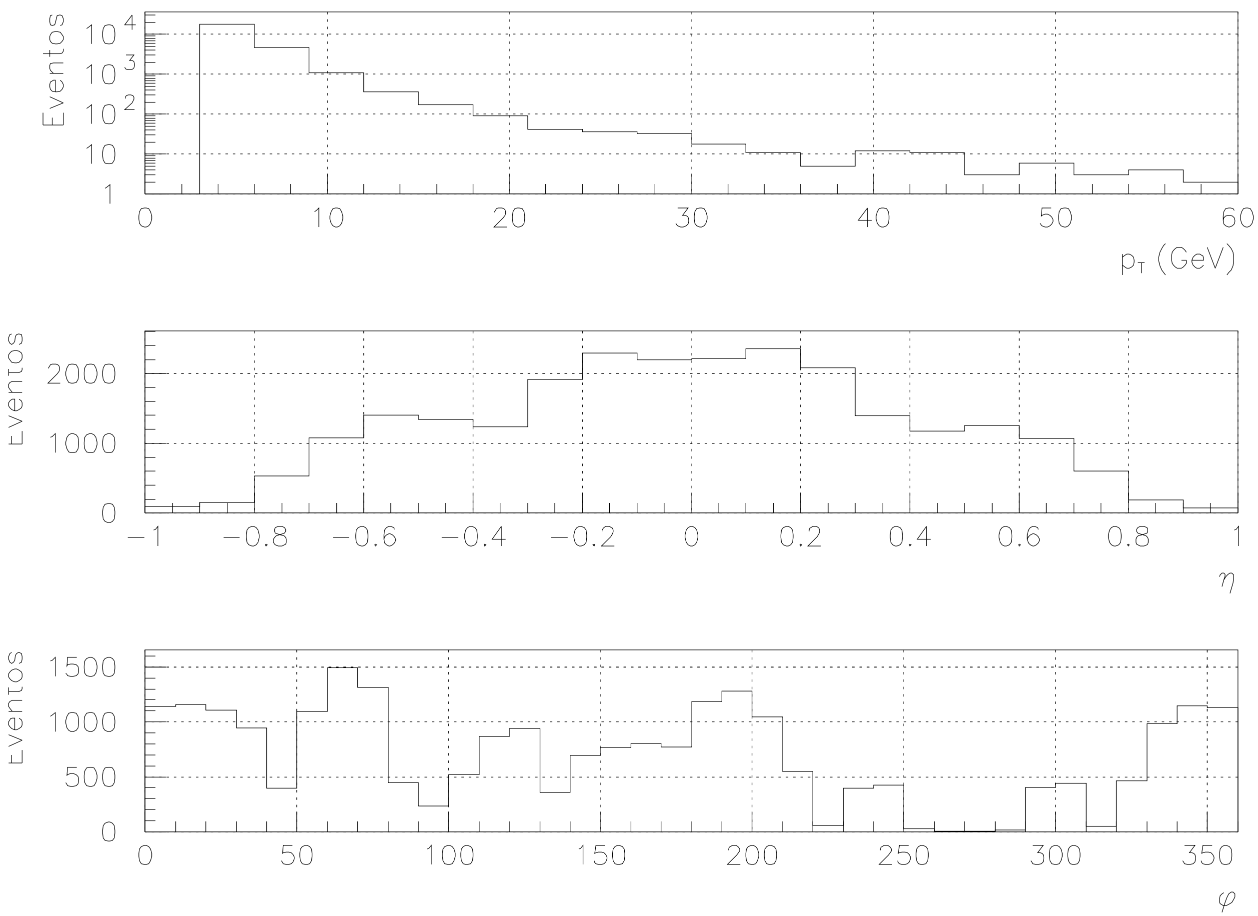
Tratamento Estatístico de Dados em Física

Distribuições Básicas

- Distribuições Experimentais
- Distribuição Binomial
- Distribuição Multinomial
- Distribuição de Poisson
- Distribuição Normal (Gaussiana)
- Distribuição de Student
- Distribuição Uniforme
- Distribuição de Breit-Wigner (Lorentz - Cauchy)
- Distribuição de Exponencial
- Distribuição de χ^2

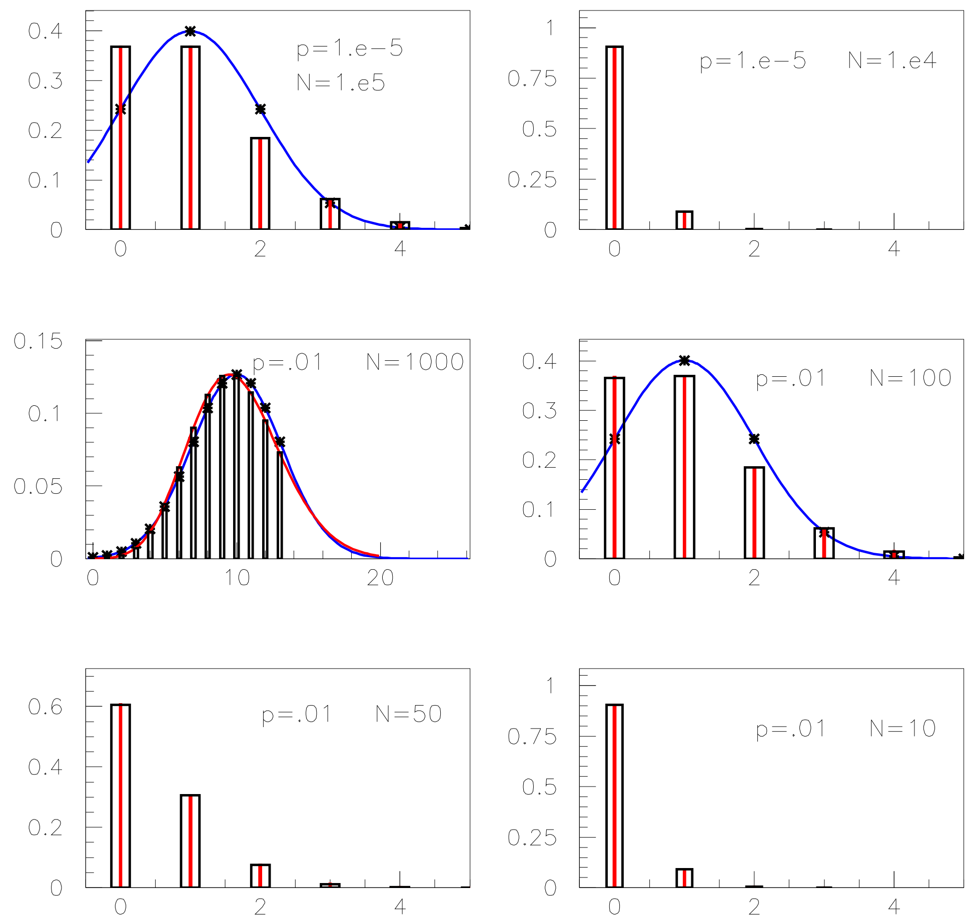
Distribuições Experimentais

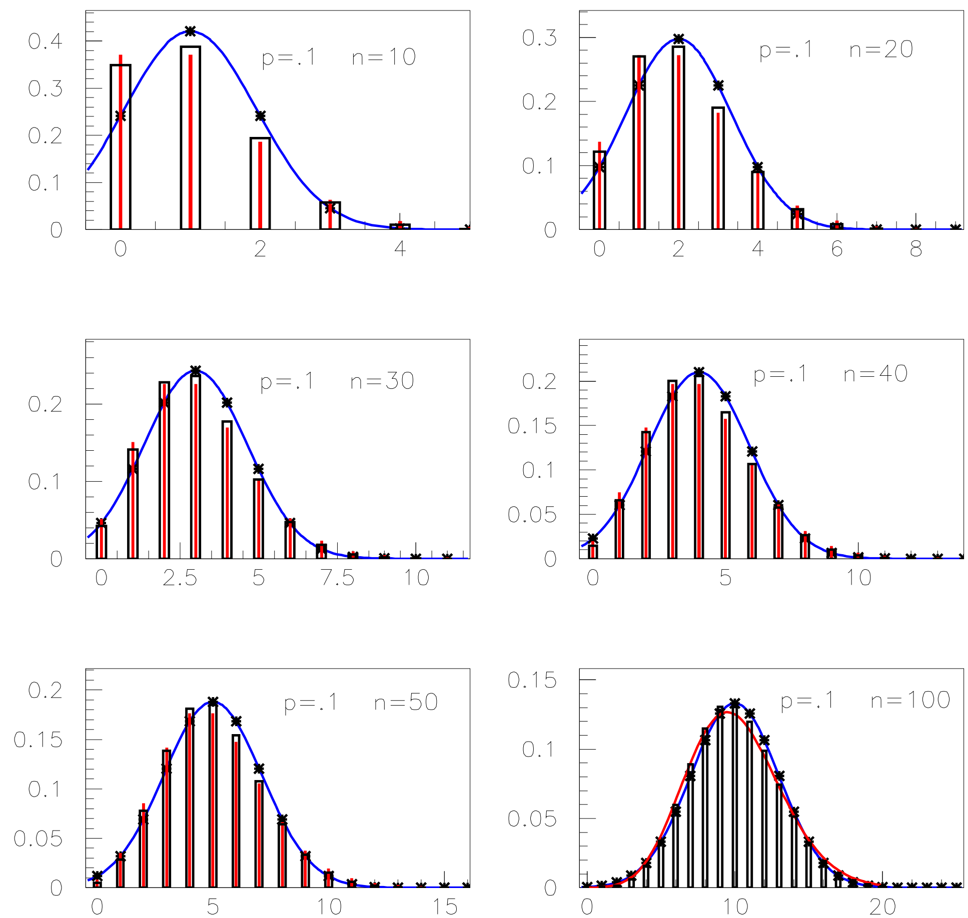


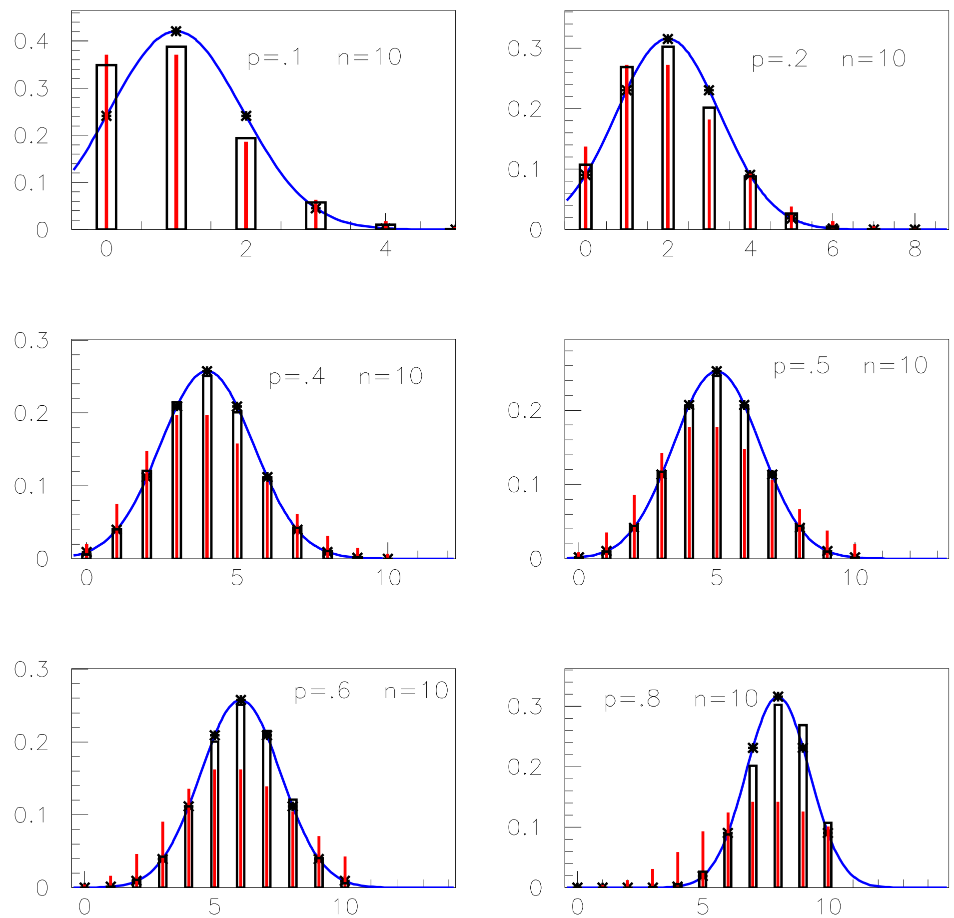
Distribuição Binomial

$$B(m|N,p) = P_m(N,p) = \frac{N!}{m!\,(N-m)!} p^m (1-p)^{N-m}$$

- $p^m (1-p)^{N-m} \quad \rightarrow \quad$ probabilidade de m ocorrências e $(N-m)$ não ocorrências de um evento em N tentativas, para uma configuração possível
- $\dfrac{N!}{m!\,(N-m)!} \quad \rightarrow \quad$ multiplicidade de cada configuração
- $\mu = Np \quad \rightarrow \quad$ valor médio de m
- $\sigma = \sqrt{Np(1-p)} \quad \rightarrow \quad$ desvio-padrão
- eventos cujas ocorrências são ou podem ser expressas por uma dicotomia.


p=1.e−5
N=1.e5
p=1.e−5 N=1.e4
p=.01 N=1000
p=.01 N=100
p=.01 N=50
p=.01 N=10


p=.1 n=10
p=.1 n=20
p=.1 n=30
p=.1 n=40
p=.1 n=50
p=.1 n=100


p=.1 n=10
p=.2 n=10
p=.4 n=10
p=.5 n=10
p=.6 n=10
p=.8 n=10

Distribuição Multinomial

A ocupação das m classes (*bins*) de um histograma pode ser descrita pela **distribuição multinomial**

$$\frac{N!}{n_1!n_2!\ldots n_m!}p_1^{n_1}p_2^{n_2}\ldots p_m^{n_m}$$

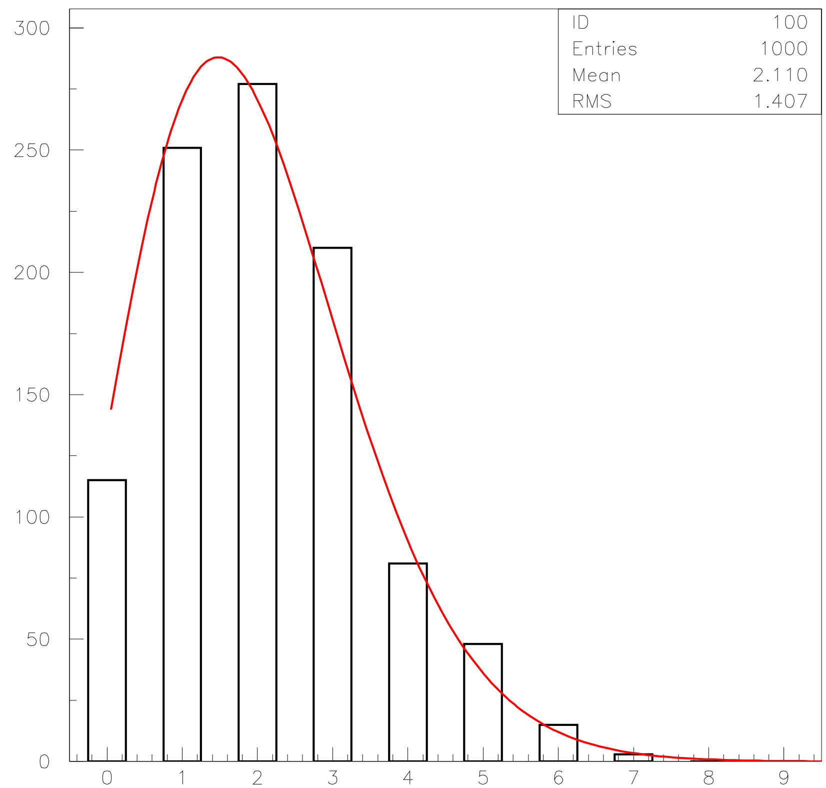
- $n_1,\ n_2,\ \ldots,\ n_m \quad \rightarrow \quad$ freqüências de cada uma das m classes
- $p_1 = n_1/N,\ p_2 = n_2/N,\ \ldots,\ p_m = n_m/N \quad \rightarrow \quad$ probabilidades de ocupação de cada uma das classes

Distribuição de Poisson

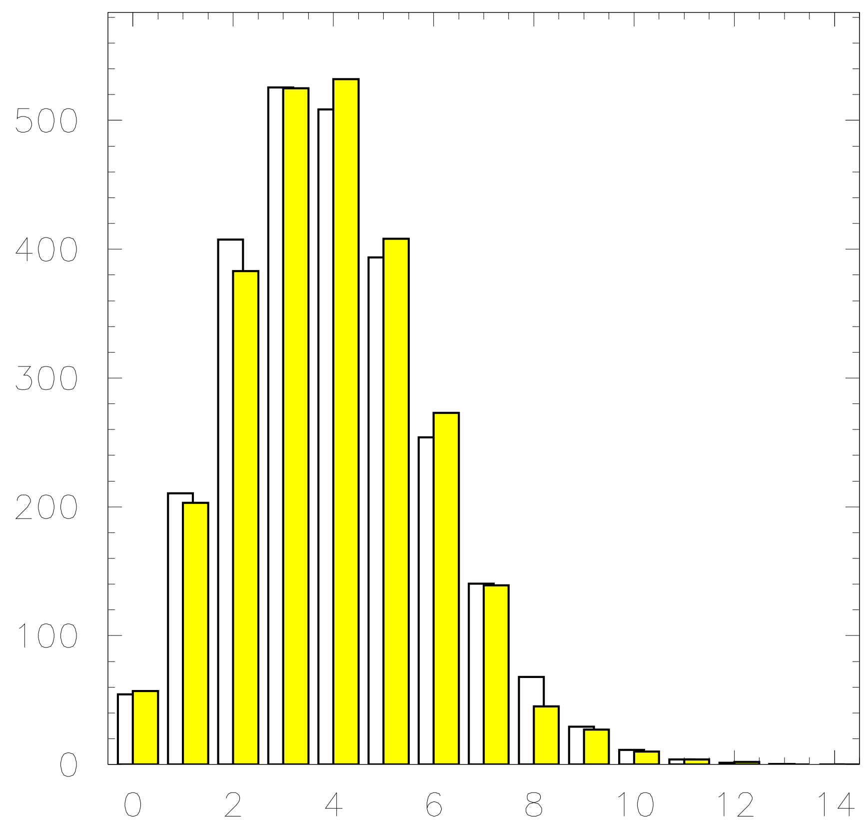
$$P_m(\mu) = \frac{\mu^m}{m!} e^{-\mu}$$

- $\mu = Np \quad (p \ll 1) \quad \rightarrow$ valor médio de m
- $\sigma = \sqrt{\mu} = \sqrt{Np} \quad (p \ll 1) \quad \rightarrow$ desvio-padrão

Simulação de um processo de Poisson



Experimento de Rutherford-Geiger

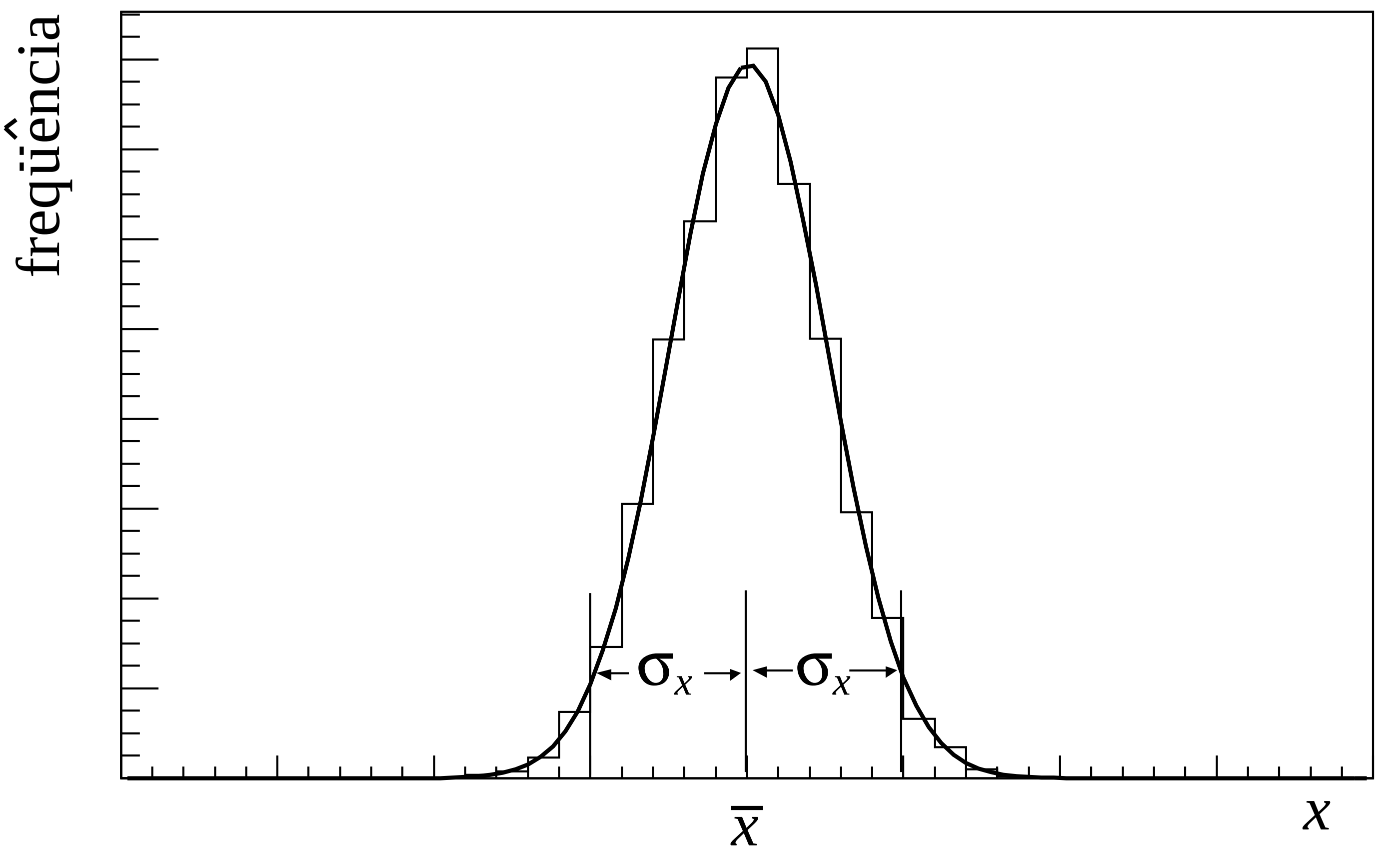


contagem de partículas α

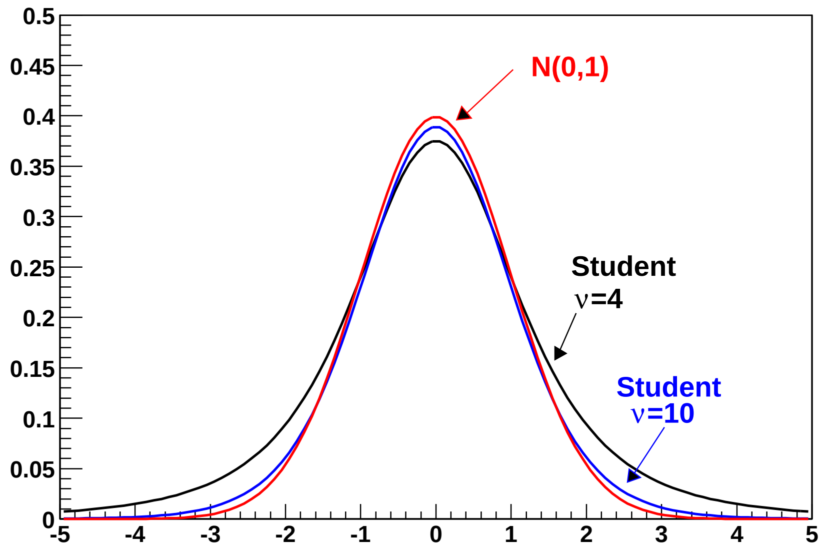
Distribuição Normal (Gaussiana)

$$\mathcal{N}_x(\mu, \sigma_x^2) = \rho(x|\mu, \sigma_x^2) = \frac{1}{\sigma_x\sqrt{2\pi}}\, e^{-\dfrac{(x-\mu)^2}{2\sigma_x^2}}$$

- $\mu \quad \rightarrow \quad$ valor médio de x
- $\sigma_x \quad \rightarrow \quad$ desvio-padrão
- $z = \dfrac{x-\mu}{\sigma_x} \quad \rightarrow \quad \mathcal{N}_z(0,1)$ (distribuição normal padrão)
- limite de várias distribuições (binomial, Poisson, Student, χ^2, medidas diretas de uma mesma grandeza)


freqüência
σ_x
σ_x
$\bar{x}$
x

Distribuição de Student



$$f(t|\nu) = A\left(1 + t^2/\nu\right)^{-(\nu+1)/2}$$

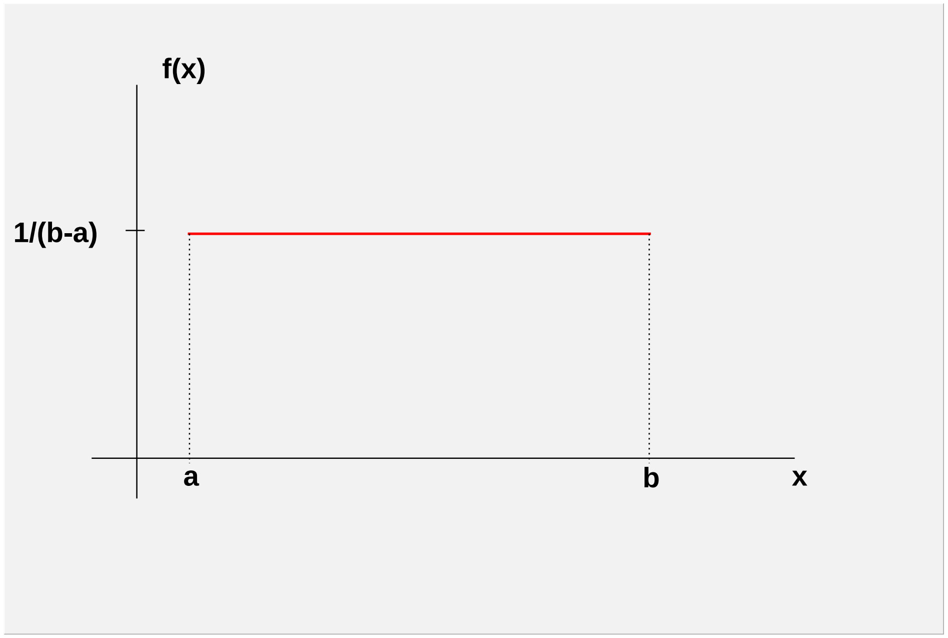
- $\sigma^2 = \dfrac{\nu}{\nu - 2} \quad \rightarrow \quad$ variância

Correção de Student

N	ν	Níveis de confiança		
		68,3%	95%	99%
2	1	1,84	12,71	63,66
3	2	1,32	4,30	9,92
5	4	1,14	2,78	4,60
8	7	1,08	2,36	3,50
10	9	1,06	2,26	3,25
15	14	1,04	2,14	2,98
20	19	1,03	2,09	2,86
∞	∞	1,00	1,96	2,58

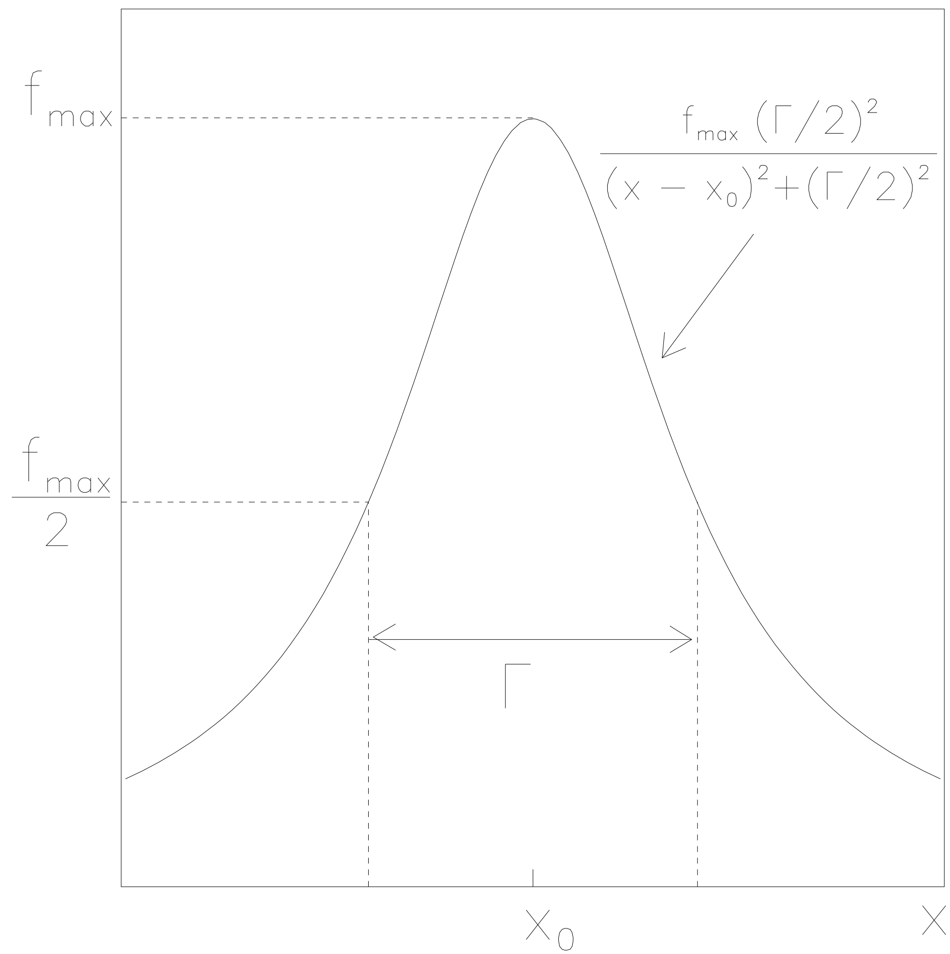
N – Número de medidas ν – Número de graus de liberdade

Distribuição Uniforme

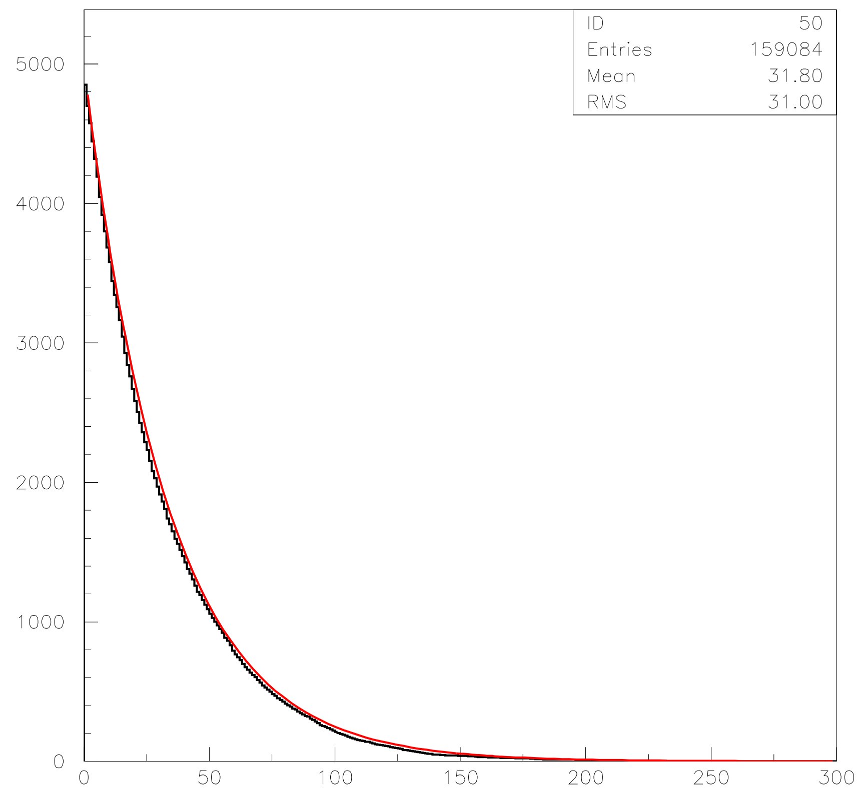


$$f(x) = \frac{1}{b-a}$$

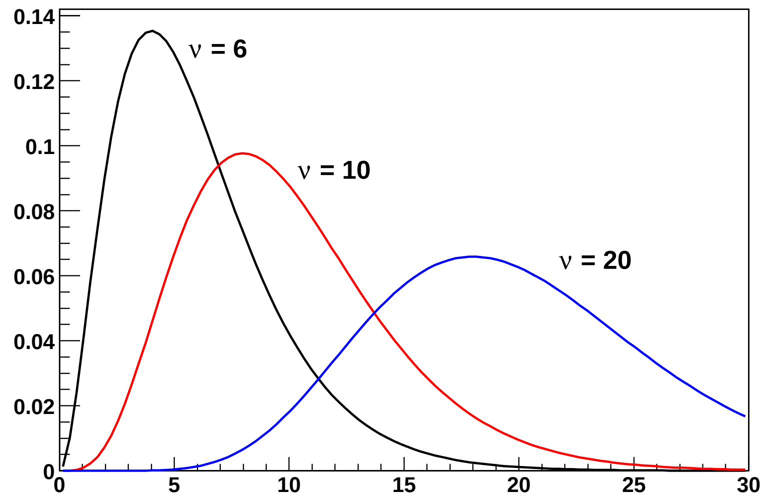
Distribuição de Breit-Wigner (Lorentz - Cauchy)



Distribuição Exponencial



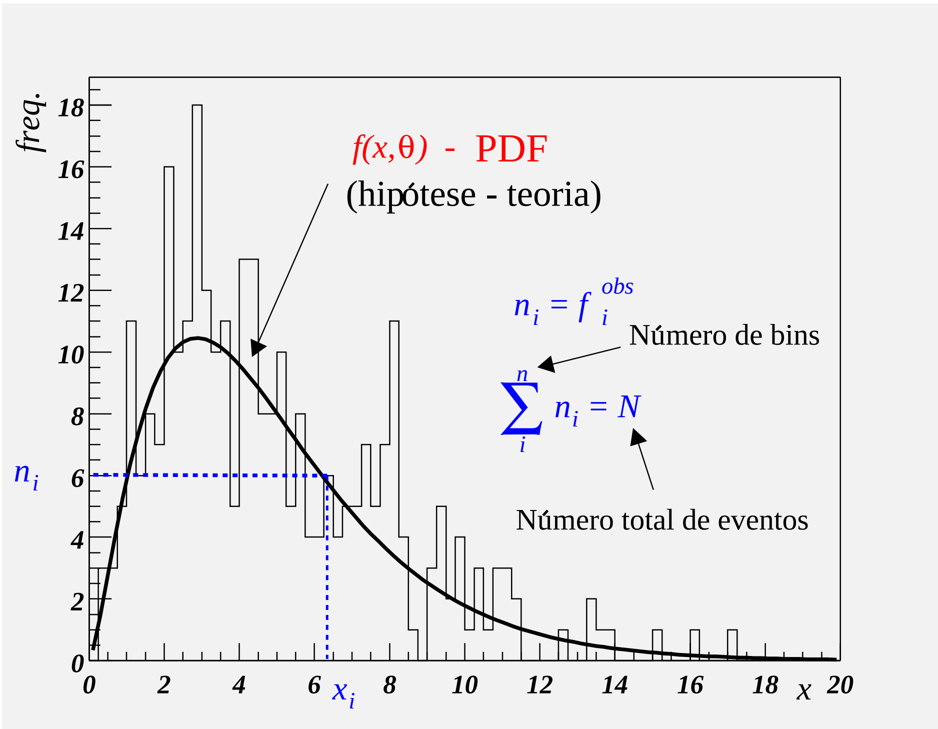
Distribuição de χ^2



$$f(\chi^2|\nu) = A\left(\frac{\chi^2}{2}\right)^{\frac{\nu}{2}-1} e^{-\frac{\chi^2}{2}}$$

- $\mu = \nu \quad \longrightarrow \quad$ valor médio
- $\sigma^2 = 2\nu \quad \longrightarrow \quad$ variância

Ajuste de Funções (PDF)



$p_i(\theta) = \int_{x_i^{\min}}^{x_i^{\max}} f(x, \theta) dx$ (probabilidade associada a cada bin)

$f_i^{\mathrm{obs}} = n_i$ (freqüência observada)

$f_i^{\mathrm{esp}} = \epsilon_i(\theta) = N p_i(\theta)$ (freqüência esperada segundo a PDF $f(x, \theta)$)

$\theta = (\theta_j) = (\theta_1, \theta_2, \ldots, \theta_p)$ (parâmetros da PDF)

- Como determinar os parâmetros?
- Quão bem a PDF se ajusta aos dados?

Método de Pearson - 1900

$$\chi^2(\theta) = \sum_{i=1}^{n} \frac{[n_i - \epsilon_i(\theta)]^2}{\epsilon_i} \approx \frac{[n_i - \epsilon_i(\theta)]^2}{n_i}$$

- Parâmetros θ_j são aqueles que minimizam χ^2

$$\frac{\partial \chi^2}{\partial \theta_j} = 0 \quad j = 1, \ldots, p, \quad \sum_{i=1}^{n} n_i = N$$

$$(p+1) \quad \text{eq. de vínculo}$$

- $\chi^2 \rightarrow (n-p-1)$ termos independentes
- $\nu = (n-p-1)$ n° de graus de liberdade

Qualidade do Ajuste

- $f_{\text{max}} \approx f(\nu)$
- Parâmetros $(\theta_j) \Longrightarrow \chi^2_{\text{calc}}$
- Se $f(\chi^2_{\text{calc}}) \approx f_{\text{max}} \Longleftrightarrow \chi^2_{\text{calc}} \approx \nu$

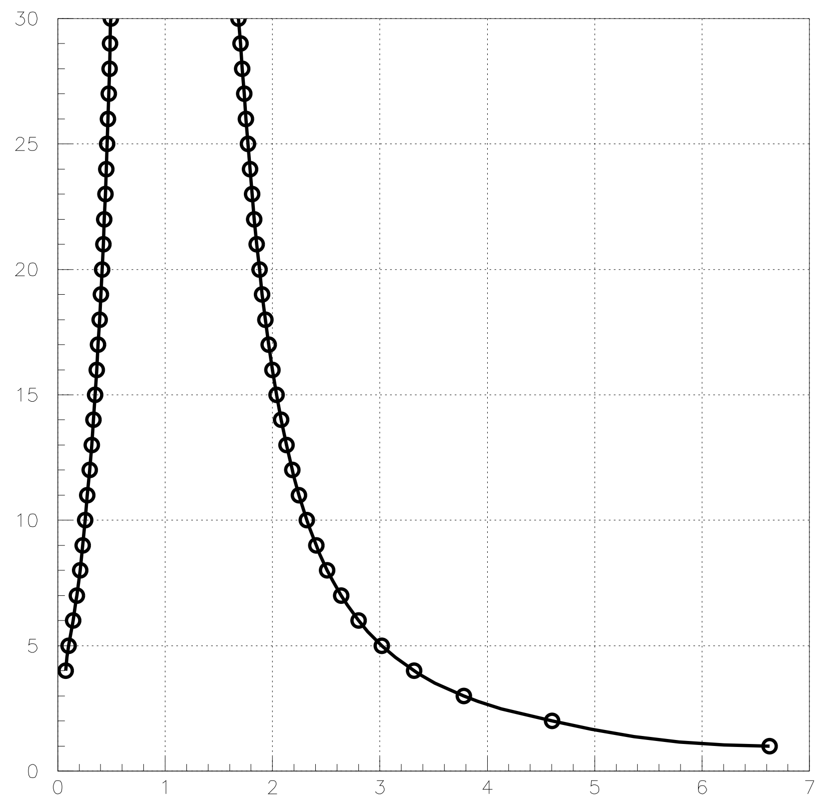
  $\boxed{\text{O ajuste é bom (??)}}$
- Se $p(\chi^2 > \chi^2_{\text{calc}}) = \int_{\chi^2_{\text{calc}}}^{\infty} f(\chi^2, \nu) d\chi^2$

  $\boxed{\text{for pequeno (??)}} \Longleftrightarrow \chi^2_{\text{calc}}(\text{grande})$

  o ajuste não é bom.
- Se $p(\chi^2 > \chi^2_{\text{calc}}) \approx 1(\chi^2_{\text{calc}} \to 0)$

  o ajuste não é compatível com os erros.

Intervalo de Confiança



Convenção: $(\chi^2_{\text{sup}} - \chi^2_{\text{inf}}) \Longleftrightarrow$ Intervalo de confiança de 90%

- $\displaystyle\int_{\chi^2_{\text{inf}}}^{\infty} f(\chi^2, \nu) d\chi^2 = 0.95$

- $\displaystyle\int_{\chi^2_{\text{sup}}}^{\infty} f(\chi^2, \nu) d\chi^2 = 0.05$

Lançamento de Dados

Ocorrência de uma face, $\Longleftrightarrow$ processo aleatório
valor ou evento (i) (probabilidade - p_i)

- a posteriori $\rightarrow p_i = \lim_{N\to\infty}\left(\frac{n_i}{N}\right)$
- a priori $\rightarrow$ $p_i = \frac{1}{6}$
- $\sum_{i=1}^{n=6} p_i = 1$ (normalização)
- $\langle i \rangle = \sum_{i=1}^{n=6} i p_i = \frac{1}{6} \underbrace{(1+2+3+4+5+6)}_{21} = 3.5$ (média)

Amostra de N lançamentos

freq. observadas $(f_i^{\text{obs}} = n_i) \quad \to \{n_1, \ldots, n_6\}$

$$\sum_{i=1}^{n=6} n_i = N, \quad p_i = \frac{1}{6}$$

freq. esperadas $(f_i^{\text{esp}} = Np_i = \epsilon_i) \quad \to \{\epsilon_1, \ldots, \epsilon_6\}$

$$\sum_{i=1}^{n=6} \epsilon_i = N \underbrace{\sum_{i=1}^{n=6} p_i}_{1} = N$$

$\boxed{N = 120}$

i	1	2	3	4	5	6
n_i	16	19	27	17	23	18
ϵ_i	20	20	20	20	20	20

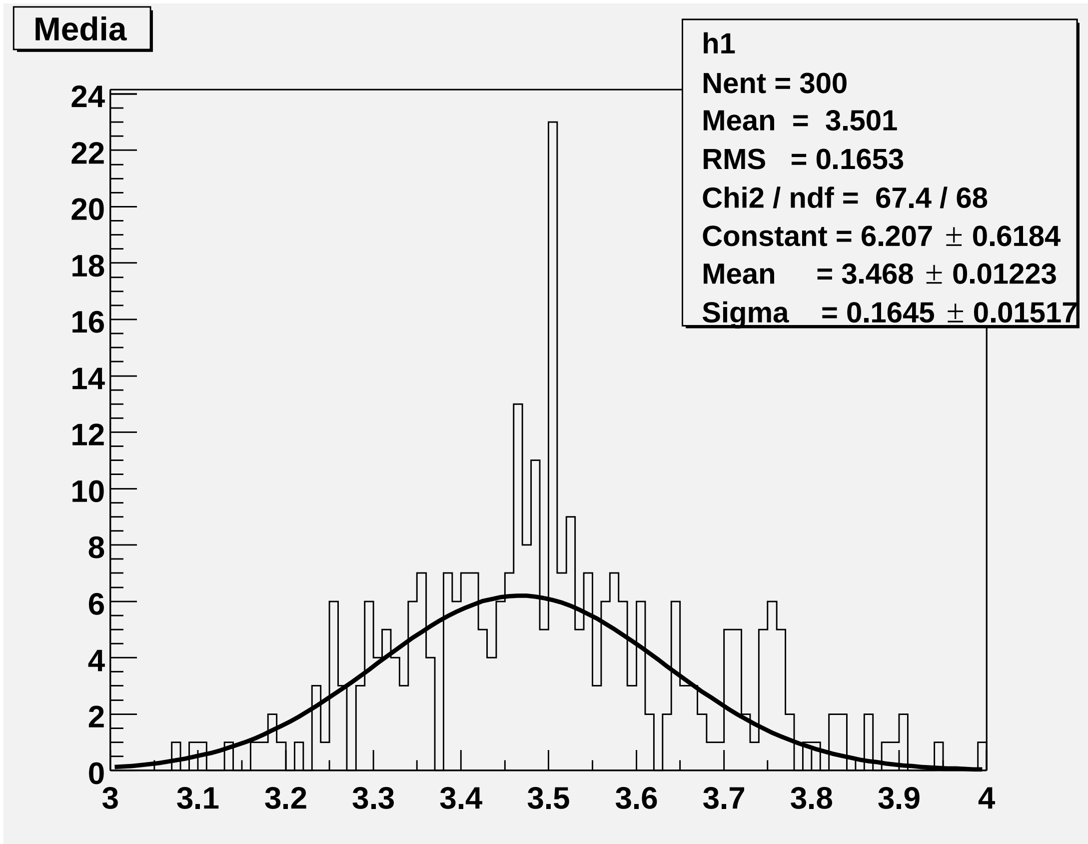
- As diferenças são significativas?
- Como caracterizar a discrepância?

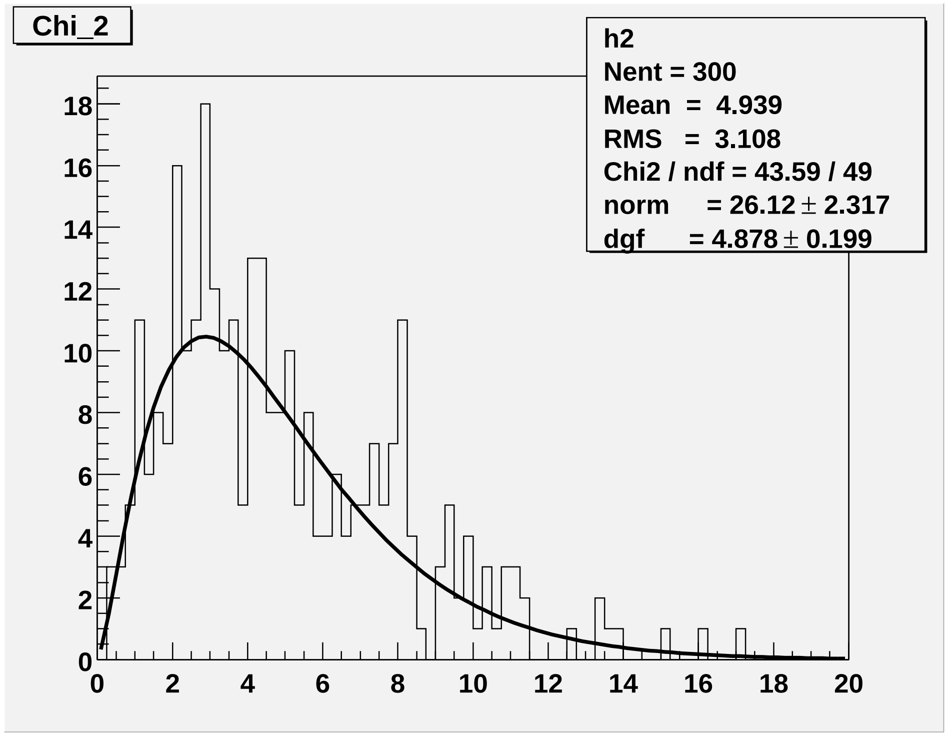
Medidas de discrepância

- $(n_i - \epsilon_i) \neq 0$ (desvio)
- $\displaystyle\sum_{i=1}^{n}(n_i - \epsilon_i) = \underbrace{\sum_{i=1}^{n} n_i}_{N} - \underbrace{\sum_{i=1}^{n} \epsilon_i}_{N} = 0$
- $\displaystyle\sum_{i=1}^{n}(n_i - \epsilon_i)^2$ (não suficiente)
- $\displaystyle\chi^2 = \sum_{i=1}^{n} \frac{(n_i - \epsilon_i)^2}{\epsilon_i} = \sum_{i=1}^{n} \frac{n_i^2}{\epsilon_i} - N$

χ^2 depende de $(n-1)$ termos independentes

$\nu = n - 1$ (número de graus de liberdade)


Media
h1
Nent = 300
Mean = 3.501
RMS = 0.1653
Chi2 / ndf = 67.4 / 68
Constant = 6.207 ± 0.6184
Mean = 3.468 ± 0.01223
Sigma = 0.1645 ± 0.01517
24
22
20
18
16
14
12
10
8
6
4
2
0
3
3.1
3.2
3.3
3.4
3.5
3.6
3.7
3.8
3.9
4



Dados

$$\begin{cases} \nu = 5 \\ \epsilon_i = N/6 \end{cases} \Longrightarrow \chi^2 = \frac{1}{N/6}\left(\sum_{i=1}^{n} n_i^2\right) - N$$

Bibliografia

- *Statistical Data Analysis*
  G. Cowan
  Oxford Press
- *Noções de Estatística, Simulações e Erros*
  A. Santoro, S. Novaes e V. Oguri
  CBPF-NT-001/01
- ROOT *Users Guide*
  http://root.cern.ch/root/
- http://pdg/lbl.gov/