# Métodos de estimação

Prof. Paulo Justiniano Ribeiro Junior

Departamento de Estatística
Universidade Federal do Paraná

# Estimação estatística

1. Estimação pontual e intervalar.
2. Determinação do tamanho da amostra.
3. Propriedades dos estimadores.
4. **Métodos de estimação**.
   - Método dos momentos.
   - Método da máxima verossimilhança.


Figura 1. Foto de Kaique Rocha no Pexels.

# Probabilidades *versus* estimação

Tabela 1. Objeto das distribuições de probabilidades e da inferência estatística.

| | Distribuição de probabilidades | Inferência Estatística |
|---|---|---|
| 1 | Distribuição **conhecida**. | Distribuição **desconhecida**. |
| 2 | Parâmetros **conhecidos**. | Parâmetros **desconhecidos**. |
| 3 | Obter **probabilidades** para valores da v.a.. | Obter **estimativas dos parâmetros** usando dados observados. |

Estimação: Especificar a distribuição e estimar parâmetros a partir dos dados observados.

# Notação e definições (relembrando)

- $\mathbf{Y} = (Y_1, \ldots, Y_n)$: v.a.'s independentes e identicamente distribuídas.
- $Y_i \sim f(\boldsymbol{\theta})$ onde $f$ denota a função densidade de probabilidade ou função de probabilidade e $\boldsymbol{\theta} = (\theta_1, \ldots, \theta_p)$ é um vetor de $p$ parâmetros populacionais.
- $\mathbf{y} = (y_1, \ldots, y_n)$ denota o vetor de valores observados da v.a. $Y$.
- Uma estatística $T(\mathbf{Y})$ pode ser um **estimador** $\hat{\theta}$ de um parâmetro $\theta$ da população.
- A distribuição de probabilidade de $T(\mathbf{Y})$ é a **distribuição amostral**. Objeto de inferência (frequentista).

- **Objetivo:** Como obter estimadores? $\rightarrow$ Métodos de estimação.

# Como o **obter** um estimador?

- **Existem estimadores "óbvios".**
  - $\overline{y} \rightarrow \mu$.
  - $\hat{p} \rightarrow p$.
  - $S^2 \rightarrow \sigma^2$.
- Existem estimadores baseados em regras físicas, geométricas, etc.

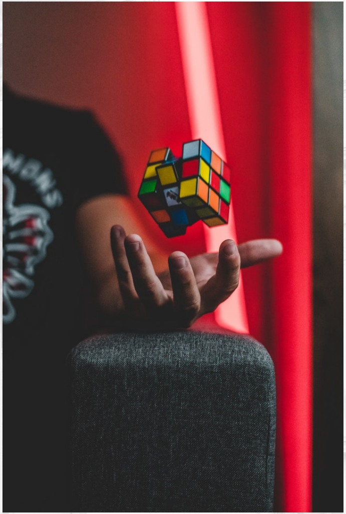


Figura 2. Foto de Marko Blazevic no Pexels.

# Como o **obter** um estimador?

- Existem estimadores "óbvios".
  - $\overline{y} \rightarrow \mu$.
  - $\hat{p} \rightarrow p$.
  - $S^2 \rightarrow \sigma^2$.
- **Existem estimadores baseados em regras físicas, geométricas, etc.**

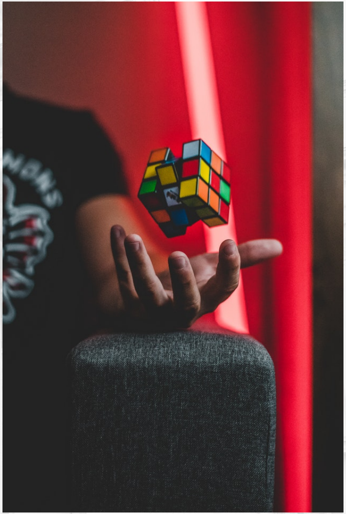


Figura 3. Foto de Marko Blazevic no Pexels.

# Exemplo: estimador do comprimento de raízes

A determinação exata do comprimento de raízes ($C$) de plantas é laborioso.

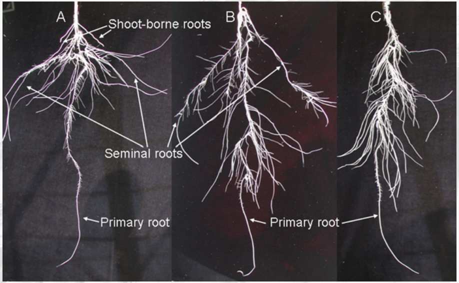



Figura 4. Extraído de Liao Chengsong no ResearchGate.

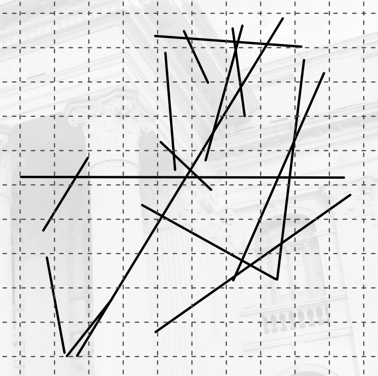


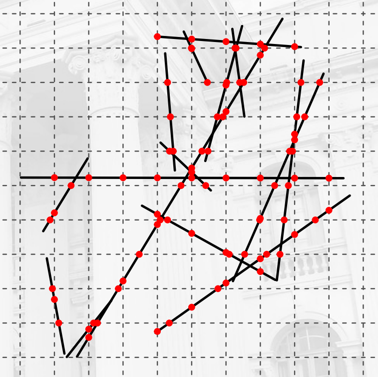
Figura 5. Exemplo da determinação de comprimento de raízes pelo método da intersecção.

# Um estimador para o comprimento das raízes

- Um **estimador** usado na prática é o seguinte

$$\hat{C} = \frac{\pi}{4} \cdot l \cdot n,$$

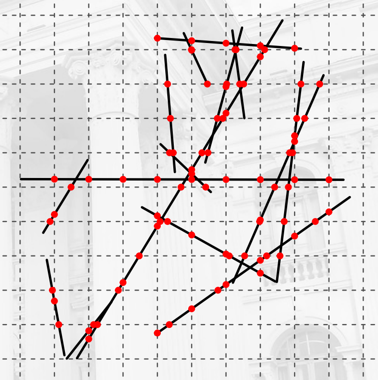
em que $n$ é o **número de interseções** das raízes (linhas sólidas) com a malha (pontos vermelhos), $l$ é a distância entre linhas da malha (linhas tracejadas).



- Para o exemplo didático ao lado, $n = 84$ e $l = 0.1$ que dá a **estimativa** $\hat{c} = 6.597$.
- O comprimento **real** é 7.072.

# Reconhecendo os componentes para a inferência

- População: as raízes.
- Parâmetro: comprimento total das raízes.
- Amostra: a forma como as raízes ficaram dispostas na malha.
- Estatística: o número de intersecções.
- Estimador: a fórmula $\hat{C} = \frac{\pi}{4} \cdot l \cdot n$
- Estimativa: o resultado de aplicar o estimador aos dados observados na amostra, no caso $\hat{c} = 6.597$.



# Como o **obter** um estimador?

- Existem estimadores “óbvios”.
  - $\overline{y} \rightarrow \mu$.
  - $\hat{p} \rightarrow p$.
  - $S^2 \rightarrow \sigma^2$.
- Existem estimadores baseados em regras físicas, geométricas, etc.
- **De forma mais geral, existem parâmetros que não possuem estimadores “imediatos”.**

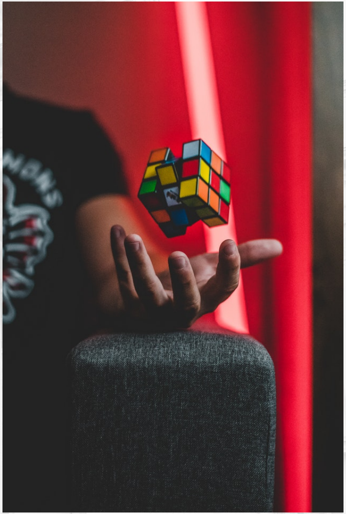


Figura 6. Foto de Marko Blazevic no Pexels.

# Método dos momentos

# Método dos momentos

- Método proposto por volta de 1887 por Pafnuty Chebyshev.
- Ideia básica é atribuída a Karl Pearson.
- **Método dos momentos**: igualar os **momentos da população**, que são definidos em termos de valores esperados, aos correspondentes **momentos da amostra**.
- Os momentos da população são funções de parâmetros desconhecidos.
- Solução da(s) equação(ões) são os estimadores dos parâmetros.

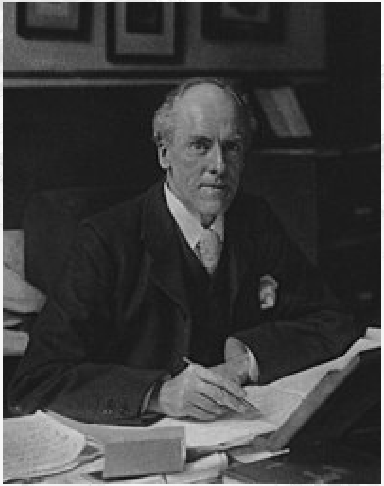


Figura 7. Karl Pearson. Retirado da Wikipedia.

# Momentos

Sejam $Y_1, Y_2, \ldots, Y_n$ v.a's com fdp ou fp $f(y; \theta)$.
O $k$-ésimo **momento da população** (ou **momento de distribuição**) é

$$
\begin{aligned}
\mathrm{E}(Y^k) &= \sum_{y \in R_y} y^k \cdot f(y; \theta) \quad \rightarrow \quad \text{v.a. discreta.} \\
&= \int_{y \in R_y} y^k \cdot f(y; \theta) \, \mathrm{d}y \quad \rightarrow \quad \text{v.a. contínua.}
\end{aligned}
$$

O correspondente $k$-ésimo **momento amostral** é

$$
m_k = \frac{1}{n} \sum_{i=1}^{n} y_i^k, \quad k \in \{1, 2, \ldots\}.
$$

# Estimador de momentos

- Sejam $Y_1, Y_2, \ldots, Y_n$ v.a.'s com fdp ou fp $f(y; \boldsymbol{\theta})$ com $p$ parâmetros $\boldsymbol{\theta} = (\theta_1, \ldots, \theta_p)$.
- Os estimadores $T_{\theta_1}, \ldots, T_{\theta_p}$ são encontrados igualando os primeiros $p$ momentos populacionais aos primeiros $p$ momentos amostrais.
- Tal procedimento resulta em um conjunto de equações que deve ser resolvido.

$$
\begin{cases}
E(Y) & = m_1 \\
E(Y^2) & = m_2 \\
E(Y^3) & = m_3 \\
\ldots
\end{cases}
$$

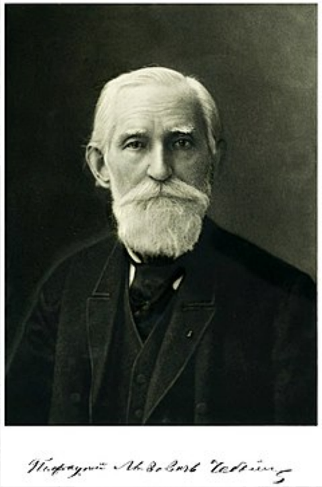


Figura 8. Pafnuty Lvovich Chebyshev. Retirado da Wikipedia.

# Exemplo: distribuição de Poisson e Exponencial

**Poisson**

$$P[Y = y] = \frac{e^{-\lambda}\lambda^{y}}{y!}$$

- Número de parâmetros: $p = 1$
- 1º momento populacional: $\mathrm{E}(Y^1) = \mu = \lambda$.
- 1º momento amostral: $m_1 = \overline{y}$.
- Dessa forma,

$$\hat{\lambda} = m_1 = \overline{Y}.$$

- A média amostral é o estimador de momentos do parâmetro $\lambda$ (média) da Poisson.

**Exponencial**

$$f(y) = \lambda \exp\{-\lambda y\}$$

- Número de parâmetros: $p = 1$
- 1º momento populacional: $\mathrm{E}(Y^1) = \mu = 1/\lambda$.
- 1º momento amostral: $m_1 = \overline{y}$.
- Dessa forma,

$$\hat{\lambda} = 1/m_1 = 1/\overline{Y}.$$

- O recíproco da média amostral é o estimador de momentos do parâmetro $\lambda$ (taxa) da Exponencial.

# Exemplo: distribuição Normal

## 1ª Equação

- Número de parâmetros: $p = 2$.
- 1º momento populacional: $\text{E}(Y^1) = \mu$.
- 1º momento amostral: $m_1 = \overline{y}$.
- Dessa forma,

$$\hat{\mu} = \overline{Y}.$$

## 2ª Equação

- 2º momento populacional: $\text{E}(Y^2) = \mu^2 + \sigma^2$.
- 2º momento amostral: $m_2 = (1/n)\sum_{i=1}^{n} y_i^2$.
- Dessa forma,

$$\mu^2 + \sigma^2 = \frac{1}{n}\sum_{i=1}^{n} Y_i^2,$$

e resolvendo em $\sigma^2$ usando o fato que $\hat{\mu} = \overline{Y}$,

$$\hat{\sigma}^2 = \frac{1}{n}\left[\sum_{i=1}^{n} Y_i^2 - n\left(\frac{1}{n}\sum_{i=1}^{n} Y_i\right)^2\right] = \frac{\sum_{i=1}^{n}(Y_i - \overline{Y})^2}{n}.$$

# Exemplo: distribuição Gama

$$f(y) = \frac{\lambda^r}{\Gamma(r)} y^{r-1} \exp\{-\lambda y\}$$

- Suponha que $Y_1, Y_2, \ldots, Y_n$ v.a's com distribuição Gama de parâmetros $r$ e $\lambda$.
- Os $p = 2$ primeiros momentos populacionais da Gama são

$$\mathrm{E}(Y) = \frac{r}{\lambda}$$
$$\mathrm{E}(Y^2) = \frac{r(r+1)}{\lambda^2}.$$

- Os estimadores de momentos são encontrados resolvendo o sistema de equações

$$\begin{cases} E(Y) & = m_1 \\ E(Y^2) & = m_2 \end{cases} \longrightarrow \begin{cases} \frac{r}{\lambda} & = \overline{y} \\ \frac{r(r+1)}{\lambda^2} & = \frac{1}{n}\sum_{i=1}^{n} y_i^2 \end{cases}$$

- Com a solução do sistema, os estimadores são

$$\hat{r} = \frac{\overline{Y}^2}{(1/n)\sum_{i=1}^{n} Y_i^2 - \overline{Y}^2} \quad \text{e} \quad \hat{\lambda} = \frac{\overline{Y}}{(1/n)\sum_{i=1}^{n} Y_i^2 - \overline{Y}^2}$$

# Exemplo: tempo de atendimento

Acredita-se que o tempo de atendimento seja uma distribuição adequada para descrever o tempo de atendimento de clientes no caixa de um supermercado. Uma amostra aleatória de $n = 20$ atendimentos foi obtida. Os tempos são os seguintes.

| | | | | |
|---|---|---|---|---|
| 2.34 | 4.03 | 4.85 | 5.68 | 7.41 |
| 2.39 | 4.04 | 4.90 | 5.76 | 7.48 |
| 2.96 | 4.17 | 5.20 | 5.88 | 7.64 |
| 3.30 | 4.61 | 5.24 | 6.62 | 8.62 |

Estime os parâmetros da distribuição Gama pelo método dos momentos.

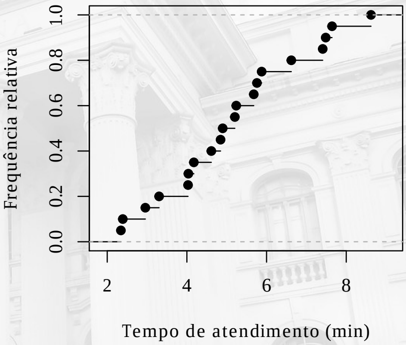



Figura 9. Gráfico de distribuição acumulada empírica com os dados de tempo de atendimento no caixa.

# Solução

Aplicando as expressões, obtém-se

- $\hat{r} = 8.9$.
- $\hat{\lambda} = 1.73$.

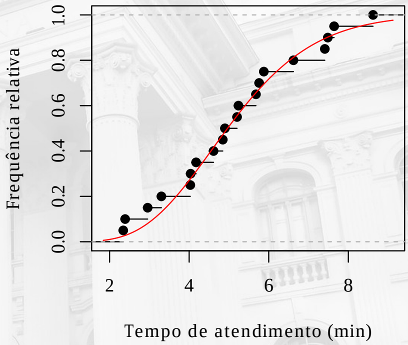



Figura 10. A curva da função de distribuição da Gama como parâmetros estimados sobreposta ao gráfico da distribuição empírica.

# Recomendações e limitações

## Vantagens

- Concepção simples e intuitiva.
- Fácil de obter (desde que os momentos populacionais estejam disponíveis).
- Em geral, oferece estimadores consistentes.
- Suposições distribucionais não são essenciais.
- Pode ser usado como guia inicial para outros métodos.
- É a base do método dos momentos generalizados.

## Desvantagens

- Difícil de expressar a incerteza associada às estimativas.
- Difícil de generalizar para modelos e/ou estruturas complexas de dados.
- Em geral, não viés não é garantido.
- Eficiência é difícil de medir e não é garantida mesmo para grandes amostras.
- Pode resultar em estimativas fora do espaço paramétrico.
- Precisa que os momentos populacionais sejam passíveis de calcular.

# Método de máxima verossimilhança

# Método de máxima verossimilhança

- Proposto por Ronald Fisher em 1922.
- É o método mais popular em estatística aplicada.
- **Ideia geral**: Encontrar o conjunto de valores para os parâmetros $\theta$ de uma distribuição de probabilidade $f(\mathbf{y}; \theta)$ que maximize a "chance" de observar a amostra de fato observada.

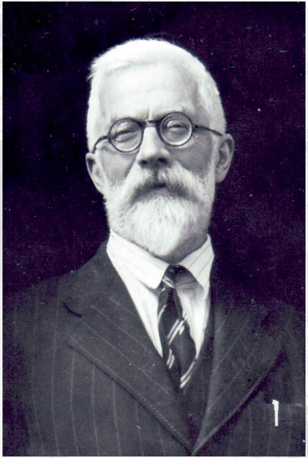


Figura 11. Ronald Fisher. Retirado do Google imagens.

# Função de verossimilhança

- **Configuração**: Sejam dados $y$ uma realização de um v.a. Y com fp ou fdp $f(\mathbf{y}; \theta)$.
- Função de **verossimilhança**

$$L(\theta) \equiv f(\mathbf{y}; \theta),$$

em que $f(\mathbf{y}; \theta)$ é a função de **distribuição conjunta** de Y.
- Supondo que as observações são independentes

$$L(\theta) \equiv \prod_{i=1}^{n} f(y_i; \theta).$$

- Notação para enfatizar que a verossimilhança é com o $\mathbf{y}$ já observado

$$L(\theta|\mathbf{y}).$$

# Estimação de população por captura e recaptura

Biólogos foram a campo, capturaram e marcaram 21 macacos ($m$) em uma reserva ecológica. Após 1 mês, eles retornaram e fizeram outra captura igual à primeira. Dos 49 ($r$) macacos capturados, 5 ($y$) apresentam a marca. Qual o tamanho da população de macacos ($m + n$)?

$$m = 21, \quad r = 49, \quad y = 5, \quad n = ?$$

Parâmetro desconhecido: $\theta = n$
E se $n$ fosse... qual a probabilidade de observar este resultado $y = 5$?

Figura 12. Foto de Pexels.

# Exemplo da estimação da população de macacos (cont)

$$L(n|y) = \frac{\binom{m}{y} \cdot \binom{n}{r-y}}{\binom{m+n}{r}}$$
$$= \frac{\binom{21}{5} \cdot \binom{n}{49-5}}{\binom{21+n}{49}}$$

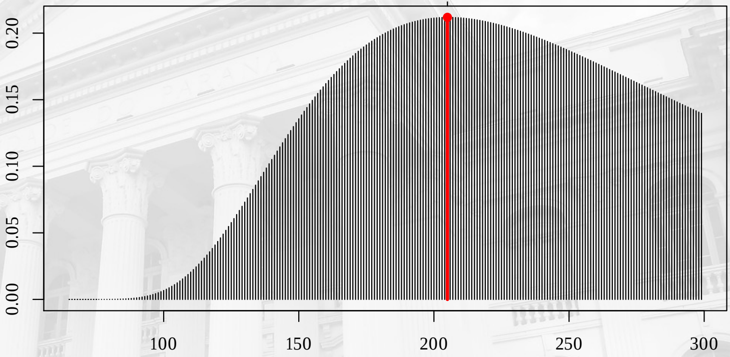



Figura 13. Função de verossimilhança para o problema de estimação do tamanho da população para o problema de captura e recaptura dos macacos.

# Função de Probabilidade *versus* de verossimilhança

## Interpretações da função conforme o argumento em uso

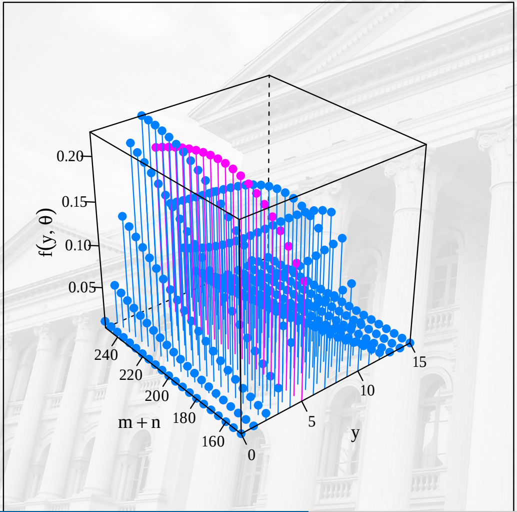



Figura 14. Função de verossimilhança com o eixo para $y$ (função de probabilidade) e um eixo para $m + n$ (função de verossimilhança).

# Função de log-verossimilhança e escore

- A função de **log-verossimilhança** é

$$l(\boldsymbol{\theta}; \mathbf{y}) = \ln \left( L(\boldsymbol{\theta}; \mathbf{y}) \right).$$

- No caso de observações independentes, tem-se

$$l(\boldsymbol{\theta}; \mathbf{y}) = \sum_{i=1}^{n} \ln \left( L(\boldsymbol{\theta}; y_i) \right).$$

- **Função escore**: Caso de observações independentes

$$U(\boldsymbol{\theta}; \mathbf{Y}) = \sum_{i=1}^{n} \frac{\partial}{\partial \theta_j} \ln f(\boldsymbol{\theta}, Y_i), \quad \text{para} \quad j \in \{1, \ldots, p\}.$$

# Estimativa e estimador de máxima verossimilhança

- **Estimativa de máxima verossimilhança**: O valor

$$\hat{\theta} = \hat{\theta}(y)$$

é a estimativa de máxima verossimilhança para $\theta$ se $L(\hat{\theta}) \geq L(\theta), \forall \theta$.

- **Estimador de máxima verossimilhança**: Se $\hat{\theta}(y)$ é a estimativa de máxima verossimilhança, então

$$\hat{\theta}(Y)$$

é o estimador de máxima verossimilhança (EMV).

# Exemplo: distribuição de Poisson

- Se $Y_i \sim P(\lambda)$, então a fp

$$f(y;\lambda) = \frac{\lambda^y \exp\{-\lambda\}}{y!}.$$

- Assumindo observações independentes, a verosssimilhança

$$L(\lambda) = \prod_{i=1}^{n} \frac{\lambda^{y_i} \exp\{-\lambda\}}{y_i!}.$$

- E, dessa forma, a função de log-verossimilhança

$$l(\lambda) = \ln(\lambda) \sum_{i=1}^{n} y_i - n\lambda - \sum_{i=1}^{n} \ln(y_i!).$$

# Exemplo: distribuição de Poisson (cont.)

- Função escore (derivada de $l(\lambda)$ em relação a $\lambda$)

$$U(\lambda) = \frac{1}{\lambda} \sum_{i=1}^{n} y_i - n.$$

- Resolvendo em $\lambda$, temos

$$\hat{\lambda} = \sum_{i=1}^{n} \frac{y_i}{n}.$$

- O estimador de máxima verossimilhança do parâmetro $\lambda$ da distribuição Poisson é a média amostral.

# Exemplo: distribuição exponencial

Se $Y_i \sim \text{Exp}(\lambda)$ e a amostra é iid, então

$$
\begin{aligned}
f(y;\lambda) &= \lambda \exp\{-\lambda y\} \\
L(\lambda) &= \prod_{i=1}^{n} \lambda \exp\{-\lambda y_i\} \\
l(\lambda) &= n \ln(\lambda) - \lambda \sum_{i=1}^{n} y_i \\
U(\lambda) &= \frac{n}{\lambda} - \sum_{i=1}^{n} y_i = 0 \\
\hat{\lambda} &= \frac{n}{\sum_{i=1}^{n} y_i} = \frac{1}{\bar{y}}.
\end{aligned}
$$

O estimador de máxima verossimilhança do parâmetro $\lambda$ é recíproco da média amostral.

# Exemplo: Distribuição Normal

- Se $Y_i \sim \mathrm{N}(\mu, \sigma^2)$, então a fdp

$$f(y; \boldsymbol{\theta} = (\mu, \sigma^2)) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left\{-\frac{(y-\mu)^2}{2\sigma^2}\right\}.$$

- Assumindo observações independentes, a verosssimilhança

$$L(\boldsymbol{\theta}) = \prod_{i=1}^{n} \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left\{-\frac{1}{2\sigma^2}(y_i-\mu)^2\right\}.$$

- Log-verossimilhança

$$l(\boldsymbol{\theta}) = -\frac{n}{2}\ln(2\pi) - \frac{n}{2}\ln(\sigma^2) - \frac{1}{2\sigma^2}\sum_{i=1}^{n}(y_i-\mu)^2.$$

# Exemplo: distribuição Normal (cont.)

- Função escore (derivada de $l(\theta)$ em relação a $\mu$)

$$U_\mu(\boldsymbol{\theta}) = \frac{1}{\sigma^2}\sum_{i=1}^{n}(y_i - \mu), \quad \text{que resolvendo} \quad \hat{\mu} = \frac{1}{n}\sum_{i=1}^{n} y_i.$$

- Função escore (derivada de $l(\theta)$ em relação a $\sigma^2$)

$$U_{\sigma^2}(\boldsymbol{\theta}) = -\frac{n}{2\sigma^2} + \frac{1}{2(\sigma^2)^2}\sum_{i=1}^{n}(y_i - \mu)^2, \quad \text{que resolvendo} \quad \hat{\sigma}^2 = \frac{1}{n}\sum_{i=1}^{n}(y_i - \mu)^2.$$

# Exemplo: distribuição Gama

- Se $Y_i \sim \text{Gama}(r, \lambda)$, então a fdp

$$f(y; \boldsymbol{\theta} = (r, \lambda)) = \frac{\lambda^r}{\Gamma(r)} y^{r-1} \exp\{-\lambda y\}.$$

- Assumindo observações independentes, a verosssimilhança

$$L(\boldsymbol{\theta}) = \prod_{i=1}^{n} \frac{\lambda^r}{\Gamma(r)} y_i^{r-1} \exp\{-\lambda y_i\}.$$

- Log-verossimilhança

$$l(\boldsymbol{\theta}) = nr\ln(\lambda) - n\ln(\Gamma(r)) - \lambda \sum_{i=1}^{n} y_i + (r-1)\sum_{i=1}^{n} \ln(y_i).$$

# Exemplo: distribuição Gama (cont.)

- Função escore (derivada de $l(\theta)$ em relação a $r$)

$$U_r(\theta) = n\ln(\lambda) - \frac{n}{\Gamma'(r)} + r\sum_{i=1}^{n}\ln(y_i) \quad \rightarrow \quad \text{Não tem solução analítica.}$$

- Função escore (derivada de $l(\theta)$ em relação a $\lambda$)

$$U_\lambda(\theta) = \frac{nr}{\lambda} - \sum_{i=1}^{n} y_i, \quad \text{que resolvendo} \quad \hat{\lambda} = \frac{nr}{\sum_{i=1}^{n} y_i}.$$

- Conclusão: não se tem **expressão analítica** para os estimadores de máxima verossimilhança da Gama.
  **Métodos numéricos** devem ser utilizados.

# Exemplo: tempo de atendimento (cont.)

Acredita-se que o tempo de atendimento seja uma distribuição adequada para descrever o tempo de atendimento de clientes no caixa de um supermercado. Uma amostra aleatória de $n = 20$ atendimentos foi obtida. Os tempos são os seguintes:

| | | | | |
|---|---|---|---|---|
| 2.34 | 4.03 | 4.85 | 5.68 | 7.41 |
| 2.39 | 4.04 | 4.90 | 5.76 | 7.48 |
| 2.96 | 4.17 | 5.20 | 5.88 | 7.64 |
| 3.30 | 4.61 | 5.24 | 6.62 | 8.62 |

Estime os parâmetros da distribuição Gama pelo método da máxima verossimilhança (use *software*).

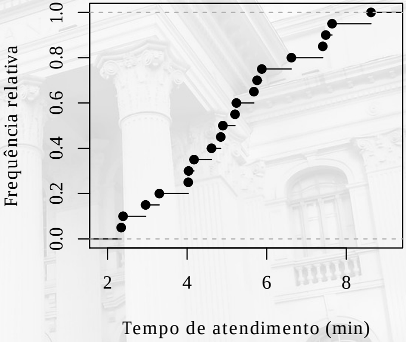



Figura 15. Gráfico de distribuição acumulada empírica com os dados de tempo de atendimento no caixa.

# Solução

Usando um algorítio numérico ou *software* estatístico, obtém-se

- $\hat{r} = 8.288015$.
- $\hat{\lambda} = 1.61$.

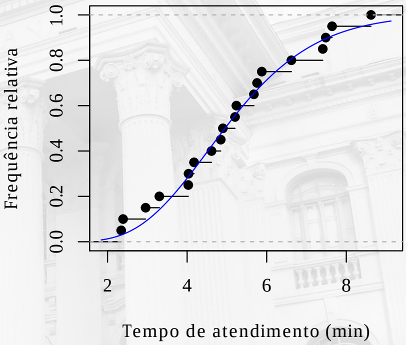



Figura 16. A curva da função de distribuição da Gama como parâmetros estimados sobreposta ao gráfico da distribuição empírica.

# Recomendações e limitações

**Vantagens**

- Concepção intuitiva.
- Propriedades assintóticas desejáveis: não-viés e eficiência.
- Estimadores consistentes.
- Metodologia completa para estimação e inferência (IC e TH).
- É o método de estimação mais popular em estatística.

**Desvantagens**

- Pode ser difícil de obter em termos práticos.
- De forma geral, requer métodos numéricos.
- Suposição explícita de uma distribuição de probabilidade.

# Inferência completa com a função de verossimilhança

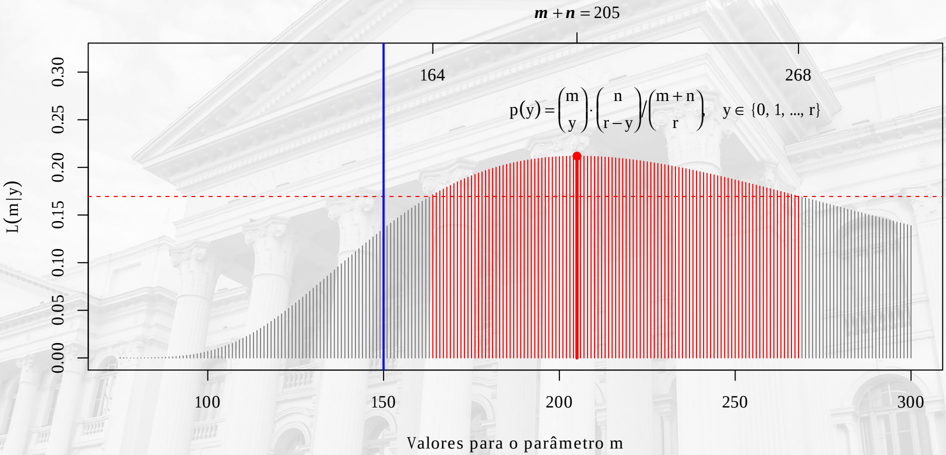



Figura 17. Função de verossimilhança usada para obter intervalo de confiança.

# Considerações finais

# Comentário finais sobre estimação estatística

- Estimação de parâmetros emprega álgebra, cálculo e métodos numéricos.
- No entanto, os métodos são conceitualmente fáceis de compreender.
  - Momentos: igualar momentos e resolver.
  - Máxima verossimilhança: maximizar a chance de observar a amostra.
- Os estimadores já foram determinados para os principais parâmetros e distribuições .
- Existem ainda **outros** métodos de estimação.
  - Método de mínimos quadrados.
  - Método da Inferência Bayesiana.
  - Métodos de estimação robustos.
  - Equações de estimação generalizadas.
  - Etc.
- Implementados em *softwares* **estatísticos**.