

# Estatística básica & e estatística em HEP

Cristiane Jahnke Fiorini da Silva



**UNICAMP**

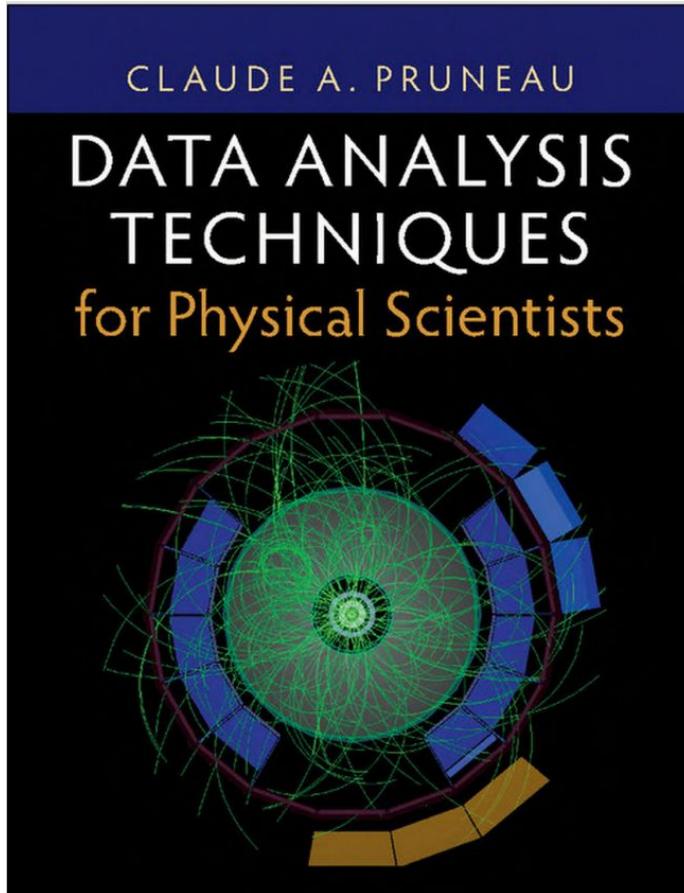


INSTITUTO DE FÍSICA  
GLEB WATAGHIN



Raios Cósmicos  
e Cronologia

# Estatística básica e estatística em High Energy Physics (HEP)



→ Capítulos 2, 3, 4, 5 e 12.

# A necessidade da probabilidade

- Medidas típicas envolvem uma grande sucessão de processos macroscópicos e microscópicos que alteram aleatoriamente seus resultados.
- Medições repetidas de uma dada quantidade física (ex: a posição, o momento ou a energia de um objeto) também produzem valores diferentes que aparentemente flutuam e adotam um padrão aleatório, o qual, na melhor das hipóteses, se agrupa em torno do valor real do observável em questão.

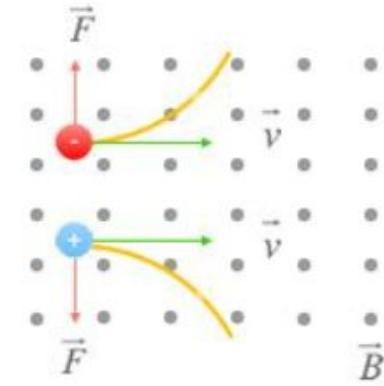


# A necessidade da probabilidade em HEP

→ Dispersões e perdas de energia em um detector afetam aleatoriamente o **momento das partículas** medidas em um espectrômetro magnético. Se uma partícula é desviada de uma forma ou de outra, sua direção e energia são ligeiramente alteradas, e o **momento determinado com base na trajetória da partícula fica ligeiramente diferente do valor real.**

→ Os valores medidos de observáveis físicos tendem a variar de uma medição para outra.

→ Medições repetidas de posições de partículas, momentos, etc devem produzir valores que variam de uma medição para outra, mas que devem **se agrupar em torno dos valores reais desses observáveis**.



# A necessidade da probabilidade

- O resultado das medições de observáveis científicos é, para todos os efeitos, um processo aleatório.
- Os resultados de uma determinada medição, ou de uma sucessão de medições, devem ser considerados variáveis aleatórias.
- **Mas** aleatoriedade não implica que todos os valores de um observável sejam igualmente prováveis: um experimento razoavelmente bem projetado e cuidadosamente realizado deve produzir valores que se agrupam próximos e em torno do valor real do observável.
- Quem já realizou uma sucessão de medições de um observável bem definido de acordo com um protocolo de medição específico pode atestar que esse agrupamento ocorre. Ex: medição do tamanho da mesa.

# Paradigmas na noção de probabilidade

Existem duas formas de abordagem estatística:

## → Abordagem Frequentista

- A estatística frequentista considera a probabilidade como uma **frequência** de eventos em experimentos repetidos.
- A probabilidade é considerada como **um valor limite da frequência relativa de um resultado** (seja um número discreto ou um intervalo de valores) quando o número de **tentativas se torna infinito**.
- Ela assume que cada parâmetro desconhecido é fixo e o objetivo é inferir esses parâmetros sem levar em conta conhecimentos prévios.

## → Abordagem Bayesiana

- Utiliza o **conhecimento prévio**, uma função de densidade de probabilidade (p.d.f.).
- Ela permite **incorporar informações prévias ao modelo e ajustá-las com os dados observados**.
- Muito útil para averiguação ou falsificação de uma determinada teoria.
- Muito utilizado para determinação de erros sistemáticos.

# Interpretações de probabilidade: frequentista vs Bayesiana

Cientistas, particularmente **físicos**, há muito **utilizam a interpretação frequentista e desenvolveram diversas ferramentas para atribuir erros às medições e realizar testes estatísticos de hipóteses científicas.**

Estudos de fenômenos em que a **noção de frequências limite** não se aplica prontamente levaram muitos cientistas a prestar mais atenção e a adotar a **interpretação Bayesiana** em seus estudos experimentais e na inferência de conclusões derivadas de seus experimentos.

A interpretação Bayesiana é agora utilizada em um número crescente de aplicações científicas.

# Interpretação frequentista

- A interpretação frequentista deriva seu nome do fato de que uma probabilidade pode, frequentemente, ser considerada como uma frequência relativa.
- Imagine um conjunto S sendo os possíveis resultados de uma medição
- Imagine A sendo um subconjunto de S.

$$p(A) = \lim_{n \rightarrow \infty} \frac{\text{number of occurrences of A}}{n}$$



Qual a probabilidade de se obter 4 ao se jogar um dado de 6 lados?  
Sabemos que a probabilidade é  $\frac{1}{6} \approx 0.1667$

Quantas vezes precisamos jogar o dado pra chegar neste valor?

[link: https://colab.research.google.com/drive/1WM5-jmsa1utldEOGgJ7RSfU2e0let9S?usp=sharing](https://colab.research.google.com/drive/1WM5-jmsa1utldEOGgJ7RSfU2e0let9S?usp=sharing)

# Interpretação Bayesiana

→ Pode ser formulado com base na teoria dos conjuntos e na probabilidade como lógica

$$p(A) = \text{degree of belief that the statement A is true}$$

- Utiliza o conhecimento prévio, uma função de densidade de probabilidade (p.d.f.).
- Ela permite incorporar informações prévias ao modelo e ajustá-las com os dados observados.

# Teorema de Bayes e inferência

- O **Teorema de Bayes** é um princípio fundamental em probabilidade que descreve como **atualizar a probabilidade** de uma hipótese **com base em novas evidências**.
- Ele fornece uma maneira de **combinar o conhecimento prévio com informações recém-observadas para fazer inferências**.

$$p(H_i|D, I) = \frac{p(D|H_i, I)p(H_i|I)}{p(D|I)}$$

- A quantidade  $p(\mathbf{D} | \mathbf{H}_i, \mathbf{I})$  representa a probabilidade de observar os dados  $\mathbf{D}$  se  $\mathbf{H}_i$  (hipóteses) e  $\mathbf{I}$  (informação prévia) forem verdadeiros.
- As quantidades  $p(\mathbf{H}_i | \mathbf{I})$  e  $p(\mathbf{H}_i | \mathbf{D}, \mathbf{I})$  representam a **probabilidade a priori** e a **probabilidade a posteriori** da hipótese  $H_i$ .
- A probabilidade  $p(\mathbf{H}_i | \mathbf{I})$  é baseada unicamente no conhecimento prévio, enquanto  $p(\mathbf{H}_i | \mathbf{D}, \mathbf{I})$  inclui tanto o conhecimento prévio quanto o novo conhecimento fornecido pela medição  $\mathbf{D}$ .
- O **denominador,  $p(\mathbf{D} | \mathbf{I})$** , corresponde à probabilidade de obter os dados  $\mathbf{D}$  dada a informação prévia disponível sobre o sistema ou fenômeno.

# Um exemplo didático de Aplicação do Teorema de Bayes

Imagine que você está tentando descobrir se uma pessoa está com gripe. Dado que:

- 10% das pessoas em uma determinada população estão gripadas.
- Existe um exame que detecta a gripe com 95% de precisão (ou seja, se a pessoa está gripada, o teste será positivo em 95% dos casos).
- Entretanto, o exame também dá um resultado positivo em 2% dos casos para pessoas que não estão gripadas (esses são os falsos positivos).

# Um exemplo didático de Aplicação do Teorema de Bayes

**Se uma pessoa faz o exame e o resultado é positivo, qual é a probabilidade de ela estar realmente gripada?**

- Probabilidade a priori ( $P(\text{Gripe})$ ): é a probabilidade de a pessoa estar gripada antes de vermos o resultado do exame. No exemplo, isso é 10% ou 0,10.
- Probabilidade condicional ( $P(\text{Exame positivo} | \text{Gripe})$ ): é a probabilidade de o exame ser positivo dado que a pessoa tem gripe. Nesse caso, é 95% ou 0,95.
- Probabilidade inversa ( $P(\text{Exame positivo} | \text{Sem gripe})$ ): é a probabilidade de o exame ser positivo mesmo que a pessoa não tenha gripe. Aqui, é 2% ou 0,02.

O **Teorema de Bayes** nos ajuda a encontrar a **probabilidade posterior**, ou seja, a probabilidade de uma pessoa estar gripada dado que o exame foi positivo. A fórmula é:

$$p(H_i|D, I) = \frac{p(D|H_i, I)p(H_i|I)}{p(D|I)}$$

$$P(\text{Gripe}|\text{Exame positivo}) = \frac{P(\text{Exame positivo}|\text{Gripe}) \cdot P(\text{Gripe})}{P(\text{Exame positivo})}$$

# Um exemplo didático de Aplicação do Teorema de Bayes

Precisamos calcular o denominador  $P(\text{Exame positivo})$ , que é a probabilidade de o exame ser positivo para qualquer pessoa. Isso pode acontecer de duas formas:

1. A pessoa está gripada e o exame dá positivo.
2. A pessoa não está gripada, mas o exame ainda assim dá positivo (falso positivo).

$$P(\text{Exame positivo}) = P(\text{Exame positivo}|\text{Gripe}) \cdot P(\text{Gripe}) + P(\text{Exame positivo}|\text{Sem gripe}) \cdot P(\text{Sem gripe})$$

$$P(\text{Gripe}) = 0,10$$

$$P(\text{Sem gripe}) = 1 - 0,10 = 0,90$$

$$P(\text{Exame positivo}|\text{Gripe}) = 0,95$$

$$P(\text{Exame positivo}|\text{Sem gripe}) = 0,02$$

$$P(\text{Exame positivo}) = (0,95 \cdot 0,10) + (0,02 \cdot 0,90)$$

$$P(\text{Exame positivo}) = 0,095 + 0,018 = 0,113$$

Aplicando o Teorema de Bayes:

$$P(\text{Gripe}|\text{Exame positivo}) = \frac{0,95 \cdot 0,10}{0,113} = \frac{0,095}{0,113} \approx 0,841$$

Mesmo com um teste positivo, a probabilidade de a pessoa estar realmente gripada é de **84,1%**. Isso mostra que, apesar do teste ser bom, a probabilidade inicial de gripe (10%) e a taxa de falsos positivos afetam muito a confiança no resultado.

# Cálculo de eficiência usando um método Bayesiano

- Quando a **eficiência** é muito próxima de zero ou de um, os cálculos de erros da eficiência apresentam problemas!
- Podem resultar em barras de erro excedendo o valor 1.

Usando Método Bayesiano:

- Prior: Eficiência está entre 0 e 1.
- Usa a probabilidade binomial de k eventos passaram na seleção.

$$P(\epsilon|k, N, I) = \frac{P(k|\epsilon, N, I)P(\epsilon|N, I)}{\mathcal{Z}}$$

## Calculating Efficiencies and Their Uncertainties

Marc Paterno  
FNAL/CD/CEPA/SLD  
paterno@fnal.gov  
FERMILAB-TM-2286-CD

May 22, 2004

### Abstract

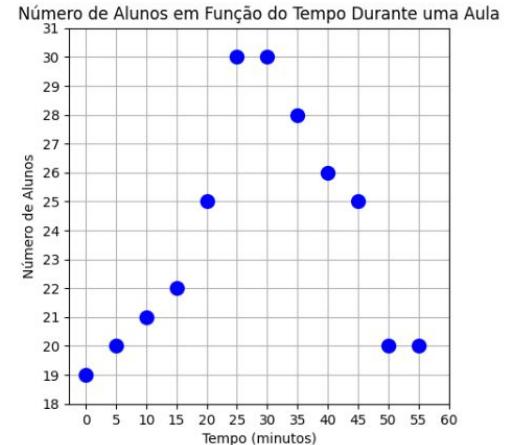
The commonly used methods for the calculation of the statistical uncertainties in cut efficiencies ("Poisson" and "binomial" errors) are both defective, as is seen in extreme cases. A method for the calculation of uncertainties based upon Bayes' Theorem is presented; this method has no problem with extreme cases. A program for the calculation of such uncertainties is also available. 

[Eficiencia\\_Bayesiana.ipynb - Colab](#)

# Distribuição de probabilidade e densidade de probabilidade

## → Variáveis Discretas

- Definição: Podem assumir apenas **valores específicos e contáveis**.
- Exemplos: Número de alunos em uma sala, quantidade de carros em um estacionamento, resultados de um dado (1, 2, 3, 4, 5 ou 6).
- Representação: São muitas vezes representadas por **números inteiros**.



## → Variáveis Contínuas

- Definição: Podem assumir **qualquer valor** em um intervalo contínuo, incluindo **frações e decimais**.
- Exemplos: Altura de uma pessoa, temperatura, peso, tempo.
- Representam medidas e podem ter **infinitos valores dentro de um intervalo**.
- Representação: Geralmente representadas por **números reais**, com casas decimais



Plots:

<https://colab.research.google.com/drive/1lfz49vjZtDXgrdETnXjh2t7se4h6bZQ?usp=sharing>

# Distribuição de probabilidade e densidade de probabilidade

Processos randômicos, ou estatísticos são descritos por uma função de densidade de probabilidade (ou função distribuição de probabilidade, **PDF**)  
→ fornece a freqüência esperada de ocorrência de cada evento.

Por exemplo, no caso de um dado, a variável randômica “x” tem uma densidade de probabilidade igual a  $P(x)=1/6$ .



A densidade de probabilidade é uma função normalizada:

$$\int P(x)dx = 1$$

$$\sum_i P(x_i) = 1$$

# PDF para variáveis discretas

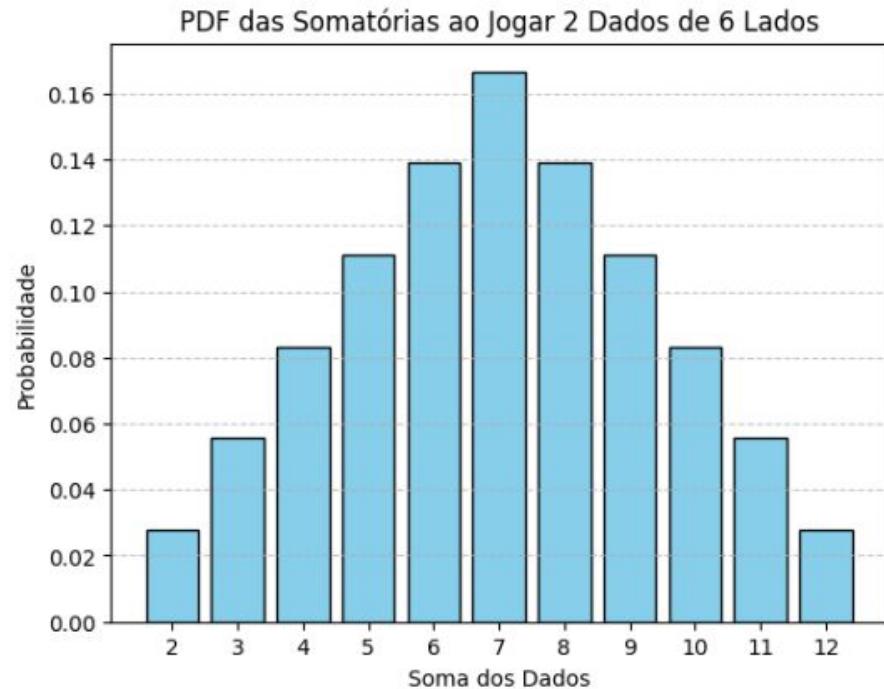
- Probabilidade de cair a face 6 de um dado?
- Probabilidade de, ao se jogar dois dados, a soma dos números ser 6? e 7?  
Qual soma é mais provável, 6 ou 7?

**Façam como exercício!**



# PDF para variáveis discretas

Probabilidade de, ao se jogar dois dados, a soma dos números ser 6? e 7?



Link: [Joga dois dados.ipynb - Colab](#)

# PDF para variáveis contínuas

Consideremos um experimento cujo resultado consiste em um único observável contínuo X.

O **espaço amostral S** associado a essa medida pode consistir em um **subconjunto de números reais**.

Neste caso **existe infinitos valores possíveis dentro de um subconjunto** e não é significativo considerar a probabilidade de um único valor x.

→ Considerar a probabilidade de que tal valor observado x seja encontrado dentro do intervalo infinitesimal **[x, x+dx]**.

→ Essa probabilidade pode ser avaliada por meio de uma função **f(x)**, conhecida como **Função Densidade de Probabilidade (PDF)**.

$$\text{Probability to observe } X \text{ in } [x, x + dx] = f(x) dx$$

→ Na **interpretação frequentista**,  $f(x) dx$  corresponde à **fração de vezes que o valor x é encontrado no intervalo [x,x+dx]**, no limite em que o número total de **medidas** é infinitamente grande.

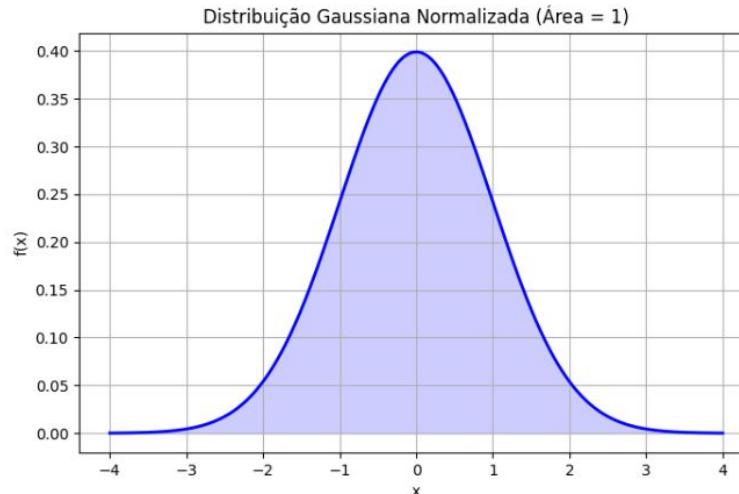
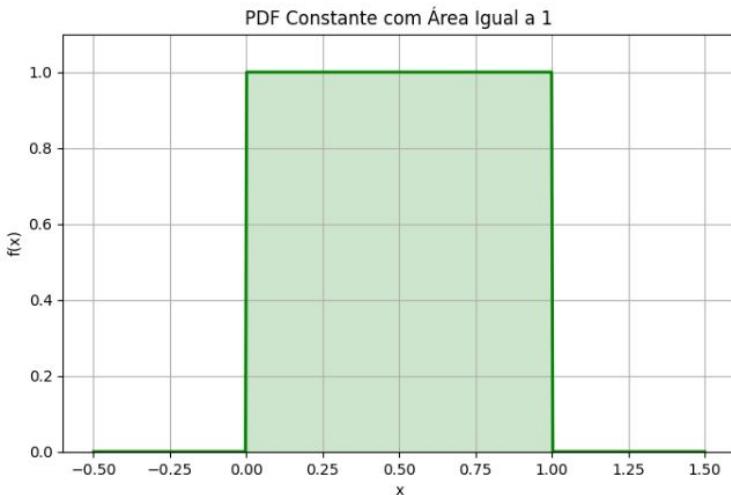
→ Na **interpretação Bayesiana**, essa quantidade fornece **o grau de crença de que um valor observado x possa estar nesse intervalo**, sem qualquer suposição específica sobre a possibilidade de o experimento ser repetido.

# PDF para variáveis contínuas

Uma PDF deve ser normalizada de forma que a probabilidade de qualquer resultado seja igual a um.

A soma das probabilidades de todos os resultados é, portanto, a integral da função  $f(x)$  sobre todo o espaço amostral  $S$  (ou espaço de hipóteses, conforme o caso) abrangido pelo observável  $X$ .

$$\int_S f(x) dx = 1$$



# Distribuição cumulativa e função densidade

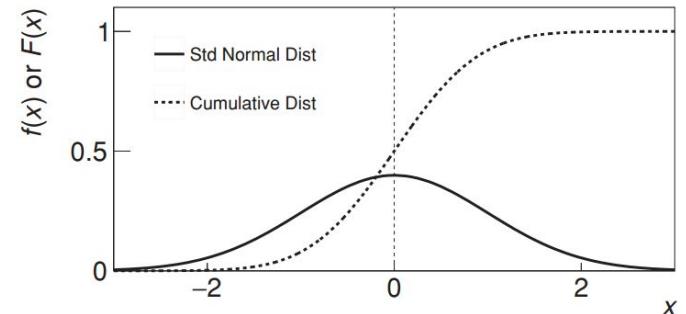
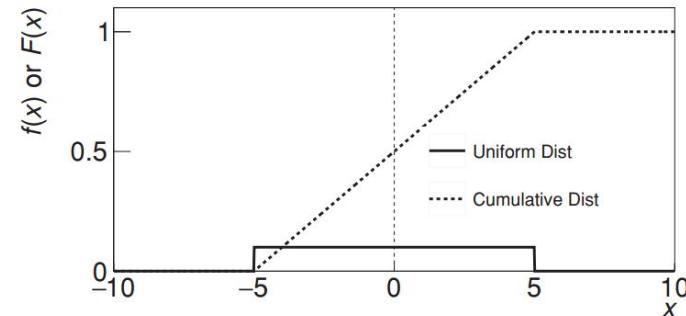
A função  $F(x)$  corresponde à probabilidade de que uma variável aleatória  $X$  assuma um valor menor ou igual a  $x$ .

$$F_n = \sum_{i=0}^n f(x_i)$$

$$F(x) = \int_{-\infty}^x f(x') dx'$$

A PDF  $f(x)$  expressa a probabilidade de que ocorrências, ou eventos, sejam observados no intervalo  $[x, x+dx]$ ; também se pode escrever:

$$f(x) \equiv \frac{1}{N} \frac{dN}{dx}$$



# Distribuição cumulativa e função densidade

Na **Distribuição de Maxwell-Boltzmann**, que descreve as velocidades das moléculas de um gás ideal, a **distribuição cumulativa** é usada para calcular a **probabilidade acumulada** de partículas terem velocidades inferiores a um certo valor.

# Distribuição cumulativa e função densidade

Na **Distribuição de Maxwell-Boltzmann**, que descreve as velocidades das moléculas de um gás ideal, a **distribuição cumulativa** é usada para calcular a **probabilidade acumulada** de partículas terem velocidades inferiores a um certo valor.

Em sistemas térmicos, é útil para determinar a proporção de partículas que têm energia abaixo de um certo limiar, o que tem implicações para reações químicas e processos de difusão.

# Distribuição cumulativa e função densidade

Na **Distribuição de Maxwell-Boltzmann**, que descreve as velocidades das moléculas de um gás ideal, a **distribuição cumulativa** é usada para calcular a **probabilidade acumulada** de partículas terem velocidades inferiores a um certo valor.

Em sistemas térmicos, é útil para determinar a **proporção de partículas que têm energia abaixo de um certo limiar**, o que tem implicações para reações químicas e processos de difusão.

Usada em **eletrônica e processamento de sinais** para **analisar o ruído em circuitos e detectores**. Ex: o ruído térmico em circuitos segue uma distribuição normal, e a **distribuição cumulativa** ajuda a entender as probabilidades de flutuações em torno da média. Isso é **crucial em experimentos que exigem sensibilidade a pequenos sinais**, como em interferometria ou sensores de alta precisão.

# Distribuição cumulativa e função densidade

Na **Distribuição de Maxwell-Boltzmann**, que descreve as velocidades das moléculas de um gás ideal, a **distribuição cumulativa** é usada para calcular a **probabilidade acumulada** de partículas terem velocidades inferiores a um certo valor.

Em sistemas térmicos, é útil para determinar a **proporção de partículas que têm energia abaixo de um certo limiar**, o que tem implicações para reações químicas e processos de difusão.

Usada em **eletrônica e processamento de sinais** para **analisar o ruído em circuitos e detectores**. Ex: o ruído térmico em circuitos segue uma distribuição normal, e a **distribuição cumulativa** ajuda a entender as probabilidades de flutuações em torno da média. Isso é **crucial em experimentos que exigem sensibilidade a pequenos sinais**, como em interferometria ou sensores de alta precisão.

**Tomada de Decisão:** Em física experimental, a **distribuição cumulativa** ajuda na tomada de decisão ao estabelecer limiares de detecção ou significância estatística, permitindo que os físicos determinem se certos valores medidos são significativos ou simplesmente parte da flutuação estatística esperada.

# Função de variável randômica

Funções de variáveis aleatórias também são variáveis aleatórias.

Frequentemente é de interesse determinar ou caracterizar a densidade de probabilidade dessas variáveis aleatórias.

$$E = \sqrt{p^2 c^2 + m^2 c^4}$$

# Caracterização de PDF

- Média
- Variância
- Mediana
- Moda
- Momentos de uma PDF
- Assimetria (Skewness)
- Curtose (Kurtosis)

# Caracterização de PDF: média

Valor esperado (média): é uma medida de tendência central que representa o "centro de massa" da distribuição, ou seja, o ponto em torno do qual os valores da variável aleatória tendem a se concentrar.

$$\bar{x} = \frac{1}{n} \sum_{i=1}^n x_i \quad \mu = \sum_{i=1}^n x_i p(x_i)$$

$$E[x] \equiv \int_{-\infty}^{\infty} x f(x) dx \equiv \mu \quad E[x] \equiv \mu \equiv \mu_x \equiv \langle x \rangle$$

# Caracterização de PDF: variância

Variância: é uma medida de dispersão que quantifica o quanto os valores de uma variável se afastam da média

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

$$\text{Var}[x] \equiv E[(x - E[x])^2] = \int_{-\infty}^{\infty} (x - \mu)^2 f(x) dx = \sigma^2$$

Usada para caracterizar a "largura" de uma distribuição e para entender a incerteza e variabilidade nos valores de uma variável aleatória contínua.

# Caracterização de PDF: mediana

A **mediana** é o valor do meio.

Para um conjunto de N dados, ordenados, a mediana é definida para

N ímpar:

$$x_{\text{mediana}} = x_{(N+1)/2}$$

$$\{1, 2, 3, 4, 5, 6, 7\} \rightarrow x_{\text{mediana}} = 4$$

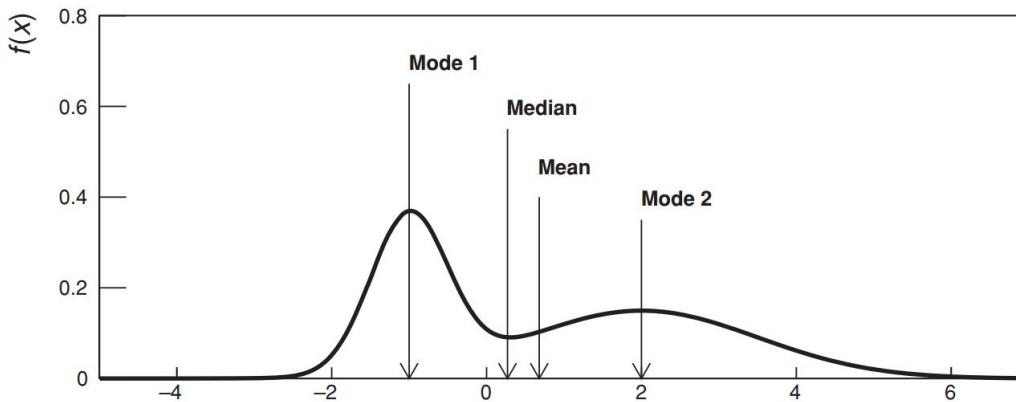
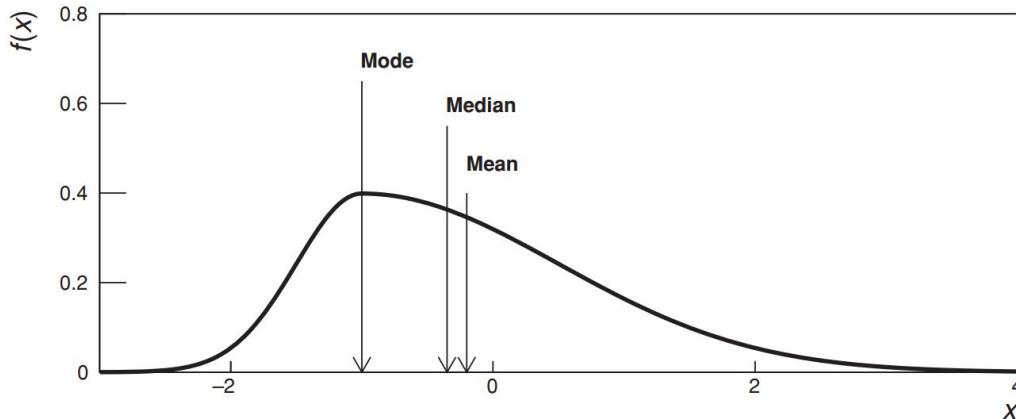
N par:

$$x_{\text{mediana}} = \frac{x_{N/2} + x_{(N+2)/2}}{2}$$

$$\{1, 2, 3, 4, 5, 6\} \rightarrow x_{\text{mediana}} = 3,5$$

# Caracterização de PDF: Moda

**Moda:** A moda de uma PDF  $f(x)$  é definida como o valor de  $x$  para o qual a PDF tem o valor de máximo



# Caracterização de PDF: momentos de uma PDF

$$\mu'_n \equiv E[x^n] = \int_{-\infty}^{\infty} x^n f(x) dx$$

$$\mu_n \equiv E[(x - E[x])^n] = \int_{-\infty}^{\infty} (x - \mu)^n f(x) dx$$

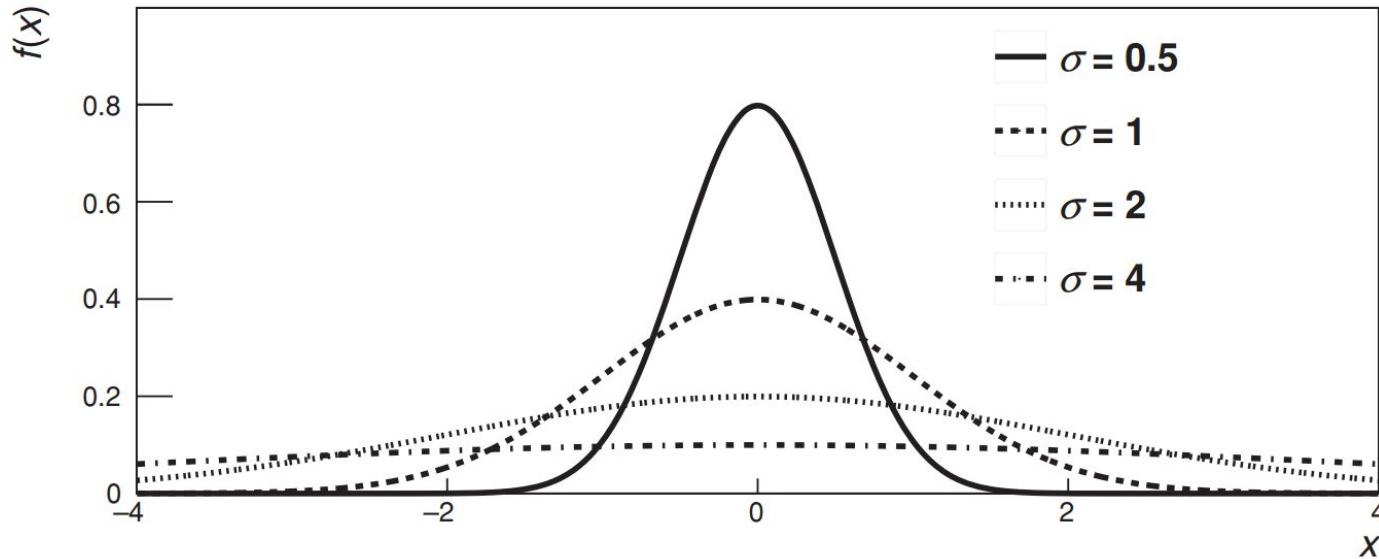
O primeiro momento é a média da PDF:

$$E[x] \equiv \int_{-\infty}^{\infty} x f(x) dx \equiv \mu$$

O segundo momento é a variância da PDF:

$$\mu_2 \equiv \text{Var}[x] \equiv E[(x - E[x])^2] = \int_{-\infty}^{\infty} (x - \mu)^2 f(x) dx = \sigma^2$$

# Ilustração do desvio padrão em uma distribuição



- PDFs amplamente distribuídas têm uma variância grande
- Distribuições estreitas têm uma variância pequena

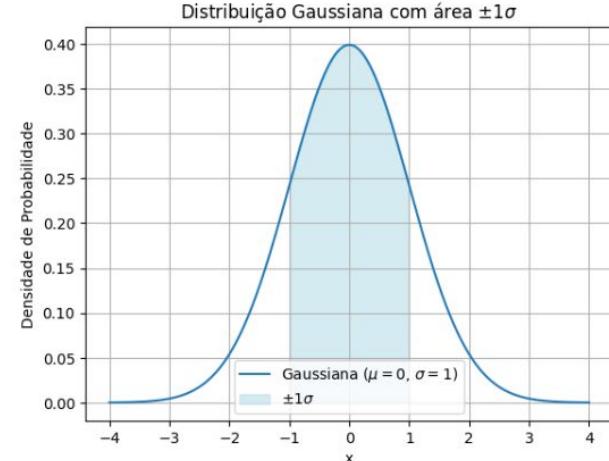
# Caracterização de PDF em HEP

Em física de altas energias, a **média** é geralmente mais utilizada do que a **mediana/moda** para caracterizar distribuições, especialmente quando a **distribuição é simétrica** e bem-comportada (como uma Gaussiana).

→ A média representa a **tendência central** e é diretamente aplicável em muitos modelos e experimentos estatísticos.

Porém, quando há **outliers** ou distribuições assimétricas (como distribuições de energia de partículas com caudas longas), a mediana pode ser preferida, **pois é menos sensível a valores extremos** e fornece uma representação mais robusta da centralidade.

Cada caso depende, então, das características específicas dos dados e do que se quer analisar.



$$f(x) = \frac{1}{\sqrt{2\pi}\sigma^2} \exp\left(-\frac{(x-\mu)^2}{2\sigma^2}\right)$$

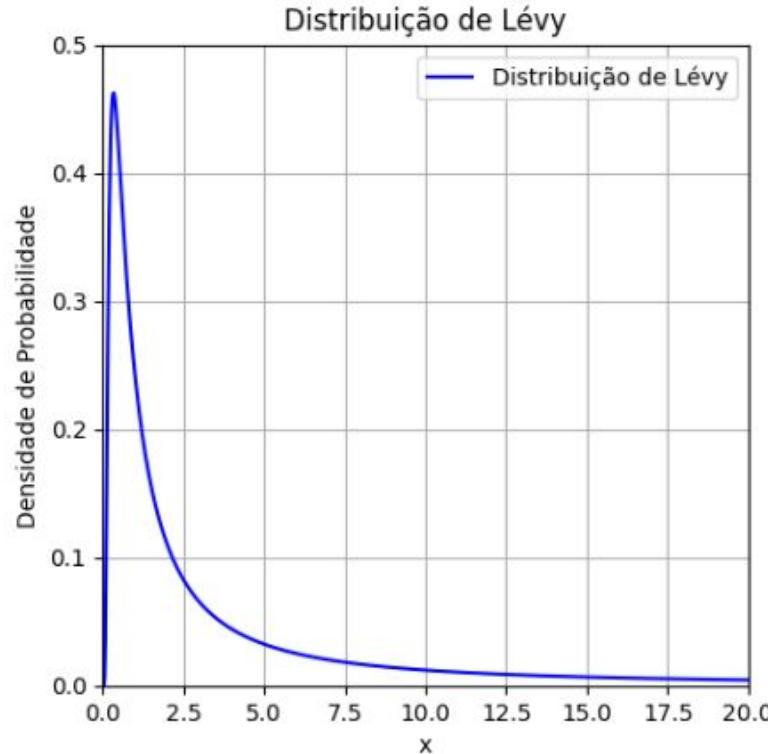
# Caracterização de PDF em HEP

Um exemplo clássico de uma **distribuição de caudas longas** é a **distribuição de Lévy**.

Características: A **distribuição de Lévy** tem uma cauda longa, o que significa que existem valores extremos que ocorrem com uma probabilidade não desprezível.

→ Isso pode levar a uma média que não representa bem a maioria dos dados, já que a presença de eventos de alta energia (os *outliers*) pode puxar a média para cima.

$$f(x; \mu, c) = \frac{1}{\sqrt{2\pi}c(x - \mu)^3} \exp\left(-\frac{c}{2(x - \mu)}\right) \quad \text{para } x \geq \mu$$



- $\mu$  é o valor mínimo para o qual a densidade de probabilidade é não nula. Se  $\mu = 0$ , a distribuição começa no zero.
- $c$  é o parâmetro de escala ("parâmetro de dispersão" ou "parâmetro de forma") e controla a largura da distribuição.

# Exemplo de cálculo de valor esperado (média):

$$p_u(x|a, b) = \begin{cases} \frac{1}{b-a} & \text{for } a \leq x \leq b \\ 0 & \text{elsewhere,} \end{cases}$$

$$\langle x \rangle \equiv \mu_x \equiv \int_{-\infty}^{\infty} p_u(x|a, b)x dx$$

**Façam como exercício!**



## Exemplo de cálculo de valor esperado (média):

$$p_u(x|a, b) = \begin{cases} \frac{1}{b-a} & \text{for } a \leq x \leq b \\ 0 & \text{elsewhere,} \end{cases}$$

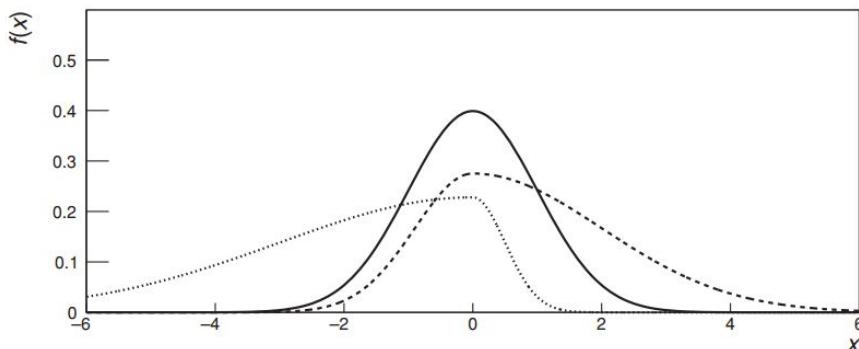
$$\begin{aligned}\langle x \rangle &= \mu_x \equiv \int_{-\infty}^{\infty} p_u(x|a, b)x dx \\ &= \frac{1}{b-a} \int_a^b x dx, \\ &= \frac{1}{b-a} \left[ \frac{x^2}{2} \right]_a^b, \\ &= \frac{a+b}{2},\end{aligned}$$

# Caracterização de uma PDF: Assimetria e Curtose

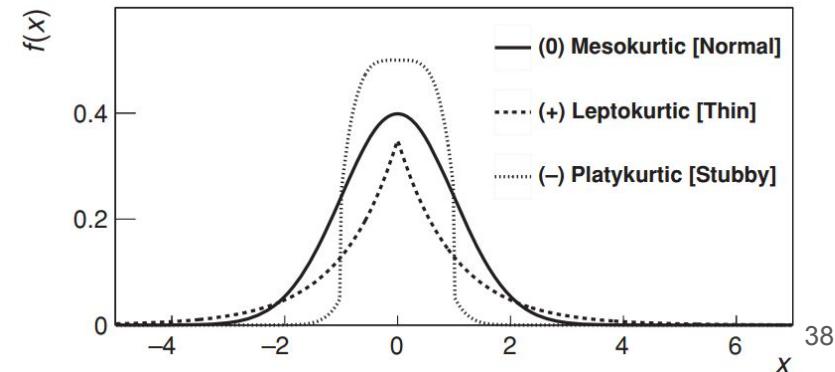
**Assimetria (Skewness):** Mede a simetria da distribuição em torno da média. Uma assimetria positiva indica uma cauda mais longa à direita (mais valores altos), enquanto uma assimetria negativa indica uma cauda mais longa à esquerda (mais valores baixos). Assimetria zero indica uma distribuição simétrica.

**Curtose:** Mede o grau de "achatamento" ou "alongamento" da distribuição em relação a uma distribuição normal. Uma curtose alta indica uma distribuição com caudas pesadas, enquanto uma curtose baixa indica uma distribuição com caudas leves.

$$\gamma_1 \equiv \text{Skew}[x] \equiv \mu_3^{\text{std}} = \frac{\mu_3}{\sigma^3}$$



$$\text{Kurt}_{\text{old}}[x] \equiv \frac{\mu_4}{\sigma^4}$$



# Intervalos de dados

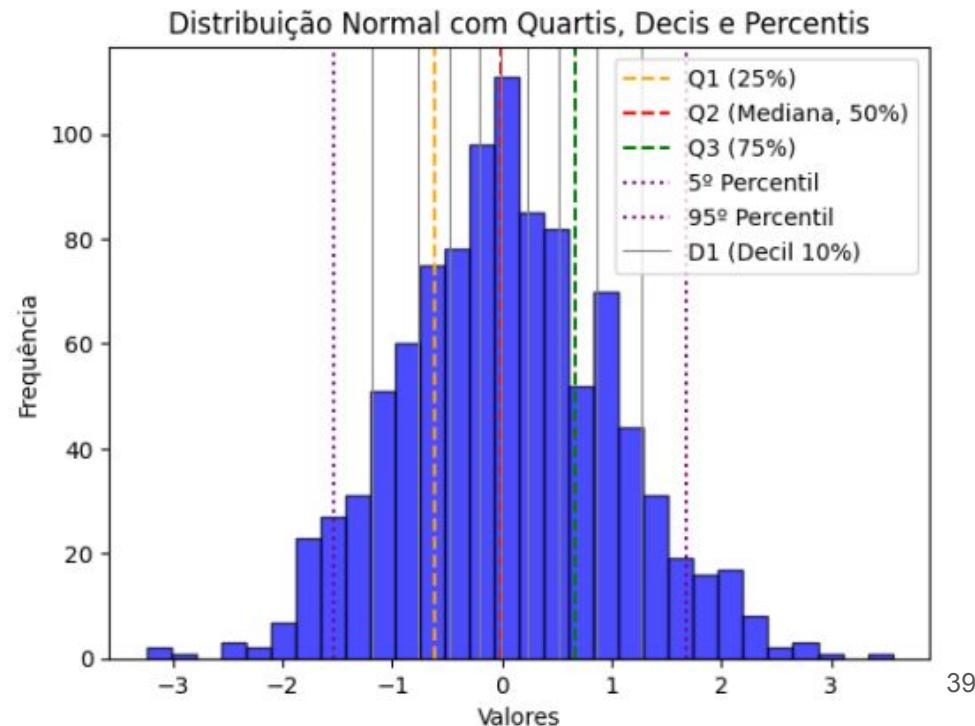
→ A maneira mais básica de caracterizar um conjunto de dados é descrever o [intervalo que ele abrange](#).

→ Para distribuições com caudas longas em direção a valores baixos ou altos (de baixa probabilidade), descrever o intervalo que abrange acaba não tendo um significado.

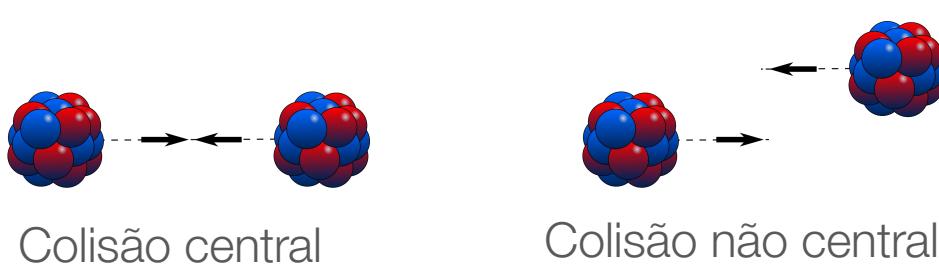
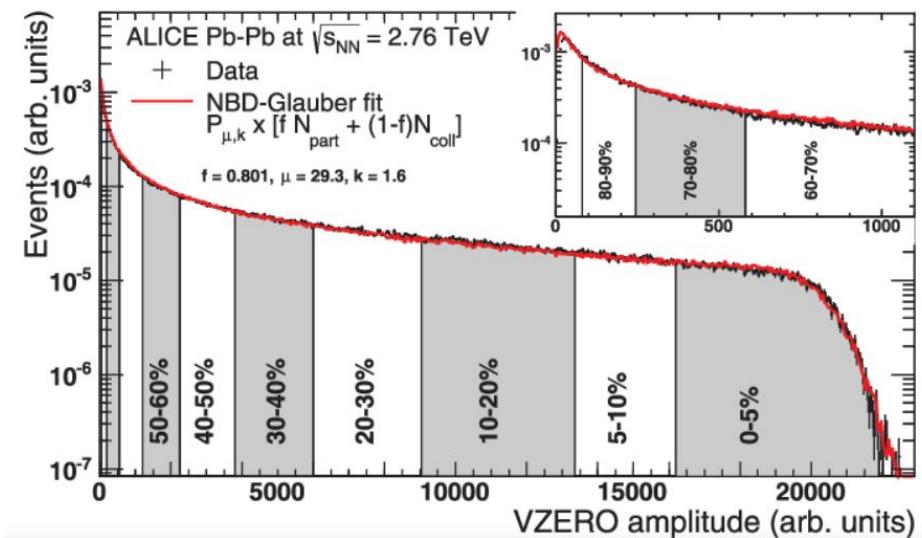
## Quartil, decil, percentil

→ A amplitude [interquartil](#) é calculada como a diferença entre os quartis superior e inferior. Os quartis inferior e superior correspondem ao ponto de 25% e ao ponto de 75% dos dados, respectivamente.

→ As noções de **decis** e **percentis** também são comumente usadas para indicar posições dentro de uma distribuição.



# Intervalos de dados: percentis em HEP



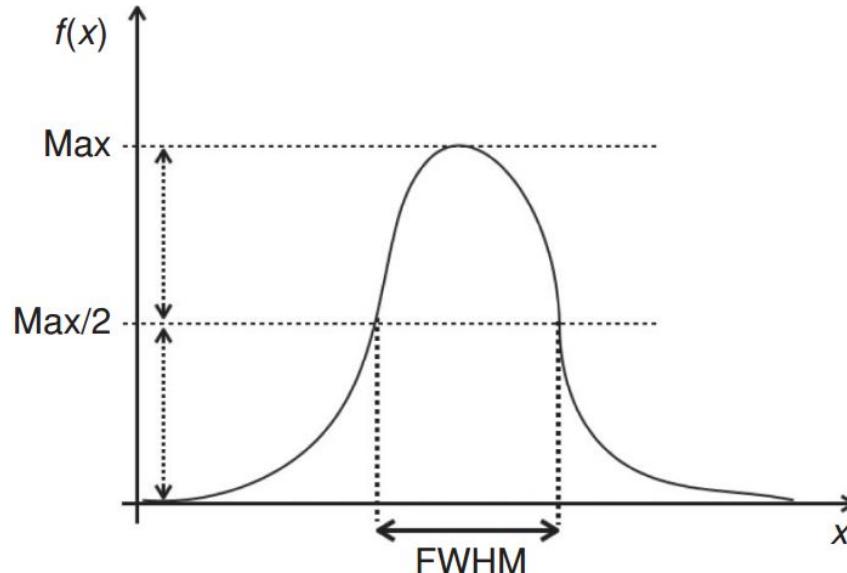
Em **colisões de Pb-Pb no LHC**, eventos podem ser selecionados de acordo com a quantidade de partículas produzidas e/ou sinal em um determinado detector.

A ideia de **percentil** é utilizada para determinar a centralidade das colisões: importante para separar eventos centrais de periféricos ou realizar análises em intervalos de centralidade.

# Intervalos de dados

Largura Total a Meia Altura (LTMA) (Full Width at Half Maximum FWHM)):

→ apresenta a vantagem, em relação ao desvio padrão, de ser relativamente imune aos efeitos associados a caudas em direção a valores baixos ou altos.



Exercício: Verifique que a [Largura Total a Meia Altura](#) de uma gaussiana é  $2.35\sigma$

**Façam como exercício!**



$$f(x) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{(x-\mu)^2}{2\sigma^2}\right)$$

Exercício: Verifique que a **Largura Total a Meia Altura** de uma gaussiana é  $2.35\sigma$

**Façam como exercício!**



$$f(x) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{(x-\mu)^2}{2\sigma^2}\right)$$

**Dica:** Para encontrar a FWHM, queremos determinar a largura entre os pontos x **onde  $f(x)$  é metade do valor máximo de  $f(x)$ .**

O valor máximo de  $f(x)$  ocorre em  $x = \mu$ , e é dado por:  $f(\mu) = \frac{1}{\sqrt{2\pi\sigma^2}}$

A meia altura é dada por:  $\frac{f(\mu)}{2} = \frac{1}{2\sqrt{2\pi\sigma^2}}$

**Resposta:** Verifique que a **Largura Total a Meia Altura** de uma gaussiana é  $2.35\sigma$

$$\frac{f(\mu)}{2} = \frac{1}{2\sqrt{2\pi\sigma^2}}$$

Na expressão da gaussiana:

$$\frac{1}{2\sqrt{2\pi\sigma^2}} = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{(x - \mu)^2}{2\sigma^2}\right)$$

Simplificando:

$$\frac{1}{2} = \exp\left(-\frac{(x - \mu)^2}{2\sigma^2}\right)$$

$$\ln\left(\frac{1}{2}\right) = -\frac{(x - \mu)^2}{2\sigma^2}$$

$$-\ln(2) = -\frac{(x - \mu)^2}{2\sigma^2}$$

$$(x - \mu)^2 = 2\sigma^2 \ln(2)$$

$$x - \mu = \pm\sigma\sqrt{2\ln(2)}$$

$$\text{FWHM} = 2|x - \mu| = 2\sigma\sqrt{2\ln(2)}$$

$$\text{FWHM} \approx 2 \times 1.177\sigma = 2.35\sigma$$

# Sistemas Multivariados

**Problemas científicos práticos raramente se limitam a medições de uma única variável (aleatoríia).**

# Sistemas Multivariados

**Problemas científicos práticos raramente se limitam a medições de uma única variável (aleatoríia).**

Em **física de partículas e nuclear**: experimentos modernos envolvem medições de um grande número de variáveis físicas simultaneamente.

# Sistemas Multivariados

**Problemas científicos práticos raramente se limitam a medições de uma única variável (aleatoríia).**

Em **física de partículas e nuclear**: experimentos modernos envolvem medições de um grande número de variáveis físicas simultaneamente.

Nos detectores, as quantidades medidas correspondem a voltagens produzidas por sensores, que podem ser interpretadas como **posições, momentos, ou energias das partículas**, etc

# Sistemas Multivariados

**Problemas científicos práticos raramente se limitam a medições de uma única variável (aleatoríia).**

Em **física de partículas e nuclear**: experimentos modernos envolvem medições de um grande número de variáveis físicas simultaneamente.

Nos detectores, as quantidades medidas correspondem a voltagens produzidas por sensores, que podem ser interpretadas como **posições, momentos, ou energias das partículas**, etc

**Colisões nucleares** produzem números variados de partículas com valores aleatórios de momento, energia ou até mesmo diferentes tipos de partículas.

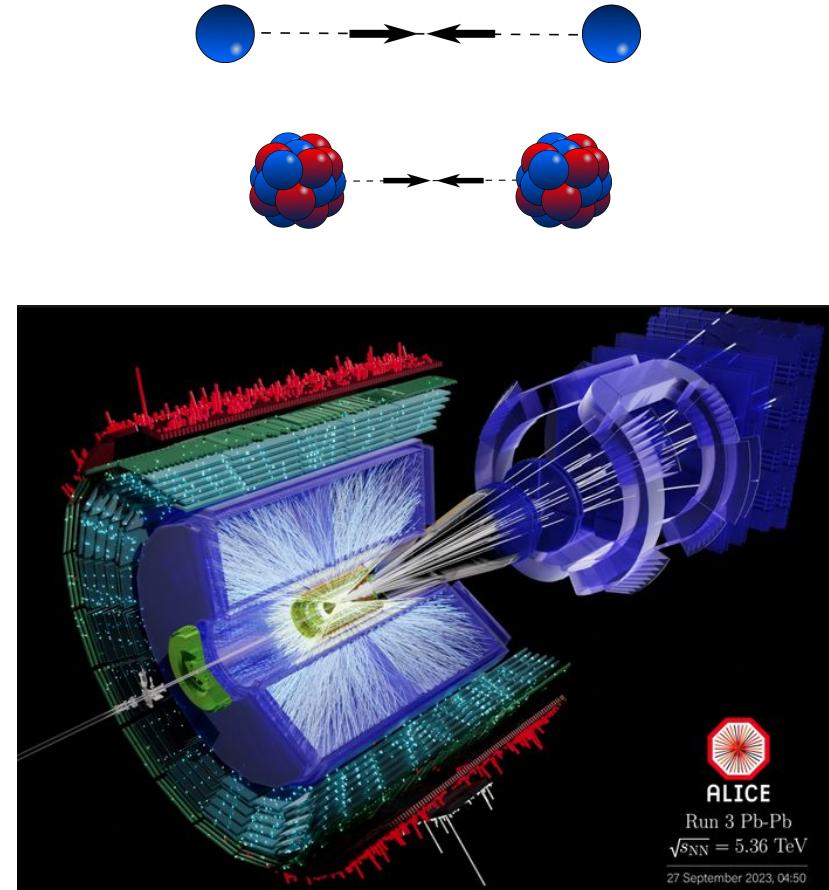
# Sistemas Multivariados

→ O número de variáveis físicas envolvidas nas interações núcleo-núcleo pode variar de algumas em interações brandas de próton-próton até **milhares** em colisões frontais de Pb-Pb no LHC.

Uma fração dessas variáveis pode estar correlacionada, enquanto outras podem ser completamente independentes.

Não se sabe a priori quais variáveis são estatisticamente independentes e quais estão correlacionadas.

→ É necessário formular a noção de densidades de probabilidade multivariadas (ou seja, envolvendo múltiplas variáveis) que englobam todas as variáveis medidas.



# Sistemas Multivariados: alguns exemplos em HEP

## Posição dos Tracks, Ângulo e Direção:

- Coordenadas espaciais ( $x, y, z$ ).
- Ângulo polar ( $\theta$ ) e azimuthal ( $\phi$ ), medidos em relação ao ponto de interação.
- Distância de menor aproximação do track ao ponto de interação primário.

## Momento:

- Momento transversal ( $p_T$ ): componente do momento perpendicular ao feixe, importante para estudos de colisões de alta energia.
- Momento total ( $p$ ): resultado da combinação vetorial dos momentos  $p_x$ ,  $p_y$  e  $p_z$

## Energia:

- Energia total da partícula.

## Identificação da Partícula:

- Tipo de partícula (elétrons, múons, píons, prótons, etc.).

## Carga das partículas:

- Carga elétrica da partícula, inferida pela curvatura da trajetória no campo magnético.

## Densidade de Ionização ( $dE/dx$ ):

- Taxa de perda de energia por unidade de comprimento, útil para a identificação da partícula em conjunto com outras variáveis.

## Multiplicidade de Partículas:

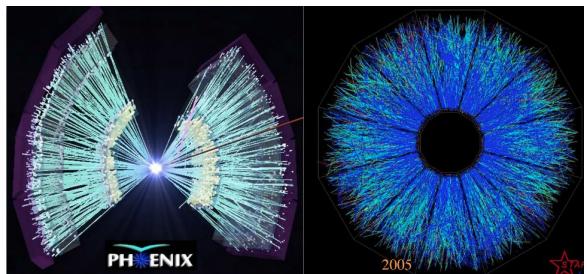
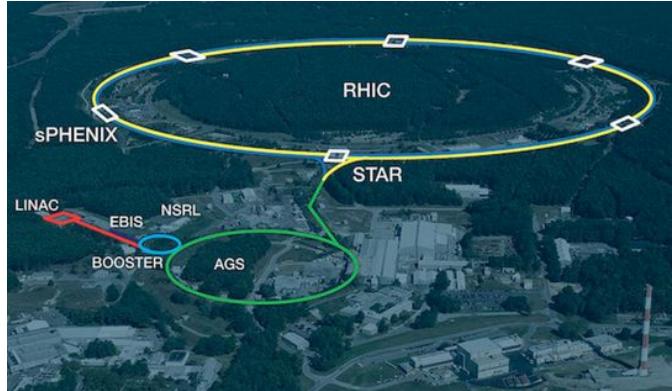
- Número de partículas detectadas em cada evento.

## Parâmetros de Vertex:

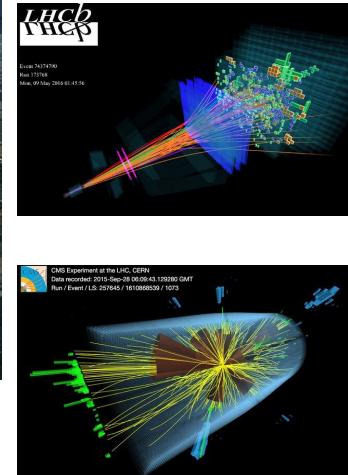
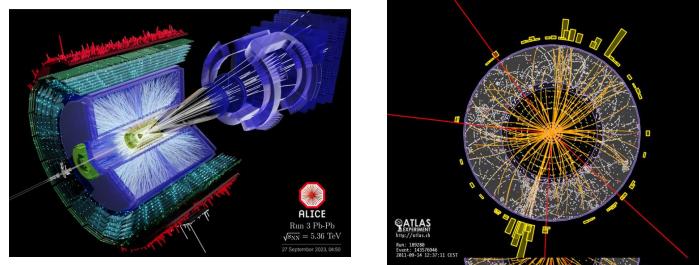
- Coordenadas do ponto de origem (ou vértice) da colisão e vértices secundários

# Os diversos detectores do RHIC e do LHC

## RHIC: PHENIX, STAR



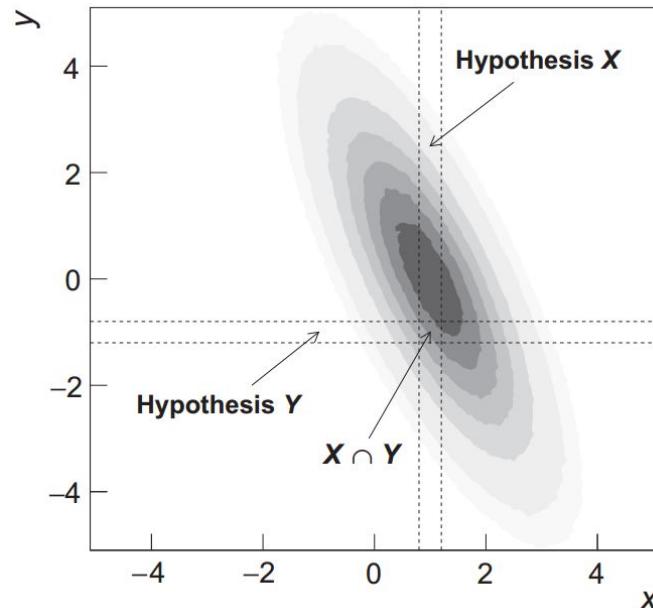
## LHC: ALICE, ATLAS, CMS, and LHCb



# Noção de densidade de probabilidade conjunta

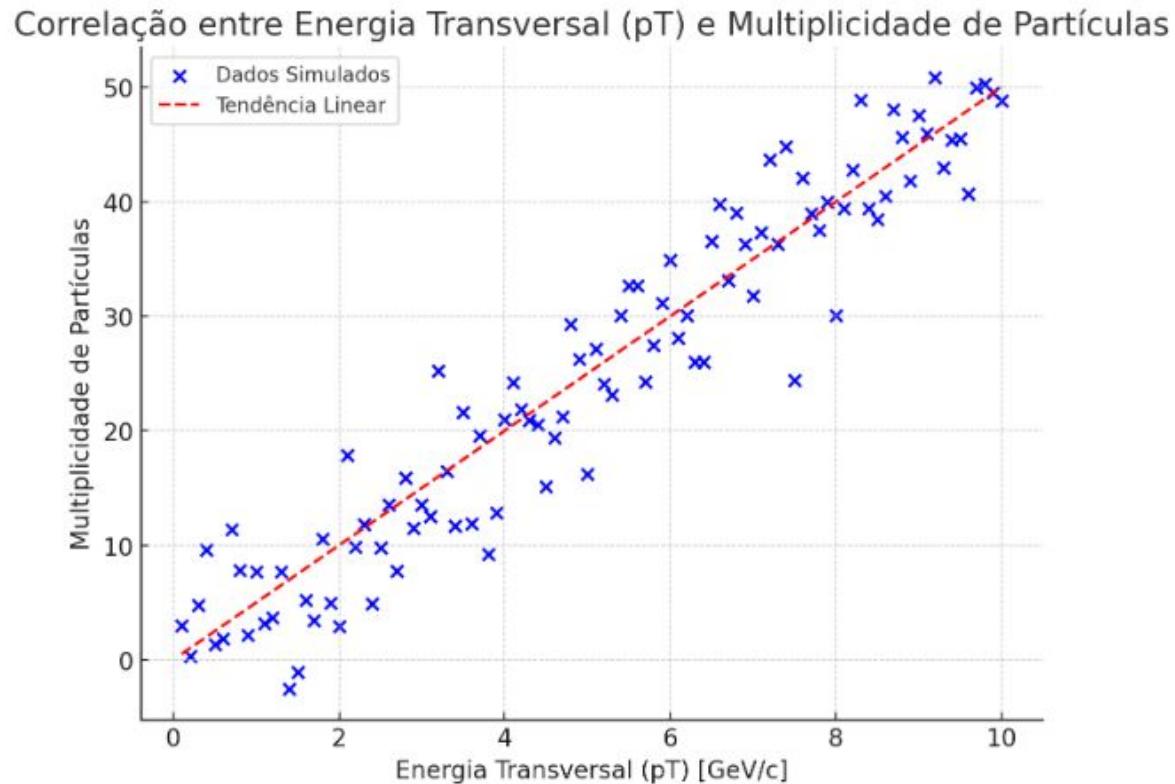
$$p(X, Y|I) = \lim_{\Delta x, \Delta y \rightarrow 0} \frac{p(x \leq X < x + \Delta x, y \leq Y < y + \Delta y|I)}{\Delta x \Delta y}$$

Nível de confiança de que as hipóteses X  
e Y possam ser verdadeiras  
conjuntamente



# Variáveis correlacionadas em análises

Exemplo de correlação entre “ $p_T$ ” e “multiplicidade das partículas”.

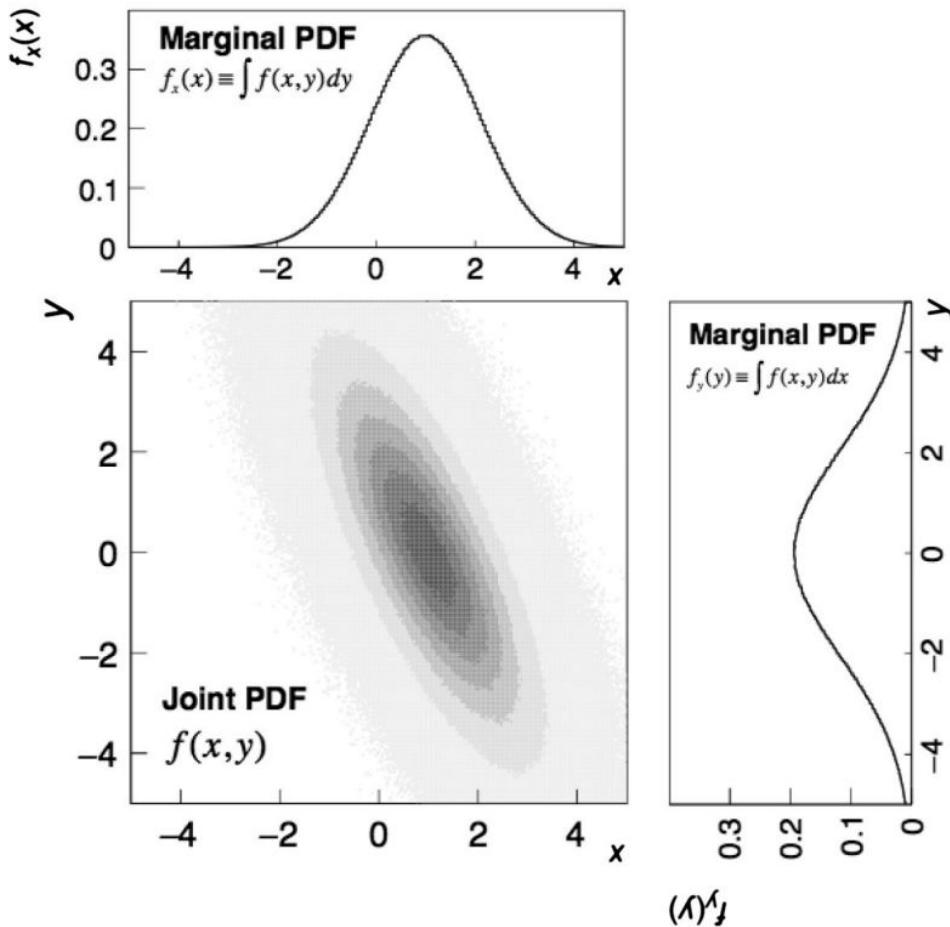


# Probabilidade marginal

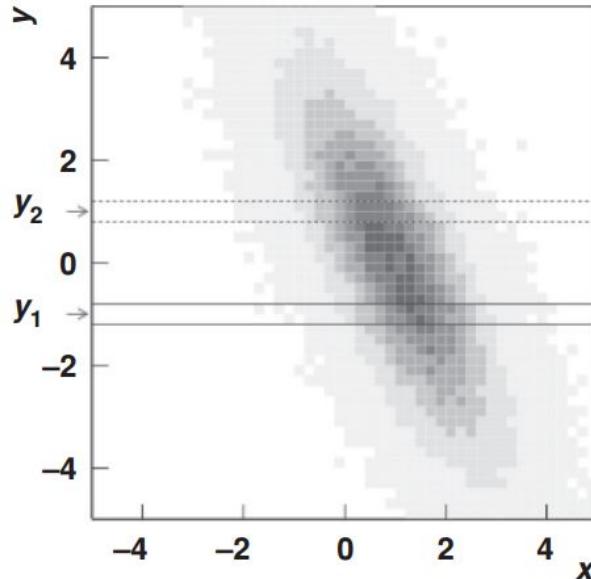
A probabilidade marginal é a probabilidade de que uma variável específica assuma um determinado valor ou intervalo, independente dos valores de outras variáveis.

Densidade de probabilidade conjunta  $x$  e  $y$ , pode-se:

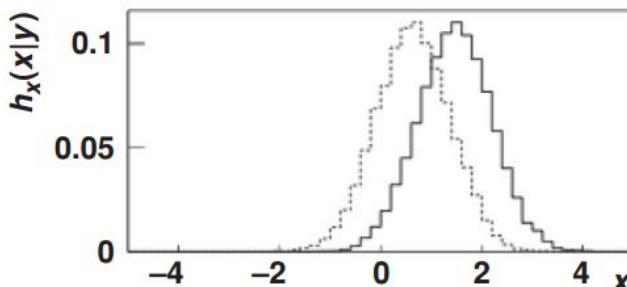
- integrar em  $y$  e obter a PDF só de  $x$
- integrar em  $x$  e obter a PDF só de  $y$



# Projeções de probabilidades



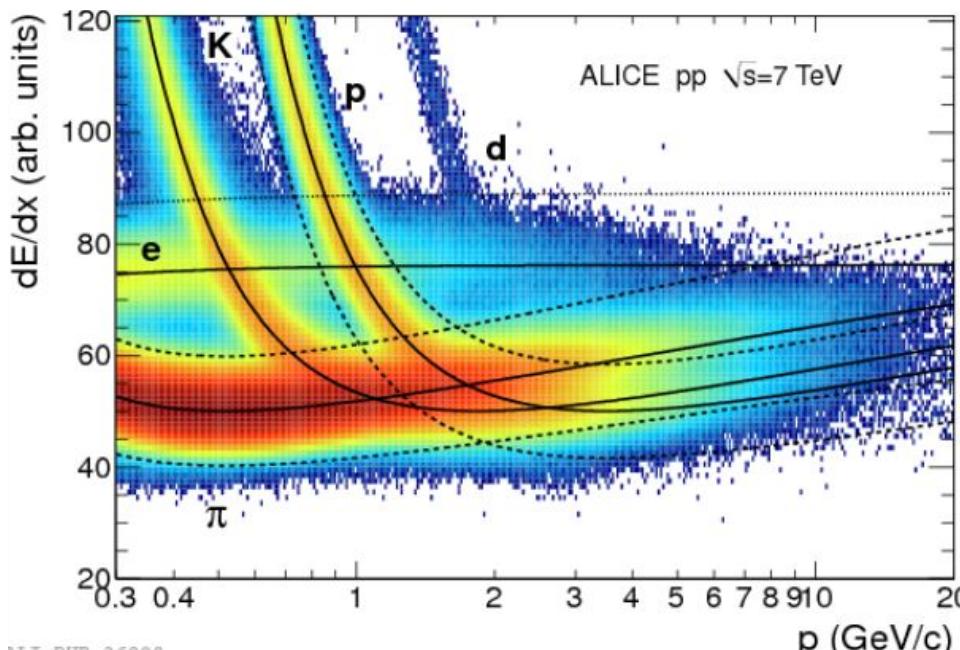
Probabilidade conjunta de  $x$  e  $y$  e as probabilidades condicionais, integrando  $y$  em duas faixas diferentes.



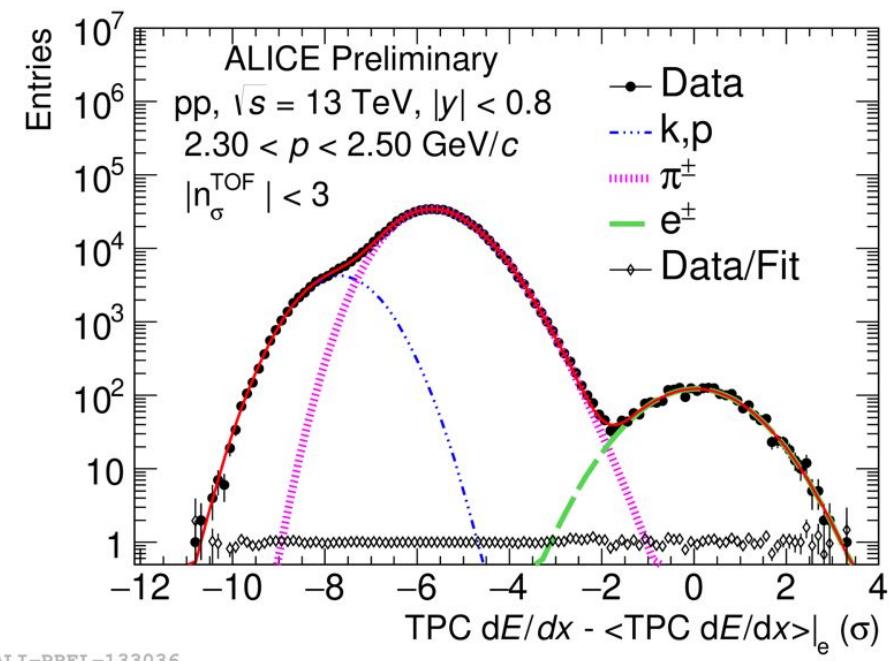
# Projeções de probabilidades em HEP

Projeções de distribuições em bins de momento

→ Projeções são cruciais para separar diferentes tipos de partículas



ALI-PUB-26808



ALI-PREL-133036

# Teorema Central do Limite

O **Teorema Central do Limite** (TCL) afirma que, para uma grande quantidade de variáveis aleatórias independentes e identicamente distribuídas, a distribuição da soma ( $X$ ) dessas variáveis tende a se aproximar de uma distribuição normal, independentemente da distribuição original das variáveis (desde que tenham momento definidos).

$$\langle X \rangle \equiv E[X] = \sum_{i=1}^n \mu_i$$

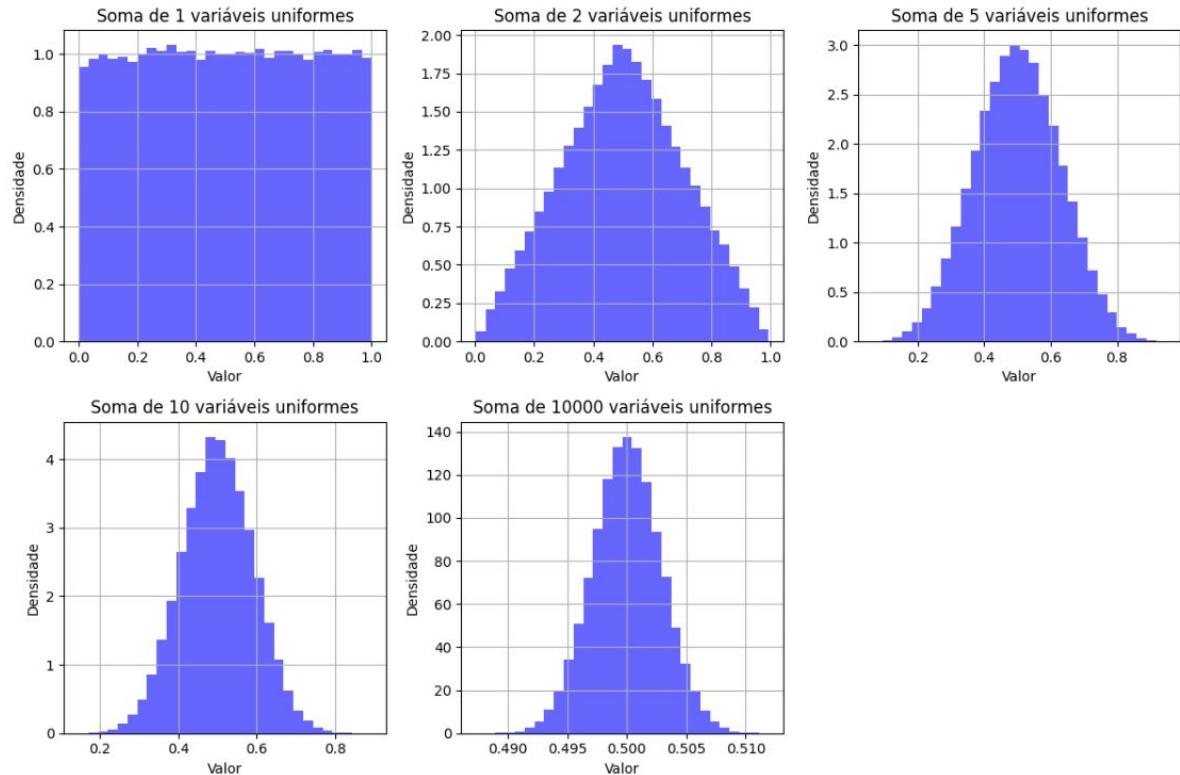
$$\langle \Delta X^2 \rangle \equiv \text{Var}[X] = \sum_{i=1}^n V_i = \sum_{i=1}^n \sigma_i^2$$

$f(X)$  becomes a Gaussian in the limit  $n \rightarrow \infty$

Sendo  $\mu_i$  e  $\sigma_i$  a média e variância das variáveis  $x_i$

# Teorema Central do Limite

Quando temos apenas 1 variável (primeiro gráfico), a distribuição é uniformemente distribuída entre 0 e 1.  
À medida que somamos mais variáveis (2, 5, 10, 10000), a distribuição das médias começa a se aproximar de uma curva normal (gaussiana), conforme previsto pelo Teorema Central do Limite.



# Teorema Central do Limite

O **Teorema Central do Limite** (TCL) encontra aplicações em praticamente todas as áreas de estudo científico.

Seja ao discutir o comportamento de sistemas complexos ou ao examinar os detalhes de medições científicas, observa-se que as variáveis medidas geralmente são influenciadas por um grande número de efeitos e processos distintos.

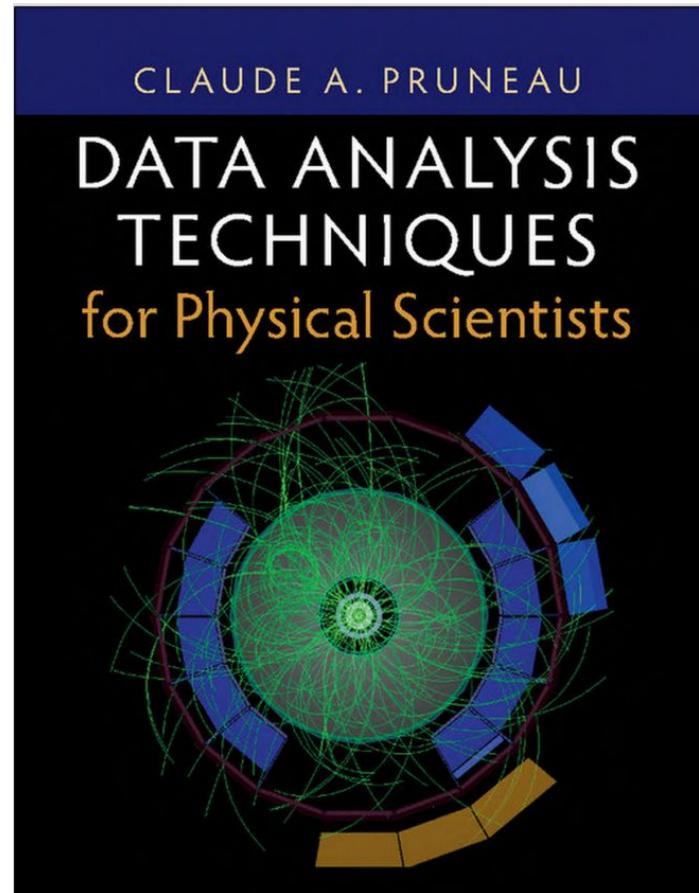
Quanto mais complexos eles são, maior é o número de efeitos e processos envolvidos. Isso implica que as flutuações e, consequentemente, os erros de medição tendem a seguir uma distribuição Gaussiana.

O TCL é muito importante e por isso é dado tanto destaque às discussões sobre distribuições Gaussianas

# Modelos de probabilidade

## Capítulo 3

- Distribuição de Bernoulli
- Distribuição Binomial
- Distribuição Binomial Negativa
- Função Lorentiziana
- Distribuição Normal
- Distribuição Gaussiana
- Distribuição de  $\chi^2$
- Distribuição de Maxwell-Boltzmann



# Distribuição de Bernoulli

Distribuição discreta que toma apenas 2 valores: 1 ou 0

$$p_{\text{Ber}}(n|\varepsilon) = \begin{cases} 1 - \varepsilon & \text{probability of failure, } n = 0, \\ \varepsilon & \text{probability of success, } n = 1. \end{cases}$$

$$p_{\text{Ber}}(n|\varepsilon) = (1 - \varepsilon)^{1-n} \varepsilon^n$$

O exemplo mais simples da distribuição de Bernoulli é o lançamento de uma única moeda, que tem uma probabilidade  $\varepsilon = 0,5$  de resultar em cara ou coroa.

O processo de detecção de partículas também pode ser considerado um "experimento de Bernoulli". Nesse caso,  $\varepsilon$  representa a probabilidade de detectar uma única partícula com um determinado detector.

# Distribuição de Bernoulli

Um exemplo do processo de detecção de partículas como um "experimento de Bernoulli" seria em um experimento de física onde estamos usando um detector de partículas para medir a passagem de um elétron.

Suponha que o detector tenha uma eficiência de 80%, ou seja, ele detecta corretamente a partícula em 80% das vezes que o elétron passa por ele. Nesse caso, a probabilidade  $\varepsilon = 0.8$  representa a chance de o detector registrar a presença do elétron.

Aqui, o experimento de Bernoulli ocorre porque, para cada elétron que passa pelo detector, há duas possíveis saídas: ou o detector registra (sucesso) ou não (fracasso), assim como no lançamento de uma moeda.

Cada detecção é um "sucesso" com probabilidade  $\varepsilon = 0.8$  ou um "fracasso" com probabilidade  $1 - \varepsilon = 0.2$

# Distribuição binomial

A distribuição binomial surge no contexto de medições, lançamentos ou amostragens idênticas e independentes, repetidas um número finito de vezes.

É uma extensão da distribuição de Bernoulli para  $n > 1$ .

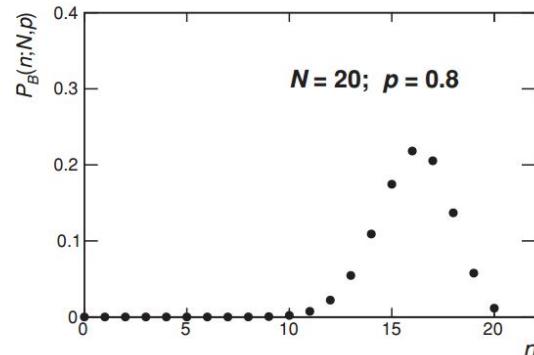
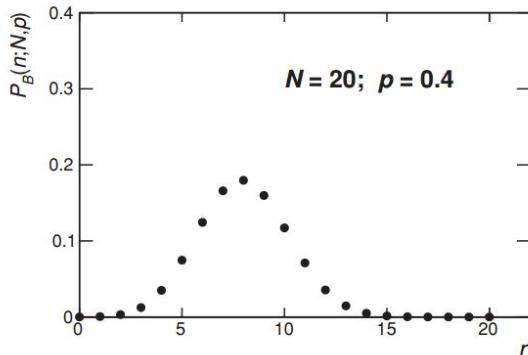
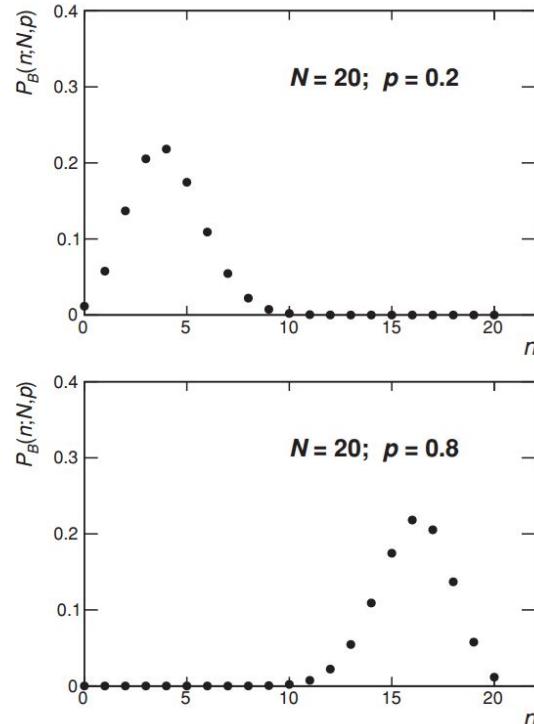
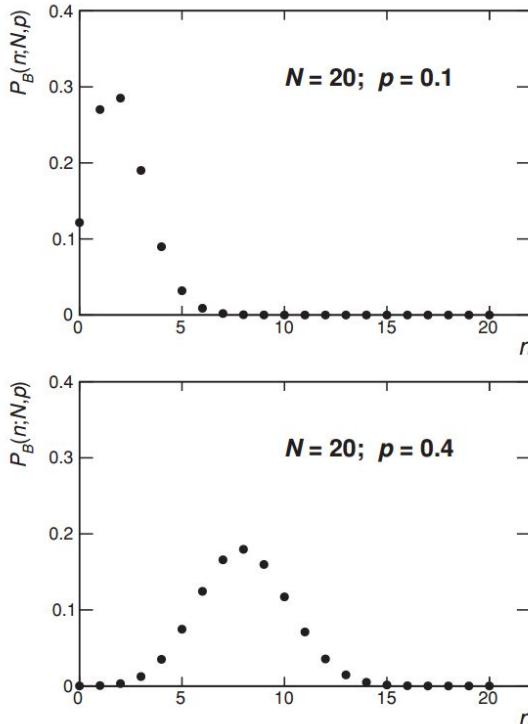
Cada medição, observação ou amostra é considerada um experimento independente, com uma probabilidade de sucesso denotada por  $\varepsilon$  ( $p = \varepsilon$ ).

Um conjunto de  $N$  medições também pode ser considerado como uma única medição, ou evento, resultando em  $n$  tentativas bem-sucedidas e  $N - n$  "falhas".

$$p_B(n|N, p) = \frac{N!}{n!(N-n)!} \varepsilon^n (1-\varepsilon)^{N-n}$$

# Distribuição binomial

$$p_B(n|N, p) = \frac{N!}{n!(N-n)!} \varepsilon^N (1-\varepsilon)^{N-n}$$



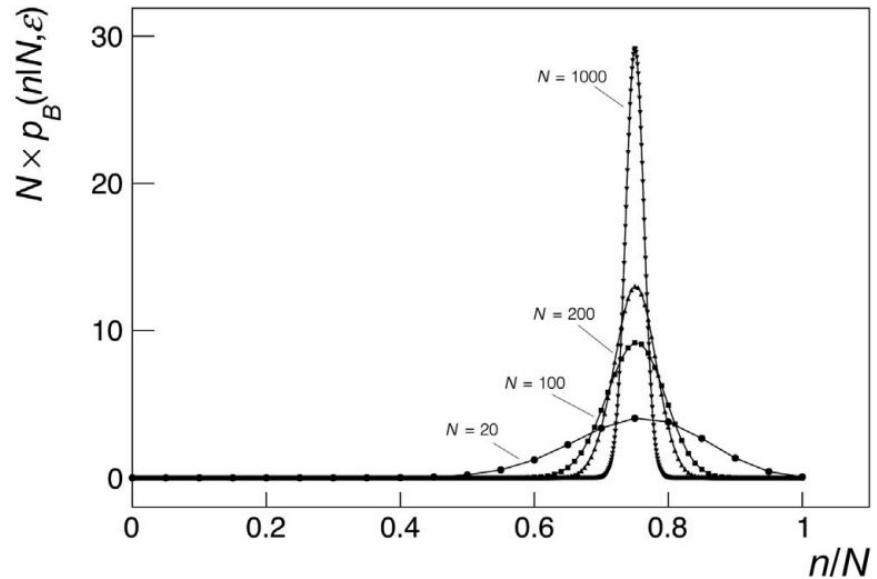
$N$  é o número de tentativas

$p$  é a probabilidade de sucesso

$n$  é o número de sucessos

Eixo y: probabilidade associada a cada número de sucessos (chance de ocorrer  $n$  sucessos em 20 tentativas)

# Distribuição binomial



Evolução da distribuição binomial ao se aumentar o número de tentativas  $N$  (para um  $\varepsilon$  fixo,  $\varepsilon = 0.75$  )

No limite  $N \rightarrow \infty$ , a distribuição se torna uma função Delta

# Distribuição Binomial Negativa

Relacionada com a distribuição binomial, mas em vez de considerar o número de sucessos dado um total de  $n$  tentativas, estuda-se o número de tentativas completadas antes de se alcançar um número prescrito de sucessos,  $k$ .

Nesse contexto, a última tentativa realizada é, por definição, um sucesso, e antes desse sucesso, deve ter havido  $k - 1$  sucessos nas  $n - 1$  tentativas anteriores.

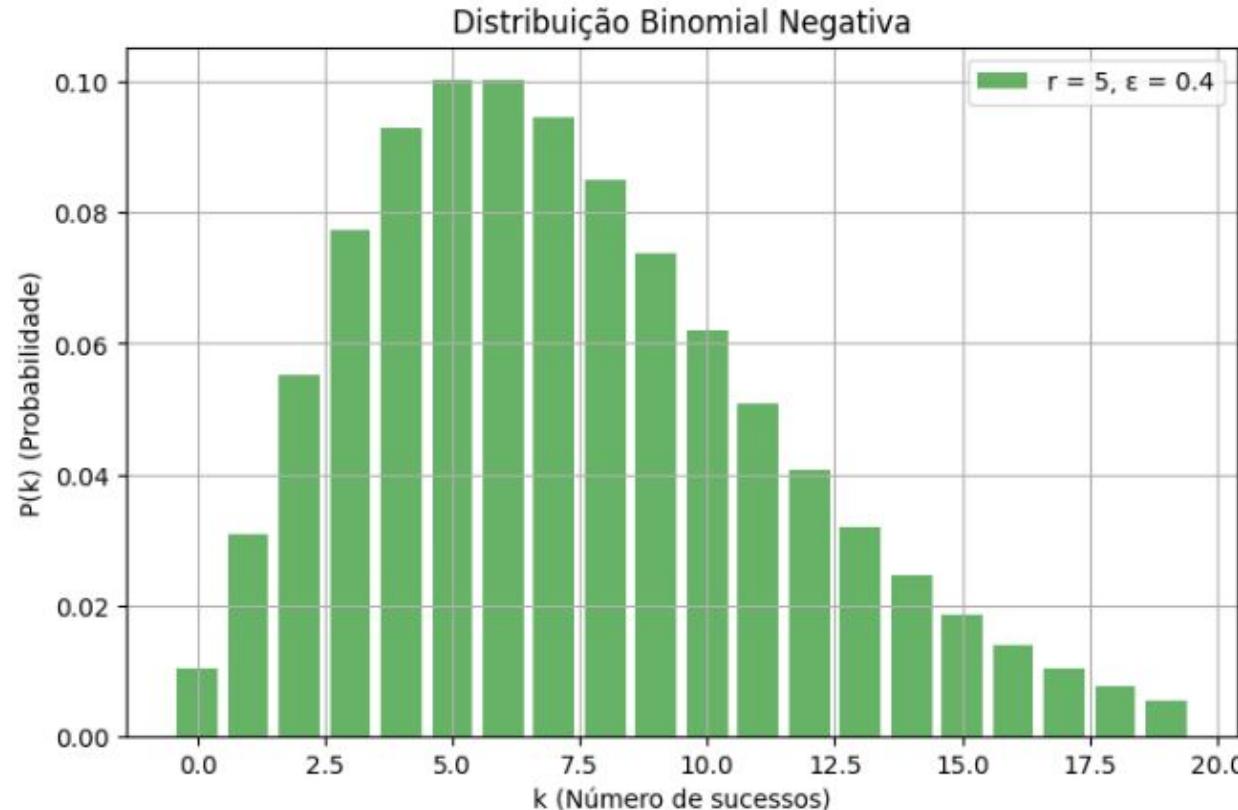
→ A probabilidade de ocorrerem  $k$  sucessos, sendo o último um sucesso, resulta na distribuição binomial negativa

$$p(n|k, \varepsilon) = \binom{n-1}{k-1} \varepsilon^k (1-\varepsilon)^{n-k}$$

$${}^N C_n = \frac{N!}{n!(N-n)!} \equiv \binom{N}{n}$$

# Distribuição Binomial Negativa

- $n$  tentativas
- $k$  sucessos
- $r$  é o número de falhas antes de  $k$  sucessos
- probabilidade de sucesso  $\epsilon$

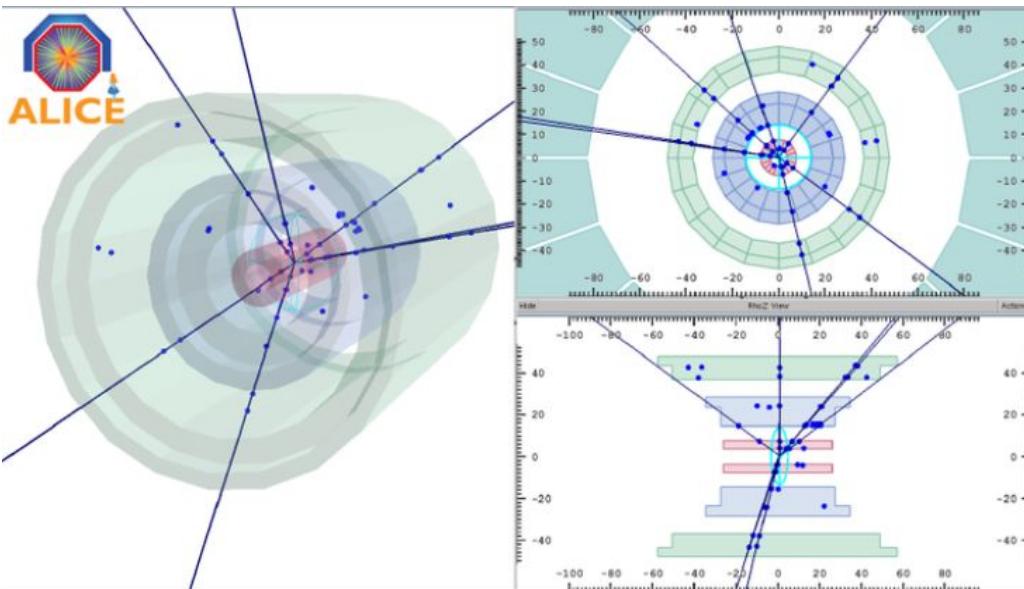


# Distribuição Binomial Negativa em HEP

Descreve distribuição de multiplicidade de partículas em HEP!

O que é “multiplicidade” em colisões de altas energias?

→ Número de partículas produzidas na colisão



Primeira tomada de dados pp do ALICE em 2009.

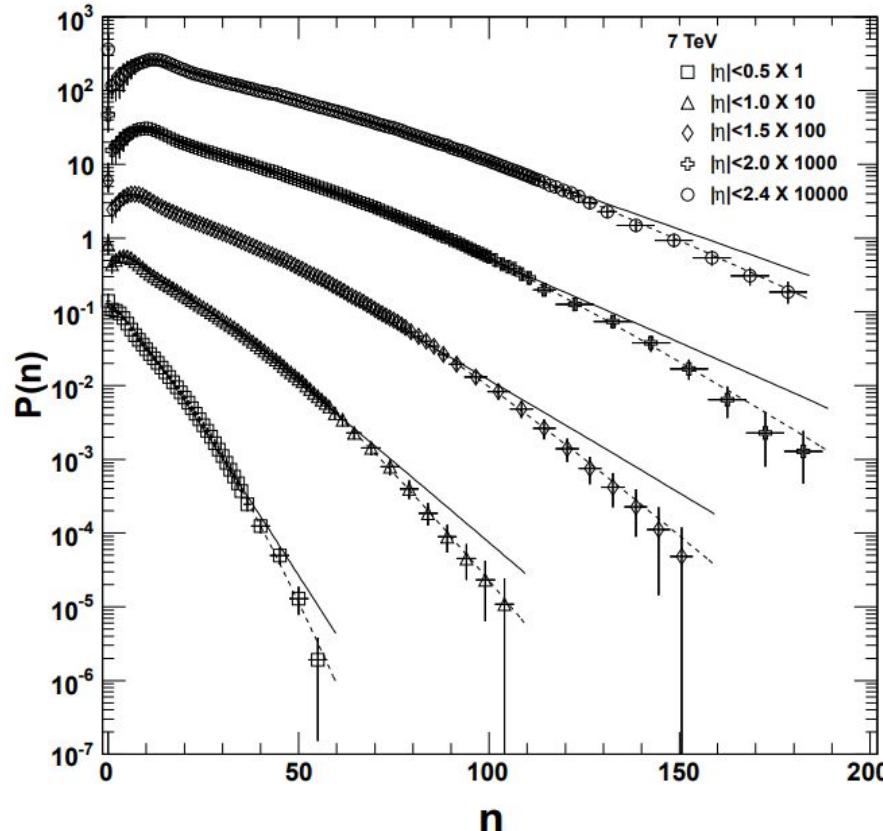
Os pontos correspondem aos impactos nos detectores de vértice de silício, e as linhas correspondem às trajetórias reconstruídas usando cortes de qualidade menos rigorosos.

A elipse desenhada no meio do detector envolve o vértice do evento reconstruído.

# Distribuição Binomial Negativa em HEP

Distribuição de **multiplicidade de hadrons carregados**.

As linhas sólidas traçadas ao longo dos pontos de dados correspondem aos respectivos ajustes de uma única **Binomial Negativa**, enquanto as linhas tracejadas correspondem aos respectivos ajustes de **duas Binomiais Negativas**.



# Distribuição Lorentziana ou de Cauchy

→ Descreve ressonâncias

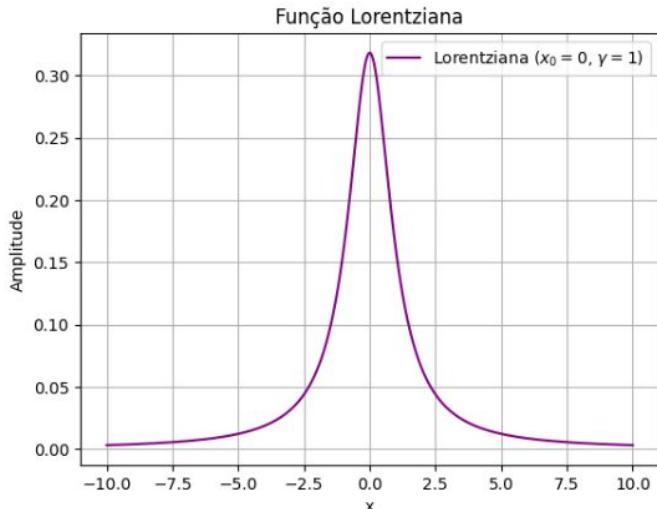
$$f(x) = \frac{1}{\pi\gamma} \left[ 1 + \left( \frac{x-x_0}{\gamma} \right)^2 \right]$$

onde:

- $x_0$  é o valor central (ou "pico") da distribuição,
- $\gamma$  é a largura à meia-altura (ou "largura de pico").

O “desvio padrão” da distribuição Lorentziana é infinito!

Use:  $\sigma^2 = \int_{-\infty}^{\infty} (x - x_0)^2 f(x) dx$



**Façam como exercício!**



# Distribuição Lorentziana ou de Cauchy

→ Descreve ressonâncias

$$f(x) = \frac{1}{\pi\gamma} \left[ 1 + \left( \frac{x-x_0}{\gamma} \right)^2 \right]$$

onde:

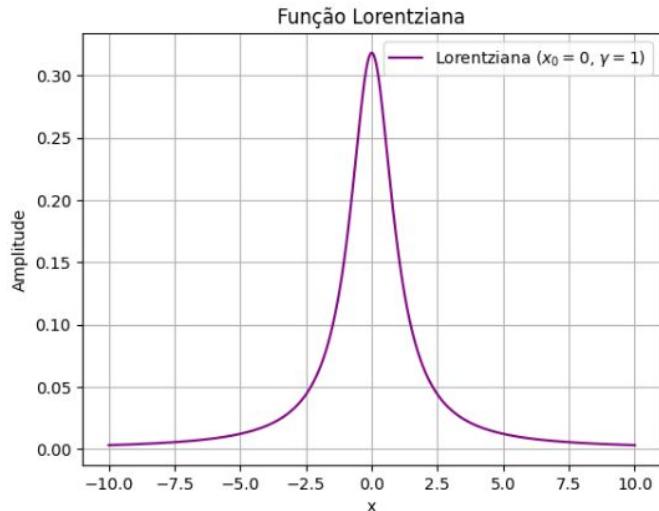
- $x_0$  é o valor central (ou "pico") da distribuição,
- $\gamma$  é a largura à meia-altura (ou "largura de pico").

O “desvio padrão” da distribuição Lorentziana é infinito!

Use:  $\sigma^2 = \int_{-\infty}^{\infty} (x - x_0)^2 f(x) dx$

**Dica: Note que o termo dominante no denominador é:**

$$\left( \frac{x-x_0}{\gamma} \right)^2$$



**Façam como exercício!**



# Resolução: desvio padrão da Função Lorentziana

$$\sigma^2 = \int_{-\infty}^{\infty} (x - x_0)^2 f(x) dx$$

$$\sigma^2 = \int_{-\infty}^{\infty} (x - x_0)^2 \cdot \frac{1}{\pi\gamma \left[ 1 + \left( \frac{x-x_0}{\gamma} \right)^2 \right]} dx$$

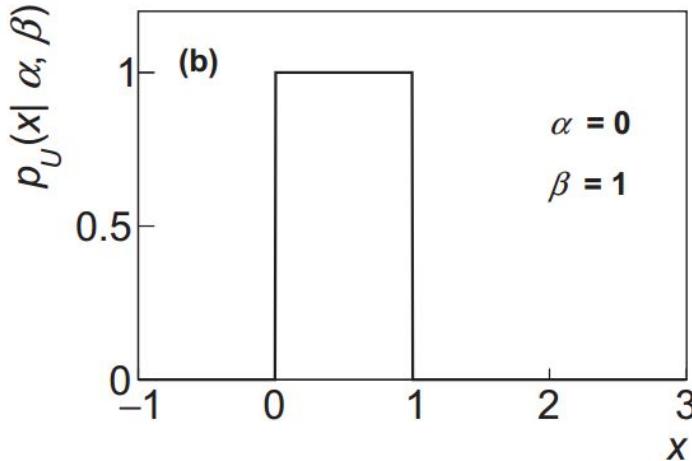
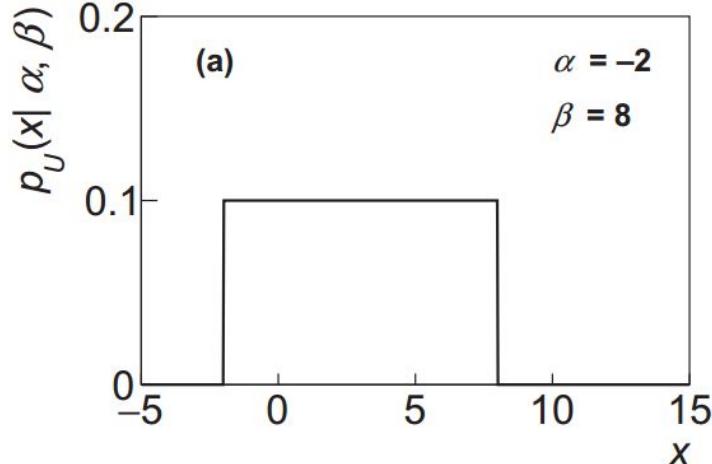
Para valores de  $|x|$  muito grandes, podemos simplificar a  $f(x)$ , pois o termo  $\left( \frac{x-x_0}{\gamma} \right)^2$  domina no denominador:

$$f(x) = \frac{1}{\pi\gamma \left[ 1 + \left( \frac{x-x_0}{\gamma} \right)^2 \right]} \quad \rightarrow \quad f(x) \approx \frac{\gamma}{\pi(x - x_0)^2}$$

$$\sigma^2 \approx \int_{-\infty}^{\infty} (x - x_0)^2 \cdot \frac{\gamma}{\pi(x - x_0)^2} dx$$

$$\sigma^2 \approx \frac{\gamma}{\pi} \int_{-\infty}^{\infty} 1 dx \quad \rightarrow \quad \boxed{\sigma^2 = \infty}$$

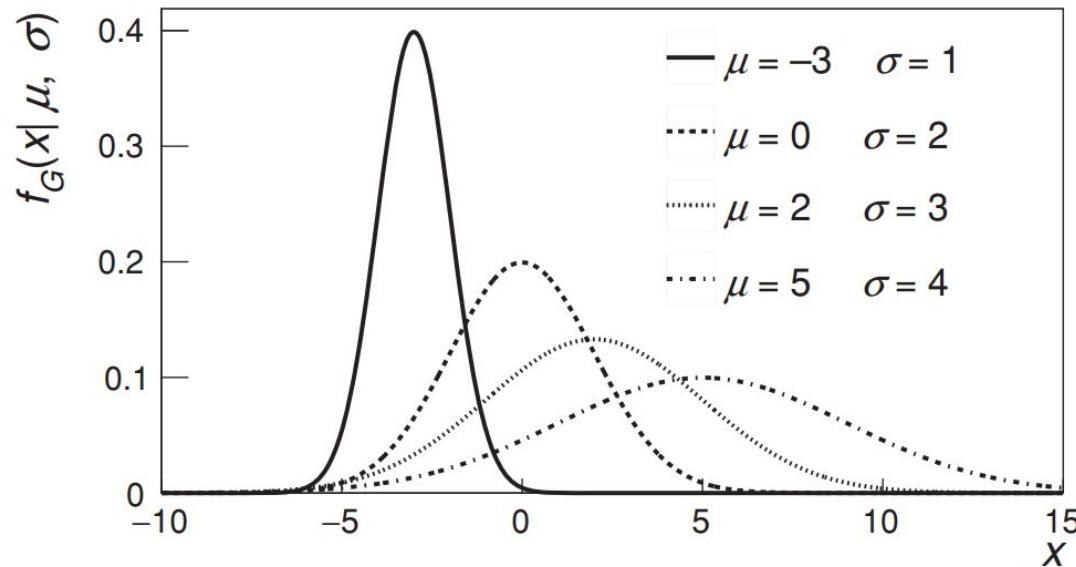
# Distribuição Uniforme



$$p_u(x|\alpha, \beta) = \begin{cases} \frac{1}{\beta-\alpha} & \alpha \leq x \leq \beta, \\ 0 & \text{otherwise.} \end{cases}$$

# Distribuição Gaussiana

Distribuição **gaussiana** ou distribuição **normal**:



→ A soma de um grande número de variáveis aleatórias independentes e identicamente distribuídas tende a seguir uma distribuição gaussiana, independentemente da distribuição original (TCL).

→ Muitos fenômenos naturais e processos de medição, quando suficientemente grandes, podem ser modelados como distribuições gaussianas.

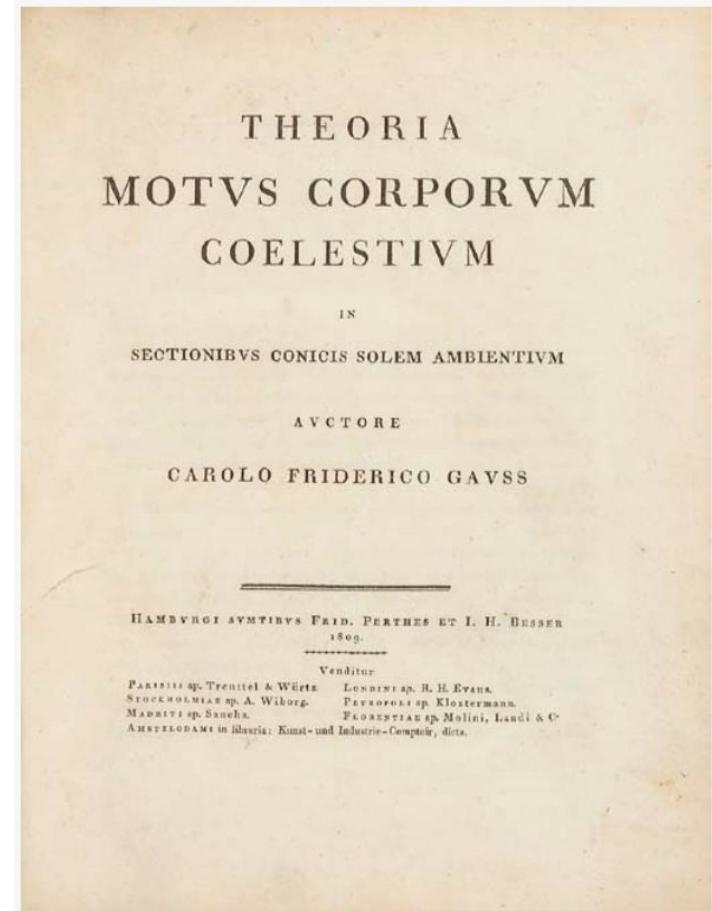
# Distribuição Gaussiana: curiosidades

Recebe o nome de **Gaussiana** por causa do **Carl Friedrich Gauss** (cientista e matemático alemão, 1777-1855).

→ Primeiro uso extensivo no **contexto científico** é atribuído a **Gauss**, que em 1809 usou esta distribuição para analisar dados de astronomia.

(Mas já havia sido “descoberta” por **De Moivre** como uma aproximação da distribuição binomial e usada por **Laplace** em 1783 para estudos de incertezas).

O estudo astronômico pelo qual Gauss é mais conhecido e onde ele utilizou a distribuição gaussiana pela primeira vez para explicar dados foi o **cálculo da órbita do asteroide Ceres**.



# Distribuição de $\chi^2$

Para  $n$  pontos  $(x_i, y_i)$  e uma função  $f$  com  $m$  parâmetros  $(\theta_1, \theta_2, \dots, \theta_m)$ , podemos avaliar a qualidade de um ajuste pela função  $\chi^2$ , definida como

$$\chi^2 = \sum_{i=1}^n \frac{[y_i - f(x_i | \vec{\theta})]^2}{\sigma_i^2}$$

Onde  $\sigma_i^2$  é o erro de  $y_i$

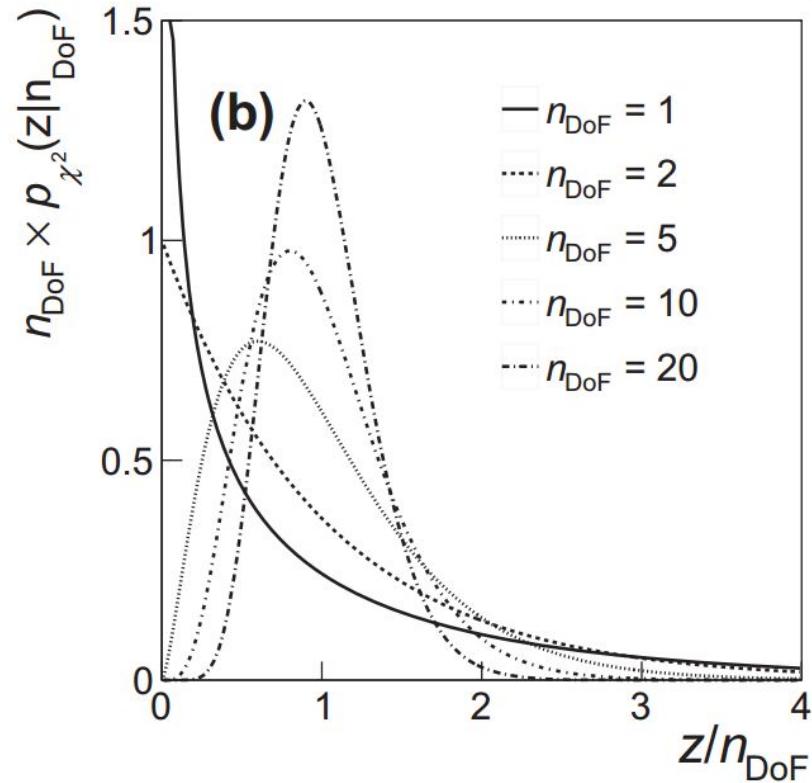
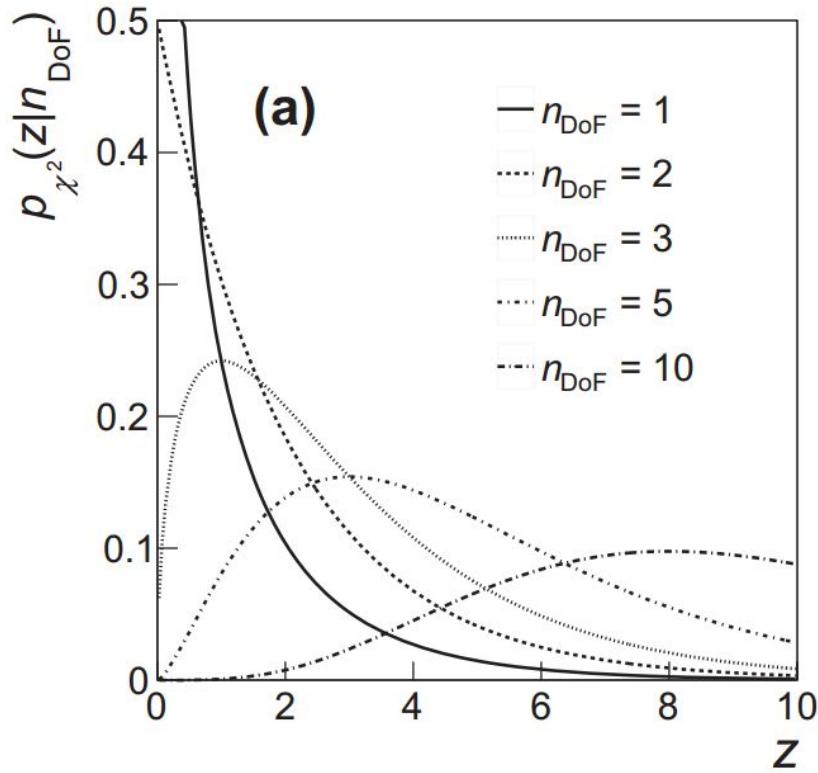
Para medidas governadas pela estatística Gaussiana (medidas repetidas de x resulta em y distribuído de forma gaussiana), a PDF dos valores de  $\chi^2$  resulta na **distribuição de  $\chi^2$** :

$$p_{\chi^2}(z|n) \equiv \frac{1}{2^{n/2}\Gamma(n/2)} z^{n/2-1} e^{-z/2} \quad \text{for } n = 1, 2, \dots$$

Sendo  $n$  o número de graus de liberdade ( $n - m$ )

Depende da função gamma

# Distribuição de $\chi^2$



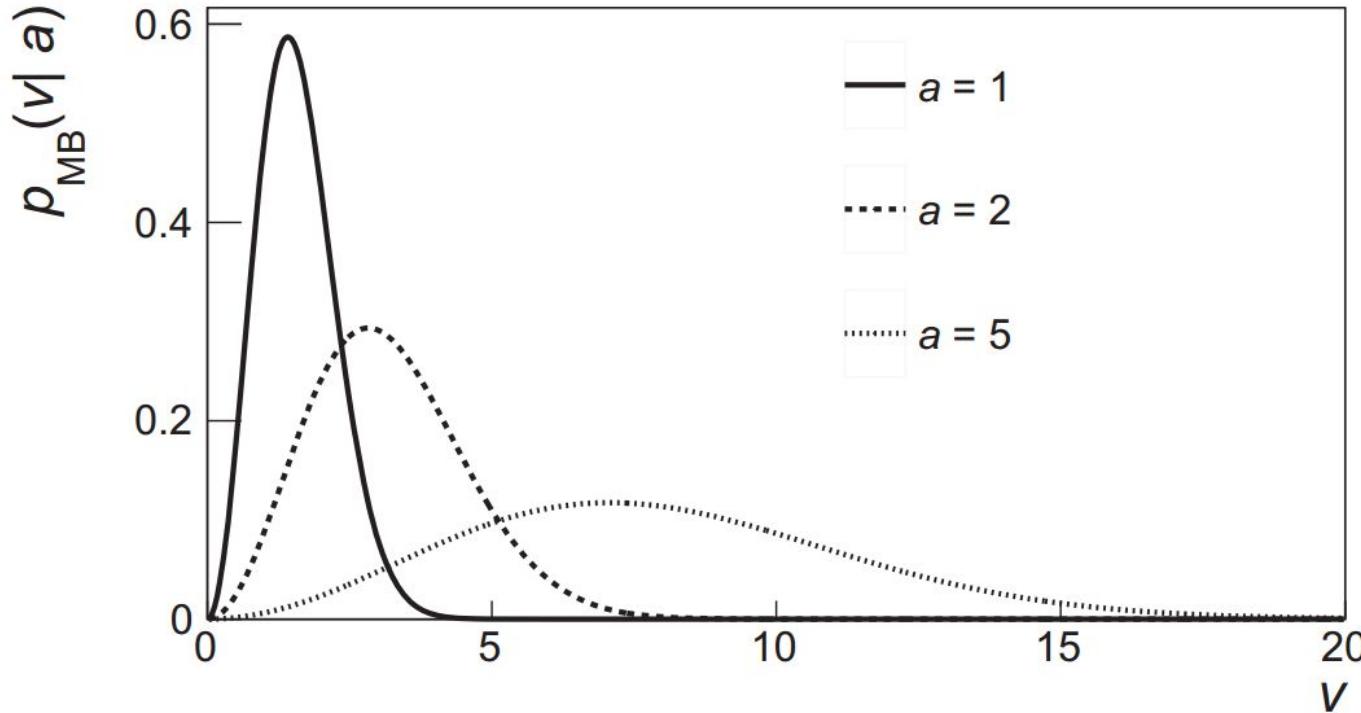
# Distribuição de Maxwell-Boltzmann

A [distribuição de Maxwell–Boltzmann](#), frequentemente chamada de distribuição de velocidade de Maxwell, é uma [densidade de probabilidade que descreve a distribuição de velocidade das moléculas de um gás ideal](#) em equilíbrio termodinâmico a uma temperatura específica T.

$$p_{\text{MB}}(v) = 4\pi \sqrt{\left(\frac{m}{2\pi kT}\right)^3} v^2 \exp\left(-\frac{mv^2}{2kT}\right)$$

m é a massa das moléculas/partículas, v é a velocidade, T a temperatura do gás, e k a constante de Boltzmann.

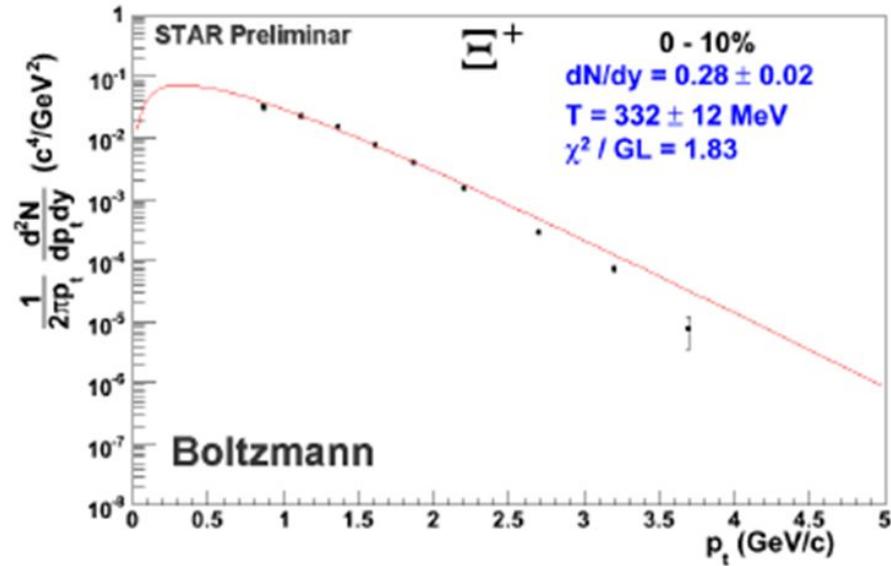
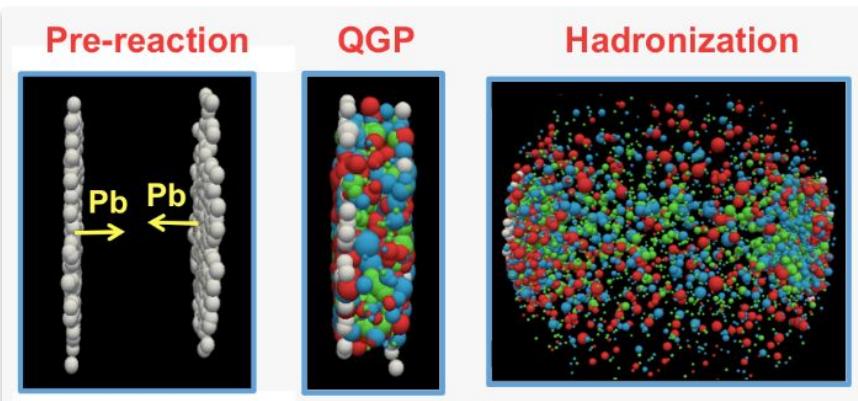
# Distribuição de Maxwell-Boltzmann



$$a = \sqrt{kT/m}.$$

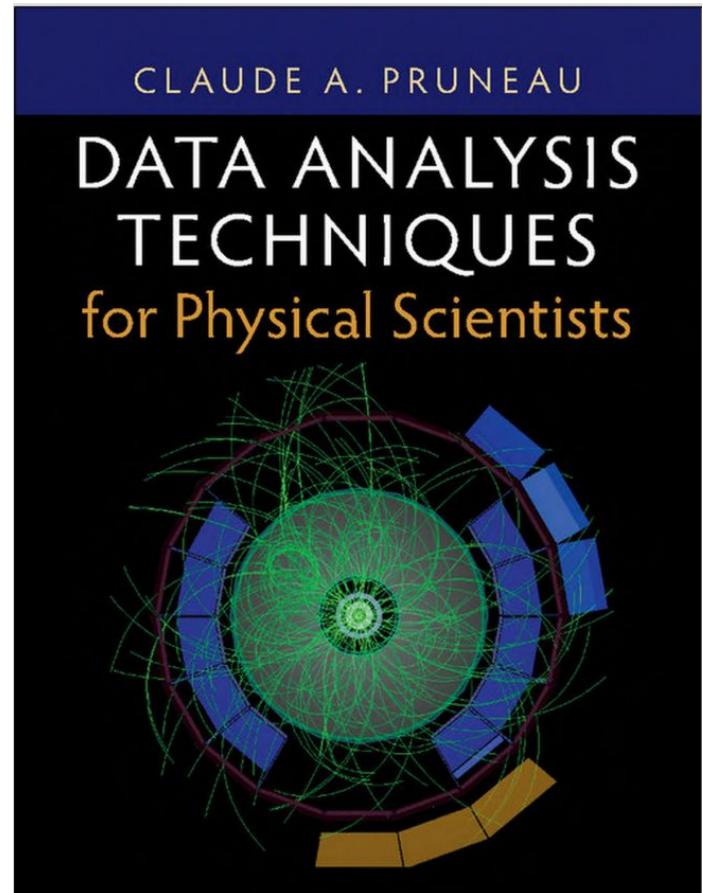
# Distribuição de Maxwell-Boltzmann

Ela também pode ser usada para descrever a distribuição de velocidade ou energia de um gás de hadrons (ou partons) idealizado, produzido em colisões de íons pesados, desde que esses possam ser considerados em equilíbrio a uma temperatura específica.



# Histogramas

## Capítulo 4



# Histogramas

Experimentalmente, [várias restrições limitam o número de vezes](#) que uma medição de um determinado observável X pode ser realizada.

No entanto, é desejável e interessante [gerar representações visuais simples dos dados adquiridos](#).

Sejam amostras de 10, 1.000, 1.000.000 ou vários bilhões de pontos de dados, é útil visualizar esses dados em um gráfico.

# Histogramas

Experimentalmente, **várias restrições limitam o número de vezes** que uma medição de um determinado observável X pode ser realizada.

No entanto, é desejável e interessante **gerar representações visuais simples dos dados adquiridos**.

Sejam amostras de 10, 1.000, 1.000.000 ou vários bilhões de pontos de dados, é útil visualizar esses dados em um gráfico.

→ Pode ser que o **observável X de interesse tenha um valor definido**, ou seja, espera-se que assuma um único valor.

- **incertezas nas medições podem levar a desvios do valor real (isto é, verdadeiro) do observável. Como um número finito de valores medidos  $\{x_i\}$ ,  $i = 1, \dots, n$ , estão distribuídos?**

# Histogramas

Experimentalmente, **várias restrições limitam o número de vezes** que uma medição de um determinado observável  $X$  pode ser realizada.

No entanto, é desejável e interessante **gerar representações visuais simples** dos dados adquiridos.

Sejam amostras de 10, 1.000, 1.000.000 ou vários bilhões de pontos de dados, é útil visualizar esses dados em um gráfico.

→ Pode ser que o **observável  $X$  de interesse tenha um valor definido**, ou seja, espera-se que assuma um único valor.

- **incertezas nas medições podem levar a desvios do valor real (isto é, verdadeiro) do observável. Como um número finito de valores medidos  $\{x_i\}$ ,  $i = 1, \dots, n$ , estão distribuídos?**

→ Os **valores de um observável contínuo  $X$**  podem ser esperados para **seguir algum modelo ou distribuição de probabilidade**. Nesse caso, é interessante comparar a densidade dos valores medidos em relação a  $x$  com o modelo ou a distribuição teórica.

→ **Plotar um gráfico com número de ocorrências ou “eventos” em função do valor  $x$  de interesse é crucial.**

# Histogramas

Na **física de partículas** é comum medir o **espectro de momento** e as **distribuições angulares** das partículas produzidas em colisões.

Embora seja **possível modelar**, pelo menos em princípio, **as distribuições gerais de momento e angulares**, não é possível prever, em uma base colisão por colisão e partícula por partícula, quais momentos e ângulos as partículas terão.

# Histogramas

Na **física de partículas** é comum medir o **espectro de momento** e as **distribuições angulares** das partículas produzidas em colisões.

Embora seja **possível modelar**, pelo menos em princípio, **as distribuições gerais de momento e angulares**, não é possível prever, em uma base colisão por colisão e partícula por partícula, quais momentos e ângulos as partículas terão.

**Os momentos e ângulos são, consequentemente, variáveis aleatórias.**

→ um número muito grande (infinito) de medições de momentos ou ângulos de produção é esperado para **gerar distribuições** dessas variáveis que sejam representativas de suas respectivas funções de densidade de probabilidade (PDFs).

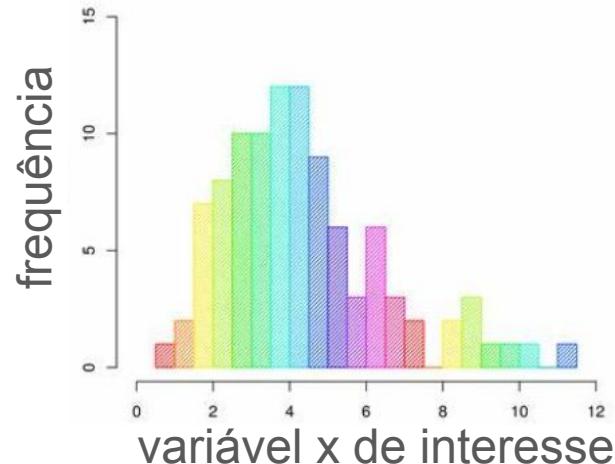
Na prática é **impossível** realizar um **número infinito** de medições.

# Histogramas

Mesmo que muitas ocorrências sejam medidas para uma variável contínua X, é de fato impossível medir a probabilidade de **um valor específico**.

No entanto, é conceitualmente e praticamente significativo determinar o número de ocorrências em um **intervalo finito de largura  $\Delta x$** .

Em princípio, realizamos medições com um valor de  $\Delta x$  tão pequeno quanto possível.

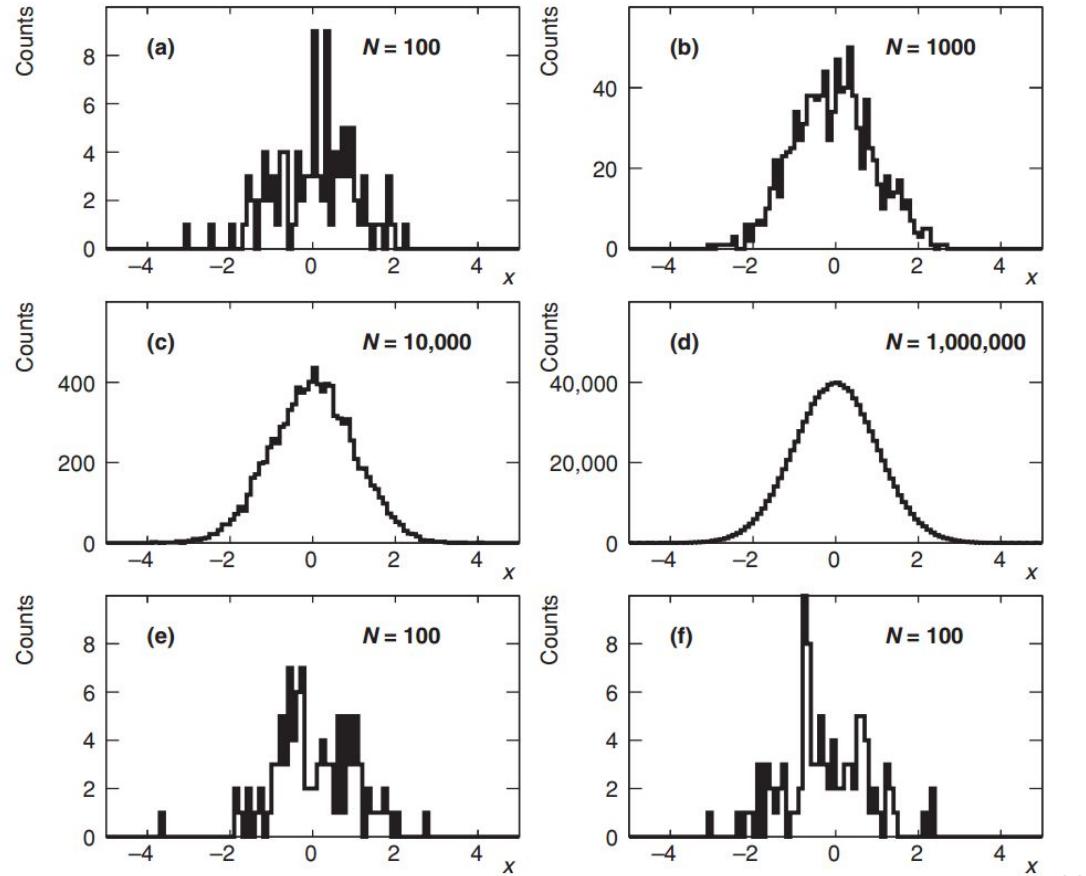


# Histogramas em uma dimensão (1D)

Histogramas obtidos com um gerador de MC (Gaussian com média 0 e sigma 1).

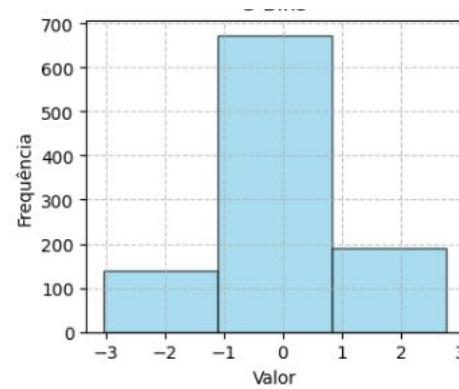
Comparando **a**, **b**, **c** e **d**, temos um aumento de dados gerados e a forma dos dados se aproxima de uma gaussiana.

Para apenas 100 pontos (casos **a**, **e** e **f**): devido a pouca estatística, a distribuição não segue a gaussiana.



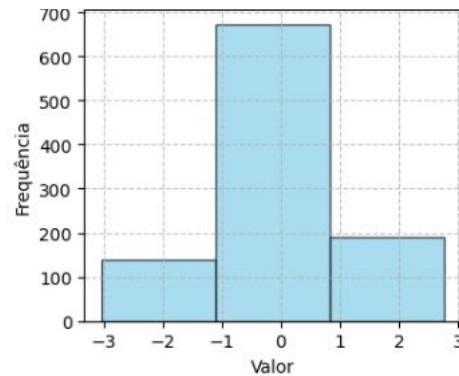
# Histogramas 1D: tamanho do intervalo (bin)

Que tipo de distribuição é esta?

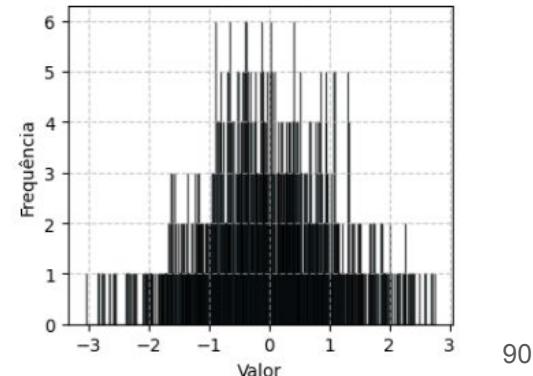


# Histogramas 1D: tamanho do intervalo (bin)

Que tipo de distribuição é esta?



E esta?

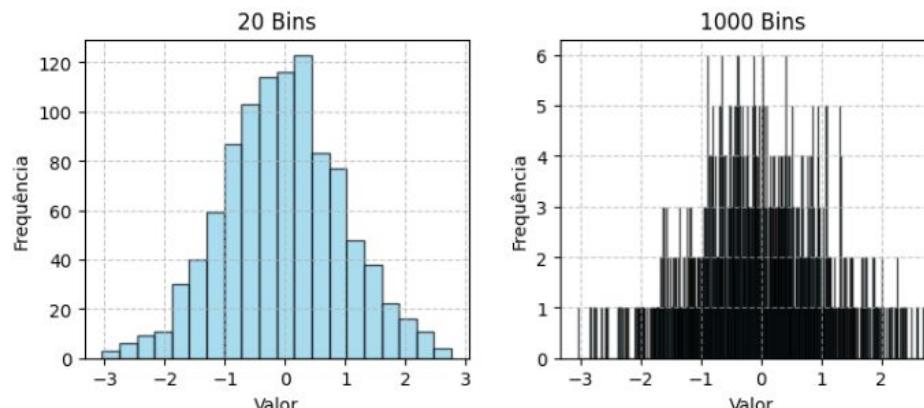
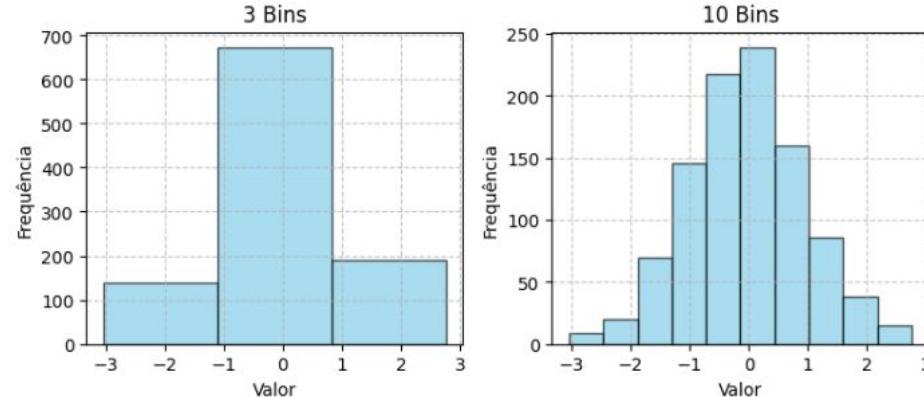


# Histogramas 1D: tamanho do intervalo (bin)

Os histogramas ao lado representam exatamente os mesmos dados.

Poucos bins: a forma da distribuição é menos precisa, pois o agrupamento exacerbado esconde detalhes da forma gaussiana.

Muitos bins: os detalhes são evidenciados, mas o **histograma fica mais ruidoso** devido à menor quantidade de dados em cada bin.



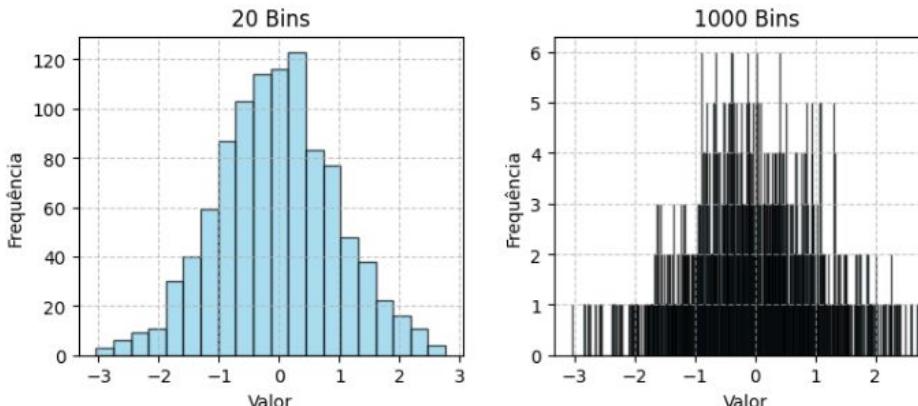
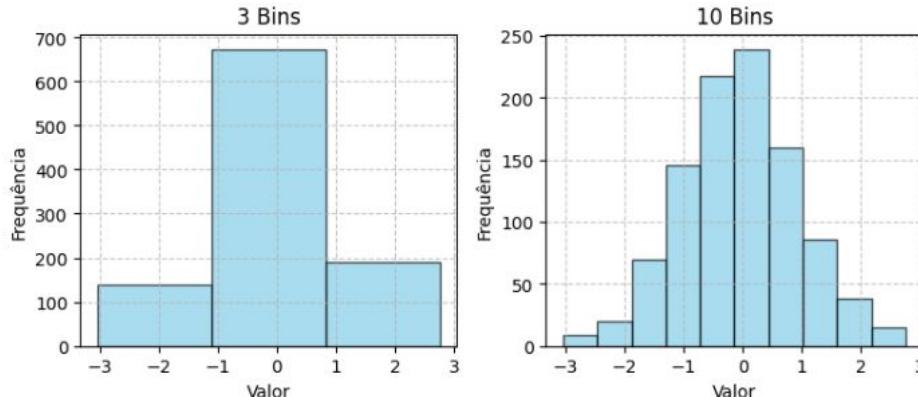
# Histogramas 1D: tamanho do intervalo (bin)

A escolha do número de bins afeta a clareza e precisão da interpretação do histograma

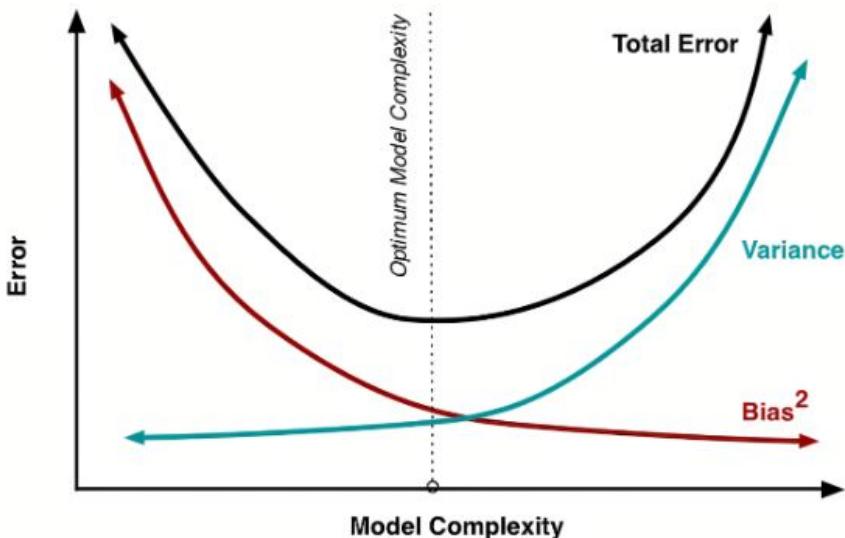
→ Bias variance tradeoff  
(compensação entre viés e variância)

Bin grande = muito bias  
Bin pequeno = muita variância no histograma

Gaussiana



# Bias variance tradeoff



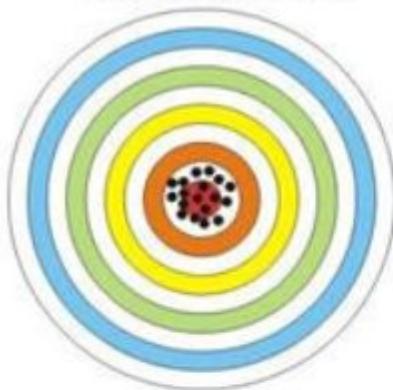
Bias variance tradeoff  
(compensação entre viés e variância)

O viés se refere à tendência de um modelo simplificar demais os dados, enquanto a variância reflete a sensibilidade do modelo a pequenas flutuações nos dados.

Encontrar o **equilíbrio ideal entre viés e variância** é crucial para alcançar um modelo com bom desempenho em dados novos.

# Acurácia e precisão

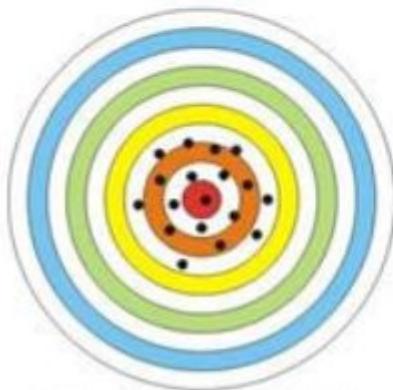
Exato e Preciso



Inexato e Preciso



Exato e Impreciso



Inexato e Impreciso

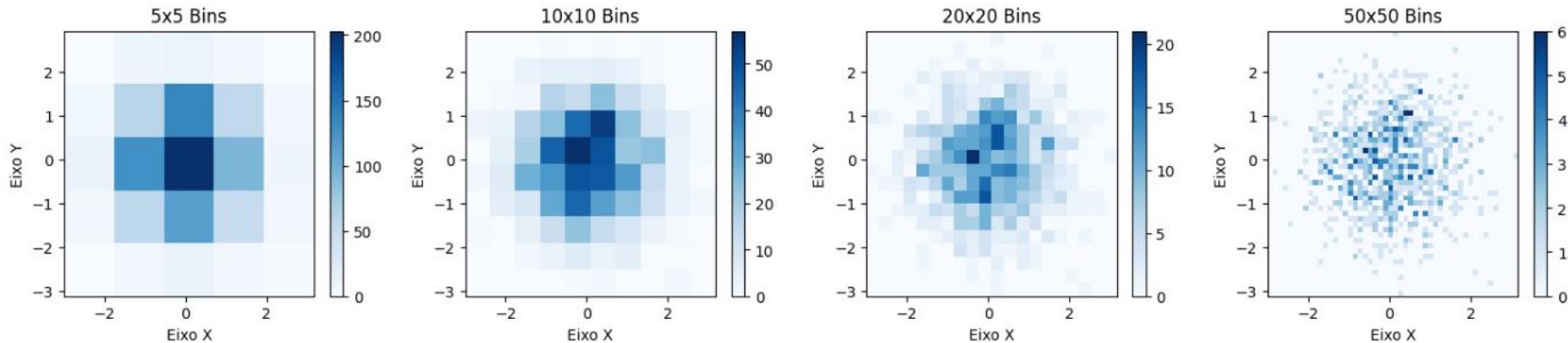
**Alta acurácia e alta precisão** (exato e preciso): Resultados [consistentes entre si](#) e próximos do valor real.

**Alta precisão, baixa acurácia** (inexato e preciso): Resultados [consistentes entre si](#), mas longe do valor real.

**Alta acurácia, baixa precisão** (exato e impreciso): Resultados próximos do valor real, mas variando entre si.

**Baixa acurácia, baixa precisão** (inexato e impreciso): resultados [variam entre si](#) e estão longe do valor real.

# Histogramas Multidimensionais (2D)



- Com **menos bins** (5x5 e 10x10), os **detalhes da distribuição são menos visíveis e mais agregados**.
- **Com um número intermediário de bins (20x20), a forma da distribuição fica mais definida.**
- Com **muitos bins** (50x50), detalhes mais finos aparecem, mas a **representação pode ficar ruidosa** devido à menor **densidade de pontos em cada célula**.

A escolha do número de bins influencia a clareza e a resolução da visualização dos dados em histogramas 2D

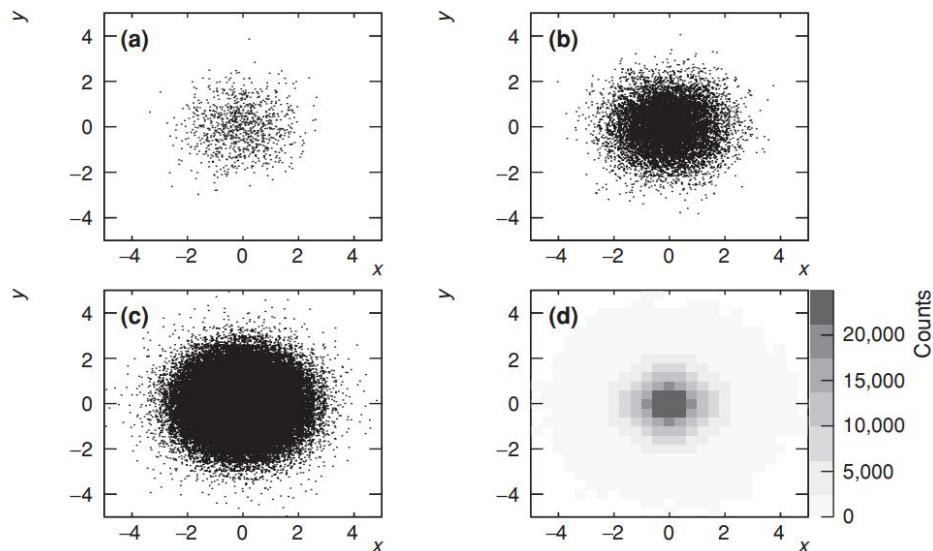
# Histogramas Multidimensionais (2D)

Tipos de gráfico 2D:

Gráfico de pontos: a, b e c

(perde informação quando vários pontos estão sobrepostos)

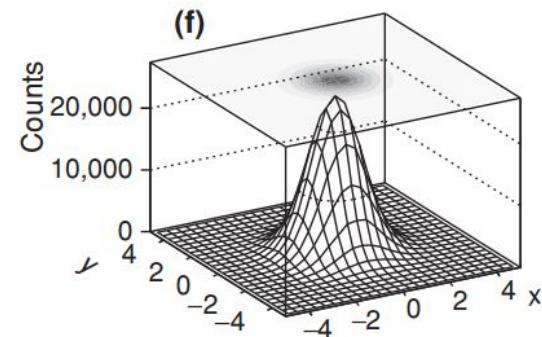
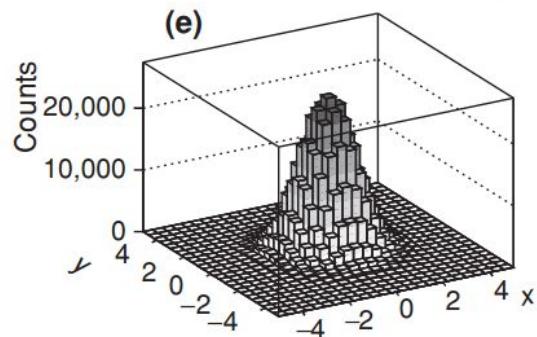
Gráfico de contorno mostrando as contagens em cada bin por uma escala de cor.



# Histogramas Multidimensionais (2D)

Tipos de gráfico 2D:

Gráfico tipo “lego” e “híbrido”

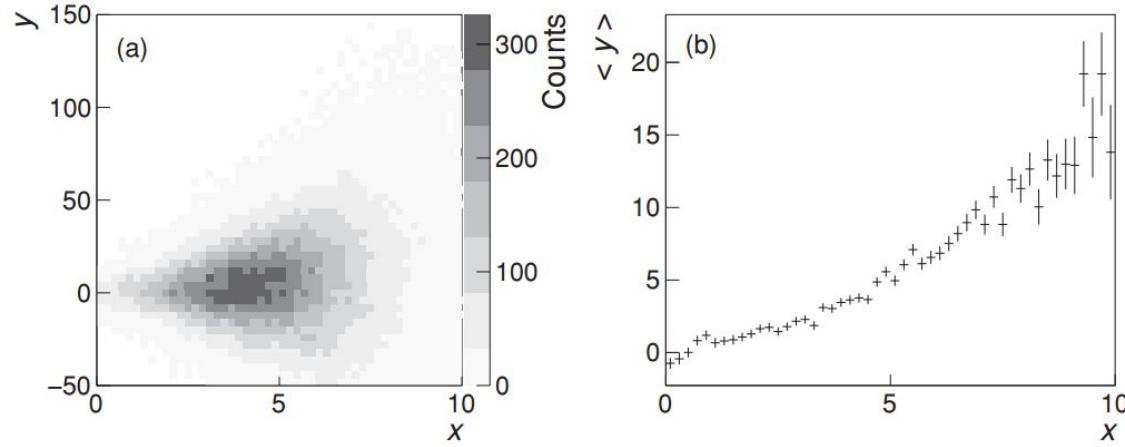


# Histogramas: valor médio

A dependência de  $y$  em  $x$   
pode ser difícil de se  
visualizada em um histograma  
2D.

Solução:

Histograma tipo “**profile**”:  
→ média de  $y$  em função de  
 $x$ .



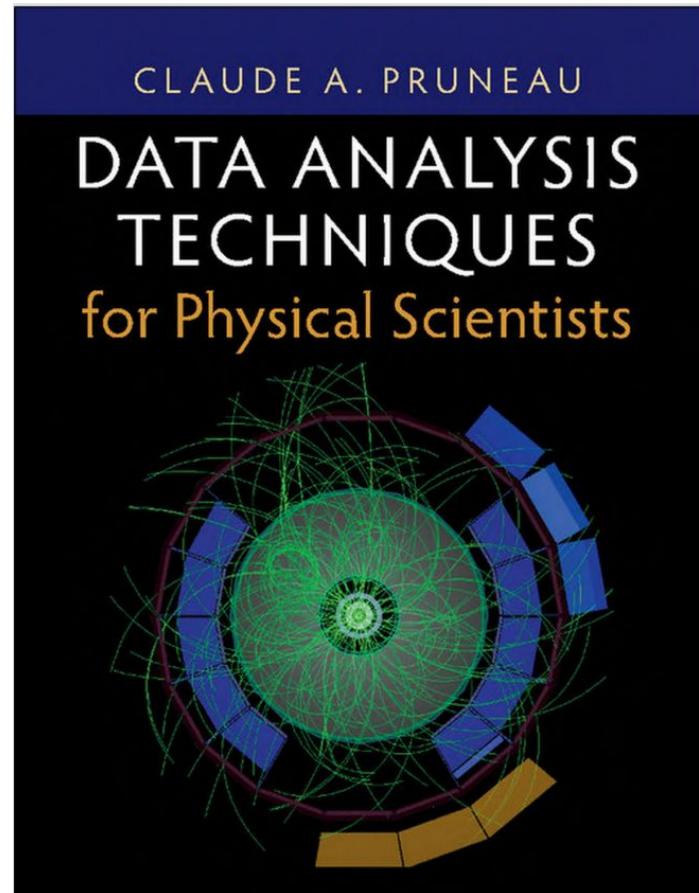
# Otimizações

## Capítulo 5

- Método da Máxima Verossimilhança
- Método dos Mínimos Quadrados
- Teste de qualidade de ajustes

## Outras fontes

- Critério de Akaike



# Otimizações

Momentos de uma PDF, embora úteis, geralmente não caracterizam completamente a distribuição de probabilidade de um fenômeno.

Muitas vezes, conhece-se a forma funcional de uma distribuição, mas os parâmetros específicos que a definem permanecem desconhecidos, como em uma distribuição log-normal com parâmetros desconhecidos  $\mu$  e  $\sigma$ .

# Otimizações

Para determinar esses parâmetros e obter o melhor ajuste aos dados, utilizam-se métodos como **máxima verossimilhança, mínimos quadrados e filtros de Kalman**.

Os métodos de máxima verossimilhança e mínimos quadrados são problemas de otimização: ambos envolvem a otimização de uma função-alvo, frequentemente chamada de função objetivo ou função mérito.

O primeiro envolve a maximização de uma função de verossimilhança, enquanto o segundo requer a minimização de uma função  $\chi^2$ .

→ **Ambas as técnicas envolvem uma busca no espaço de parâmetros (do modelo) por um valor ótimo, ou seja, um extremo de uma função objetivo.**

# Método da Máxima Verossimilhança

Uma função de verossimilhança mede a probabilidade de observar um conjunto específico de dados, dado um conjunto de parâmetros ( $\theta$ ) de um modelo probabilístico. Em outras palavras, ela expressa a plausibilidade de diferentes valores dos parâmetros para explicar os dados observados ( $x_1, x_2, \dots$ ).

$$p(x_1|\vec{\theta}) dx_1 \times p(x_2|\vec{\theta}) dx_2 \times \cdots \times p(x_n|\vec{\theta}) dx_n = \prod_{i=1}^n p(x_i|\vec{\theta}) dx_i$$

# Método da Máxima Verossimilhança

A função de verossimilhança é usada no **método de máxima verossimilhança** para encontrar os parâmetros que **maximizam essa função**, identificando os valores que tornam os dados observados os mais prováveis possíveis sob o modelo

$$L(\vec{x}|\vec{\theta}) \equiv \prod_{i=1}^n f(x_i|\vec{\theta}) \quad \frac{\partial L(\vec{\theta})}{\partial \theta_i} = 0 \quad \text{for } i = 1, \dots, m,$$

Busca-se os valores de  $\theta$  que maximizam a verossimilhança dados os pontos de dados; em outras palavras, procura-se os valores dos **parâmetros  $\theta$**  para que os pontos de dados medidos sejam os mais prováveis.

# Método da Máxima Verossimilhança: aplicação I

Aplicando em uma distribuição de Poisson:

Supernovas são fenômenos relativamente raros que ocorrem aleatoriamente nesta e em outras galáxias. Com um telescópio de grande abertura e uma câmera de alta resolução e eficiência, pode-se caracterizar a taxa de explosões de supernovas conforme o tipo de galáxia onde ocorrem. Por serem raras, o número de observações  $n$  por período de tempo (por exemplo, por noite) pode ser modelado por uma distribuição de Poisson.

$$p(n|\mu) = \frac{e^{-\mu} \mu^n}{n!}$$

onde  $\mu$  representa a taxa média de explosões (por noite).

**Façam como exercício!**

Para simplificar, vamos supor que é possível observar a mesma região do céu e pela mesma duração exata durante  **$N = 100$  noites**, com condições de observação iguais. **O número de observações feitas a cada noite é denotado por  $n_i$ ,  $i = 1, 2, \dots, 100$ .**



# Método da Máxima Verossimilhança: aplicação I

$$\begin{aligned} L(\mu) &= \prod_{i=1}^N \frac{e^{-\mu} \mu^{n_i}}{n_i!}, \\ &= \left( \prod_{i=1}^N n_i! \right)^{-1} \exp(-N\mu) \mu^{(\sum_{i=1}^N n_i)} \end{aligned}$$

Dados os fatores exponenciais, podemos maximizar o logaritmo da verossimilhança:

$$\ln L(\mu) = -N\mu + \left( \sum_{i=1}^N n_i \right) \ln \mu - \sum_{i=1}^N \ln n_i$$

# Método da Máxima Verossimilhança: aplicação I

$$\ln L(\mu) = -N\mu + \left( \sum_{i=1}^N n_i \right) \ln \mu - \sum_{i=1}^N \ln n_i$$

$$0 = \frac{\partial \ln L}{\partial \mu} = -N + \frac{\sum_{i=1}^N n_i}{\mu}$$

$$\hat{\mu} = \frac{1}{N} \sum_{i=1}^N n_i$$

# Método da Máxima Verossimilhança: aplicação II

Determine os estimadores da média e da variância de uma função de densidade de probabilidade Gaussiana.

Suponhamos que  $n$  valores medidos  $x_i$  estejam distribuídos de acordo com uma função de densidade de probabilidade Gaussiana com média  $\mu$  e desvio padrão  $\sigma$  desconhecidos.

$$\begin{aligned}\ln L(\mu, \sigma^2) &= \ln \left( \prod_{i=1}^n p(x_i; \mu, \sigma^2) \right), \\ &= \ln \left( \prod_{i=1}^n \frac{1}{\sqrt{2\pi}\sigma} \exp \left( -\frac{(x_i - \mu)^2}{2\sigma^2} \right) \right)\end{aligned}$$

**Façam como exercício!**



# Método da Máxima Verossimilhança: aplicação II

$$\begin{aligned}\ln L(\mu, \sigma^2) &= \sum_{i=1}^n \left( -\ln \sqrt{2\pi} - \frac{1}{2} \ln \sigma^2 - \frac{(x_i - \mu)^2}{2\sigma^2} \right), \\ &= -n \ln \sqrt{2\pi} - \frac{n}{2} \ln \sigma^2 - \sum_{i=1}^n \frac{(x_i - \mu)^2}{2\sigma^2}.\end{aligned}$$

Derivando em **relação a média** e igualando a zero:

$$0 = \frac{\partial \ln L(\mu, \sigma^2)}{\partial \mu} = - \sum_{i=1}^n \frac{\partial}{\partial \mu} \left( \frac{(x_i - \mu)^2}{2\sigma^2} \right) \quad \hat{\mu} = \frac{1}{n} \sum_{i=1}^n x_i$$

# Método da Máxima Verossimilhança: aplicação II

Derivando em [relação a  \$\sigma^2\$](#)

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

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

# Método da Máxima Verossimilhança em HEP

$$\ln\{L\} = \sum_{i=1}^N \ln\{F(x^i, m^i)\}$$

$$F(x, m) = f_{\text{sig}} \cdot F_{\text{sig}}(x) \cdot M_{\text{sig}}(m) + (1 - f_{\text{sig}}) F_{\text{Bkg}}(x) \cdot M_{\text{Bkg}}(m)$$

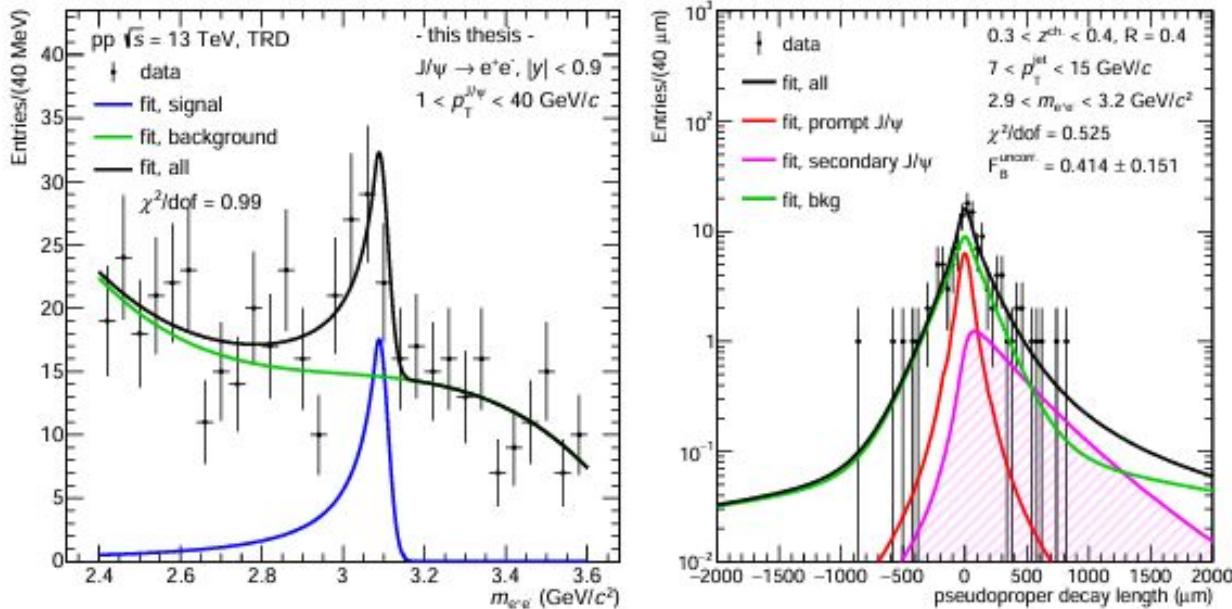
$$F_{\text{sig}}(x) = f'_B \cdot F_B(x) + (1 - f'_B) \cdot F_{\text{prompt}}(x)$$

$$F_{\text{prompt}}(x) = R(x)$$

$$F_B = \chi_B(x) \otimes R(x)$$

- Use log-likelihood fit
  - Blue: PDFs from MC - weight to get distribution in each  $z^{\text{ch.}}$  bin
  - Red: PDFs from data - fit in bins of  $z^{\text{ch.}}$

# Método da Máxima Verossimilhança em HEP



A separação de  $J/\psi$  (um méson formado pelos quarks c cbar) do tipo prompt e não prompt é feita **maximizando o logaritmo da função de verossimilhança** nas variáveis massa e “pseudoproper decay length”.

# Método dos Mínimos Quadrados

O método dos mínimos quadrados é uma técnica matemática usada para ajustar um modelo a um conjunto de dados, de forma que o modelo seja o mais próximo possível dos dados observados.

A ideia central é minimizar a soma dos quadrados das diferenças (ou "resíduos") entre os valores observados e os valores previstos pelo modelo.

# Método dos Mínimos Quadrados

Se as variáveis  $y_i$  são distribuídas de forma gaussiana (ou seja, dado  $x$ , diversas medidas de  $y$  geram uma distribuição gaussiana):

$$L(\vec{y}|\vec{x}, \vec{\sigma}, \vec{\theta}) = \prod_{i=1}^N \frac{1}{\sqrt{2\pi\sigma_i^2}} \exp\left(-\frac{(y_i - f(x_i|\vec{\theta}))^2}{2\sigma_i^2}\right)$$

$$\ln L = -\frac{1}{2} \sum_{i=1}^N \ln(2\pi\sigma_i^2) - \frac{1}{2} \sum_{i=1}^N \frac{(y_i - f(x_i|\vec{\theta}))^2}{\sigma_i^2}$$

Maximizar o **In L** implica em minimizar o  $\chi^2$  (por que?)

$$\chi^2(\vec{\theta}) = \sum_{i=1}^N \frac{[y_i - f(x_i|\vec{\theta})]^2}{\sigma_i^2}$$

# Método dos Mínimos Quadrados: aplicação - reta

$$y = f(x|a_0, a_1) = a_0 + a_1 x$$

$$\chi^2 = \sum_{i=1}^N \frac{[y_i - f(x_i|a_0, a_1)]^2}{\sigma_i^2}$$

$$\chi^2(a_0, a_1) = \sum_{i=1}^N \frac{(y_i - a_0 - a_1 x_i)^2}{\sigma_i^2}$$

$$\frac{\partial \chi^2(a_0, a_1)}{\partial a_0} = 0$$

$$\frac{\partial \chi^2(a_0, a_1)}{\partial a_1} = 0$$

**Façam como exercício!**



# Método dos Mínimos Quadrados: aplicação

$$0 = \frac{\partial \chi^2(a_0, a_1)}{\partial a_0} = -2 \sum_{i=1}^N \frac{(y_i - a_0 - a_1 x_i)}{\sigma_i^2},$$

$$0 = \frac{\partial \chi^2(a_0, a_1)}{\partial a_1} = -2 \sum_{i=1}^N \frac{(y_i - a_0 - a_1 x_i) x_i}{\sigma_i^2}$$

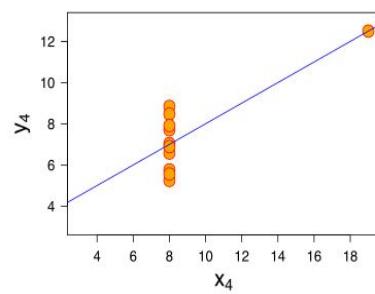
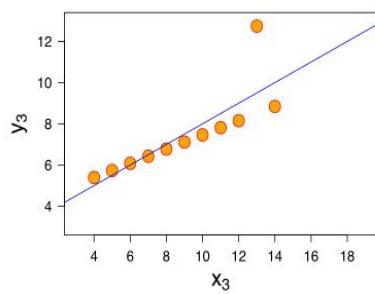
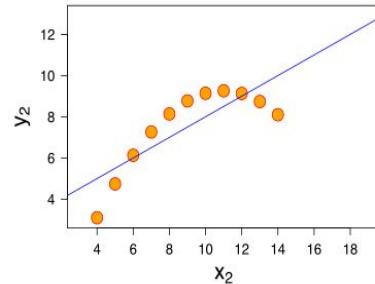
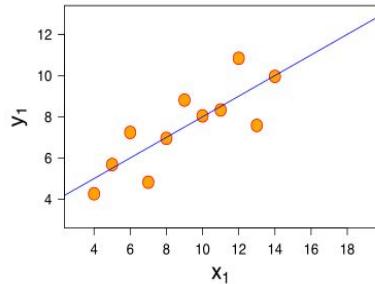
$$a_0 \sum_{i=1}^N \frac{1}{\sigma_i^2} + a_1 \sum_{i=1}^N \frac{x_i}{\sigma_i^2} = \sum_{i=1}^N \frac{y_i}{\sigma_i^2},$$

$$a_0 \sum_{i=1}^N \frac{x_i}{\sigma_i^2} + a_1 \sum_{i=1}^N \frac{x_i^2}{\sigma_i^2} = \sum_{i=1}^N \frac{x_i y_i}{\sigma_i^2}.$$

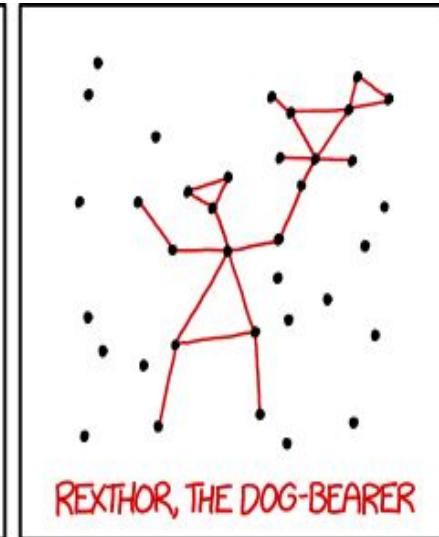
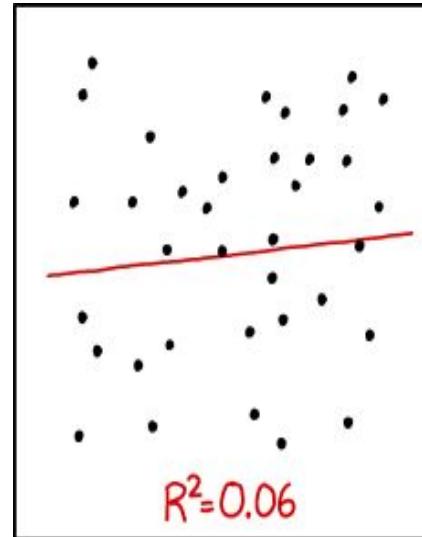
$$\begin{aligned} S &\equiv \sum_{i=1}^N \frac{1}{\sigma_i^2} & S_x &\equiv \sum_{i=1}^N \frac{x_i}{\sigma_i^2} \\ S_{xx} &\equiv \sum_{i=1}^N \frac{x_i^2}{\sigma_i^2} & S_y &\equiv \sum_{i=1}^N \frac{y_i}{\sigma_i^2} \\ S_{xy} &\equiv \sum_{i=1}^N \frac{x_i y_i}{\sigma_i^2} & \Delta &\equiv S_{xx} S - S_x^2 \end{aligned}$$

$$\begin{aligned} \hat{a}_0 &= \frac{1}{\Delta} (S_y S_{xx} - S_x S_{xy}), \\ \hat{a}_1 &= \frac{1}{\Delta} (S S_{xy} - S_x S_y). \end{aligned}$$

# Método dos Mínimos Quadrados



Quatro conjuntos de dados para os quais algumas das propriedades estatísticas usuais (média, variância, correlação e regressão linear) são as mesmas, embora os conjuntos de dados sejam diferentes. Anscombe, Francis J. (1973) Graphs in statistical analysis. American Statistician, 27, 17–21. Imagem: Wikimedia Commons.



I DON'T TRUST LINEAR REGRESSIONS WHEN IT'S HARDER TO GUESS THE DIRECTION OF THE CORRELATION FROM THE SCATTER PLOT THAN TO FIND NEW CONSTELLATIONS ON IT.

# Teste de qualidade de ajustes

Desvio normalizado entre os valores medidos  $\mathbf{y}_i$  e os estimados ( $\boldsymbol{\mu}$ ) por um dado modelo/ajuste de acordo com parâmetros  $\boldsymbol{\theta}$  e relativo ao desvio padrão  $\sigma_i$

$$z_i(\vec{\theta}) = (y_i - \mu_i(\hat{\theta})) / \sigma_i$$

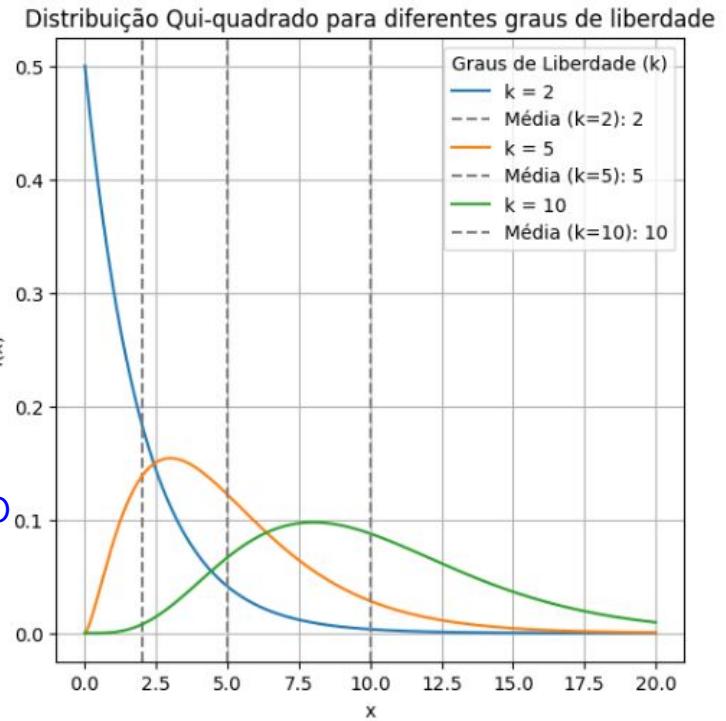
- Fornece uma medida de compatibilidade entre os dados e o modelo (obtido pelo ajuste) relativo às incertezas.
- $z_i$  terá uma distribuição gaussiana se os dados  $y_i$  forem distribuídos de forma gaussiana
- Se a função  $\mu_i(\theta)$  for linear nos parâmetros  $\theta$ , e representar apropriadamente os dados, esperamos o mínimo de  $\chi^2$  obtido pelo MMQ seja distribuído como uma PDF de  $\chi^2$  com  $N - m$  graus de liberdade (sendo  $N$  o número de pontos e  $m$  o número de parâmetros no fit).

$$\chi^2 = \sum_i z_i^2$$

# Teste de qualidade de ajustes: $\chi^2$

$$\chi^2 = \sum_{i=1}^n \frac{[y_i - f(x_i|\vec{\theta})]^2}{\sigma_i^2}$$

O valor esperado (média) de uma variável aleatória com distribuição de  $\chi^2$  é igual ao número de graus de liberdade.



$$p_{\chi^2}(z|n) \equiv \frac{1}{2^{n/2}\Gamma(n/2)} z^{n/2-1} e^{-z/2} \quad \text{for } n = 1, 2, \dots$$

# Critério de Akaike

O **critério de informação de Akaike (AIC)** é uma métrica que mensura a qualidade de um modelo estatístico visando também a sua simplicidade.

→ métrica para **comparação e seleção de modelos**, em que menores valores de AIC representam uma maior qualidade e simplicidade.

$$\text{AIC} = 2k - 2 \ln(\mathcal{L})$$

Dado uma coleção de modelos/funções candidatos para se explicar dados experimentais, o **modelo com menor AIC é o escolhido**.

AIC **bonifica a qualidade de ajuste** (altos valores para a função de verossimilhança) e **penaliza a quantidade de parâmetros** do modelo.

→ não fornece uma medida de **qualidade do modelo global**, apenas relativa no que diz respeito à **comparação entre modelos candidatos**

# Critério de Akaike

Aplicação do critério de Akaike no ajuste de dados com erros normalmente distribuídos

$$\text{AIC} = 2k - 2 \ln(\mathcal{L})$$

$$\mathcal{L} = \prod_{i=1}^n \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{(y_i - \hat{y}_i)^2}{2\sigma^2}\right)$$

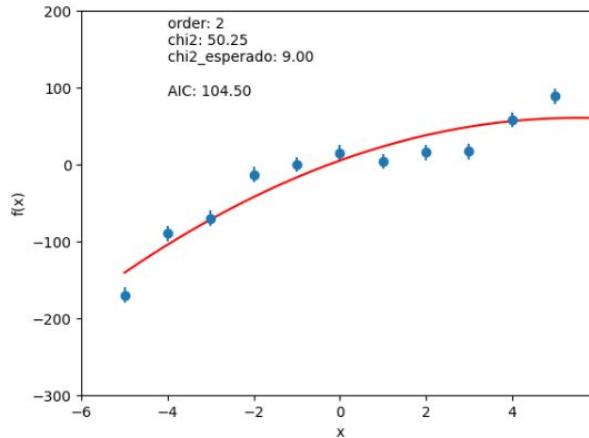
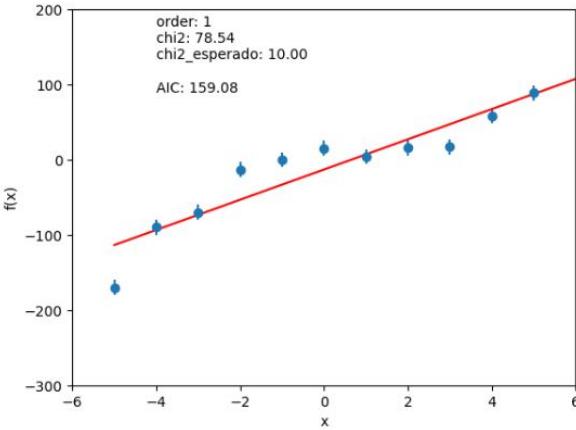
$$\ln(\mathcal{L}) = \sum_{i=1}^n \left( -\frac{1}{2} \ln(2\pi\sigma^2) - \frac{(y_i - \hat{y}_i)^2}{2\sigma^2} \right)$$

$$\ln(\mathcal{L}) = -\frac{n}{2} \ln(2\pi\sigma^2) - \frac{1}{2\sigma^2} \sum_{i=1}^n (y_i - \hat{y}_i)^2$$

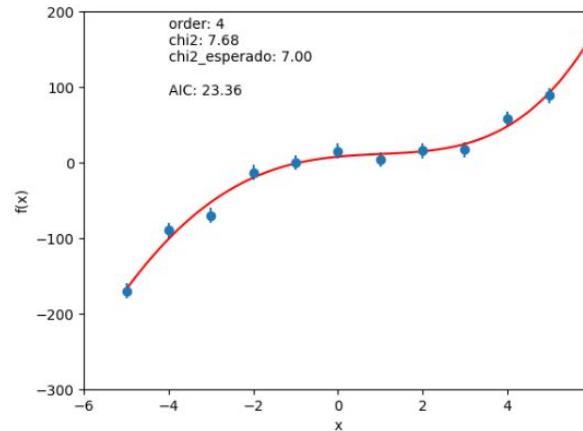
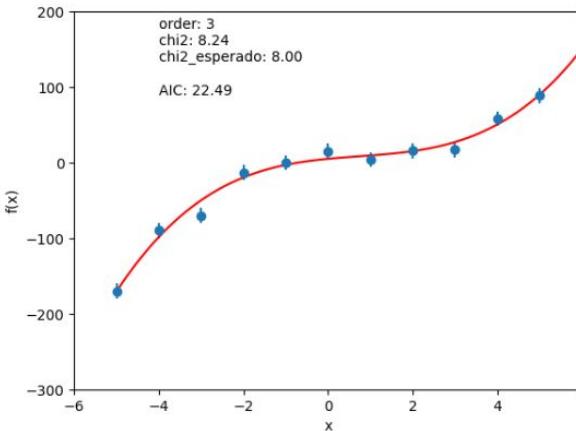
$$\ln(\mathcal{L}) = -\frac{n}{2} \ln(2\pi\sigma^2) - \frac{\chi^2}{2\sigma^2}$$

$$\boxed{\text{AIC} = 2k + 2\chi^2}$$

# Critério de Akaike: exemplo



$$AIC = 2k + 2\chi^2$$



link: [Akaike.ipynb - Colab](#)

# Filtro de Kalman

- O **filtro de Kalman** é um algoritmo que **combina medições ruidosas de um sistema com um modelo matemático** para estimar de forma precisa o estado real desse sistema (como posição, velocidade, etc.) ao longo do tempo.
- Ele é amplamente usado para **rastrear a posição de objetos em movimento e filtrar ruídos em sistemas dinâmicos, aproveitando as informações atuais e passadas para minimizar o erro da estimativa.**
- **Aplicações:** processamento de sinais, ajuste de sinais, reconhecimento de padrões, navegação e controle.
- Por design, um filtro de Kalman opera **recursivamente** em uma ou várias sequências de dados de entrada ruidosos para produzir estimativas estatisticamente ótimas do estado subjacente de um sistema físico.

# Filtro de Kalman

Filtros de Kalman possuem as seguintes características atrativas que os tornam **preferíveis** aos métodos globais de mínimos quadrados em certas circunstâncias:

- Um **filtro de Kalman é recursivo**, sendo, portanto, adequado para o processamento progressivo de sinais, especialmente na **busca e ajuste de trajetórias em sistemas de detecção grandes e complexos**.
- Um filtro de Kalman facilita a resolução eficiente e **remoção de pontos discrepantes**.
- Ao contrário dos métodos de mínimos quadrados, **um filtro de Kalman não envolve a manipulação ou inversão de grandes matrizes**.

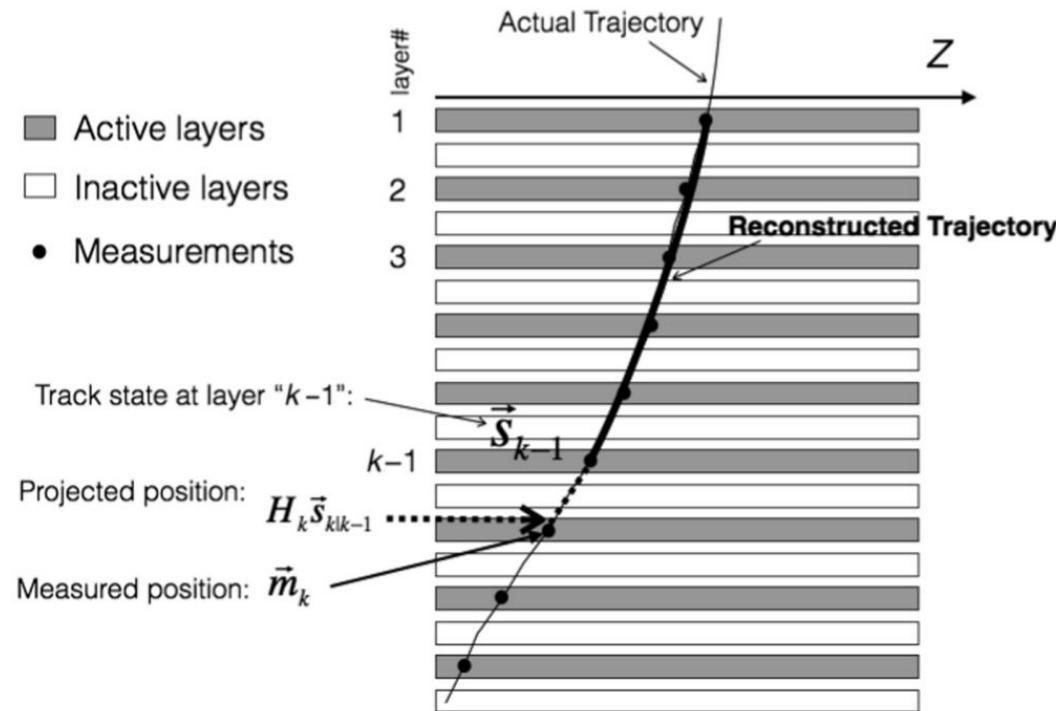
Filtros de Kalman básicos assumem que o conhecimento do estado na etapa k, denotado por  $s_{k|k-1}$  e conhecido como priori, pode ser previsto com base no conhecimento do estado em  $t_{k-1}$  de acordo com um modelo linear:

$$\vec{s}_{k|k-1} = \mathbf{F}_k \vec{s}_{k-1} + \vec{w}_k$$

$\mathbf{F}$  é uma função linear que descreve a evolução do estado  
 $w$  corresponde ao ruído do processo

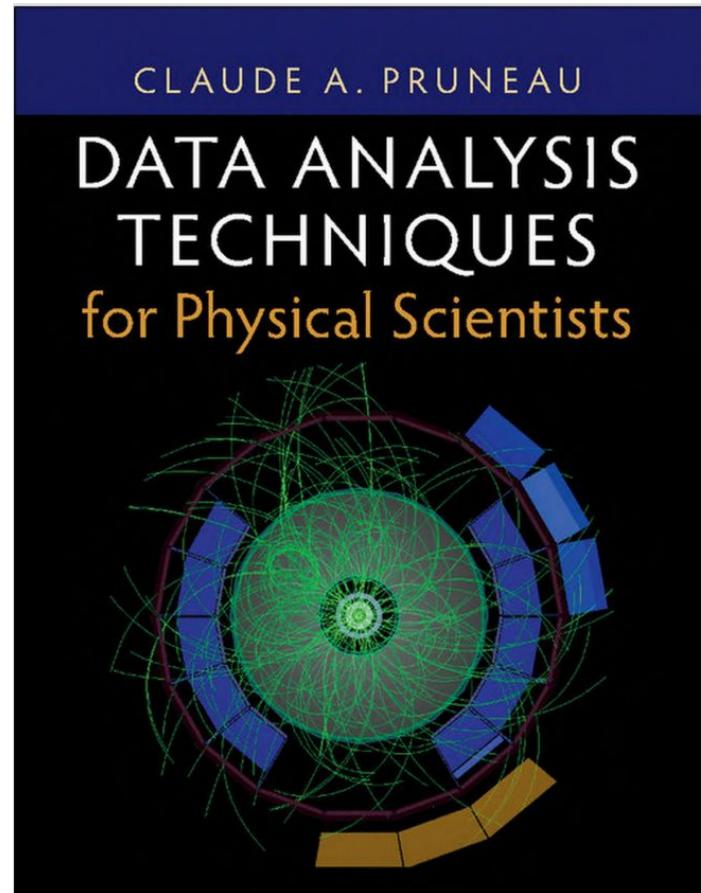
# Filtro de Kalman: aplicação em HEP

Método progressivo para ajustar **trajetórias de partículas**



# Erros experimentais e Unfolding

Capítulo 12



# Erros experimentais

Qual a diferença de erro estatístico e erro sistemático?

O **erro estatístico** está ligado à natureza probabilística dos dados coletados e à quantidade de eventos observados.

→ O erro estatístico diminui conforme se aumenta a coleta de dados, ou seja, com mais tempo de tomada de dados ou com uma taxa de colisão mais alta.

O **erro sistemático**, refere-se a desvios que não são causados pelo acaso, mas por imperfeições no próprio experimento ou no método de análise.

→ Mas se a fonte do erro sistemático for conhecida, e for possível avaliar o mesmo, deverá ser corrigida e portanto, não deverá consistir de uma incerteza.

# Erros experimentais

O **erro sistemático** consiste de um erro que é igual a todos os “n” dados medidos, **devido a efeitos desconhecidos.**

O erro sistemático não pode ser reduzido ou avaliado simplesmente aumentando o número de repetições.

Este erro surge devido a fatores que afetam de maneira consistente os resultados, como **limitações dos detectores, calibrações imprecisas, métodos de análise imperfeitos ou condições ambientais que alteram as medições.**

Exemplo: um detector desalinhado afeta as medições de maneira constante.

# Unfolding

Unfolding é o processo de **correção de dados experimentais** para compensar os **efeitos de resolução limitada** e outras **imperfeições dos detectores**.

→ Permite recuperar a distribuição verdadeira dos dados, ou seja, a distribuição que seria observada em um detector perfeito.

→ **Exemplo:** medição da distribuição de momento transversal ( $p_T$ ) de partículas em colisões no LHC.

→ Quando partículas são detectadas, a **medição de seu momento transversal não é exata**, devido a **limitações do detector**, como a **resolução finita** e **efeitos de ruído**.

→ Esse processo **distorce a distribuição observada de  $p_T$** , fazendo-a diferir da distribuição verdadeira.

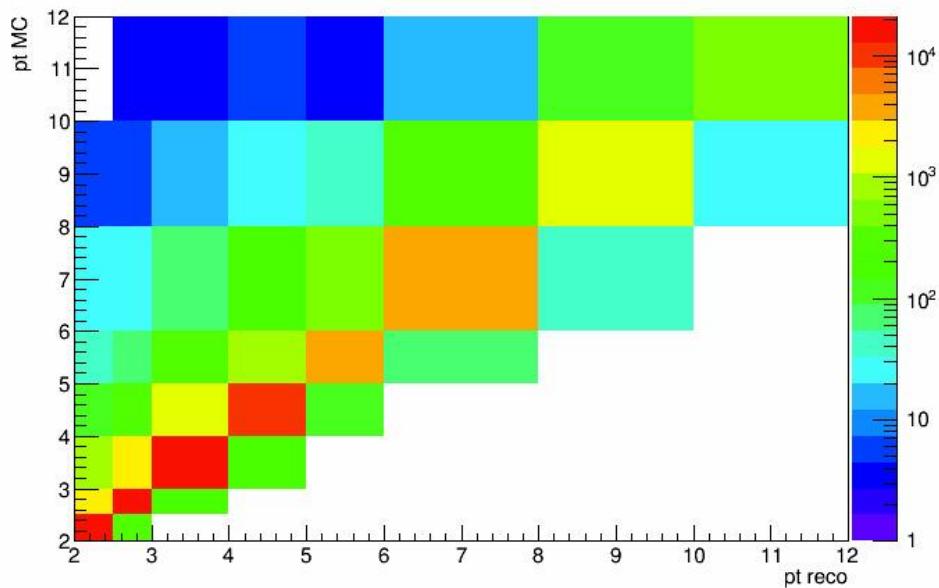
# Unfolding

Passos para fazer o unfolding:

**Simulação com Detector:** Uma simulação de MC é realizada, onde se **modela a resposta do detector para a distribuição teórica de  $p_T$** . Isso gera uma **matriz de resposta** que descreve como a distribuição verdadeira se transforma na distribuição observada.

**Aplicação do Unfolding:** Técnicas de unfolding utilizam a **matriz de resposta** para "inverter" o efeito do detector, aproximando a distribuição verdadeira a partir dos dados observados.

$$p_T^{true} = C^{-1} p_T^{rec}$$



Matriz de correlação ou Smearing matrix

**Mais em:**

Data Analysis Techniques for Physical Scientists

Hands-on

