

Princípio de Máxima Verossimilhança

O **Princípio de Máxima Verossimilhança** é uma maneira de estimar parâmetros desconhecidos de um modelo a partir dos dados observados.

Escolher o valor do parâmetro que torna os dados que observamos mais plausíveis segundo o modelo.

Por exemplo: depois de lançar uma moeda várias vezes, qual probabilidade de sair cara melhor explica os resultados?

1. Exemplo: uma moeda possivelmente viciada

Suponha que lançamos uma moeda **10 vezes**, obtendo **7 caras e 3 coroas**.

Chamaremos de p a probabilidade desconhecida de sair cara. A probabilidade de sair coroa será $1 - p$.

Assumindo lançamentos independentes, a probabilidade da **sequência específica observada** é:

$$L(p) = \underbrace{p \times \cdots \times p}_{7 \text{ caras}} \times \underbrace{(1 - p) \times \cdots \times (1 - p)}_{3 \text{ coroas}}$$

Portanto:

$$L(p) = p^7(1 - p)^3$$

Essa é a **função de verossimilhança**. Mantemos os resultados fixos e experimentamos diferentes valores de p .

Valor candidato de p	Verossimilhança $L(p)$
0,20	0,00000655
0,50	0,00097656
0,70	0,00222357
0,90	0,00047830

Observe que $p = 0,90$ favorece bastante as caras, mas torna as **três coroas** pouco plausíveis. Precisamos explicar o conjunto dos resultados.

O ponto mais alto da curva ocorre em:

$$\hat{p} = \frac{7}{10} = 0,70$$

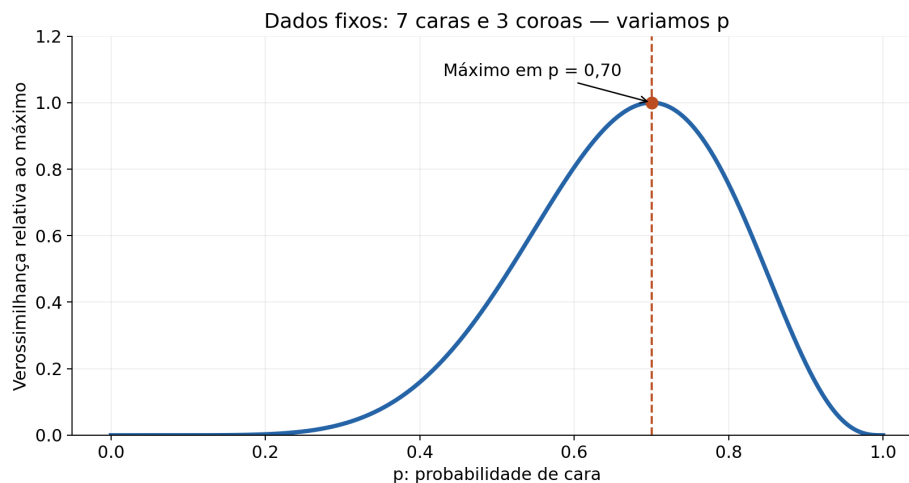


Figure 1: Curva de verossimilhança para sete caras e três coroas, com máximo em p igual a 0,70.

O símbolo $\hat{p}$, lido como “p chapéu”, indica uma **estimativa**.

Isso não prova que a moeda tenha exatamente 70% de probabilidade de cara. Significa que **0,70 é o valor que maximiza a verossimilhança desses dados**.

Observação: se considerarmos apenas a contagem “7 caras em 10 lançamentos”, sem uma ordem específica, usamos $\binom{10}{7}p^7(1-p)^3$. Esse fator constante não altera o ponto de máximo.

2. Probabilidade e verossimilhança: o que muda?

Usamos a mesma expressão matemática, mas mudamos o que está fixo e o que queremos investigar.

Perspectiva	O que está fixo?	Pergunta
Probabilidade	O parâmetro p	“Se $p = 0,50$, quais resultados podem ocorrer?”
Verossimilhança	Os dados observados	“Obtive 7 caras e 3 coroas: quais valores de p explicam melhor esses dados?”

Verossimilhança não é a probabilidade de o parâmetro ser verdadeiro. Por exemplo, $L(0,70)$ não representa a probabilidade de que $p = 0,70$.

No gráfico, a curva foi dividida por seu valor máximo para facilitar a comparação. A altura 1 significa “maior verossimilhança”, e não “100% de certeza”.

3. Exemplo: estimar a média de uma distribuição normal

Considere três medidas independentes:

8, 10, 12

Suponha que elas venham de uma distribuição normal com **média desconhecida** μ e **desvio padrão conhecido** $\sigma = 2$.

Vamos comparar três médias candidatas: 6, 10 e 14.

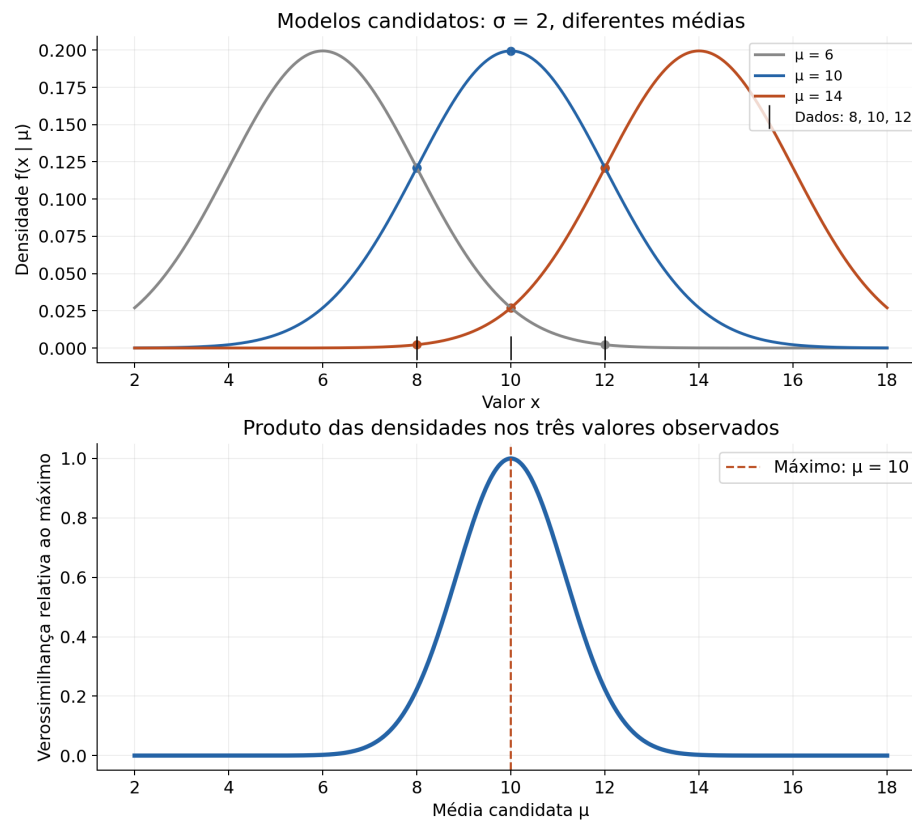


Figure 2: Distribuições normais com médias candidatas 6, 10 e 14, e curva de verossimilhança com máximo em média 10.

Na parte superior, cada curva representa um modelo candidato. Os pontos sobre as curvas mostram a **densidade atribuída a cada observação**.

Para avaliar o conjunto dos dados, multiplicamos essas densidades:

$$L(\mu) = f(8 \mid \mu) f(10 \mid \mu) f(12 \mid \mu)$$

A parte inferior mostra esse produto enquanto variamos a média candidata.

O máximo ocorre em:

$$\hat{\mu} = \bar{x} = \frac{8 + 10 + 12}{3} = 10$$

Para uma amostra de uma distribuição normal, a estimativa de máxima verossimilhança da média é a média aritmética dos dados.

Como estamos trabalhando com uma variável contínua, usamos **densidades**, e não probabilidades de valores exatos. Na distribuição normal, a probabilidade de observar exatamente um único valor é zero; probabilidades são calculadas para intervalos.

4. Como encontramos o máximo matematicamente?

Em geral, para observações independentes $x_1, \dots, x_n$, temos:

$$L(\theta) = \prod_{i=1}^n f(x_i \mid \theta)$$

Aqui, θ representa o parâmetro desconhecido. Queremos encontrar:

$$\hat{\theta} = \arg \max_{\theta} L(\theta)$$

“Arg max” significa **o valor do parâmetro no qual a função atinge seu máximo**.

É comum usar o logaritmo da verossimilhança:

$$\ell(\theta) = \log L(\theta) = \sum_{i=1}^n \log f(x_i \mid \theta)$$

Isso transforma produtos em somas. Como o logaritmo é crescente, **maximizar L ou $\log L$ produz o mesmo resultado**.

Para a moeda:

$$\ell(p) = 7 \log p + 3 \log(1 - p)$$

Derivando e igualando a zero:

$$\frac{7}{p} - \frac{3}{1-p} = 0 \implies 7(1-p) = 3p \implies \boxed{\hat{p} = 0,70}$$

5. Um exemplo simples em Python

Podemos encontrar aproximadamente o máximo testando muitos valores candidatos:

```
import numpy as np
import matplotlib.pyplot as plt

caras = 7
coroas = 3

# Valores candidatos para a probabilidade de cara
p = np.linspace(0, 1, 1001)

# Dados fixos; parâmetro p variável
L = p**caras * (1 - p)**coroas

# Índice do maior valor de verossimilhança
indice = np.argmax(L)
p_estimado = p[indice]

print(f"Estimativa de máxima verossimilhança: {p_estimado:.2f}")

plt.plot(p, L)
plt.axvline(p_estimado, color="red", linestyle="--")
plt.xlabel("Probabilidade de cara (p)")
plt.ylabel("Verossimilhança L(p)")
plt.title("7 caras e 3 coroas")
plt.grid(alpha=0.3)
plt.show()
```

Resultado:

Estimativa de máxima verossimilhança: 0.70

As fórmulas usam notação LaTeX e requerem um visualizador Markdown com suporte a matemática.