Propriedades dos estimadores

Prof. Paulo Justiniano R. Jr

Departamento de Estatística
Universidade Federal do Paraná

Estimação estatística

Falar sobre **população** $Y \sim \text{Dist.}_y(\theta)$ a partir da observação da **amostra** $\hat{\theta}(y_1, \ldots, y_n) \sim \text{Dist.Am.}_{\hat{\theta}}(\theta)$.

1. Como expressar incerteza?
   Estimação pontual e intervalar.
2. Amostra? De qual tamanho?
   Determinação do tamanho da amostra.
3. O que é "estimar bem"?
   Propriedades dos estimadores.
4. Como estimar?
   Métodos de estimação.

Figura 1. Distribuição amostral de diferentes estimadores de um parâmetro.

Estimação estatística

1. Estimação pontual e intervalar.
2. Determinação do tamanho da amostra.
3. **Propriedades dos estimadores**.
    - Vício (ou não tendenciosidade).
    - Variância.
    - Erro quadrático médio.
    - Consistência.
4. Métodos de estimação.

Figura 2. Foto de Karolina Grabowska no Pexels.

Notação e definições

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 $\boldsymbol{\theta}$ da população.
- A distribuição de probabilidade de $T(\mathbf{Y})$ é a **distribuição amostral**. Objeto de inferência (frequentista).

- **Objetivo:** O que caracteriza **bons** estimadores? $\rightarrow$ Propriedades dos estimadores.

Questões

- O que torna um estimador "bom" em termos práticos?
- Existe "erro" na estimação? Como medir?
- Quais as propriedades desejáveis de um estimador?
- Como comparar dois (ou mais) estimadores?

Figura 3. Foto de cottonbro no Pexels.

Vício de um estimador

Estimadores não viciados

- Um estimador deve fornecer **valores próximos** do valor verdadeiro do parâmetro que está sendo estimado.
- Um estimador é **não viciado** para θ se o valor esperado de $\hat{\theta}$ for igual a θ.
- Isso quer dizer que a **média da distribuição amostral** de $\hat{\theta}$ é θ.
- Em certos casos, é possivel determinar o vício de um estimador de forma **analítica**.
- Em situações mais complexas, pode-se determinar de forma **computacional**, por meio de simulação.

Figura 4. Foto de icono.com no Pexels.

Definição

O estimador pontual $\hat{\theta} = T_\theta(y_1, y_2, \ldots, y_n)$ é um estimador **não viciado** para o parâmetro θ se

$$\mathrm{E}(\hat{\theta}) = \theta.$$

Se o estimador for viciado ou tendencioso, então a diferença

$$\mathrm{B}(\hat{\theta}) = \mathrm{E}(\hat{\theta}) - \theta$$

é chamada de **vício** (*bias*) do estimador $\hat{\theta}$.

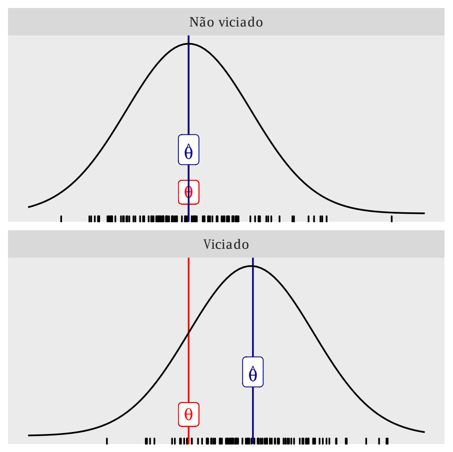

Figura 5. Exemplo de estimador viciado e não viciado.

Exemplo: vício da variância amostral (analítico)

$$
\begin{aligned}
\mathrm{E}(S^2) &= \mathrm{E}\left(\frac{\sum_{i=1}^{n}(Y_i-\overline{Y})^2}{n-1}\right) = \frac{1}{n-1}\cdot \mathrm{E}\left(\sum_{i=1}^{n}(Y_i-\overline{Y})^2\right)\\
&= \frac{1}{n-1}\cdot \mathrm{E}\left(\sum_{i=1}^{n}(Y_i^2-2Y_i\overline{Y}+\overline{Y}^2)\right) = \frac{1}{n-1}\cdot \mathrm{E}\left(\sum_{i=1}^{n}Y_i^2-n\overline{Y}^2\right)\\
&= \frac{1}{n-1}\left(\sum_{i=1}^{n}\mathrm{E}(Y_i^2)-n\mathrm{E}(\overline{Y}^2)\right)\\
&= \frac{1}{n-1}\left(\sum_{i=1}^{n}(\mu^2+\sigma^2)-n(\mu^2+\sigma^2/n)\right)\\
&= \frac{1}{n-1}\left(n\mu^2+n\sigma^2-n\mu^2-\sigma^2\right) = \frac{(n-1)\sigma^2}{n-1} = \sigma^2.
\end{aligned}
$$

Sobre os estimadores da variância

- Em Estatística Descritiva, apresentou-se duas **expressões para a variância**

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

- O denominador $n - 1$ na expressão da variância amostral (S^2) é o que o torna um **estimador não viciado** para σ^2.
- Portanto, $\hat{\sigma}^2$ é um estimador viciado.
- Neste caso, o vício depende do tamanho da amostra

$$\mathrm{B}(\hat{\sigma}^2) = \frac{n-1}{n} \cdot \sigma^2 - \sigma^2 = -\frac{1}{n} \cdot \sigma^2.$$

- Pode não ser possível determinar o vício do estimador de forma **analítica**.

Exemplo: vício de estimadores da variância (computacional)

Também é possível determinar o vício de estimadores por **simulação**.
Neste caso, procedeu-se da seguinte forma:

1. Extraiu-se $k = 100$ amostras de tamanho $n = 6$ de uma população com variância conhecida $\sigma^2 = 1$.
2. Para cada amostra $i \in \{1, \ldots, k\}$, determinou-se a estimativa da variância usando: S^2 e $\hat{\sigma}^2$.
3. Obteve-se a média das estimativas ao longo das k amostras que é o equivalente computacional da esperança do estimador.

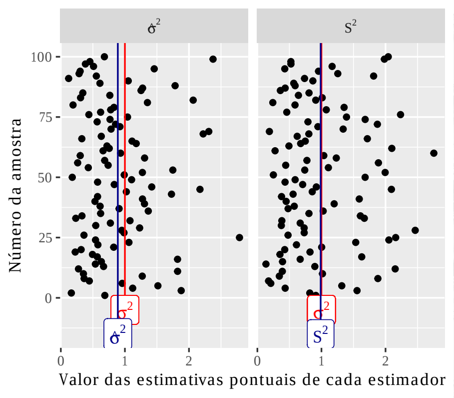

Figura 6. Simulação computacional para o vício dos estimadores da variância.

Exemplo: estimador da média com mediana e média aparada

- A distribuição Normal é simétrica, então média e mediana coincidem.
- A mediana e a média aparada são menos influenciadas por valores extremos.
- Seriam então bons estimadores?
- Qual o vício de cada um destes estimadores?
- Ao lado, simulação computacional com $k = 500$ amostras aleatórias de tamanho $n = 20$.

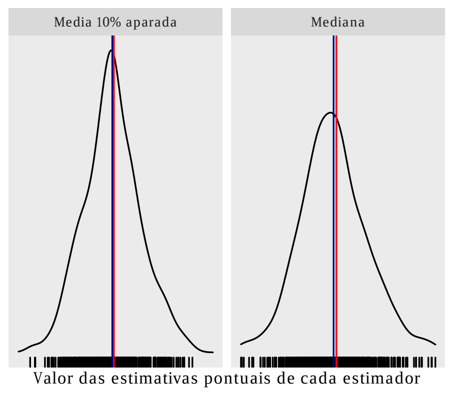

Figura 7. Simulação computacional para o vício dos estimadores da média: mediana e média 10% aparada.

Variância de um estimador

Variância de um estimador

- Sejam $\hat{\theta}_1$ e $\hat{\theta}_2$ estimadores não viciados de θ.
- Então, $E(\hat{\theta}_1) = E(\hat{\theta}_2) = \theta$.
- No entanto, as **variâncias** destas distribuições amostrais podem ser diferentes.
- É razoável escolher o estimador que apresente a **menor variância**.

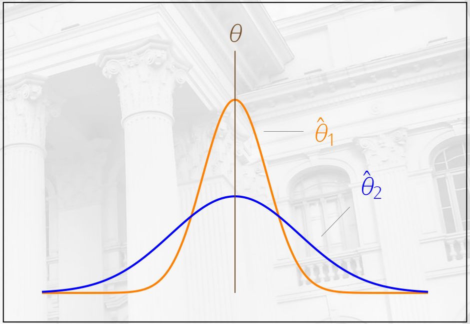

Figura 8. Distribuição amostral de dois estimadores não viciados.

Exemplo: estimador da média com mediana e média aparada

- Na simulação computacional, $Y \sim \mathrm{N}(\mu = 0, \sigma^2 = 1)$ e amostras de tamanho $n = 20$.
- Os estimadores **média 10% aparada** ($\hat{\theta}_1$) e **mediana** ($\hat{\theta}_2$) para μ tem variâncias

$$\mathrm{V}(\hat{\theta}_1) = 0.0477 \quad \text{e} \quad \mathrm{V}(\hat{\theta}_2) = 0.0673.$$

- Na simulação obteve-se $\mathrm{V}(\overline{y}) = 0.0465$.
- A variância teórica da média amostral é

$$\mathrm{V}(\overline{Y}) = \frac{\sigma^2}{n} = \frac{1}{20} = 0.05.$$

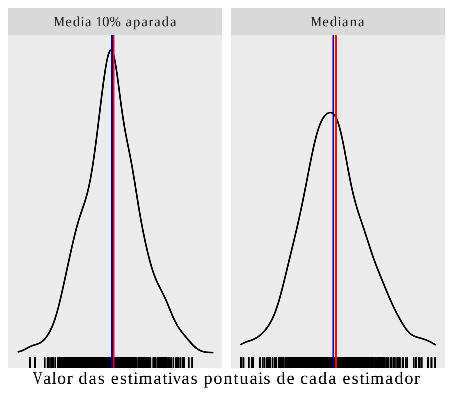

Figura 9. Simulação computacional para o vício dos estimadores da média: mediana e média 10% aparada.

Exemplo: estimadores da média para v.a. Uniforme Contínua

- Nem sempre o estimador "óbvio" é o melhor estimador.
- O melhor estimador pode depender da distribuição da v.a.
- Estimadores concorrentes:
  - Média amostral

$$\hat{\theta}_1 = \overline{Y} = \frac{1}{n}\sum_{i=1}^{n} Y_i.$$

  - Valor médio

$$\hat{\theta}_2 = \frac{Y_{(1)} + Y_{(n)}}{2}.$$

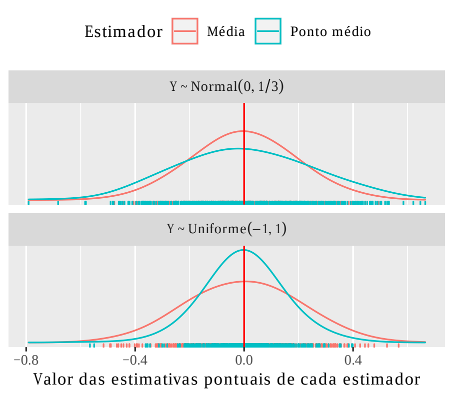

Figura 10. Simulação computacional para a variância de estimadores da média.

Erro-padrão

É o desvio-padrão da distribuição amostral do estimador.

- O **erro-padrão** de um estimador é a raiz quadrada da variância do estimador.

$$\text{v.a.: } Y \leftrightarrow \text{Variância: } \mathrm{V}(Y) \leftrightarrow \text{Desvio-padrão: } \sqrt{\mathrm{V}(Y)}$$

$$\text{Estimador: } \hat{\theta} \leftrightarrow \text{Variância: } \mathrm{V}(\hat{\theta}) \leftrightarrow \text{Erro-padrão: } \sqrt{\mathrm{V}(\hat{\theta})}$$

- É frequente reportado acompanhando estimativas pontuais para representar sua incerteza.
- Como visto, quando $\hat{\theta} \overset{\text{aprox}}{\sim}$ Normal, o erro-padrão multiplicado por um fator (por exemplo ± 2) define o intervalo de confiança.

Table 1: Dependent Variable: Enrollment

Variable	Model 1	Model 2	Model 3
frl	-0.012	-0.052	-0.156**
	(0.04)	(0.06)	(0.06)
minority		0.034	0.097*
		(0.05)	(0.04)
disting			0.439***
			(0.06)
cons	0.529***	0.530***	0.518***
	(0.03)	(0.03)	(0.02)
N	196	196	196
R^2	0.001	0.003	0.235

$^{***}p < 0.01$, $^{**}p < 0.05$, $^{*}p < 0.1$

Figura 11. Tabela reportando estimativas e seus erros-padrões. Extraído de stackoverflow.com.

Combinando propriedades

Estimador não viciado de variância mínima (ENVVM)

Se considerarmos todos os estimadores não viciados de θ, aquele com a menor variância será chamado de **estimador não viciado de variância mínima** (ENVVM).

A média amostral

Se $Y_1, Y_2, \ldots, Y_n$ for uma amostra aleatória de tamanho n, proveniente de uma v.a. aleatória de distribuição Normal, então a média amostral $\overline{Y} = \sum_{i=1}^{n} Y_i/n$ é um ENVVM para μ.

Erro quadrático médio

Erro quadrático médio

- Nem sempre se dispõe de estimadores não viciados.
- Há situações em que estimadores viciados têm distribuição amostral com menor variância.
- Como escolher o estimador neste caso conciliando ambos aspectos, vício e variância?

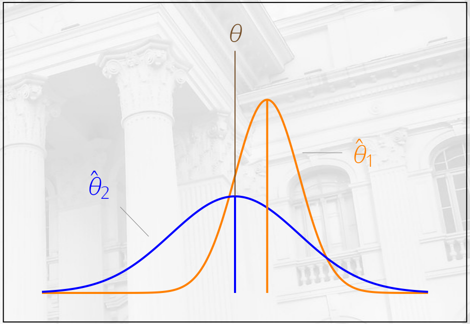

Figura 12. Distribuição amostral de dois estimadores.

Decomposição em vício e variância

- O **erro quadrático médio** (EQM) é uma medida que concilia vício e variância.
- O EQM de um estimador $\hat{\theta}$ do parâmetro θ é definido como

$$\mathrm{EQM}(\hat{\theta}) = \mathrm{E}(\hat{\theta} - \theta)^2.$$

- Ele pode ser reescrito como função da **variância** e **vício**

$$\begin{aligned} \mathrm{EQM}(\hat{\theta}) &= \mathrm{E}[\hat{\theta} - \mathrm{E}(\hat{\theta})]^2 + [\mathrm{E}(\hat{\theta}) - \theta]^2 \\ &= \mathrm{V}(\hat{\theta}) + \mathrm{B}^2(\hat{\theta}). \end{aligned}$$

- Portanto, o EQM de um estimador não viciado é a própria variância.

Analogia do tiro ao alvo

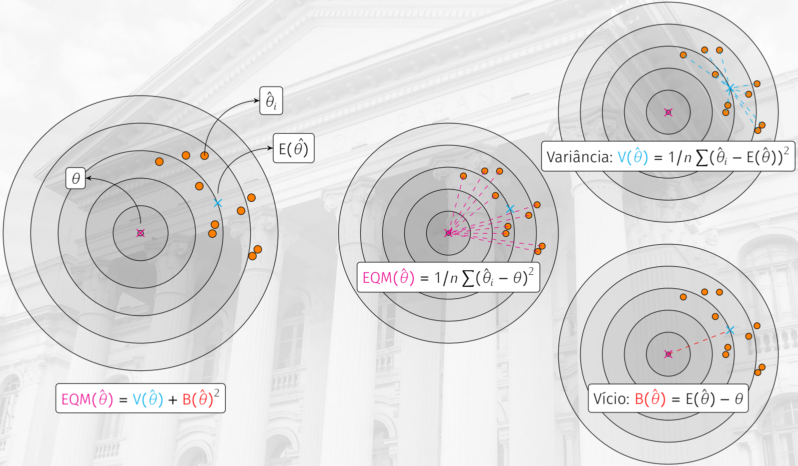

Figura 13. Analogia do tiro ao alvo para o erro quadrático médio e sua decomposição.

Ilustração da decomposição do erro quadrático médio

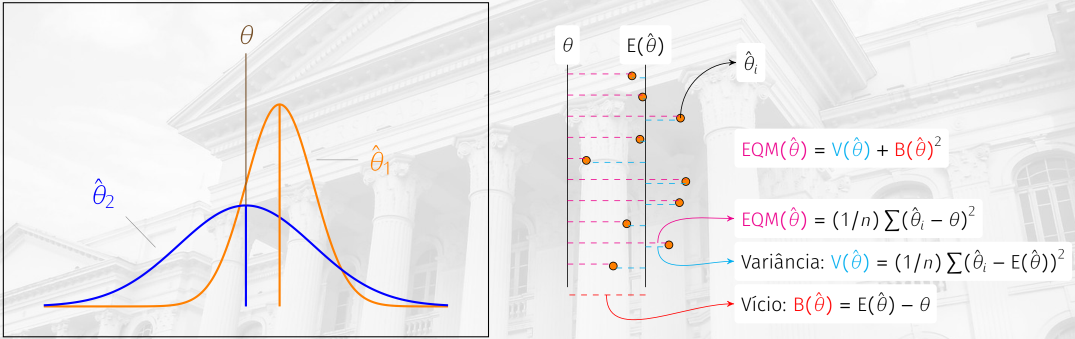

Figura 14. A decomposição do erro quadrático médio.

Eficiência relativa de um estimador

- O **erro quadrático médio** é uma métrica importante para comparar estimadores.
- Ele é usado para definir a **eficiência relativa** de um estimador comparado a outro,

$$\text{Efr}(\hat{\theta}_1, \hat{\theta}_2) = \frac{\text{EQM}(\hat{\theta}_1)}{\text{EQM}(\hat{\theta}_2)}.$$

- Se a $\text{Efr}(\hat{\theta}_1, \hat{\theta}_2) < 1$, conclui-se que $\hat{\theta}_1$ é um estimador superior a $\hat{\theta}_2$ e vice-versa.

Consistência de um estimador

Consistência de um estimador

- Não viés é uma propriedade desejável.
- Pode ser restrita em situações mais gerais.
- O viés de um estimador pode "sumir" quando a amostra aumenta de tamanho.
- Consistência é uma propriedade mais geral.
- Verifica o que acontece com o estimador quando a amostra aumenta de tamanho.

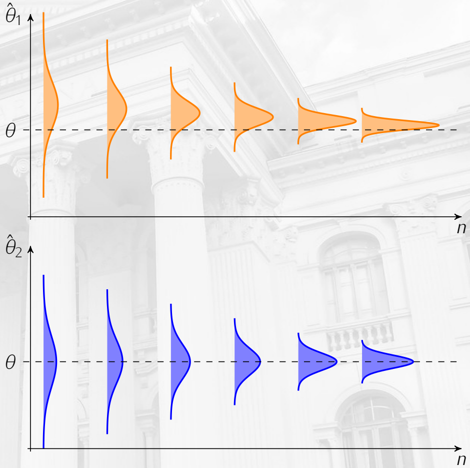

Figura 15. Consistência para dois estimadores.

Consistência de um estimador

- Verificar a consistência de um estimador não é trivial.
- Precisam da idéia de **convergência** de v.a.
- Um estimador é **consistente em probabilidade** se

$$P(|\hat{\theta} - \theta| \geq \epsilon) \rightarrow 0 \quad \text{quando} \quad n \rightarrow \infty.$$

- Um estimador é **erro quadrático médio consistente** se

$$E(\hat{\theta} - \theta)^2 \rightarrow 0 \quad \text{quando} \quad n \rightarrow \infty.$$

- Para consistência em probabilidade, a **Desigualdade de Chebyshev** permite dizer que

$$V(\hat{\theta}) \rightarrow 0, \quad \text{para} \quad n \rightarrow \infty,$$

então $\hat{\theta}$ é consistente em probabilidade para θ.
- Existem outras formas de consistência $\rightarrow$ *Fisher consistency*.
- Mais detalhes estão fora do escopo deste curso.

Consistência do estimador $\hat{\sigma}^2$ da variância

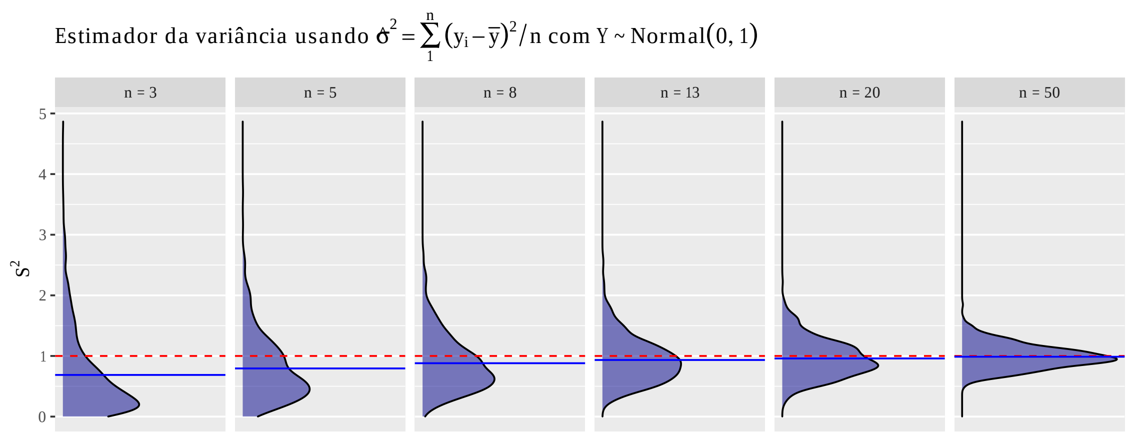

Figura 16. Ilustração por simulação computacional da consistência para o estimador $\hat{\sigma}^2$ da variância.

Inconsistência do estimador $\tilde{\sigma}$ do desvio-padrão

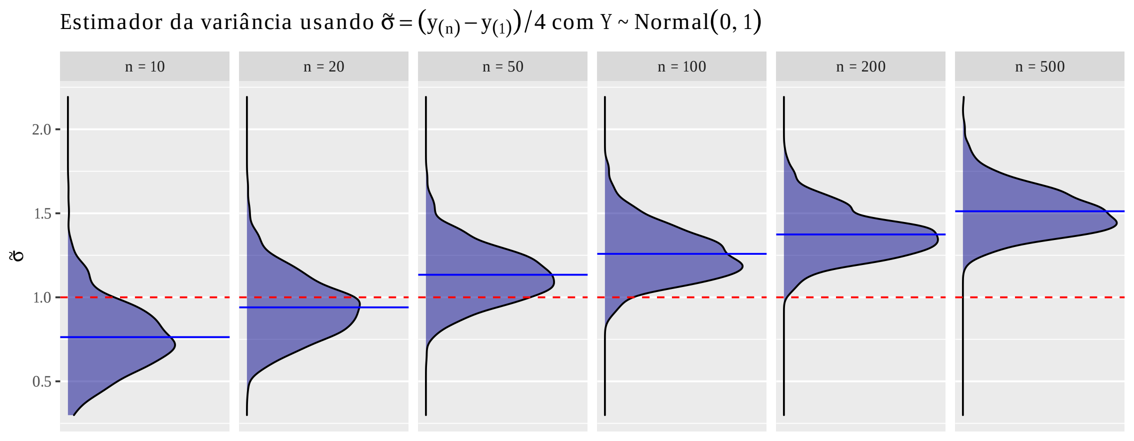

Figura 17. Ilustração por simulação computacional da inconsistência do estimador $\tilde{\sigma}$ do desvio-padrão baseado na regra empírica da amplitude.

Considerações finais

Notas finais

- O **estimador ideal** é aquele que captura a informação da amostra da forma mais eficiente.
- Deseja-se que seja não viciado, com a menor variância possível e consistente.
- A maioria dos estimadores vistos aqui apresentam tais características.
- Estimadores “empíricos” podem não apresentá-las.
- Há situações em que estimadores “óbvios” são superados por outros devidamente formulados.

Figura 18. Distribuição amostral de diferentes estimadores de um parâmetro.