum_{k=0}^{n_{up}(r_S)} P(k \mid r_S) > \alpha$$

n_{up} **is discrete $\Rightarrow$ exact coverage for $C.L. = \alpha$ not possible**

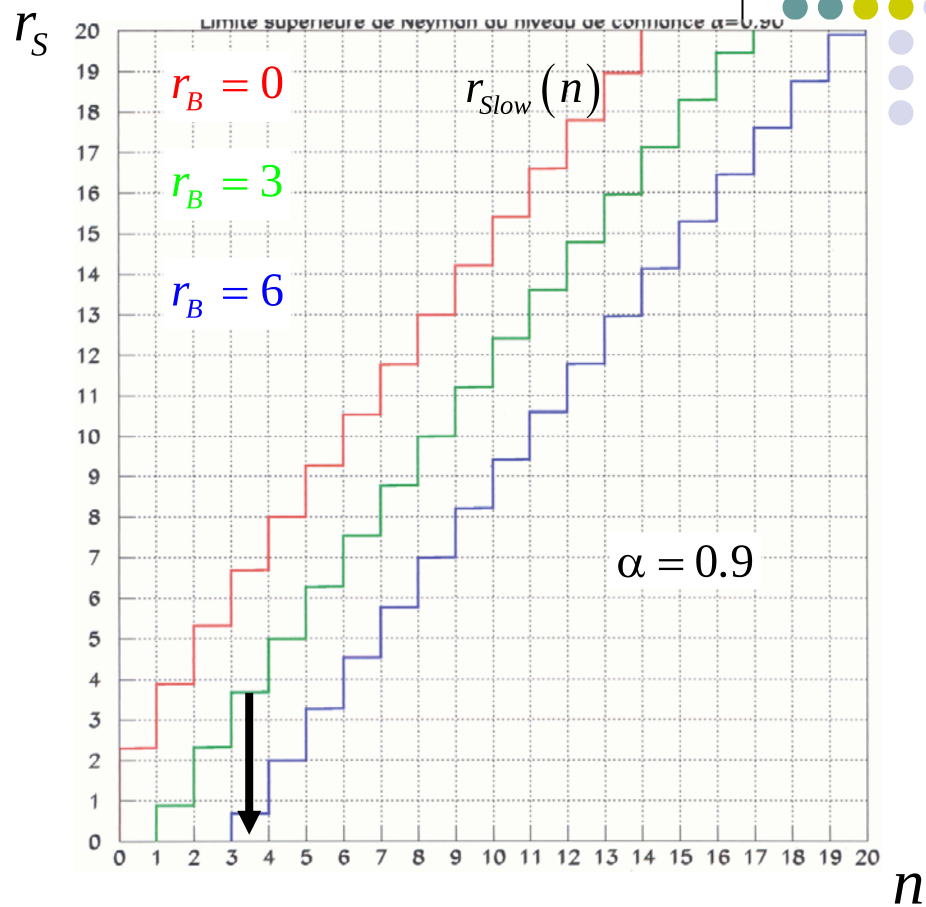
$P\left(r_{S0} < r_{Slow}(\hat{n})\right) \ge \alpha$ **with** $r_{Slow}(n) = n_{up}^{-1}\left(n_{up}(r_S)\right)$

Small event count with background : Neyman belts

Centred belts



Upper belt



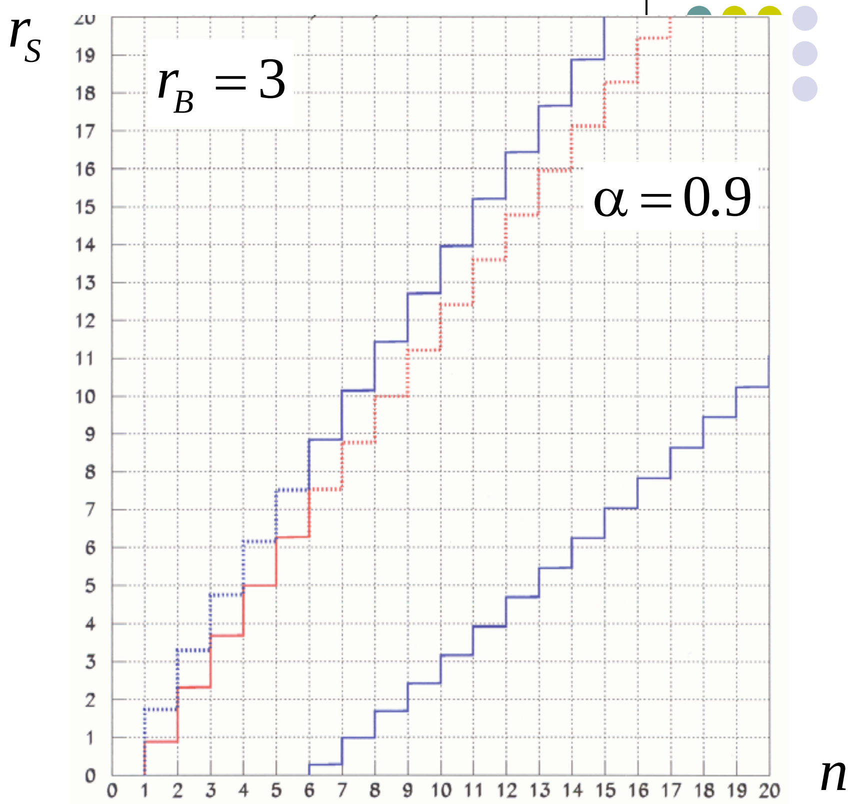
If $r_B = 3$ and $n = 9 \Rightarrow P\left(r_s \in [2.4, 12.8]\right) = 0.9$

If $r_B = 3$ and $n = 3 \Rightarrow P\left(r_s < 3.7\right) = 0.9$

If $n \ll r_B$ - large negative fluctuation on r_B :

$r_{SI}(n) = 0 \Leftrightarrow$ the confidence interval is void

Small event count with background : Neyman hybrid belts



Choosing belts a priori is correct from the statistical point of view but the choice may be meaningless:

- **compute a pair of limits on the signal when the measurement is clearly compatible with the expected background.**

- **compute un upper limit on the signal when the measurement is clearly larger than the expected background.**

Why not make the choice a posteriori ? Statistically incorrect: the coverage is incorrect.

To make the choice a posteriori is statistically incorrect.

The coverage does not correspond $\alpha = 0.9$

The Feldman – Cousins prescription

Unified approach to the classical analysis of small signals.
G. Feldman and R. Cousins
Phys. Rev. D57 (1998) 3873

Purpose : define an a priori unique set of belts with proper coverage.

This paper is a first of a long series that treat the questions of small signals above small background and measurements with standard errors extending over a physical limit from the frequentist point of view with prescriptions that insure correct coverage.

Other prescriptions are used
Some improvements to i.e.
• take the systematic errors into account,
• take the error on the background into account,
•

Th. Junk
NIM A434 (1999) 435
C. Junti
arXiv:hep-ex/9901015
B. Roe and M. Woodroofe
arXiv:physics/9812036v3
G. Punzi
arXiv:hep-ex/9912048

Gaussian error near physical limit: Feldman–Cousins prescription

Ordering prescriptions

1 • By definition : $P\left(t \in \left[t_{min}(\theta), t_{max}(\theta)\right]\right) = \int_{t_{min}(\theta)}^{t_{mac}(\theta)} f(t \mid \theta)\, dt = \alpha$

2 • The likelihood ratio $R(t) = \dfrac{P(t \mid \theta)}{P\left(t \mid \theta^*(t)\right)}$ is maximum

$\theta^*(t)$ is the physically allowed value of θ that maximises $P(t \mid \theta)$ given t

Applying to prescription to measurements with standard errors extending over a physical limit.

If $\hat{\theta} \geq 0 \;:\; \theta^* = \hat{\theta}$ $\quad \Rightarrow N(t \mid t,1) = \dfrac{1}{\sqrt{2\pi}} \Rightarrow \quad R(t) = \dfrac{\dfrac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}(t-\theta)_2}}{\sqrt{2\pi}} = e^{-\frac{1}{2}(t-\theta)^2}$

If $\hat{\theta} \leq 0 \;:\; \theta^* = 0$ $\quad \Rightarrow N(t \mid 0,1) = \dfrac{1}{\sqrt{2\pi}} e^{-\frac{t^2}{2}} \Rightarrow R(t) = \dfrac{\dfrac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}(t-\theta)_2}}{\dfrac{1}{\sqrt{2\pi}} e^{-\frac{t^2}{2}}} = e^{-\frac{1}{2}\left(\theta^2 - 2t\theta\right)}$

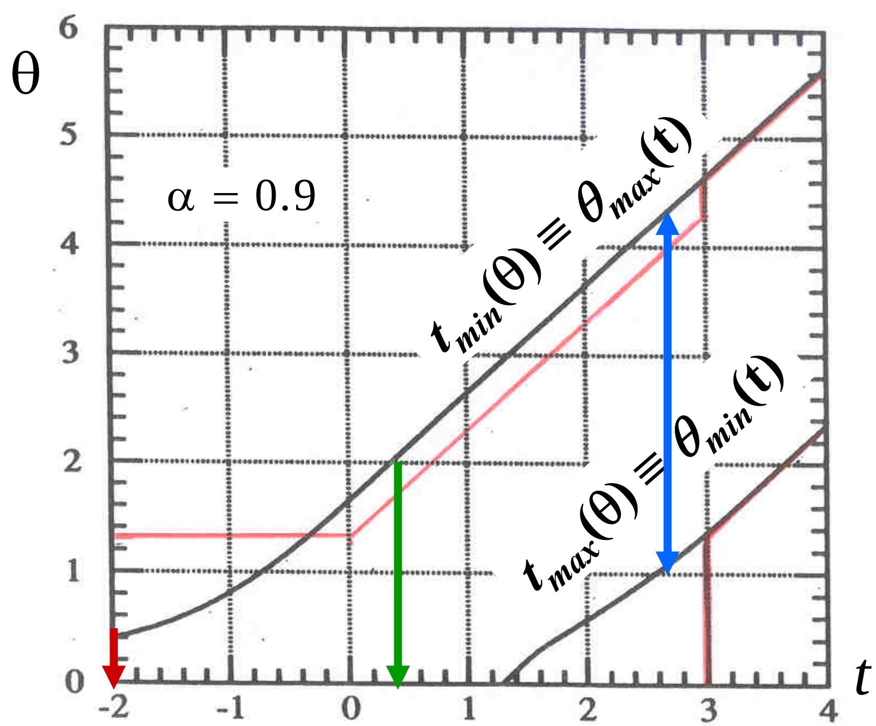
Gaussian error near physical limit: Feldman–Cousins belts

For a finite values of θ : solve numerically the system

$$\int_{t_{\min}(\theta)}^{t_{max}(\theta)} N(t\,|\,\theta,1) = \alpha$$
$$R\left(t_{min}(\theta)\,|\,\theta\right) = R\left(t_{max}(\theta)\,|\,\theta\right)$$

with

$$R(t) = e^{-\frac{1}{2}(t-\theta)_2} \qquad \textbf{if } t \geq 0$$
$$R(t) = \frac{e^{-\frac{1}{2}(t-\theta)_2}}{e^{-\frac{t^2}{2}}} = e^{-\frac{1}{2}\left(\theta^2 - 2t\theta\right)} \qquad \textbf{if } t \leq 0$$



Feldman – Cousins belt

Neyman hybrid belt

If $\hat{\theta} = 2.6, P\left(\theta_0 \in [1.0, 4.4]\right) = 0.9$

If $\hat{\theta} = 0.4, P\left(\theta_0 < 2.0\right) = 0.9$

For large values of t : identical belts

$$\lim_{t \to -\infty} \theta_b(t) = 0$$

but the confidence interval is never null

If $\hat{\theta} = -2, P\left(\theta_0 < .0\right) = 0.9$

Small event count with background : Feldman–Cousins prescription

Ordering prescriptions

1• By definition : $P\left(n \in \left[n_{min}\left(r_s\right), n_{max}\left(r_S\right)\right]\right) = \sum_{k=n_{min}(n_S)}^{n_{max}(n_S)} P\left(k \mid r_S\right) \approx \alpha$

2• The likelihood ratio $R\left(n\right) = \dfrac{P\left(n \mid r_S\right)}{P\left(n \mid r_S^*\left(n\right)\right)}$ is maximum

$r_S^*\left(n\right)$ is the value of r_S that maximises $P\left(n \mid r_S\right)$ given n and $r_S^* \geq 0$

Applying to prescription to search for a small signals above small background

$$r_S^*\left(n\right) = Max\left(0, n - r_B\right)$$

$$P\left(n \mid r_S, r_B\right) = \frac{\left(r_S + r_B\right)^n}{n!} e^{-\left(r_S + r_B\right)}$$

$$P\left(n \mid r_S^*\left(n\right), r_B\right) = \frac{\left(r_S^*\left(n\right) + r_B\right)^n}{n!} e^{-\left(r_S^*\left(n\right) + r_B\right)},$$

$$R\left(n\right) = \frac{P\left(n \mid r_S, r_B\right)}{P\left(n \mid r_S^*\left(n\right), r_B\right)}$$

Small event count with background : Feldman–Cousins belts (1)

For a finite values of r_S in the sensible range:

Select the values of n in decreasing values of $R(n)$ until $\sum P(n \mid r_S) \geq \alpha$

Use the two extreme values of n as limits

Example: Compute limits at $r_S = 1$ for $r_B = 3$ and $\alpha = 0.9$

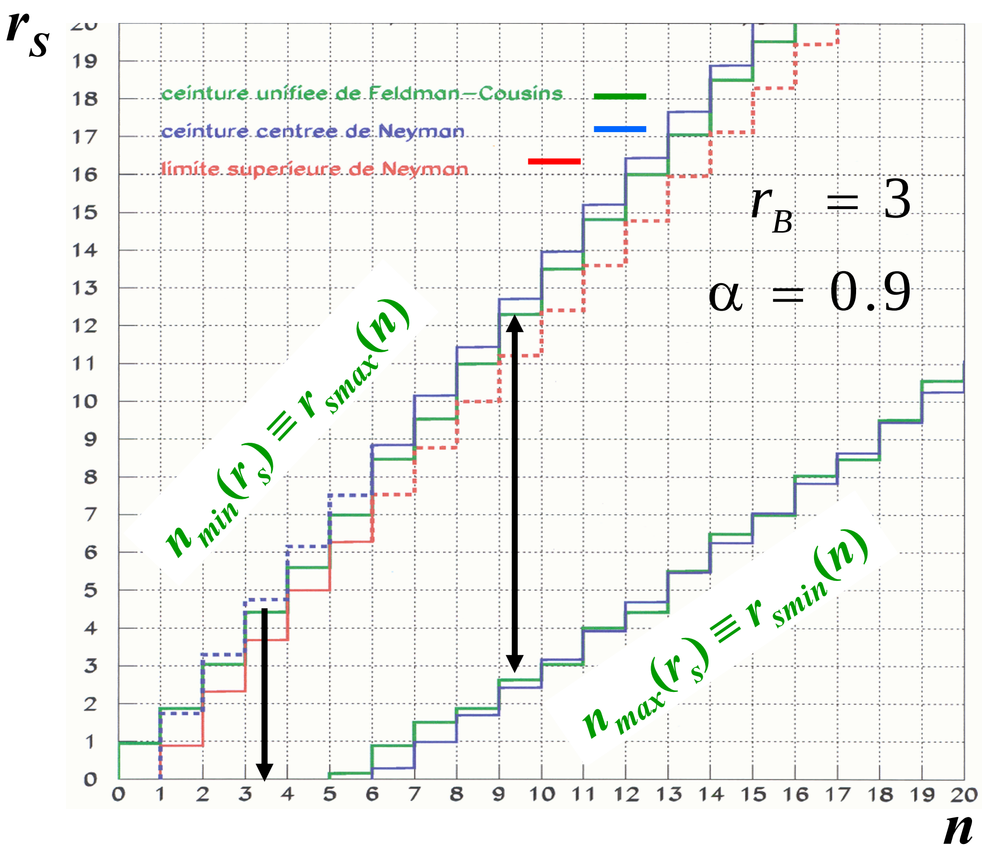
n	$P(n \mid r_S)$	$r_S^*(n)$	$P(n \mid r_S^*(n))$	$R(n)$	$rank$
0	0.030	0.0	0.050	0.607	6
1	0.106	0.0	0.149	0.708	5
2	0.185	0.0	0.224	0.826	3
3	0.216	0.0	0.224	0.963	2
4	0.189	1.0	0.195	0.966	1
5	0.132	2.0	0.175	0.753	4
6	0.077	3.0	0.161	0.480	7
7	0.039	4.0	0.149	0.259	
8	0.017	5.0	0.140	0.121	
9	0.007	6.0	0.132	0.050	
10	0.002	7.0	0.125	0.018	
11	0.001	8.0	0.119	0.006	

$$\sum_{k=1,}^{6} P_k(n \mid r_S = 1) = 0.858 < 0.9 < \sum_{k=1,}^{7} P_k(n \mid r_S = 1) = 0.935$$

$$\Rightarrow n_{min}(r_S = 1 \mid n_B = 3, \alpha = 0.9) = 0$$

$$\Rightarrow n_{max}(r_S = 1 \mid n_B = 3, \alpha = 0.9) = 6$$

Small event count with background : Feldman–Cousins belts (2)



If $r_B = 3$ and $n = 9$:

$P\left(r_{S0} \in [2.6, 12.3]\right) \geq 0.9$

If $r_B = 3$ and $n = 3$:

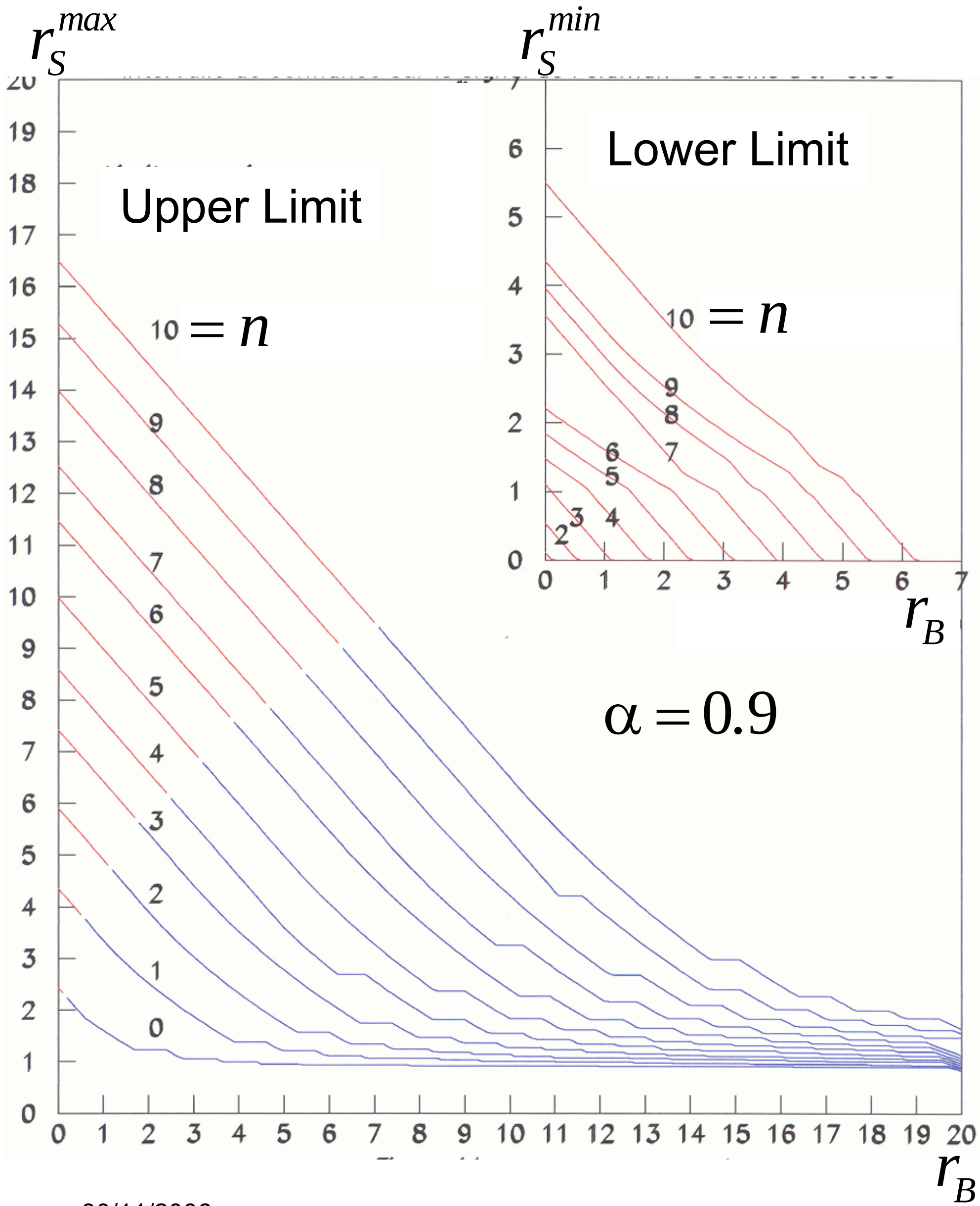
$P\left(r_{S0} < 4.5\right) \geq 0.9$

Because n is integer: coverage cannot be strictly exact

Confidence interval never void, though $lim\ r_{S\,max} \xrightarrow[n \ll r_B]{} 0$

if $n = r_B - 3 = 0$: $P\left(r_s \leq 1\right) = 0.$

Summary of Feldman–Cousins prescription for counting experiment



Examples :

If $r_B = 4.7$ **and** $n = 8$ **:** $P\left(r_{S0} \in [0.7, 9.3]\right) = 0.9$

If $r_B = 2.5$ **and** $n = 3$ **:** $P\left(r_{S0} < 4.9\right) = 0.9$

Application to the search for neutrino oscillation

The confidence interval of parametre θ of a model that predicts $r_S(\theta)$ is:

$P\left(\theta_0 \in [\theta_{min}, \theta_{max}]\right) = \alpha$ **with**

$\theta_{min} \Rightarrow r_S(\theta_{min}) = r_{S\,min}$

$\theta_{max} \Rightarrow r_S(\theta_{max}) = r_{S\,max}$

Alternatively, θ may replace r_S as ordinate of all plots and the physical bounds on θ used.

The procedure may be extended to a set of parametre $\underline{\theta}$ of a model that predicts $r_S(\underline{\theta})$

$$\begin{array}{l} |\nu_\ell\rangle = \cos\theta|\nu_1\rangle + \sin\theta|\nu_2\rangle \\ |\nu_{\ell'}\rangle = \cos\theta|\nu_2\rangle - \sin\theta|\nu_1\rangle \end{array} \Rightarrow P\left(\nu_\ell \to \nu_{\ell' \neq \ell}\right) = \sin^2(2\theta)\sin^2\left(1.27\frac{\Delta m^2\left[\left(eV/c^2\right)^2\right]L[km]}{E[GeV]}\right)$$

e.g. $P(\nu_\mu \to \nu_\tau)$ **in a ν_μ beam**

Face the two pathological problems:

- **if mixing $\sin^2(2\theta)$ small : small signal**
- **the combination of the true value Δm^2 and values of L/E accessible to experiment such that small signal even for large mixing**
- **physical bounds : $0 \leq \sin^2(2\theta) \leq 1$**

+ background : ν_μ misidentified as ν_τ

Application to the search for neutrino oscillation : outcome of one experiment

$$P\left(\nu_\mu \to \nu_\tau\right) = sin^2(2\theta)\,sin^2\left(1.27\frac{\Delta m^2\left[\left(eV/c^2\right)^2\right]L[km]}{E[GeV]}\right)$$

For simplicity, assume that $L \gg$ fluctuations on L

$P\left(\nu_\mu \to \nu_\tau\right) = P(E) \Rightarrow$ divide in N energy bins (the better the energy resolution, the more bins)

In each bin i :

define a set of ν_τ selection criteria $\Rightarrow log\,\lambda = log\frac{\mathcal{L}(\nu_\tau)}{\mathcal{L}(\nu_\mu)}$ is large

compute the expected background : r_{Bi}

compute for a discrete set of pairs of values of $\left(\Delta m^2, sin^2(2\theta)\right)$ the expected signal :

$$r_{Si}\left(\Delta m^2, sin^2(2\theta)\right)$$

computation are not analytic and require Monte-Carlo simulation.

Outcome of one experiment :

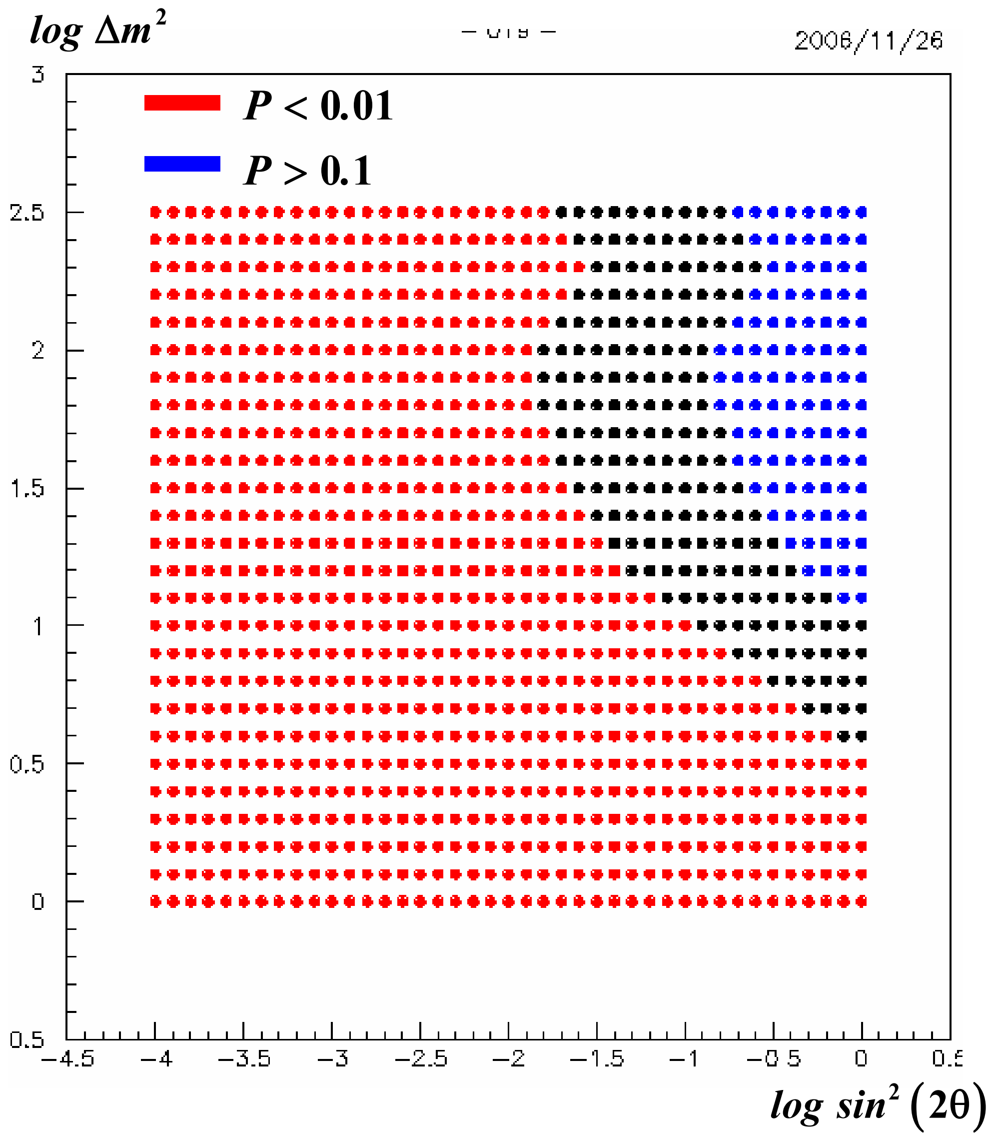
$$\left.\begin{array}{l} r_{Bi} \\ r_{Si}\left(\Delta m^2, sin^2(2\theta)\right) \\ \text{one particular set } n_i \text{ of observed events} \end{array}\right\} i = 1, N$$

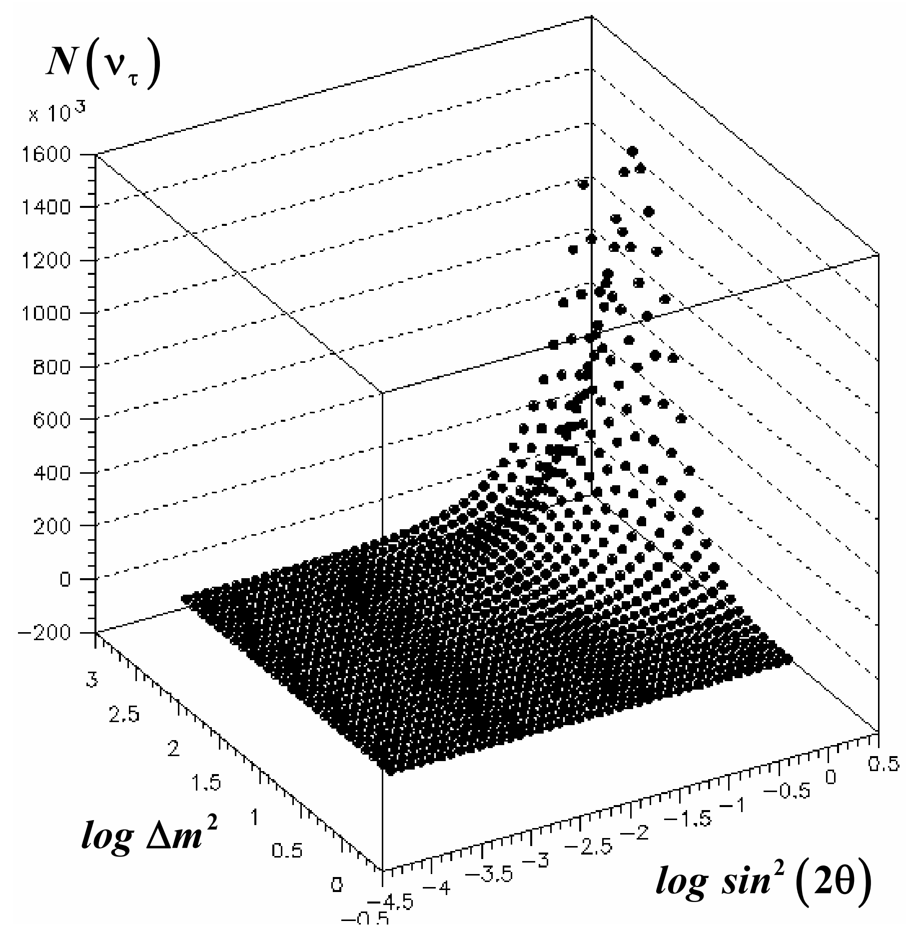
Application to the search for neutrino oscillation : oscillation probability

E uniform on $[10.,60.]\,GeV$

$L = 0.6\,km$

$N(\nu_\mu) = 2\,000\,000$





Application to the search for neutrino oscillation : belts

Definition of the belts at $C.L. = \alpha$ in the $\left(\Delta m^2, \sin^2(2\theta)\right)$ plane

For each discrete pair of values of $\left(\Delta m^2, \sin^2(2\theta)\right) \Rightarrow \underline{r}_S\left(\Delta m^2, \sin^2(2\theta)\right)$

The possible outcomes from a large number of simulation experiments

is a large number of sets $\underline{n} = (n_1, \ldots n_N)$

resulting from fluctuations on the expected background $(\underline{r}_B)$ and signal $\underline{r}_S\left(\Delta m^2, \sin^2(2\theta)\right)$

For each set, compute

$$R\left(\underline{n} \mid \Delta m^2, \sin^2(2\theta)\right) = \frac{P\left(\underline{n} \mid \underline{r}_S\left(\Delta m^2, \sin^2(2\theta)\right), \underline{r}_B\right)}{P\left(\underline{n} \mid \underline{r}_S^*\left(\Delta m^2, \sin^2(2\theta), \underline{r}_B \mid \underline{n}\right)\right)} = \prod_{i=1}^{N} \frac{P\left(n_i \mid r_{Si}\left(\Delta m^2, \sin^2(2\theta)\right), r_{Bi}\right)}{P\left(n_i \mid r_{Si}^*\left(n_i; \Delta m^2, \sin^2(2\theta), r_{Bi} \mid \underline{n}\right)\right)}$$

where $\underline{r}_S^*\left(\underline{n}; \Delta m^2, \sin^2(2\theta)\right)$ maximizes $P\left(\underline{n} \mid \underline{r}_S\left(\Delta m^2, \sin^2(2\theta)\right)\right)$ in the physical domain given $\underline{n}$.

In regions where P_{oscl}^{true} is small, $\underline{r}_S <\approx \underline{r}_B$: values of $\underline{n} < \underline{r}_B \;\Rightarrow\; P_{osc} < 0$ or $\sin^2(2\theta) < 0$ may occur.

In regions where $P_{oscl}^{true} \approx 1$: $N(\nu_\tau) > N(\nu_\mu) \Rightarrow P_{osc} > 1$ or $\sin^2(2\theta)>1$ may occur.

Select the value $R_\alpha\left(\Delta m^2, \sin^2(2\theta)\right)$ such that a fraction α of the sets $(\underline{n})$ have $R(\underline{n}) < R_\alpha$

Application to the search for neutrino oscillation: confidence domain

Confidence domain set by one experiment at $C.L. = \alpha$ in the $\left(\Delta m^2, sin^2(2\theta)\right)$ plane

One experiment $\Leftrightarrow$ one particular set $(\underline{\hat{n}})$

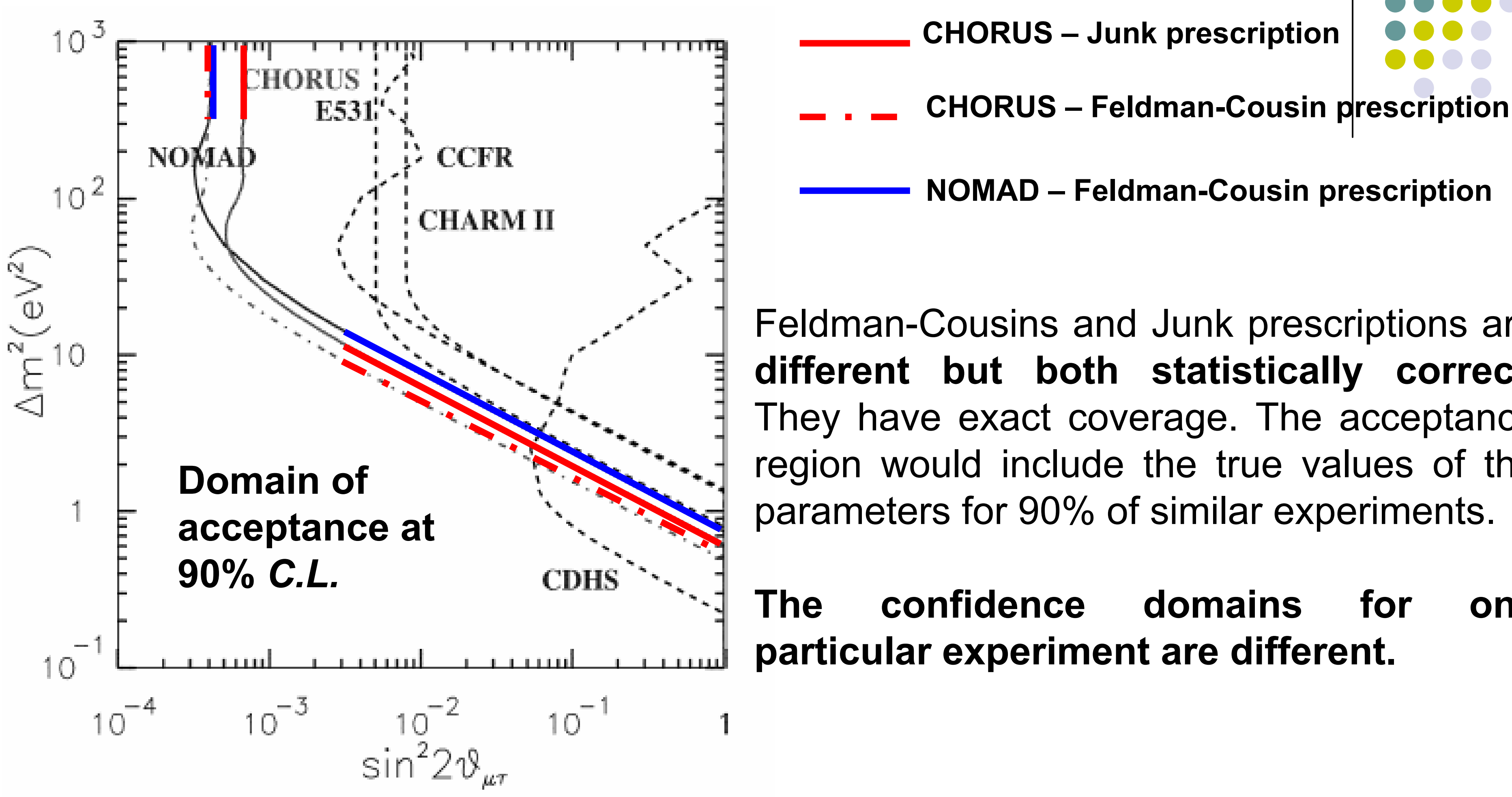
For each discrete pair of values of $\left(\Delta m^2, sin^2(2\theta)\right) \Rightarrow \underline{r}_S\left(\Delta m^2, sin^2(2\theta)\right)$

compute $R\left(\underline{\hat{n}} \mid \Delta m^2, sin^2(2\theta)\right) = \dfrac{P\left(\underline{\hat{n}} \mid \underline{r}_S\left(\Delta m^2, sin^2(2\theta)\right), \underline{r}_B\right)}{P\left(\underline{\hat{n}} \mid \underline{r}_S^*\left(\underline{n}; \Delta m^2, sin^2(2\theta), \underline{r}_B\right)\right)}$

Confidence domain : Pairs of values $\left(\Delta m^2, sin^2(2\theta)\right)$

such that $R\left(\underline{\hat{n}} \mid \Delta m^2, sin^2(2\theta)\right) < R_\alpha\left(\Delta m^2, sin^2(2\theta)\right)$

Application to the search for neutrino oscillation: CHORUS and NOMAD rejection domains and the use of different prescriptions



Feldman-Cousins and Junk prescriptions are **different but both statistically correct**. They have exact coverage. The acceptance region would include the true values of the parameters for 90% of similar experiments.

The confidence domains for one particular experiment are different.

Application to the search for neutrino oscillation : gain in binning

Net gain in binning events according to toplogy/kinematics in bins of low background and high S/B

Bin A :	$S = 3.0$	$S + B = 24.9$
Bin B :	$S = 2.2$	$S + B = 2.5$
A+B :	$S = 5.2$	$S + B = 26.4$

Statistically easier to detect a signal above background in bin B than in the whole sample dominated by bin A.

Bin B brings most of the information.

Bin A brings small but not null additionnal information without altering the information provided by bin B.

Bins A and B are two independent experiments

X – Monte-Carlo Simulation

Uniform pseudo-random numbers ξ generators on $[0,\ 1]$

Pseudo-random : result of an algorithm that produce identical numbers when repeated
Random : no way a posteriori to distinguish from genuine random numbers
Good generator:

- Randomness and uniformity.
- Length of the sequence before repetition and disjoint sub-sequences.
- Reproducibility.
- Portability.
- Speed.

The first generator : von Neumann generator

- X_0 made of r digits,
- Y_1, from the $r/2$ central digits of X_0
- $X_1 = Y_1 \times Y_1$
- Y_2, from the $r/2$ central digits of X_1
- ...

If $X_n = X_{n-1}$: period fixed to 1

RANMAR, mix the bits of two generators (G. Marsaglia, A. Zaman and W.-W Tsang, Stat. Prob. Lett 9 (1990) 35)

1. Differed Fibonacci sequence : $X_i = (X_{i-p} - X_{i-q} + 1.)\ mod(1.)\ , \quad q < p$
2. Arithmetic sequence : $0 < c,d < 1$

$Y_i = Y_{i-1} - c$

if $Y_i < 0. \Rightarrow Y_i = Y_i + d$

3. Combibation of the two sequences

$Z_i = (X_i - Y_i + 1.)\ mod(1.)$

Period of 10^{43}
Large number of sub-sequences (30000×30000) of mean periodicity 10^{31}

Non-uniform discrete random numbers generators

Random number on $[a,b]$

ξ on $[0,\ 1] \Rightarrow x = a + (b - a)\,\xi.$

- x may take a finite set of N values $(x_1, \ldots, x_N)$ with probabilities $(p_1, \ldots, p_N)$

binomial distribution

Define the distribution function $P_n = \sum_{i=1}^{n} p_i \ , n = 1, N$

$P_N = 1;$ call $P_0 = 0$

Accept value x_n for x if $P_{n-1} < \xi \le P_n$

- x may take an infinite set of values

Poisson distribution

Define the distribution function $P_n = \sum_{i=1}^{n} p_i \ , n = 1, N$

for a finite number of values until $P_N \simeq 1.$

If $\xi > P_N$: compute $P_{N+1}, \ldots, P_{N'}$ until $\xi \le P_{N'}$

Non-uniform continuous random numbers generators : cumulative method

Generate value of x following PDF $f(x)$

$$F(x) = \int_a^x f(x')\,dx' \quad ; \; 0 \le F(x) \le 1$$

$$dF(x) = f(x)\,dx$$

$$x = F^{-1}(\xi) \text{ follows } f(x)$$

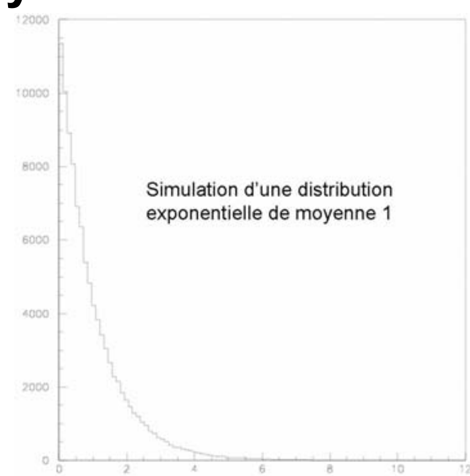
Demonstration : $PDF(x) = PDF(\xi)\dfrac{d\xi}{dx} = 1 \times \dfrac{dF(x)}{dx} = \dfrac{f(x)\,dx}{dx} = f(x)$

Method limited to PDF where $F(x)$ and $F^{-1}(\xi)$ are analytical

Exponential PDF : $f(x) = \dfrac{1}{\beta} e^{-\frac{x}{\beta}}$

$$F(x) = 1 - e^{-\frac{x}{\beta}} = \xi$$

$$x = -\beta \; log(1 - \xi)$$



χ^2_ν PDF : If $\nu = 2n$ $\quad x = -2 \; log(\xi_1 \xi_2 \ldots \xi_n)$

If $\nu = 2n+1$ $\quad x = -2 \; log(\xi_1 \xi_2 \ldots \xi_n) - 2 \; log\,\xi_{n+1} \, cos^2(2\pi \xi_{n+2})$

Non-uniform continuous random numbers generators : cumulative method

Isosceles trapezoidal and triangular distributions

$2a < 2b$ the two trapeze bases

c the minimum value of x: $x \in [c, c+2b]$

$$x = c + (b+a)\,\xi_1 + (b-a)\,\xi_2$$

If $a = 0$: triangular distribution

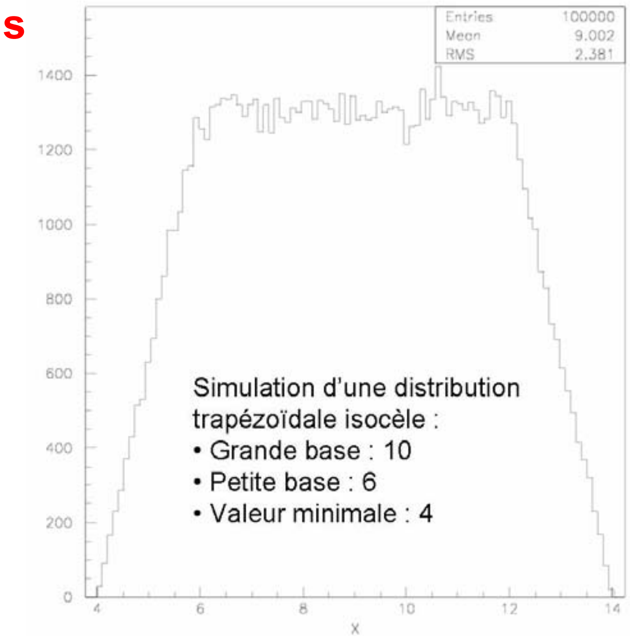
$$x = c + b(\xi_1 + \xi_2)$$

Left half-trapeze or half-triangle

If $x > c+b : x = (c+b) - (x - (c+b))$

Right half-trapeze or half-triangle

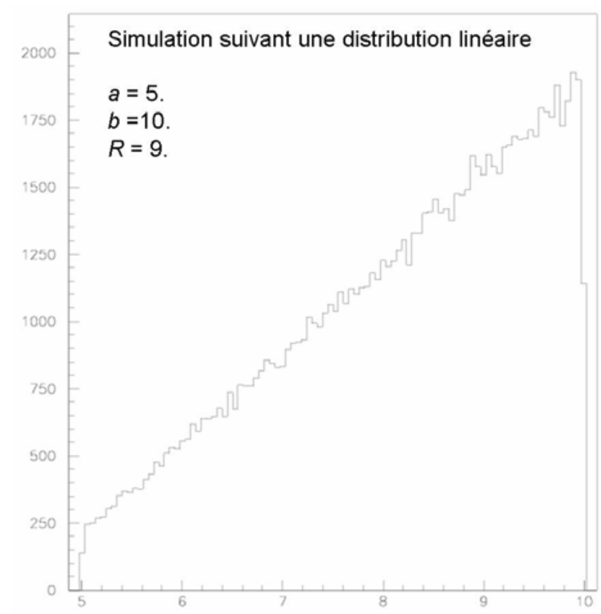
If $x < c+b : x = (c+b) + ((c+b) - x)$



Linear distribution $f(x) = ax + b$ on $[0,1]$

$$\text{Define } \alpha,\beta: \left\{ \begin{array}{l} r = \dfrac{f(1)}{f(0)} = \dfrac{a+b}{b} \\ \displaystyle\int_0^1 (ax+b)\,dx = \frac{a}{2} + b = 1 \end{array} \right\} \Rightarrow \left\{ \begin{array}{l} b = \dfrac{2}{r+1} \\ a = b(r-1) \end{array} \right.$$

$$F(x) = \xi = \frac{a}{2}x^2 + bx \Rightarrow x = \frac{-b + \sqrt{b^2 + 2a\xi}}{a}$$



Non-uniform continuous random numbers generators : gaussian distribution

Exact method of Box - Muller

Take

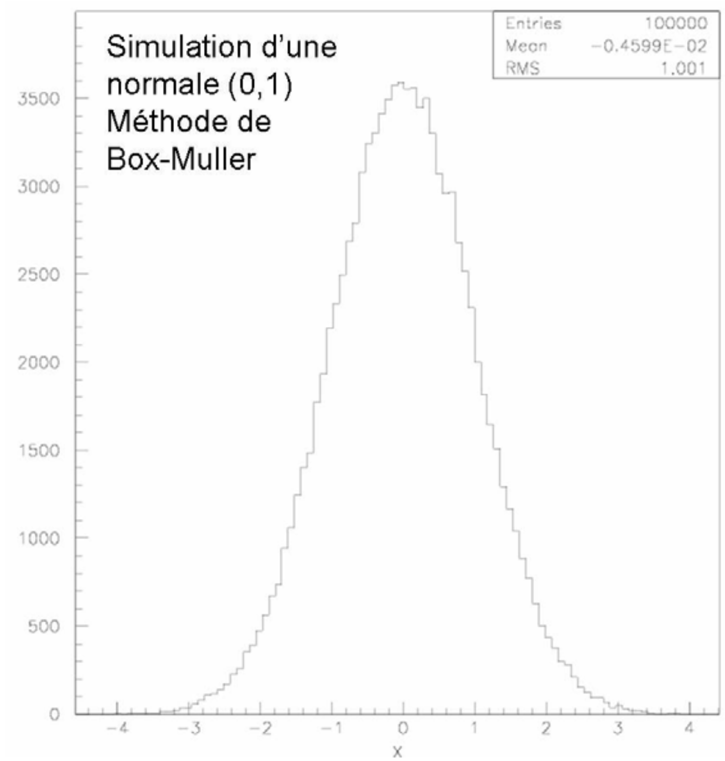
$$x_1 = \sqrt{-2\log \xi_1}\cos 2\pi\xi_2$$

$$x_2 = \sqrt{-2\log \xi_1}\sin 2\pi\xi_2$$

or

$$\xi_1 = e^{-\frac{\left(x_1^2+x_2^2\right)}{2}}$$

$$\xi_2 = \frac{\arg \mathrm{tg}\left(x_2 / x_1\right)}{2\pi}$$



$$f(x_1,x_2) = \begin{vmatrix} \frac{\partial \xi_1}{\partial x_1} & \frac{\partial \xi_1}{\partial x_2} \\ \frac{\partial \xi_2}{\partial x_1} & \frac{\partial \xi_2}{\partial x_2} \end{vmatrix} f(\xi_1,\xi_2) \quad \text{with} \quad f(\xi_1,\xi_2) = f(\xi_1)f(\xi_2) = 1$$

$$f(x_1,x_2) = \frac{1}{2\pi}e^{-\frac{1}{2}\left(x_1^2+x_2^2\right)} = \frac{1}{\sqrt{2\pi}}e^{-\frac{1}{2}x_1^2}\frac{1}{\sqrt{2\pi}}e^{-\frac{1}{2}x_2^2} = f\left(x_1\right)f\left(x_2\right)$$

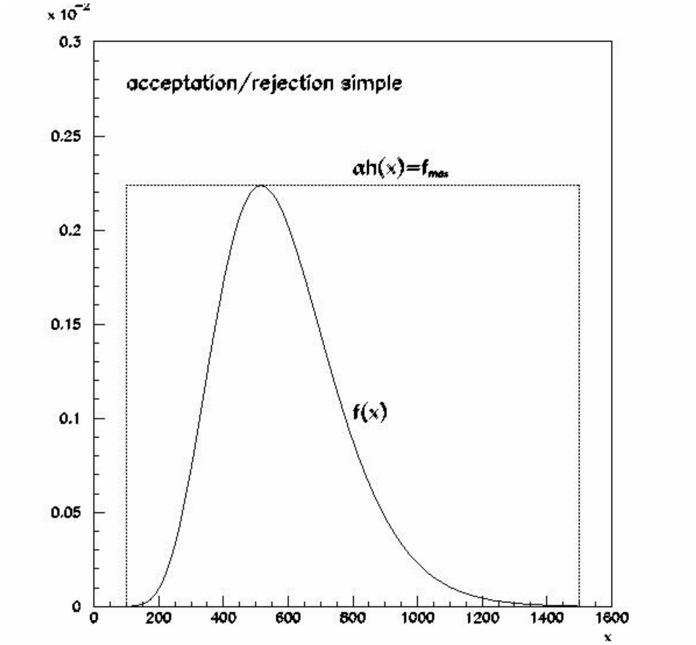
x_1 and x_2 are independent and distributed following $N\left(0,1\right)$

Non-uniform continuous random numbers generators : simple acceptation/rejection

Generate x following $f(x)$ on $[a,b]$

$h(x) = 1/(b-a)$ uniform on $[a,b]$

$\alpha = f_{max} \times (b-a) \ \Rightarrow \alpha h(x) = f_{max} \geq f(x)$ on $[a,b]$



Algorithm:

- generate ξ_1 and ξ_2 on [0, 1]
- $x = a + \xi_1 \times (b-a)$
- accept x if $\xi_2 < \dfrac{f(x)}{\alpha h(x)} = \dfrac{f(x)}{f_{max}}$

Efficiency (number of ξ to get one x) $\varepsilon = 2\dfrac{\int_a^b \alpha h(x)\,dx}{\int_a^b f(x)\,dx} = 2\alpha$

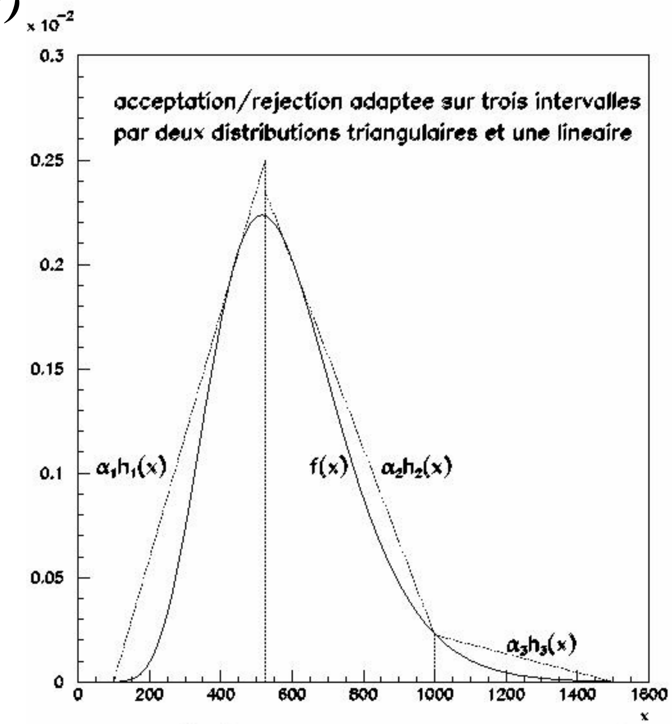
ε may become very small if $f(x)$ has long tails.

If $a,b = \pm\infty$: tails must be cut

$f(x)$ limited on a series of K intervals $[a_{k-1}, a_k]$ by $\alpha_k h_k(x)$

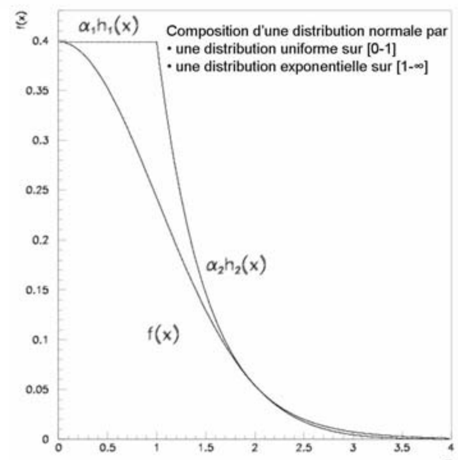
with $\alpha_k > 1 \Rightarrow \alpha_k h_k(x) \geq f(x)$ on $[a_{k-1}, a_k]$

$h_k(x)$ a PDF allowing cumulative method



Algorithm:

- $p_j = \dfrac{\alpha_j}{\sum_{i=1}^{n} \alpha_i}, \quad P_j = \dfrac{\sum_{i=1}^{j} \alpha_i}{\sum_{i=1}^{n} \alpha_i} \qquad \forall j = 1, n$
- generate ξ_1, ξ_2 on [0,1]
- select interval $k = [a_{k-1}, a_k] \Rightarrow P_{k-1} \leq \xi < P_k$
- generate x sur $[a_{k-1}, a_k]$ following $h_k(x)$
- accept x si $\xi \leq f(x) / \alpha_k h_k(x)$



Example: gaussian distribution

$$f(x) = \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}} \leq \alpha_k h_k(x) \quad k = 1,2$$

k	$[x]$	α_i	$h_i(x)$
1	$[0-1]$	$\frac{1}{\sqrt{2\pi}}$	1
2	$[1-\infty]$	$\frac{1}{2\sqrt{2\pi}}$	$2e^{-2(x-1)}$