

# Tratamento Estatístico de Dados em Física Experimental

Prof. Zwinglio Guimarães

2º semestre de 2018

Tópico 2 – 2ª semana – 18 e 20/09

# Principais parâmetros de uma função densidade de probabilidade (revisão)

- O valor médio (verdadeiro),  $x_0$ , que é obtido por:

$$x_0 = \langle x \rangle = \int_{-\infty}^{+\infty} x f(x) dx$$

- E o desvio-padrão (verdadeiro),  $\sigma_0$ , que é obtido por:

$$\sigma_0^2 = \langle (x - x_0)^2 \rangle = \int_{-\infty}^{+\infty} (x - x_0)^2 f(x) dx$$

- **Consideração prática:**  $\langle (x - x_0)^2 \rangle = \langle x^2 \rangle - x_0^2$

# Outros parâmetros de localização de funções densidade de probabilidade

- **Moda,  $x_{mp}$** : valor de  $x$  em que  $f(x)$  é máximo
- **Mediana,  $x_M$** : valor de  $x$  tal que a probabilidade de se obter um dado com  $x \leq x_M$  é igual ao de  $x \geq x_M$ . Ou seja:

$$\int_{-\infty}^{x_M} f(x)dx = \int_{x_M}^{+\infty} f(x)dx = 0.5$$

# Momentos de funções densidade de probabilidade

O momento de ordem  $n$ ,  $\mu_n$ , é dado por:

$$\mu_n = \langle x^n \rangle = \int_{-\infty}^{+\infty} x^n f(x) dx$$

O momento central de ordem  $n$ ,  $\mu_n^0$  é dado por:

$$\mu_n^0 = \langle (x - x_0)^n \rangle = \int_{-\infty}^{+\infty} (x - x_0)^n f(x) dx$$

# Outros parâmetros úteis para caracterizar funções densidade de probabilidade:

## Momentos centrais normalizados

- Obliquidade ou Assimetria (“*skewness*”):

$$S = \frac{\mu_3^0}{\sigma^3} = \frac{\langle (x - x_0)^3 \rangle}{\sigma^3}$$

- Curtose (“*kurtosis*”):

$$K = \frac{\mu_4^0}{\sigma^4} = \frac{\langle (x - x_0)^4 \rangle}{\sigma^4}$$

# A função de distribuição acumulada

- A função de distribuição acumulada,  $g(x)$ , corresponde à integral da função densidade de probabilidade,  $f(x)$ , desde  $-\infty$  até  $x$ :

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

Obs: os pacotes estatísticos costumam conter funções de distribuição acumulada para as funções densidade de probabilidade mais comuns. Usadas em testes estatísticos para avaliar se o valor de  $x$  não é muito pequeno ( $g(x) \sim 0$ ) nem muito grande ( $g(x) \sim 1$ ).

# A função densidade de probabilidade gaussiana

- A função densidade de probabilidade gaussiana é dada por:

$$f(x) = \frac{1}{\sigma \sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{x-x_0}{\sigma}\right)^2}$$

*"...lei em que todos creem. Os experimentais pensam que é um teorema matemático e os matemáticos, que é um fato experimental."*

(citado em J.-P. Benzécri, *Histoire et Prehistoire de l'Analyse des Données*. Paris, Bordas, 1982)

# Resultados importantes da gaussiana

## (1) - A Gaussiana Padrão

- Uma mudança de variáveis muito comum consiste em escrever a gaussiana em termos do erro normalizado:

$$z = \left( \frac{x - x_0}{\sigma} \right)$$

$$f(z) = \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}z^2}$$

A função  $f(z)$  é conhecida como **Gaussiana Padrão**.



# Resultados importantes sobre a gaussiana

## (2) - Intervalos de confiança

- Alguns intervalos de confiança (I.C.) correspondentes à integral da gaussiana padrão são muito conhecidos:

$$a) P(|z| \leq 1) = \int_{-1}^1 f(z) dz = 0,683 \rightarrow \text{I.C. de 68,3 \%}$$

$$b) P(|z| \leq 2) = \int_{-2}^2 f(z) dz = 0,954 \rightarrow \text{I.C. de 95,4 \%}$$

$$c) P(|z| \leq 3) = \int_{-3}^3 f(z) dz = 0,997 \rightarrow \text{I.C. de 99,7 \%}$$

Usualmente esses resultados são apresentados em tabelas de integrais da gaussiana padrão entre 0 e  $z$ :  $I(z) = \int_0^z f(z') dz'$

Obs: É comum haver uma função que corresponde a integral da gaussiana padrão entre  $-\infty$  e  $z$ :  $C(z) = \int_{-\infty}^z f(z') dz'$

# O Teorema Central do Limite

- A soma,  $S$ , de variáveis aleatórias independentes  $x_i$ , cada uma obedecendo à uma função densidade de probabilidade  $f_i(x)$  com valores médios verdadeiros  $\mu_i$  e variâncias  $\sigma_i^2$  finitas, tende a uma gaussiana quando o número de variáveis na somatória,  $N$ , tende ao infinito.
- A variável  $S$  terá média igual à soma das médias e a variância é igual à soma das variâncias:

$$S = \sum_{i=1}^N x_i$$

$$\mu_0 = \langle S \rangle = \sum_{i=1}^N \mu_i$$

$$\sigma_0^2 = \sum_{i=1}^N \sigma_i^2$$

# Outra forma de escrever o Teorema Central do Limite

- A variável soma normalizada:

$$S^* = \frac{S - \mu_0}{\sigma_0} = \frac{\sum_{i=1}^N x_i - \sum_{i=1}^N \mu_i}{\sqrt{\sum_{i=1}^N \sigma_i^2}}$$

$$S = \sum_{i=1}^N x_i$$
$$\mu_0 = \sum_{i=1}^N \mu_i$$
$$\sigma_0^2 = \sum_{i=1}^N \sigma_i^2$$

de  $N$  variáveis aleatórias independentes,  $x_i$ , que obedecem funções densidade de probabilidade  $f_i(x)$  com médias  $\mu_i$  e variâncias  $\sigma_i^2$  finitas tem a seguinte propriedade:

$$\lim_{N \rightarrow \infty} P(S^* \leq a) = \int_{-\infty}^a \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}z^2} dz$$

# No caso de variáveis discretas

$$\int \phi(x) f(x) dx \longrightarrow \sum_i \phi(X_i) F(X_i)$$

onde  $F(X)$  é a função de probabilidade de obter em uma medição o valor  $X$ .

- No caso de comparações com histogramas as expressões para variáveis discretas são usadas juntamente com a aproximação

$$F(X) \cong f(x) \Delta x$$

onde  $\Delta x$  é a largura de cada canal do histograma.