FGV EMAp

João Pedro Jerônimo

# Inferência Estatística

## Revisão para A2

Rio de Janeiro

2025

# Conteúdo

# Distribuição Amostral de Estimadores

Quando temos um estimador $\delta$, e note que estamos falando de um **estimador** e não de uma **estimativa**, temos que, como ele é função de variáveis aleatórias, ele próprio é uma variável aleatória, que possui sua própria distribuição, seus próprios parâmetros, média, variância, etc. Essa distribuição própria da estimativa é o que chamamos de **Distribuição Amostral do Estimador**

> **Definição 1.1** (Distribuição Amostral do Estimador): Dadas as variáveis aleatórias $\underline{X} = (X_1, ..., X_n)$ e $T = r(\underline{X})$ um estimador, onde $\underline{X}$ tem uma distribuição indexada pelo parâmetro $\theta$, então a distribuição de $T|\theta$ chamada de Distribuição Amostral de $T$. ($\mathbb{E}_\theta[T]$ é a média de $T$ na distribuição amostral)

O nome vem do fato que $T$ depende da amostra $\underline{X}$. Na maioria das vezes, $T$ não depende de $\theta$. Mas por que essa distribuição me é interessante?

Vamos supor que eu tenho um estimador $\hat{\theta}$ de $\theta$, pode me ocorrer de eu querer saber a chance de o meu estimador estar próximo do meu $\theta$ de verdade, por exemplo, qual a chance de a distância entre meu estimador e meu $\theta$ ser de só 0.1 medidas? Então podemos querer calcular:

$$\mathbb{P}\left(|\hat{\theta} - \theta| < 0.1\right) \tag{1}$$

Pela lei da probabilidade total, temos também:

$$\mathbb{P}\left(|\hat{\theta} - \theta| < 0.1\right) = \mathbb{E}\left[\mathbb{P}\left(|\hat{\theta} - \theta| < 0.1|\theta\right)\right] \tag{2}$$

Outro uso que podemos derivar para a distribuição amostral é escolher entre vários experimentos qual será performado para obter o melhor estimador de $\theta$. Por exemplo, podemos querer saber qual a quantidade de amostras necessárias para atingir um objetivo em específico

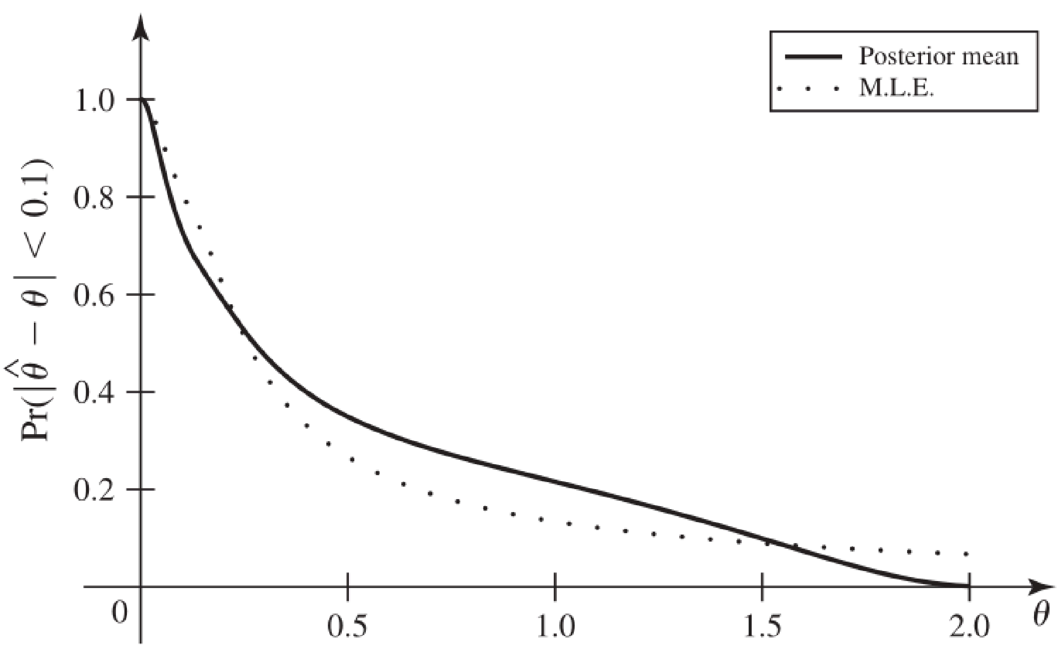


Figura 1: Imagem que representa $\mathbb{P}\left(|\hat{\theta} - \theta\right)| < 0.1)$ em função da quantidade de amostras em um dos exemplos do livro, mostrando que dependendo da nossa situação, podemos quere escolher estimadores diferentes

Uma outra medida interessante que foi apresentada em um dos exemplos do livro é uma distância relativa:

$$\mathbb{P}\left(|\frac{\hat{\theta}}{\theta} - 1| < 0.1\right) \tag{3}$$

Ou seja, a probabilidade de que meu estimador esteja a pelo menos $10\%$ de $\theta$ de distância de $\theta$

*Exemplo 1.1*: Vamos tentar condensar tudo o que vimos em um exemplo. Vamos supor que temos uma clínica que está a fim de identificar ou prever pacientes candidatos a um remédio específico para tratamento da depressão. Então podemos modelar a variável aleatória de um paciente usar ou não esse remédio como uma Bernoulli com $\theta$ de chance de utilizar o remédio ($\mathbb{P}(X = 1) = \theta$). Sabemos por capítulos anteriores que $T = \frac{1}{n} \sum_{i=1}^{n} X_i$ (A proporção de pacientes que vão utilizar o remédio) é uma estatística suficiente e também é o EVM (Estimador de Máxima Verossimilhança) de $\theta$

Porém, $T$ também é uma variável aleatória com distribuição própria, então ela pode assumir vários valores, mas queremos que ela seja o mais próximo possível de $\theta$. Então, que tal calcularmos:

$$\mathbb{P}(|T - \theta| < 0.1) \tag{4}$$

Para isso temos que saber exatamente a distribuição de $T$. Não é muito dificil, na verdade! Sabemos que $T = \frac{1}{n}Y$ com $Y = \sum_{i=1}^{n} X_i$ e $Y|\theta \sim \text{Bin}(n, \theta)$. Então sabemos que:

$$\begin{aligned} \mathbb{P}(T = t|\theta) &= \mathbb{P}\left(\frac{1}{n}Y = t|\theta\right) = \mathbb{P}(Y = nt|\theta) \\ &= \binom{n}{nt}\theta^{nt}(1-\theta)^{n-nt} \end{aligned} \tag{5}$$

Assim, encontramos a nossa Distribuição Amostral do estimador $T$. Então agora poderíamos calcular a equação mencionada anteriormente

$$\mathbb{P}(|T - \theta| < 0.1) = \mathbb{P}(-0.1 < T - \theta < 0.1) = \mathbb{P}(\theta - 0.1 < T < \theta + 0.1) \tag{6}$$

Podemos utilizar a distribuição amostral de $T$ que encontramos anteriormente, porém, por questões de praticidade, vou fazer um pouco diferente:

$$\mathbb{P}(\theta - 0.1 < T < \theta + 0.1) = \mathbb{P}(\lceil(\theta - 0.1)n\rceil \leq Y \leq \lfloor(\theta + 0.1)n\rfloor) \tag{7}$$

Eu adicionei o piso e o teto por conta que $Y$ assume apenas valores inteiros. Então temos que:

$$\mathbb{P}(|T - \theta| < 0.1) = \sum_{k=\lceil(\theta-0.1)n\rceil}^{\lfloor(\theta+0.1)n\rfloor} \binom{n}{k}\theta^k(1-\theta)^{n-k} \tag{8}$$

# Distribuição Chi-Quadrado

Essa distribuição é muito utilizada dentro da estatística serve como base para uma outra distribuição que veremos posteriormente. Ela vai ser útil no próximo capítulo também pois ela é a distribuição do estimador de máxima verossimilhança da variância de uma Normal com média $\mu$ conhecida

$$\widehat{\sigma^2} = \frac{1}{n} \sum_{i=1}^{n} (X_i - \mu)^2 \tag{9}$$

**Definição 2.1**: $\forall m \in \mathbb{R}$, uma distribuição Gamma$\left(\frac{m}{2}, \frac{1}{2}\right)$ é também chamada de $\mathrm{X}_m^2$ (Chi quadrado com $m$ graus de liberadade). Ou seja, se $X \sim \mathrm{X}_m^2$:

$$f_X(x) = \frac{\left(\frac{1}{2}\right)^{\frac{m}{2}}}{\Gamma\left(\frac{m}{2}\right)} x^{\frac{m}{2}-1} e^{-\frac{1}{2}x} \tag{10}$$

## 2.1 Propriedades

**Teorema 2.1.1**: Se $X \sim \mathrm{X}_m^2$ então:

$$\mathbb{E}[X] = m \qquad \mathbb{V}[X] = 2m \tag{11}$$

*Demonstração*: A esperança de uma Gamma$(\alpha, \beta)$ é $\frac{\alpha}{\beta}$, logo:

$$\mathbb{E}[X] = \frac{\frac{m}{2}}{\frac{1}{2}} = m \tag{12}$$

E a variância é $\frac{\alpha}{\beta^2}$, logo:

$$\mathbb{V}[X] = \frac{\frac{m}{2}}{\frac{1}{4}} = 2m \tag{13}$$

□

**Teorema 2.1.2**: A função geradora de momentos de uma $\mathrm{X}_m^2$ é dada por

$$\psi(t) = \left(\frac{1}{1-2t}\right)^{m/2} \qquad \left(t < \frac{1}{2}\right) \tag{14}$$

**Teorema 2.1.3**: Se $X_1, ..., X_k$ são iid e $X_i \sim \mathrm{X}_{m_i}^2$, então $Y = \sum_{j=1}^{k} X_j \sim \mathrm{X}_{\sum_{j=1}^{k} m_j}^2$

*Demonstração*: Sabemos que, dado a FGM de $X$ ($\psi_X$) e de $Y$ ($\psi_Y$) onde $X$ e $Y$ são iid, então a FGM de $X + Y$ é $\psi_X \psi_Y$. Sabendo disso, calculamos a FGM de $X_1 + ... + X_k$:

$$\begin{aligned} \psi_Y(t) &= \prod_{j=1}^{k} \left(\frac{1}{1-2t}\right)^{m_j/2} \qquad \left(t < \frac{1}{2}\right) \\ &= \left(\frac{1}{1-2t}\right)^{\frac{1}{2}\sum_{j=1}^{k} m_j} \qquad \left(t < \frac{1}{2}\right) \end{aligned} \tag{15}$$

□

**Teorema 2.1.4**: Se $X \sim N(0,1)$, então $Y = X^2 \sim \mathrm{X}_1^2$

*Demonstração*: Sabemos que se $X$ tem pdf $f_X(x)$ e $Z = h(X)$, então a pdf de $Z$ é

$$f_Z(z) = \frac{f_X(x)}{|h'(x)|} \tag{16}$$

Então, considerando $h(x) = x^2$ e $f_X(x) = \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}}$, temos que

$$f_Z(z) = \frac{\frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}}}{2x} = \frac{1}{2\sqrt{\pi z}} e^{-\frac{1}{2}z} \tag{17}$$

Perceba que isso é a densidade de uma $\mathrm{X}_1^2$, veja:

$$\frac{\left(\frac{1}{2}\right)^{\frac{1}{2}}}{\Gamma\left(\frac{1}{2}\right)} z^{\frac{1}{2}-1} e^{-\frac{1}{2}z} = \frac{\frac{1}{\sqrt{2}}}{\sqrt{\pi}} \frac{1}{\sqrt{z}} e^{-\frac{1}{2}z} = \frac{1}{\sqrt{2\pi z}} e^{-\frac{1}{2}z} \tag{18}$$

Logo, $Z \sim \mathrm{X}_1^2$. A função $f_Z(z)$ possui um termo $\frac{1}{2}$ pois preciamos lembrar que a normal vai de $[-\infty, \infty]$, então para que $h(x)$ seja monótona, precisamos restringir em $[-\infty, 0]$ e $[0, \infty]$ então obtemos duas funções que integram $\frac{1}{2}$ em cada intervalo, logo o resultado integra 1 □

**Corolário 2.1.4.1**: Se $X_1, ..., X_n \sim N(0,1)$, então:

$$X_1^2 + ... + X_n^2 \sim \mathrm{X}_m^2 \tag{19}$$

# Distribuição Conjunta da Média e Variância Amostral

Vamos supor que estamos fazendo um experimento e coletamos amostras de uma distribuição normal com média $\mu$ e variância $\sigma^2$, porém, não sabemos nenhum dos dois, então queremos estimá-los! Podemos escolher vários estimadores, por exemplo, o EVM:

$$\hat{\mu} = \frac{1}{n} \sum_{i=1}^{n} X_i \qquad \widehat{\sigma^2} = \frac{1}{n} \sum_{i=1}^{n} (X_i - \hat{\mu})^2 \tag{20}$$

Porém, como vimos na primeira parte, eles são variáveis aleatórias com distribuições próprias, então podemos utilizar técnicas para saber o quão bem eles aproximam $\mu$ e $\sigma^2$, mas antes, temos que descobrir suas distribuições amostrais! Acontece que a distribuição de $\hat{\mu}$ depende de $\sigma^2$, porém, veremos que a distribuição conjunta de $\hat{\mu}$ e $\widehat{\sigma^2}$ nos permite inferir $\mu$ sem referenciar $\sigma$

## 3.1 Independência da Média e Variância Amostrais

**Teorema 3.1.1**: Suponha que as variáveis $X_1, ..., X_n$ são iid, onde $X_j \sim N(0, 1)$. Seja $Q$ uma matriz ortogonal $n \times n$ e $\underline{Y} = Q\underline{X}$ onde $\underline{X} = (X_1, ..., X_n)^T$, então $Y_j \sim N(0, 1)$ e $Y_1, ..., Y_n$ são iid, além de que $\sum X_i^2 = \sum Y_i^2$

*Demonstração*: A distribuição conjunta de $X_1, ..., X_n$ é dada por:

$$f_X(\underline{x}) = \frac{1}{(2\pi)^{\frac{n}{2}}} \exp\left\{-\frac{1}{2} \sum_{i=1}^{n} x_i^2\right\} \tag{21}$$

Sabemos que, se $Z = h(Y)$ com $h$ monótona, então

$$f_Z(z) = \frac{f_Y(y)}{|h'(y)|} \tag{22}$$

Se considerarmos $h(x) = Qx$, então $\underline{Y} = h(\underline{X})$, logo:

$$f_Y(y) = \frac{f_X(Q^T y)}{|\det(Q)|} \tag{23}$$

Porém, $\det(Q) = 1$ pois $Q$ é ortogonal. Além disso, sabemos que $\|x\|^2 = \|Qx\|^2$, então:

$$f_Y(y) = \frac{1}{(2\pi)^{\frac{n}{2}}} \exp\left\{-\frac{1}{2} \sum_{i=1}^{n} y_i^2\right\} \tag{24}$$

Logo, chegamos que $Y_j \sim N(0, 1)$ e são iid $\square$

Agora vamos demonstrar um dos teoremas mais surpreendentes da estatística

**Teorema 3.1.2** (Independência da Média e Variância Amostral): Sejam $X_1, ..., X_n \sim N(\mu, \sigma^2)$ e dados os estimadores:

$$\hat{\mu} = \frac{1}{n}\sum_{i=1}^{n} X_i \qquad \widehat{\sigma^2} = \frac{1}{n}\sum_{i=1}^{n} (X_i - \hat{\mu})^2 \tag{25}$$

As variáveis aleatórias $\hat{\mu}$ e $\widehat{\sigma^2}$ são **INDEPENDENTES** entre si. Junto desse teorema, também mostraremos que:

$$\frac{1}{\sigma^2}\sum_{i=1}^{n} (X_i - \hat{\mu})^2 \sim \mathrm{X}^2_{n-1} \tag{26}$$

*Demonstração*: Primeiro vamos provar considerando que $\mu = 0$ e $\sigma^2 = 1$, de forma que utilizaremos esse resultado para generalizar posteriormente.

Vamos primeiro definir $u = \left(\frac{1}{\sqrt{n}} \ \cdots \ \frac{1}{\sqrt{n}}\right)^T$, então construímos uma matriz $Q$ utilizando Gram-Schmidt de forma que $u$ seja a primeira linha dessa matriz. Então definimos:

$$\begin{pmatrix} Y_1 \\ \vdots \\ Y_n \end{pmatrix} = Q \begin{pmatrix} X_1 \\ \vdots \\ X_n \end{pmatrix} \tag{27}$$

Vimos pelo Teorema 3.1.1 que $Y_j$ são iid e são normais padrão também. Guarde essa informação! Não é difícil ver que:

$$Y_1 = \frac{1}{\sqrt{n}}\sum_{i=1}^{n} X_i = \hat{\mu}\sqrt{n} \tag{28}$$

Como $\sum_{i=1}^{n} X_i^2 = \sum_{i=1}^{n} Y_i^2$, então:

$$\sum_{i=2}^{n} Y_i^2 = \sum_{i=1}^{n} Y_i^2 - Y_1^2 = \sum_{i=1}^{n} X_i^2 - n(\hat{\mu})^2 = \sum_{i=1}^{n} (X_i - \hat{\mu})^2 \tag{29}$$

Ou seja, $\hat{\mu}$ e $\widehat{\sigma^2}$ são independentes! Dado esse resultado, consideremos agora média e variância não-padrões. Então vamos definir:

$$Z_i = \frac{X_i - \mu}{\sigma} \tag{30}$$

Então $Z_1, ..., Z_n$ são iid. Sabemos que $\overline{Z}_n$ e $\sum\left(Z_i - \overline{Z}_n\right)$ são independentes. Perceba também que, como $\sum\left(Z_i - \overline{Z}_n\right) = \sum_{i=2}^{n} Y_i$, então $\sum\left(Z_i - \overline{Z}_n\right) \sim \mathrm{X}^2_{n-1}$. Porém, sabemos também que $\overline{X}_n \sim N\left(\mu, \frac{\sigma^2}{n}\right)$, então

$$\begin{gathered} \overline{Z}_n = \frac{\overline{X}_n - \mu}{\sigma} \\ \Rightarrow \sum\left(Z_i - \overline{Z}_n\right)^2 = \frac{1}{\sigma^2}\sum\left(X_i - \overline{X}_n\right)^2 \end{gathered} \tag{31}$$

Logo, $\overline{X}_n$ e $\frac{1}{n}\sum\left(X_i - \overline{X}_n\right)^2$ são independentes e $\frac{1}{\sigma^2}\sum\left(X_i - \overline{X}_n\right)^2 \sim \mathrm{X}^2_{n-1}$ □

Esse resultado é interessante pois, em certas ocasiões, podemos querer saber a seguinte probabilidade:

$$\mathbb{P}\left(|\hat{\mu}-\mu| \leq \frac{1}{5}\sigma, |\hat{\sigma}-\sigma| \leq \frac{1}{5}\sigma\right) \geq \frac{1}{2} \tag{32}$$

Já que ela indica uma probabilidade de proximidade entre meus estimadores e meus parâmetros. Porém, pelo Teorema 3.1.2, podemos separar essa probabilidade em:

$$\underbrace{\mathbb{P}\left(|\hat{\mu}-\mu| \leq \frac{1}{5}\sigma\right)}_{p_1} \underbrace{\mathbb{P}\left(|\hat{\sigma}-\sigma| \leq \frac{1}{5}\sigma\right)}_{p_2} \geq \frac{1}{2} \tag{33}$$

e isso simplifica bastante nossas contas! Vamos definir $U \sim N(0,1)$, então podemos reescrever as probabilidades como:

$$p_1 = \mathbb{P}\left(\frac{\sqrt{n}}{\sigma}\ |\hat{\mu}-\mu| < \frac{1}{5}\sqrt{n}\right) = \mathbb{P}\left(|U| < \frac{1}{5}\sqrt{n}\right) \tag{34}$$

definindo $V = \frac{n}{\sigma^2}\widehat{\sigma^2}$, sabemos que $V \sim \mathrm{X}^2_{n-1}$, então

$$\begin{aligned} p_2 &= \mathbb{P}\left(-\frac{1}{5}\sigma \leq \hat{\sigma}-\sigma \leq \frac{1}{5}\sigma\right) = \mathbb{P}\left(\frac{4}{5}\sigma \leq \hat{\sigma} \leq \frac{6}{5}\sigma\right) \\ &= \mathbb{P}\left(\frac{16}{25}\sigma^2 \leq \hat{\sigma}^2 \leq \frac{36}{25}\sigma^2\right) = \mathbb{P}\left(\frac{16}{25}n \leq V \leq \frac{36}{25}n\right) \end{aligned} \tag{35}$$

e como $V \sim \mathrm{X}^2_{n-1}$, basta consultar uma tabela ou um software para descobrir esses quantis

# Distribuições $t$

Outra família interessante de distribuições! Imagine a situação: Dado um experimento com amostras aleatórias $N(\mu, \sigma^2)$, sabemos que $\overline{X}_n \sim N\left(\mu, \frac{\sigma^2}{n}\right)$, logo:

$$\frac{\sqrt{n}\left(\overline{X}_n - \mu\right)}{\sigma} \sim N(0,1) \tag{36}$$

Porém, podemos não saber $\sigma$ e queremos substituir, por exemplo, por seu EMV, então qual seria a distribuição de

$$\frac{\sqrt{n}\left(\overline{X}_n - \mu\right)}{\hat{\sigma}} \tag{37}$$

**Definição 4.1** (Distribuição $t$ com $m$ graus de liberdade): Se $Z \sim N(0,1)$ e $Y \sim \mathrm{X}^2_m$, então

$$X = \frac{Z}{\sqrt{\frac{Y}{m}}} \sim t_m \tag{38}$$

onde dizemos $t$ com $m$ graus de liberdade

**Teorema 4.1** (PDF): A pdf de $X \sim t_m$ é:

$$f_X(x) = \frac{\Gamma\left(\frac{m+1}{2}\right)}{(m\pi)^{\frac{1}{2}}\Gamma\left(\frac{m}{2}\right)}\left(1 + \frac{x^2}{m}\right)^{-(m+1)/2} \tag{39}$$

*Demonstração*: Lembrando: $Y \sim \mathrm{X}^2_m$ e $Z \sim N(0,1)$. Vamos aplicar as transformações! Sabemos que:

$$f_{XW}(x,w) = f_{YZ}(y,z) \mid \frac{\partial(y,z)}{\partial(x,w)} \mid \tag{40}$$

então denotando $W = Y$, temos:

$$Z = X\left(\frac{W}{m}\right)^{\frac{1}{2}} \qquad Y = W \tag{41}$$

Então vamos ter que:

$$\begin{gathered} \frac{\partial y}{\partial x} = 0 \qquad \frac{\partial y}{\partial w} = 1 \\ \frac{\partial z}{\partial x} = \left(\frac{w}{m}\right)^{\frac{1}{2}} \end{gathered} \tag{42}$$

então vamos ter que

$$\mid \frac{\partial(y,z)}{\partial(x,w)} \mid = -\left(\frac{w}{m}\right)^{\frac{1}{2}} \tag{43}$$

$$
\begin{aligned}
f_{XW}(x,w) &= f_{WZ}(w,z)\left(\frac{w}{m}\right)^{\frac{1}{2}} \\
&= f_W(w) f_Z(z)\left(\frac{w}{m}\right)^{\frac{1}{2}} \qquad \text{(Independência de Y e Z)} \\
&= f_W(w) f_Z\left(x\left(\frac{w}{m}\right)^{\frac{1}{2}}\right)\left(\frac{w}{m}\right)^{\frac{1}{2}} \\
&= \underbrace{\frac{\left(\frac{1}{2}\right)^{\frac{m}{2}}}{\Gamma\left(\frac{m}{2}\right)} w^{\frac{m}{2}-1} e^{-\frac{1}{2}w}}_{f_W(w)} \underbrace{\frac{1}{\sqrt{2\pi}} e^{\frac{-x^2 w}{2m}}}_{f_Z(z)} \left(\frac{w}{m}\right)^{\frac{1}{2}}
\end{aligned}
\tag{44}
$$

reescrevendo para ficar algo mais limpo e unir os termos comuns:

$$
f_{XW}(x,w) = \frac{\left(\frac{1}{2}\right)^{\frac{m}{2}}}{\Gamma\left(\frac{m}{2}\right)\sqrt{2\pi m}} w^{\frac{m-1}{2}} \exp\left\{-\frac{1}{2}\left(1+\frac{x^2}{m}\right)w\right\} \tag{45}
$$

Agora, para obtermos a marginal de $X$, precisamos integrar isso tudo com relação a $w$. Porém, basta integrarmos aquilo que é em função apenas de $w$, as constantes nós podemos adicionar novamente depois, então:

$$
f_X(x) \propto \int_0^\infty w^{\frac{m-1}{2}} \exp\left\{-\frac{1}{2}\left(1+\frac{x^2}{m}\right)w\right\} \mathrm{d}w \tag{46}
$$

Se definirmos $\alpha = \frac{m+1}{2}$ e $\beta = \frac{1}{2}\left(1+\frac{x^2}{m}\right)$, então a integral equivale a:

$$
f_X(x) \propto \int_0^\infty w^{\alpha-1} e^{-\beta w} \,\mathrm{d}w = \frac{\Gamma(\alpha)}{\beta^\alpha} = \frac{\Gamma\left(\frac{m+1}{2}\right)}{\left(\frac{1}{2}\left[1+\frac{x^2}{m}\right]\right)^{(m+1)/2}} \tag{47}
$$

$$
\Rightarrow f_X(x) = \frac{\left(\frac{1}{2}\right)^{\frac{m}{2}}}{\Gamma\left(\frac{m}{2}\right)\sqrt{2\pi m}} \Gamma\left(\frac{m+1}{2}\right)\left(\frac{1}{2}\left[1+\frac{x^2}{m}\right]\right)^{-(m+1)/2} \tag{48}
$$

juntando tudo, temos:

$$
f_X(x) = \frac{\Gamma\left(\frac{m+1}{2}\right)}{\sqrt{m\pi}\;\Gamma\left(\frac{m}{2}\right)}\left(1+\frac{x^2}{m}\right)^{-(m+1)/2} \tag{49}
$$

□

## 4.1 Propriedades

**Teorema 4.1.1**: Se $T \sim t_m$, então:

$$
\begin{aligned}
\mathbb{E}[T] &= 0 \qquad (m > 1) \\
\mathbb{V}[T] &= \frac{m}{m-2} \qquad (m > 2)
\end{aligned}
\tag{50}
$$

*Demonstração*: Seja $Z \sim N(0,1)$ e $W \sim \mathrm{X}^2_m$, sabemos que

$$T = \frac{Z}{\sqrt{\frac{W}{m}}} \sim t_m \tag{51}$$

porém, temos que $T|W = w \sim N\left(0, \frac{m}{w}\right)$, logo:

$$\mathbb{E}[T|W = w] = 0 \tag{52}$$

pela lei de adão:

$$\mathbb{E}[\mathbb{E}[T|W = w]] = \mathbb{E}[T] = \mathbb{E}[0] = 0 \tag{53}$$

No mesmo raciocínio, lembrando a lei de EVA

$$\begin{aligned} \mathbb{V}[X] = \mathbb{E}[\mathbb{V}[X|Y]] + \mathbb{V}[\mathbb{E}[X|Y]] \\ \Rightarrow \mathbb{V}[T] = \mathbb{E}[\mathbb{V}[T|W]] + \mathbb{V}[\mathbb{E}[T|W]] \end{aligned} \tag{54}$$

sabemos que $\mathbb{E}[T|W] = 0$, então basta calcularmos $\mathbb{V}[T|W]$ que, como vimos antes, vai ser $\frac{W}{m}$, então:

$$\mathbb{V}[T] = \mathbb{E}\left[\frac{m}{W}\right] \tag{55}$$

Sabemos que $W$ é uma $\Gamma\left(\frac{m}{2}, \frac{1}{2}\right)$, então $W^{-1}$ é uma Gamma Inversa, logo, sua média vai ser:

$$\mathbb{E}[W^{-1}] = \frac{\frac{1}{2}}{\frac{m}{2} - 1} = \frac{\frac{1}{2}}{\frac{m-2}{2}} = \frac{1}{m - 2} \tag{56}$$

logo:

$$\mathbb{V}[T] = \frac{m}{m - 2} \tag{57}$$

□

**Teorema 4.1.2** (Divergência dos Momentos): Se $T \sim t_m$, então $\mathbb{E}[\![T|^p]$ diverge se $p \geq m$. Se $m$ é inteiro, então apenas os $m - 1$ primeiros momentos existem

**Teorema 4.1.3**: Sejam $X_1, ..., X_n \sim N(\mu, \sigma^2)$ e $\sigma' = \left(\frac{1}{n-1} \sum \left(X_i - \overline{X}_n\right)^2\right)^{\frac{1}{2}}$, então

$$\frac{\sqrt{n}\left(\overline{X}_n - \mu\right)}{\sigma'} \sim t_{n-1} \tag{58}$$

*Demonstração*: Defina $S_n^2 = \sum \left(X_i - \overline{X}_n\right)^2$, $Z - \sqrt{n}\left(\overline{X}_n - \mu\right)/\sigma$ e $Y = S_n^2/\sigma^2$. Sabemos que $Y$ e $Z$ são independentes e $Y \sim \mathrm{X}_{n-1}^2$. Definimos então:

$$U = \frac{Z}{\sqrt{\frac{Y}{n-1}}} \tag{59}$$

que é uma $t_{n-1}$ por definição. Porém, perceba que:

$$U = \frac{\frac{\sqrt{n}\left(\overline{X}_n - \mu\right)}{\sigma}}{\frac{1}{\sigma}\sqrt{\frac{S_n^2}{n-1}}} = \frac{\sqrt{n}\left(\overline{X}_n - \mu\right)}{\sigma'} \tag{60}$$

$\square$

Essa propriedade é interessante, pois saímos de variáveis que dependiam diretamente de $\sigma$ para uma variável que tem distribuição que **não depende** de $\sigma$

**Teorema 4.1.4**: Uma distribuição $t_1$ é equivalente a uma distribuição de *Cauchy*

**Definição 4.1.1** (Distribuição de Cauchy): Se $X \sim \text{Cauchy}(x_0, \gamma)$, então temos:

$$f_X(x) = \frac{1}{\pi\gamma\left[1 + \left(\frac{x - x_0}{\gamma}\right)^2\right]} \tag{61}$$

$$F_X(x) = \frac{1}{\pi}\arctan\left(\frac{x - x_0}{\gamma}\right) + \frac{1}{2} \tag{62}$$

Além do fato que:

$$\mathbb{E}[\![X|^k]\!] = \infty \qquad \forall k \tag{63}$$

# Intervalos de Confiança

Essa é uma parte que costuma gerar bastante confusão! Intervalos de confiança adicionam mais informações a um estimador $\hat{\theta}$ quando não conhecemos o $\theta$, de forma que podemos encontrar um intervalo $[A, B]$ com alta probabilidade de conter $\theta$

## 5.1 Intervalo de Confiança para a média de uma Normal

Dada a amostra $X_1, ..., X_n \sim N(\mu, \sigma^2)$, sabemos que $U = \sqrt{n}\left(\overline{X} - \mu\right)/\sigma' \sim t_{n-1}$. Eu gostaria de achar um intervalo no qual eu tenho uma chance boa de encontrar minha média, então eu gostaria de algo do tipo:

$$\mathbb{P}(-c < U < c) = \gamma \tag{64}$$

O método mais comum é calcular diretamente o $c$ que torna a equação (64) verdadeira. Isso é equivalente a dizer:

$$\mathbb{P}\left(\overline{X}_n - \frac{c\sigma'}{\sqrt{n}} < \mu < \overline{X}_n + \frac{c\sigma'}{\sqrt{n}}\right) = \gamma \tag{65}$$

Vale ressaltar que essa probabilidade é referente a distribuição conjunta de $\overline{X}_n$ e $\sigma'$ para valores **fixos** de $\mu$ e $\sigma$ (Independentemente de sabermos eles ou não). Então vamos tentar achar o $c$ que satisfaz isso

$$\mathbb{P}(-c < U < c) = \gamma \Leftrightarrow T_{n-1}(c) - T_{n-1}(-c) = \gamma \tag{66}$$

Pela simetria de $t$ em 0, posso reescrever como:

$$2T_{n-1}(c) - 1 = \gamma \Rightarrow c = T_{n-1}^{-1}\left(\frac{1+\gamma}{2}\right) \tag{67}$$

Então, depois que descobrimos $c$, nosso intervalo de confiança vira:

$$\begin{aligned} A &= \overline{X}_n - \frac{\sigma'}{\sqrt{n}}\, T_{n-1}^{-1}\left(\frac{1+\gamma}{2}\right) \\ B &= \overline{X}_n + \frac{\sigma'}{\sqrt{n}}\, T_{n-1}^{-1}\left(\frac{1+\gamma}{2}\right) \end{aligned} \tag{68}$$

Dada essa noção inicial, vamos definir formalmente esses intervalos:

**Definição 5.1.1** (Intervalo de confiança): Seja $\underline{X} = (X_1, ..., X_n)$ uma amostra de uma distribuição indexada pelo(s) parâmetro(s) $\theta$ e seja $g(\theta) : \Omega \to \mathbb{R}$. Seja também $A$ e $B$ duas estatísticas $(A \leq B)$ que satisfazem:

$$\mathbb{P}(A < g(\theta) < B) \geq \gamma \tag{69}$$

O intervalo aleatório $(A, B)$ é chamado de intervalo de confiança $\gamma$ para $g(\theta)$ ou de intervalo de confiança $100\gamma\%$ para $g(\theta)$. Depois que $X_1, ..., X_n$ foi observado e o intervalo $A = a$ e $B = b$ foi computado, chamamos o valor observado do intervalo de **valor observado do intervalo de confiança**. Se a equação (69) vale a igualdade $\forall c \in (A, B)$, então chamamos esse intervalo de **exato**

Aqui eu vou definir melhor a interpretação com relação a essa definição, que pode ser um pouco confusa. A interpretação do intervalo $A, B$ em si é bem direta, representa um intervalo **aleatório** que tem probabilidade $\gamma$ de conter $g(\theta)$. Porém, ao observamos as amostras e calcularmos $A = a$ e $B = b$, o intervalo $(a, b)$ **não necessariamente contém $g(\theta)$ com probabilidade $\gamma$**, como assim? Lembra que $(A, B)$ é um intervalo **aleatório**, enquanto $(a, b)$ é uma das muitas possíveis ocorrências desse intervalo! A interpretação correta é, que quanto mais repetimos o experimento e computamos $(a, b)$ e armazenamos esses valores observados de intervalo, uma fração $\gamma$ deles contém $g(\theta)$, porém, **não sabemos dizer quais contém e quais não contém**

**Teorema 5.1.1**: Dada a amostra $X_1, ..., X_n \sim N(\mu, \sigma^2)$. $\forall \gamma \in [0, 1]$, o intervalo $(A, B)$ com seguintes pontos:

$$\begin{aligned} A &= \overline{X}_n - \frac{\sigma'}{\sqrt{n}} \, T_{n-1}^{-1}\left(\frac{1+\gamma}{2}\right) \\ B &= \overline{X}_n + \frac{\sigma'}{\sqrt{n}} \, T_{n-1}^{-1}\left(\frac{1+\gamma}{2}\right) \end{aligned} \tag{70}$$

é um intervalo de confiança $\gamma$-exato

## 5.2 Intervalos de Confiança Unilaterais

Nós vimos como encontrar intervalos aleatórios $(A, B)$ que tem probabilidade $\gamma$ de conter o parâmetro $\theta$, porém, podem acontecer situações que apenas obter um limite superior ou inferior seja suficiente para nós

Dados $\gamma_1$ e $\gamma_2$ com $\gamma_2 > \gamma_1$ e $\gamma_2 - \gamma_1 = \gamma$, então:

$$\mathbb{P}(T_{n-1}^{-1}(\gamma_1) < U < T_{n-1}^{-1}(\gamma_1)) = \gamma \tag{71}$$

E então obtemos que, perante todos os intervalos aleatórios possíveis, o intervalo de confiança $\gamma$ com o menor tamanho é o simétrico

$$\gamma_1 = 1 - \gamma_2 \tag{72}$$

Porém, há casos que um intervalo não-simétrico é útil (Como mencionei o caso anterior de apenas limites superiores ou intefiores)

**Definição 5.2.1** (Intervalo de Confiança Generalizado): Seja $\underline{X} = (X_1, ..., X_n)$ uma amostra de uma distribuição parametrizada por $\theta$. Seja $g(\theta) : \Omega \to \mathbb{R}$ e seja $A$ uma estatística tal que:

$$\mathbb{P}(A < g(\theta)) \geq \gamma \qquad \forall \theta \tag{73}$$

Então o intervalo aleatório $(A, +\infty)$ é chamado de intervalo de confiança unilateral $\gamma$ de limite inferior $A$. A mesma definição vale para a estatística $B$ tal que:

$$\mathbb{P}(g(\theta) < B) \geq \gamma \tag{74}$$

Então o intervalo aleatório $(-\infty, B)$ é chamado de intervalo de confiança unilateral $\gamma$ de limite superior $B$. Se a desigualdade "$\geq \gamma$" é uma igualdade para todo $\theta$, então tanto o intervalo quanto os limites são chamados de exatos

**Teorema 5.2.1** (Intervalo unilateral da média da normal): Seja $X_1, ..., X_n \sim N(\mu, \sigma^2)$, as estatísticas a seguir são, respectivamente, limites inferior e superior exatos com coeficiente $\gamma$ para $\mu$:

$$\begin{aligned} A &= \overline{X}_n - T_{n-1}^{-1}(\gamma)\frac{\sigma'}{\sqrt{n}} \\ B &= \overline{X}_n + T_{n-1}^{-1}(\gamma)\frac{\sigma'}{\sqrt{n}} \end{aligned} \quad (75)$$

## 5.3 Intervalo de confiança para outros parâmetros

Até agora, só vimos a aplicação de intervalos de confiança para a distribuição normal, mas por quê? Pois a normal possui propriedades que tornam encontrar os intervalos de confiança mais fáceis, como por exemplo, encontrarmos estatísticas (Por exemplo $T = \sqrt{n}\left(\overline{X}_n - \mu\right)/\sigma'$) que não dependem do parâmetro que queremos estimar, e isso na verdade é uma definição útil que pode nos ajudar:

**Definição 5.3.1** (Pivô): Seja $\underline{X} = (X_1, ..., X_n)$ uma amostra de uma distribuição parametrizada por $\theta$ e $V(\theta, \underline{X})$ uma variável aleatória tal que sua distribuição **não depende de $\boldsymbol{\theta}$** e é a mesma $\forall \theta$, então chamamos $V(\theta, \underline{X})$ de **quantidade pivotal** ou **pivô**

Podemos então utilizar dessa definição para construir intervalos de confiança. Porém, para isso, precisamos de uma "função inversa" desse $V$, algo do tipo:

$$r(V(\theta, \underline{X}), \underline{X}) = g(\theta) \quad (76)$$

**Teorema 5.3.1**: Seja $\underline{X} = (X_1, ..., X_n)$ uma amostra de uma distribuição parametrizada por $\theta$. Suponha que existe um pivô $V$. Seja $F_V(v)$ a CDF de $V$ e contínua. Assuma também que a função $r$ tal qual a equação (76) existe e é estritamente crescente em $v$ para cada $\underline{x}$. Seja $\gamma \in (0, 1)$ e $\gamma_2 > \gamma_1$ tal que $\gamma_2 - \gamma_1 = \gamma$, então as seguintes estatísticas são endpoints de um invervalo de confiança $\gamma$-exato para $g(\theta)$

$$\begin{aligned} A &= r\left(F_V^{-1}(\gamma_1), \underline{x}\right) \\ B &= r\left(F_V^{-1}(\gamma_2), \underline{x}\right) \end{aligned} \quad (77)$$

Se $r$ é estritamente decrescente, então invertemos $A$ e $B$

*Demonstração*: Se $r(\theta, \underline{x})$ é estritamente crescente em $v$ para todo $\underline{x}$, então:

$$V(\theta, \underline{X}) < c \Leftrightarrow g(\theta) < r(c, \underline{X}) \quad (78)$$

Defina $c = F_V^{-1}(\gamma_i)$ para $i = 1, 2$, então obtemos:

$$\begin{aligned} \mathbb{P}(g(\theta) < A) &= \gamma_1 \\ \mathbb{P}(g(\theta) < B) &= \gamma_2 \end{aligned} \quad (79)$$

Como $V$ tem distribuição contínua e $r$ é estritamente crescente, então:

$$\mathbb{P}(A = g(\theta)) = \mathbb{P}\left(V(\theta, \underline{X}) = F_V^{-1}(\gamma_1)\right) = 0 \quad (80)$$

Similarmente com $\mathbb{P}(B = g(\theta))$, então combinamos as duas equações na equação (79) para obter $\mathbb{P}(A < g(\theta) < B) = \gamma$ $\square$

*Exemplo 5.3.1*: Seja $X_1, ..., X_n$ uma amostra aleatória de uma distribuição normal com média $\mu$ e variância $\sigma^2$ desconhecidos. Vimos anteriormente que:

$$V(\theta, \underline{X}) = \frac{1}{\sigma^2} \sum_{i=1}^{n} \left(X_i - \overline{X}_n\right)^2 \sim \mathrm{X}^2_{n-1} \qquad \forall \theta = (\mu, \sigma^2) \tag{81}$$

Logo, $V$ é um pivô, de forma que conseguimos utilizá-lo para achar intervalos de confiança para $\sigma^2$

Porém é bem comum que o pivô não exista em casos discretos

*Exemplo 5.3.2*: Seja $\theta$ a proporção de sucessos em uma população muito grande de pacientes tratados com imipramina. Suponha que os clínicos desejem uma variável aleatória $A$ tal que, para todo $\theta$, tenhamos

$$\Pr(A < \theta) \geq 0.9 \tag{82}$$

Isto é, eles querem ter **90% de confiança** de que a proporção de sucesso seja **pelo menos $A$**. Os dados observáveis consistem no número $X$ de sucessos em uma amostra aleatória de $n = 40$ pacientes. Nenhuma variável pivotal existe neste exemplo, e os intervalos de confiança são mais difíceis de construir