APRENDIZADO DE MÁQUINA PARA FÍSICA

Aula 1

Revisão de Probabilidad e Estatística

André Sznajder
Clemencia Mora Herrera
Helena Malbouisson

INFORMAÇÃO DA DISCIPLINA

- Repositório GitHub https://github.com/clemencia/ML4PPGF_UERJ
- Twiki : http://dfnae.fis.uerj.br/twiki/bin/view/DFNAE/ML4PPGF
    - Ementa
- Referências Parte I Probabilidade e Estatística
    - Vitor Oguri, Métodos Estatísticos em Física Experimental, Editora Livraria da Física. 2017
    - G. Cowan, Statistical Data Analysis, Oxford, 1998.
    - F. James, Statistical Methods in Experimental Physics, 2nd ed., World Scientific, 2006
    - S. Brandt, Statistical and Computational Methods in Data Analysis, Springer, 1998
    - L. Lista, Statistical Methods for Data Analysis in Particle Physics, Springer, 2017

PROBABILIDADE vs. ESTATÍSTICA

Probabilidade

Rama da matemática que estuda *modelos matemáticos* de distribuições de probabilidades (ou densidade) e a descrição das suas propriedades usando *parâmetros* como média, variância, correlações etc.

Através de conceitos estatísticos podemos quantificar a correspondência entre predições da teoria e observações experimentais.

Estatística

Conjunto de métodos para tratar da coleta, análise, interpretação e apresentação de dados. É uma prática indutiva e ***empírica***. A estatística quer achar as limitações que os ***dados impõem*** aos parâmetros θ de um modelo (Point and Interval Estimation). Também serve para decidir qual ***modelo*** se ajusta melhor ao conjunto de dados (Hypothesis Test).

DEFINIÇÕES DE PROBABILIDADE

A probabilidade pode ser definida de diversas formas:

- **Frequentista**: razão do número de ocorrências para o número total de experimentos (no limite de N muito grande). Válido para experimentos "repetíveis" ***P a posteriori***

- **Bayesiana**: grau de crença que uma ocorrência seja "certa". (apostaria no bolão?). Pode ser aplicada a eventos desconhecidos. ***P a priori***

- **Kolmogorov** * : definição ***axiomática,*** independente da interpretação. O conjunto de números Pi associado a um conjunto de elementos é um conjunto de probabilidades se cumpre as propriedades matemáticas das probabilidades (Pi positivo, Σ Pi=1, P(a U b) = Pa+Pb se a e b excludentes). Formalismo aplicável às interpretações frequentista e bayesiana.

KOLMOGOROV

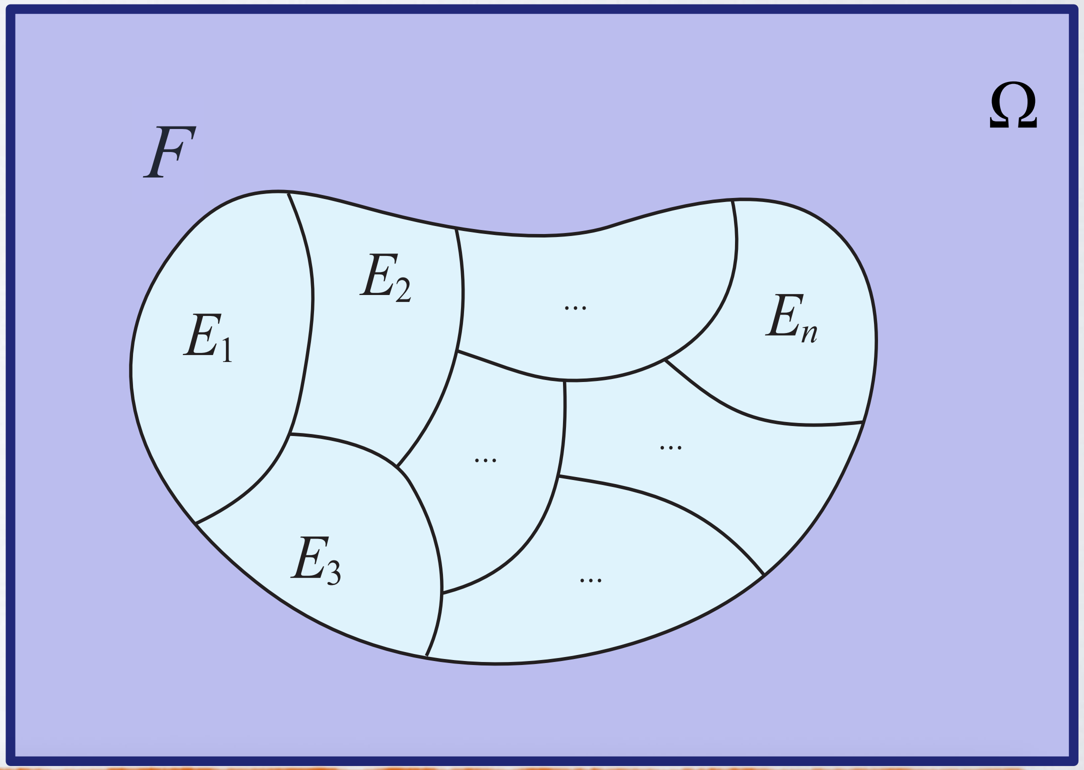
Definição axiomática:

- Terminologia: Ω = espaço da amostra, F = espaço de eventos ($F \subset \Omega$) e P = medida da probabilidade

- Seja $(\Omega, F \subseteq \Omega, P)$ espaço de medida que satisfaz:
  - $P(E) \geq 0 \;\; \forall E \in F$
  - $P(\Omega) = 1$ (normalização)
  - $\forall (E_1, E_2, \ldots, E_n) \in F : E_i \cap E_j = \varnothing$

    (i.e., eventos excludentes)

    $$\Rightarrow P(\cup_{i=1,\ldots,n} E_i) = \sum_{i=1,\ldots,n} P(E_i)$$



Válida tanto para interpretações frequentistas quanto bayesianas

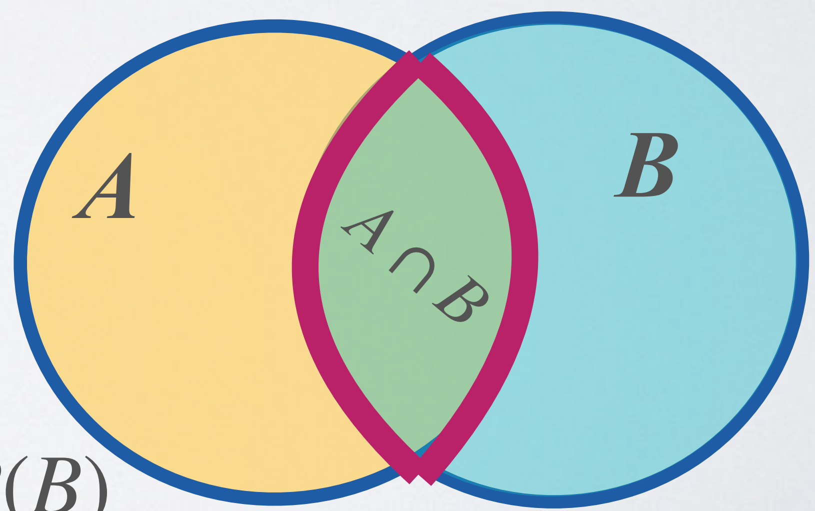
TEOREMAS

notação

$$A + \bar{A} = \Omega \;\Rightarrow\; P(A) + P(\bar{A}) = 1$$

$$P(A \text{ ou } B) \equiv P(A \cup B)$$
$$P(A \text{ e } B) \equiv P(A \cap B)$$

- $P(\bar{A}) = 1 - P(A)$
- $P(A \cup \bar{A}) = 1$
- $P(\varnothing) = 0$
    - Se A e B são excludentes, i.e. $A \cap B = \varnothing \Rightarrow P(A \cap B) = 0$
- Se $A \subset B \Rightarrow P(A) \leq P(B)$
- $P(A \cup B) = P(A) + P(B) - P(A \cap B)$
- A e B são independentes se $P(A \cap B) = P(A) \times P(B)$



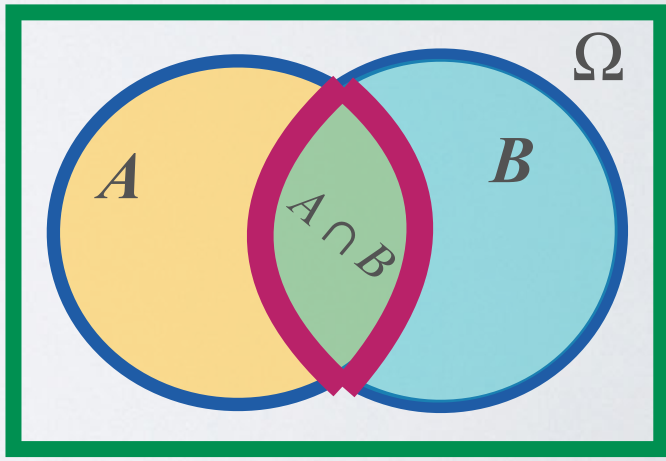
PROBABILIDADE CONDICIONAL

- A ***probabilidade condicional*** de A relativa a B é denotada $P(A|B)$
  - ➡ equivale a considerar B como o "universo" dos resultados possíveis de A
- Está dada pela razão da probabilidade de ocorrências conjuntas em relação a probabilidade de ocorrência do B: $P(A|B) = \dfrac{P(A \cap B)}{P(B)}$

➡ para eventos excludentes: $P(A|B) = P(B|A) = 0$

➡ para eventos independentes: $P(A \cap B) = P(A) \times P(B)$ $\Rightarrow P(A \cap B) = P(A|B) \times P(B) = P(B|A) \times P(A)$

então $P(A|B) = P(A)$ e $P(B|A) = P(B)$

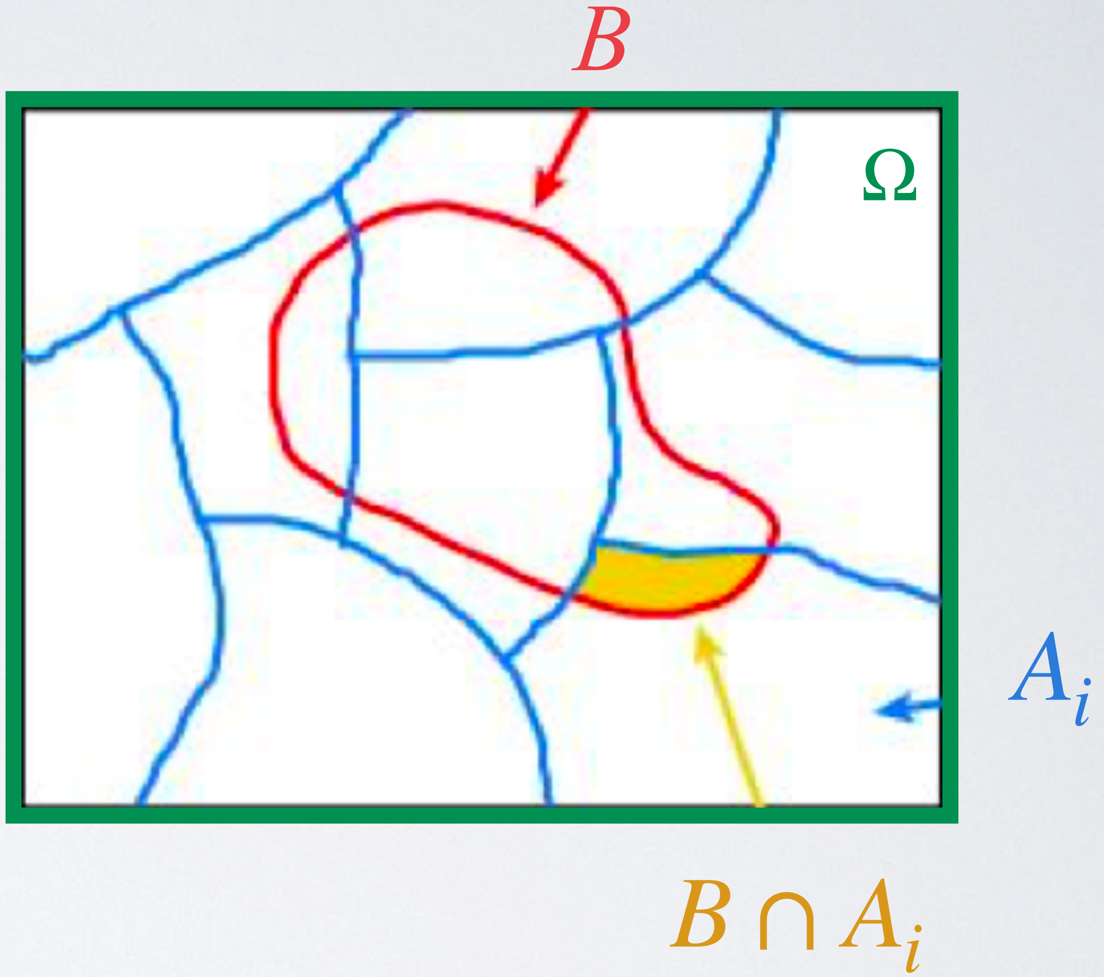


- A fórmula de Bayes mais geral $P(B|A) = \dfrac{P(A|B) \times P(B)}{P(A)}$
- Alternativamente $P(B|A) = \dfrac{P(A|B) \times P(B)}{P(A|B) \times P(B) + P(A|\bar{B}) \times P(\bar{B})}$

LEI DA PROBABILIDADE TOTAL

- A_i partição do espaço de amostra em subconjuntos excludentes, e onde $\cup_i A_i = \Omega$ teremos

$$P(B) = \sum_i P(B|A_i) \times P(A_i)$$



- Agora podemos reescrever o teorema de Bayes como

$$P(A|B) = \frac{P(B|A) \times P(A)}{\sum_i P(B|A_i) \times P(A_i)}$$

EXERCÍCIOS

1. Um teste de detecção de um determinado vírus tem eficiência de quase 100%, ou seja, todo paciente infectado (I) ontem um positivo P, ou seja $p(P|I)=1$. No entanto o teste pode entregar resultados positivos a pacientes saudáveis (falso positivo) com uma taxa de 1 em mil $p(P|S)=10^{-3}$. Numa pequena cidade, as autoridades estimam que 1 a cada 10 mil moradores estejam infectados pelo vírus, i.e. $p(I)=10^{-4}$. Qual a probabilidade de uma pessoa estar saudável tendo recebido um resultado positivo do teste?

2. Numa cidadezinha, 15% dos táxis são azuis e os outros são verdes. Certa noite um táxi atropela uma pessoa e foge. Uma testemunha identifica a cor do táxi envolvido no acidente como azul. A polícia verifica que, em circunstâncias similares as da noite do acidente, a testemunha consegue identificar a cor corretamente 80% das vezes, e confunde as cores 20% das vezes. Qual a probabilidade de ter sido realmente azul o táxi do atropelamento?

3. Teste gravidez : Se o teste de farmácia tem uma taxa $<10^{-6}$ de falso positivo (praticamente 0), e uma taxa de falso negativo de 0.1%. Qual a probabilidade de uma pessoa estar grávida se o teste deu negativo? (considere que, sem proteção, a taxa de "falha" na anticoncepção é ~85%)

4. Tem 3 urnas A, B e C. A urna A contém 5 bolas brancas e 6 pretas, a urna B tem 4 bolas brancas e 5 pretas e a urna C tem 4 brancas e 4 pretas. Após escolher por acaso uma das urnas (P(A)=P(B)=P(C)) e se retirar uma bola preta, determine a probabilidade de que a bola sorteada tenha vindo da urna C.

DENSIDADE DE PROBABILIDADE

- Uma variável aleatória pode ser discreta (números nos dados, cara ou cruz) ou contínua (medidas de uma grandeza x). A probabilidade de achar um valor no intervalo $[x, x+dx]$:

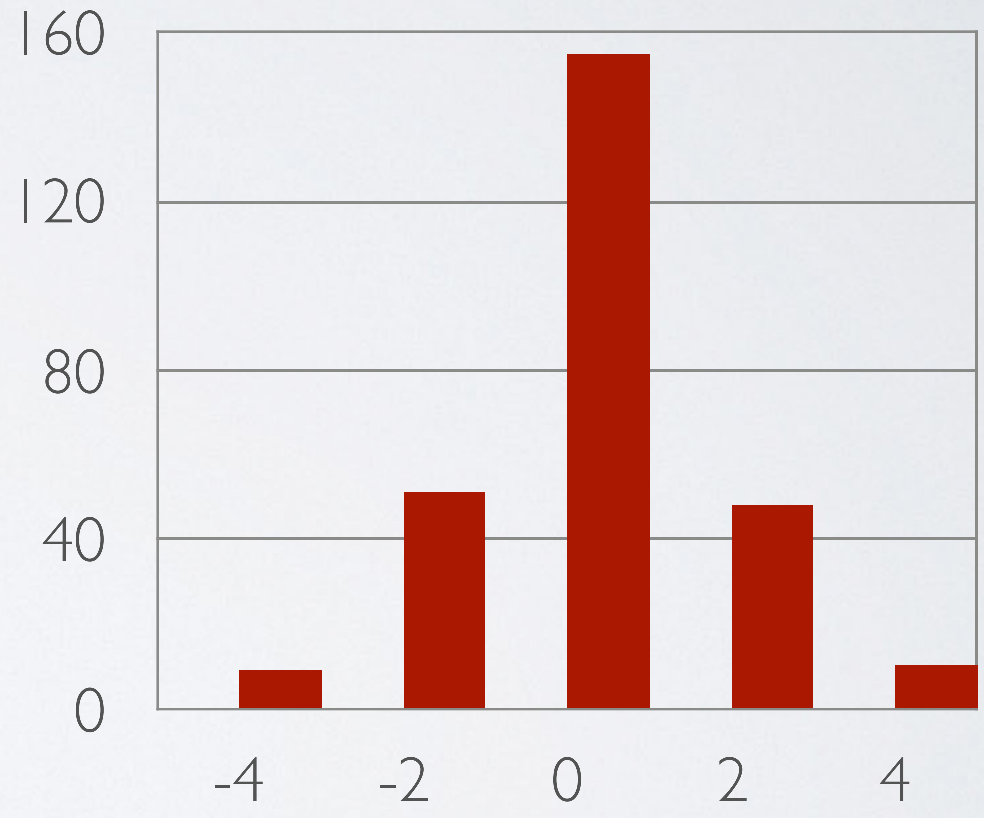
$$P(x_{obs} \in [x, x+dx]) = f(x)dx \quad \text{ou} \quad \frac{\mathrm{d}P(x)}{\mathrm{d}x} = f(x)$$

- onde $f(x)$ é chamada de função de densidade de probabilidade (PDF)
  - Na interpretação frequentista, $f(x)$ seria a fração de vezes que a variável seja achada no intervalo $[x, x+dx]$
- A condição de normalização é $\int_{-\infty}^{\infty} f(x)dx = P(x_{obs} \in [-\infty, \infty]) = 1$

HISTOGRAMAS

- O histograma é uma função que conta as frequências (número de observações) que acontecem em categorias disjuntas (*bins* ou intervalos). Entrega uma visualização da forma, localização e dispersão dos dados

Bin Low	Bin High	Freq.
-4	-2.01	9
-2	-0.01	51
0	1.99	155
2	3.99	48
4	5.99	10



Na interpretação frequentista: uma PDF é um histograma no limite de infinitas medidas, categorias de largura zero e normalizado à unidade

PARÁMETROS E VALORES ESPERADOS

- o valor esperado da variável é a média, definida como: $\mu = E[x] = \int x\, f(x) dx$
    - para outras funções $\mu_a = E[a(x)] = \int a(x) f(x) dx$
- $\sigma^2 = E[(x-\mu)^2] = \int (x-\mu)^2 f(x) dx = E[x^2] - \mu^2$ é a variância
    - o desvio padrão $\sigma = \sqrt{E[x^2] - \mu^2}$
- Covariância (para mais de uma variável aleatória)

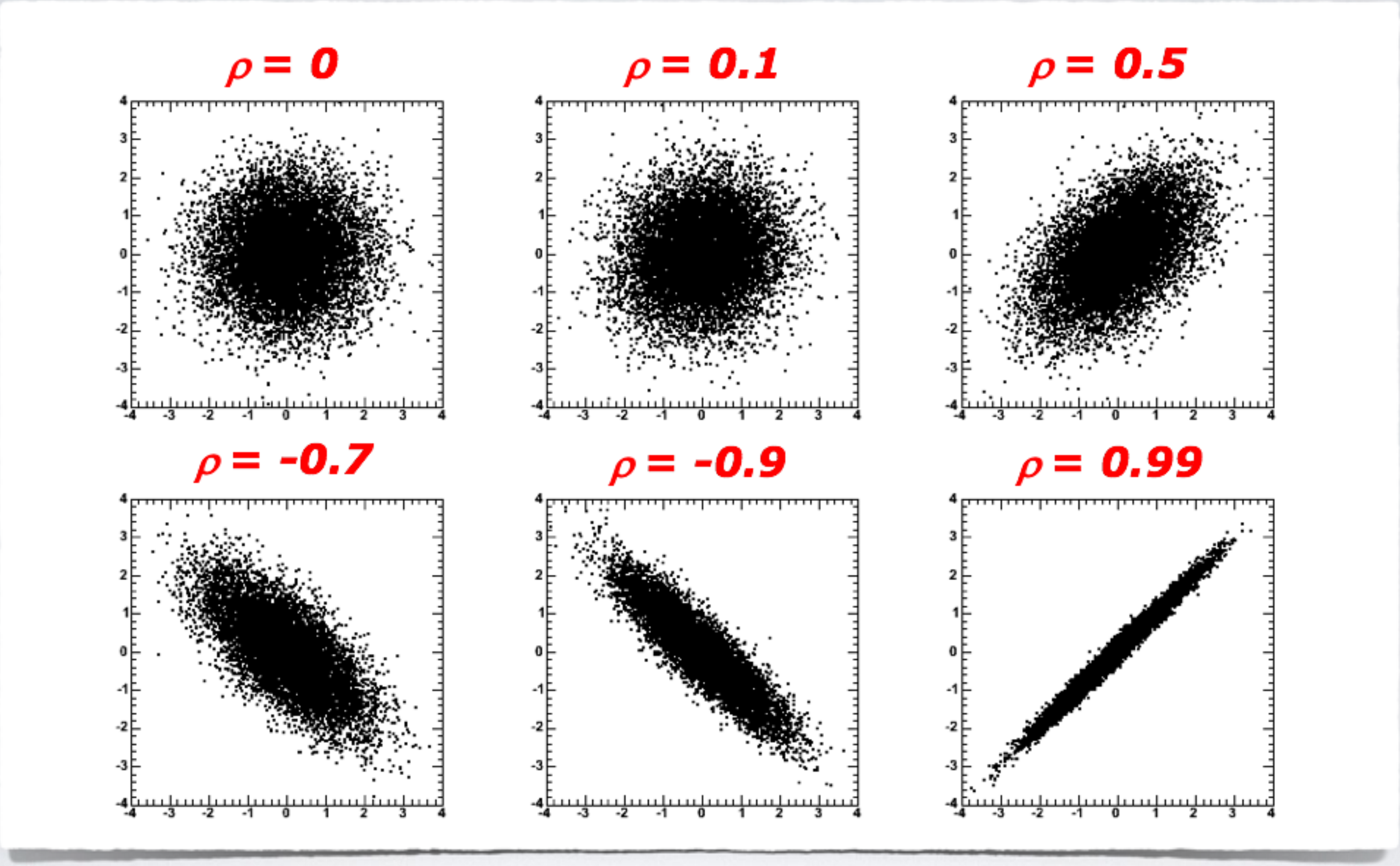
$$V_{xy} = E[(x-\mu_x)(y-\mu_y)] = \int\int (x-\mu_x)(y-\mu_y) f(x,y)\, dx\, dy = E[xy] - \mu_x \mu_y$$

- Se forem independentes a pdf pode ser fatorizada $f(x,y) = g(x)\, h(y)$

$$E[xy] = \int x\, g(x) dx \int y\, h(y) dy = \mu_x \mu_y \Rightarrow V_{xy} = 0$$

- Coeficiente de correlação $\rho_{xy} = \dfrac{V_{xy}}{\sigma_x \sigma_y}$

VALORES DE CORRELAÇÃO

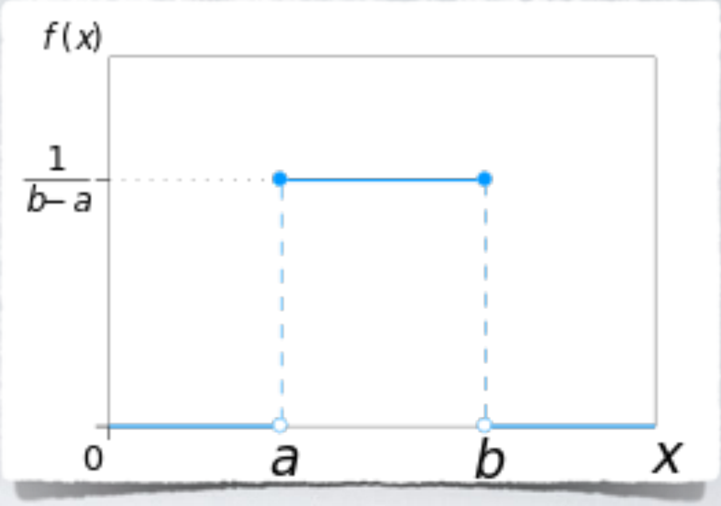


exercício em jupyter/colab

FUNÇÕES DE DISTRIBUIÇÃO DE PROBABILIDADE

Distribuição uniforme:

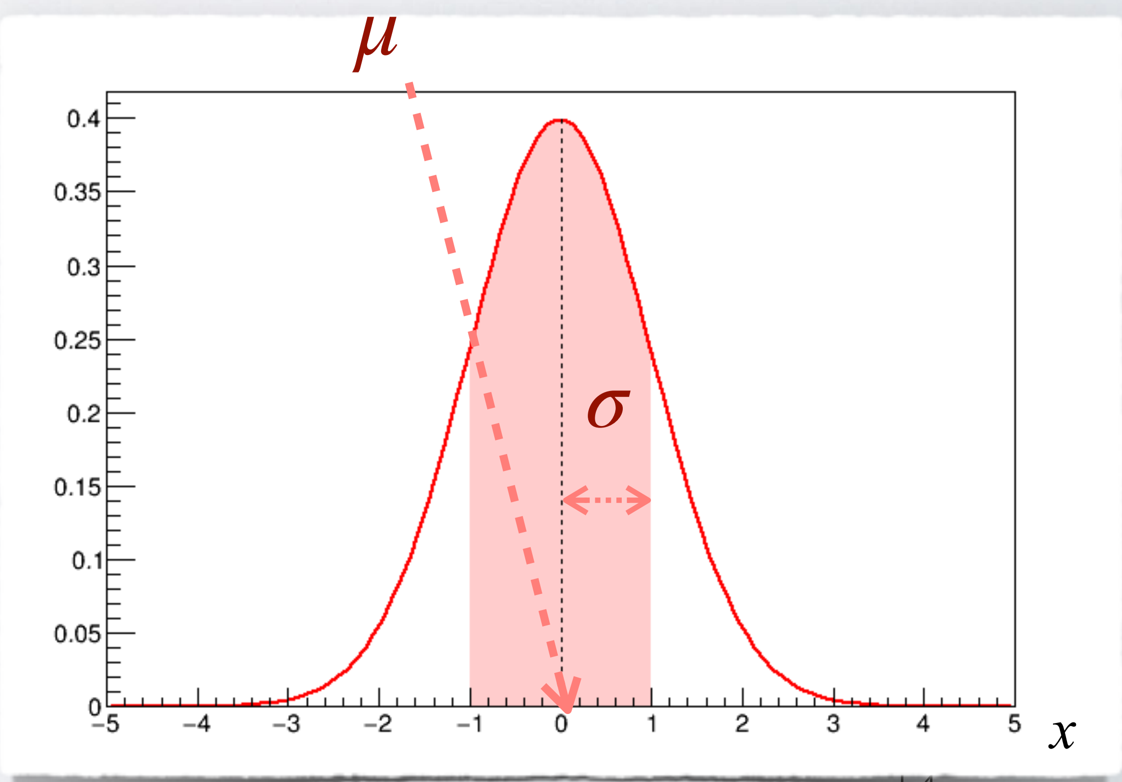
$$f(x) = \begin{cases} \dfrac{1}{b-a} & a \leq x \leq b \\ 0 & x < a \text{ ou } x > b \end{cases}$$



- distribuição básica de ***geradores pseudo-aleatórios***

Distribuição Gaussiana ou Normal:

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



TEOREMA DO LIMITE CENTRAL (LEI DOS ERROS)

- A soma de N variáveis aleatórias independentes, mas que obedeçam a mesma distribuição de probabilidades se aproxima de uma distribuição normal a medida que N $\rightarrow \infty$
  - a distribuição de probabilidade das variáveis aleatórias precisa ter valor esperado $E[x_i] = \mu$ e variância $V(x_i) = \sigma^2$ ***definidas***
- A média de N variáveis aleatórias é uma variável aleatória, e sua distribuição de probabilidades converge a uma Gaussiana ***independentemente*** da PDF original das variáveis !
  - a média $\bar{x} = \sum x_i / N$ tem valor esperado $E[\bar{x}] = \mu$ e tem variância $V(\bar{x}) = \sigma^2 / N$

Teorema do limite central : ver notebook