

Inferência Estatística II

Prof. Fernando de Souza Bastos
fernando.bastos@ufv.br

Departamento de Estatística
Programa de Pós-Graduação em Estatística Aplicada e Biometria
Universidade Federal de Viçosa
Campus UFV - Viçosa



Sumário

- 1 Preliminares
- 2 Revisão Sobre Estimação Pontual
- 3 Estimadores Pontuais
 - Estimadores Não Viesados
- 4 Distribuições Limites
 - Convergência em Probabilidade
 - Desigualdades de Markov e Tchebychev
 - Lei Fraca dos Grandes Números
 - Propriedades de Convergência em Probabilidades

A Inferência Estatística é um ramo da Estatística que se dedica a tirar conclusões sobre uma população com base em uma amostra representativa dessa população. Surgida no início do século XX, essa área se desenvolveu como uma resposta à necessidade de generalizar os resultados obtidos a partir de dados amostrais, permitindo a extrapolação de informações de forma rigorosa e fundamentada.

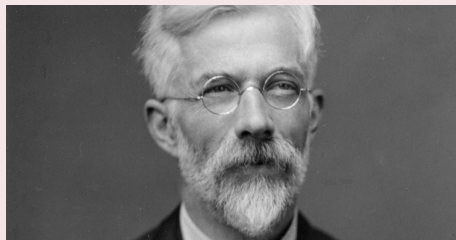
Dentre os principais pesquisadores, destacam-se Ronald Fisher, Jerzy Neyman e Egon Pearson, que foram responsáveis por desenvolver a Teoria da Estimação e a Teoria dos Testes de Hipóteses. Além desses, outros autores importantes incluem William Gosset (conhecido como Student), que criou o teste t de Student, e Harold Hotelling, que desenvolveu a Teoria da Distribuição Multivariada.

Dentre os principais pesquisadores, destacam-se Ronald Fisher, Jerzy Neyman e Egon Pearson, que foram responsáveis por desenvolver a Teoria da Estimção e a Teoria dos Testes de Hipóteses. Além desses, outros autores importantes incluem William Gosset (conhecido como Student), que criou o teste t de Student, e Harold Hotelling, que desenvolveu a Teoria da Distribuição Multivariada.

Sugestão de Leitura:

Uma Senhora Toma Chá...: Como a Estatística Revolucionou a Ciência no Século XX. (SALSBURG, 2009)

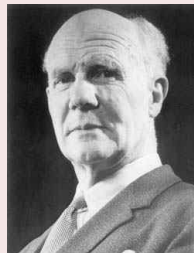
Ronald A. Fisher (1890-1962) estatístico e geneticista britânico, desenvolveu a análise de variância (ANOVA) e o teste de significância, e foi um pioneiro no uso de experimentos controlados para validar hipóteses científicas.



Karl Pearson (1857-1936) foi um estatístico e biométrico britânico, conhecido por desenvolver o conceito de correlação e regressão, além de fundar o campo da biometria. Ele também fez importantes contribuições para a teoria da probabilidade e a inferência estatística.



Egon Pearson (1895-1980) foi um estatístico britânico e filho de Karl Pearson. Ele é mais conhecido por suas contribuições à teoria estatística, especialmente no desenvolvimento da teoria dos testes de hipóteses. Juntamente com Ronald A. Fisher, ele desenvolveu o conceito de erro tipo I e erro tipo II em testes de hipóteses.



Jerzy Neyman (1894-1981) polonês de renome internacional, conhecido por suas contribuições fundamentais à teoria estatística. Neyman é famoso pelo desenvolvimento do método de intervalos de confiança e pelo conceito de testes de hipóteses que incluem o teste de significância. Também desempenhou um papel crucial na fundação do campo da estatística matemática e na educação estatística.



William Sealy Gosset (1876-1937) foi um estatístico britânico que trabalhou sob o pseudônimo "Student". Gosset desenvolveu o teste t enquanto trabalhava para a cervejaria Guinness, onde lidava com amostras pequenas e necessitava de métodos estatísticos apropriados para análise de dados limitados.

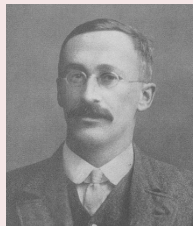
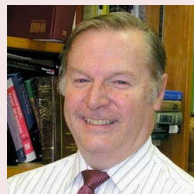


Fig. 1. "Student" (William Sealy Gosset) in 1908. (Reproduced

Harold Hotelling (1895-1973) foi um estatístico e economista americano, conhecido por suas contribuições significativas à estatística multivariada e à econometria. Entre suas realizações notáveis estão o desenvolvimento do teste de Hotelling, que é uma generalização do teste t para múltiplas variáveis. Suas ideias e métodos continuam a influenciar áreas como Análise Estatística, Economia e Ciência de dados.



Revisão de Estimação Pontual

Em um problema Estatístico típico, temos uma variável aleatória X de interesse, mas sua função de densidade de probabilidade $f(x)$ ou sua função de massa de probabilidade $p(x)$ não é conhecida. Nossa ignorância pode ser classificada de maneira geral de duas formas:

Revisão de Estimação Pontual

Em um problema Estatístico típico, temos uma variável aleatória X de interesse, mas sua função de densidade de probabilidade $f(x)$ ou sua função de massa de probabilidade $p(x)$ não é conhecida. Nossa ignorância pode ser classificada de maneira geral de duas formas:

Classificações da Ignorância

- 1 $f(x)$ ou $p(x)$ é completamente desconhecida.
- 2 A forma de $f(x)$ ou $p(x)$ é conhecida, mas um parâmetro θ é desconhecido, onde θ pode ser um vetor.

Vamos trabalhar, em especial, com a segunda classificação, embora algumas de nossas discussões se apliquem também à primeira classificação.

Alguns exemplos são os seguintes:

- (a) X tem uma distribuição exponencial, $Exp(\theta)$, onde θ é desconhecido.

Alguns exemplos são os seguintes:

- (a) X tem uma distribuição exponencial, $Exp(\theta)$, onde θ é desconhecido.
- (b) X tem uma distribuição binomial $b(n, p)$, onde n é conhecido, mas p é desconhecido.

Alguns exemplos são os seguintes:

- (a) X tem uma distribuição exponencial, $Exp(\theta)$, onde θ é desconhecido.
- (b) X tem uma distribuição binomial $b(n, p)$, onde n é conhecido, mas p é desconhecido.
- (c) X tem uma distribuição gama $\Gamma(\alpha, \beta)$, onde tanto α quanto β são desconhecidos.

Alguns exemplos são os seguintes:

- (a) X tem uma distribuição exponencial, $Exp(\theta)$, onde θ é desconhecido.
- (b) X tem uma distribuição binomial $b(n, p)$, onde n é conhecido, mas p é desconhecido.
- (c) X tem uma distribuição gama $\Gamma(\alpha, \beta)$, onde tanto α quanto β são desconhecidos.
- (d) X tem uma distribuição normal $N(\mu, \sigma^2)$, onde tanto a média μ quanto a variância σ^2 de X são desconhecidas.

Alguns exemplos são os seguintes:

- (a) X tem uma distribuição exponencial, $Exp(\theta)$, onde θ é desconhecido.
- (b) X tem uma distribuição binomial $b(n, p)$, onde n é conhecido, mas p é desconhecido.
- (c) X tem uma distribuição gama $\Gamma(\alpha, \beta)$, onde tanto α quanto β são desconhecidos.
- (d) X tem uma distribuição normal $N(\mu, \sigma^2)$, onde tanto a média μ quanto a variância σ^2 de X são desconhecidas.
- (e) Se a variável X segue uma distribuição normal com média μ desconhecida e variância σ^2 conhecida, a média amostral \bar{X} é uma estimativa pontual para μ .

Função de Densidade e Parâmetros

Descrição

Muitas vezes denotamos este problema dizendo que a variável aleatória X tem uma função de densidade ou massa da forma $f(x; \theta)$ ou $p(x; \theta)$, onde $\theta \in \Omega$ para um conjunto especificado Ω . Por exemplo, se $\Omega = \{\theta \mid \theta > 0\}$, então θ é um parâmetro positivo da distribuição. Queremos estimar θ com base em uma amostra.

Função de Densidade e Parâmetros

Descrição

Muitas vezes denotamos este problema dizendo que a variável aleatória X tem uma função de densidade ou massa da forma $f(x; \theta)$ ou $p(x; \theta)$, onde $\theta \in \Omega$ para um conjunto especificado Ω . Por exemplo, se $\Omega = \{\theta \mid \theta > 0\}$, então θ é um parâmetro positivo da distribuição. Queremos estimar θ com base em uma amostra.

As observações da amostra têm a mesma distribuição que X e as denotamos como as variáveis aleatórias X_1, X_2, \dots, X_n , onde n denota o tamanho da amostra. Quando a amostra é realmente retirada, usamos letras minúsculas x_1, x_2, \dots, x_n como os valores ou realizações da amostra. Muitas vezes, assumimos que as observações da amostra X_1, X_2, \dots, X_n também são mutuamente independentes, caso em que chamamos a amostra de amostra aleatória.

Definição 1

Se as variáveis aleatórias X_1, X_2, \dots, X_n são independentes e identicamente distribuídas (iid), então essas variáveis constituem uma amostra aleatória de tamanho n da distribuição comum.

Definições de Amostra e Estatística

Definição 1

Se as variáveis aleatórias X_1, X_2, \dots, X_n são independentes e identicamente distribuídas (iid), então essas variáveis constituem uma amostra aleatória de tamanho n da distribuição comum.

Definição 2

Seja X_1, X_2, \dots, X_n uma amostra de uma variável aleatória X . Seja $T = T(X_1, X_2, \dots, X_n)$ uma função da amostra. Então T é chamado de Estatística. Uma vez que a amostra é retirada, então t é chamado de realização de T , onde $t = T(x_1, x_2, \dots, x_n)$ e x_1, x_2, \dots, x_n são as realizações da amostra.

Exemplos

- A média amostral $\bar{X} = \frac{1}{n} \sum_{i=1}^n X_i$ é uma estatística que estima a média populacional.
- A variância amostral é uma função dos dados da amostra que estima a variância populacional.

Estimadores Não Viesados

Sejam X_1, X_2, \dots, X_n uma amostra aleatória de uma variável aleatória X com uma função de densidade ou massa da forma $f(x; \theta)$ ou $p(x; \theta)$, onde $\theta \in \Omega$ para um conjunto especificado Ω . Nessa situação, faz sentido considerar uma estatística T , como um estimador de θ . Mais formalmente, T é chamado de estimador pontual de θ . Embora chamemos T de estimador de θ , chamamos sua realização t de estimativa de θ .

Estimadores Não Viesados

Sejam X_1, X_2, \dots, X_n uma amostra aleatória de uma variável aleatória X com uma função de densidade ou massa da forma $f(x; \theta)$ ou $p(x; \theta)$, onde $\theta \in \Omega$ para um conjunto especificado Ω . Nessa situação, faz sentido considerar uma estatística T , como um estimador de θ . Mais formalmente, T é chamado de estimador pontual de θ . Embora chamemos T de estimador de θ , chamamos sua realização t de estimativa de θ .

Observem que para ser um estimador pontual basta ser uma Estatística. Não fazemos nenhuma menção a qualquer correspondência entre o estimador e o parâmetro a ser estimado. No entanto, existem várias propriedades dos estimadores pontuais que discutiremos e algumas técnicas úteis que ajudam a encontrar os melhores candidatos, tais como Viés, Consistência, Eficiência, entre outras.

Observações:

- Se $E(T) \neq \theta$, dizemos que T é um estimador viesado de θ .

Observações:

- Se $E(T) \neq \theta$, dizemos que T é um estimador viesado de θ .
- Se $\lim_{n \rightarrow \infty} E(T) = \theta$, dizemos que T é um estimador assintoticamente não viesado para θ .

Exemplos

Considere X_1, \dots, X_n uma amostra aleatória de uma distribuição com média μ e variância $\sigma^2 < \infty$.

- $\bar{X}_n = \frac{\sum_{i=1}^n X_i}{n}$ é um estimador não viesado para μ ;
- Além disso, $\text{Var}(\bar{X}_n) = \frac{\sigma^2}{n} \xrightarrow{n \rightarrow \infty} 0$;
- $S^2 = \frac{1}{n-1} \sum_{i=1}^n (X_i - \bar{X}_n)^2$ é um estimador não viesado para σ^2 ;

Exemplos

Considere X_1, \dots, X_n uma amostra aleatória de uma distribuição com média μ e variância $\sigma^2 < \infty$.

- $\bar{X}_n = \frac{\sum_{i=1}^n X_i}{n}$ é um estimador não viesado para μ ;
- Além disso, $Var(\bar{X}_n) = \frac{\sigma^2}{n} \xrightarrow{n \rightarrow \infty} 0$;
- $S^2 = \frac{1}{n-1} \sum_{i=1}^n (X_i - \bar{X}_n)^2$ é um estimador não viesado para σ^2 ;

Vejam alguns exemplos práticos, [cliquem aqui!](#)

Distribuição do Máximo de uma Uniforme

Considere $X_i \stackrel{\text{iid}}{\sim} U(0, \theta), i = 1, \dots, n$. Seja $Y_n = \max(X_1, \dots, X_n)$.

$$F_{Y_n}(y) = P(Y_n \leq y) = P(\max(X_1, \dots, X_n) \leq y)$$

Distribuição do Máximo de uma Uniforme

Considere $X_i \stackrel{\text{iid}}{\sim} U(0, \theta)$, $i = 1, \dots, n$. Seja $Y_n = \max(X_1, \dots, X_n)$.

$$\begin{aligned} F_{Y_n}(y) &= P(Y_n \leq y) = P(\max(X_1, \dots, X_n) \leq y) \\ &= P(X_1 \leq y, X_2 \leq y, \dots, X_n \leq y) \end{aligned}$$

Distribuição do Máximo de uma Uniforme

Considere $X_i \stackrel{\text{iid}}{\sim} U(0, \theta)$, $i = 1, \dots, n$. Seja $Y_n = \max(X_1, \dots, X_n)$.

$$\begin{aligned} F_{Y_n}(y) &= P(Y_n \leq y) = P(\max(X_1, \dots, X_n) \leq y) \\ &= P(X_1 \leq y, X_2 \leq y, \dots, X_n \leq y) \\ &= \prod_{i=1}^n P(X_i \leq y) = [P(X_1 \leq y)]^n \end{aligned}$$

Distribuição do Máximo de uma Uniforme

Considere $X_i \stackrel{\text{iid}}{\sim} U(0, \theta)$, $i = 1, \dots, n$. Seja $Y_n = \max(X_1, \dots, X_n)$.

$$\begin{aligned} F_{Y_n}(y) &= P(Y_n \leq y) = P(\max(X_1, \dots, X_n) \leq y) \\ &= P(X_1 \leq y, X_2 \leq y, \dots, X_n \leq y) \\ &= \prod_{i=1}^n P(X_i \leq y) = [P(X_1 \leq y)]^n \\ &= \left(\frac{y}{\theta}\right)^n, \text{ para } y \in (0, \theta). \end{aligned}$$

Exemplos

Logo, a função densidade de Y_n é

$$f_{Y_n(y)} = \frac{dF_{Y_n}(y)}{dy} = \frac{ny^{n-1}}{\theta^n}.$$

Segue que,

$$E(Y_n) = \int_0^\theta yf(y)dy = \frac{n}{\theta^n} \int_0^\theta y^n dy = \frac{n}{n+1}\theta$$

Exemplos

Logo, a função densidade de Y_n é

$$f_{Y_n(y)} = \frac{dF_{Y_n}(y)}{dy} = \frac{ny^{n-1}}{\theta^n}.$$

Segue que,

$$E(Y_n) = \int_0^\theta yf(y)dy = \frac{n}{\theta^n} \int_0^\theta y^n dy = \frac{n}{n+1}\theta$$

Logo,

- Y_n é um estimador viesado para θ ;

Exemplos

Logo, a função densidade de Y_n é

$$f_{Y_n}(y) = \frac{dF_{Y_n}(y)}{dy} = \frac{ny^{n-1}}{\theta^n}.$$

Segue que,

$$E(Y_n) = \int_0^\theta yf(y)dy = \frac{n}{\theta^n} \int_0^\theta y^n dy = \frac{n}{n+1}\theta$$

Logo,

- Y_n é um estimador viesado para θ ;
- Note que Y_n é assintoticamente não viesado para θ e que $\frac{n+1}{n}Y_n$ é um estimador não viesado para θ .

Definição 3

Dizemos que uma sequência de variáveis aleatórias $\{X_n\}_{n \geq 1}$ converge para uma variável aleatória X se, para todo $\varepsilon > 0$,

$$\lim_{n \rightarrow \infty} P(|X_n - X| \geq \varepsilon) = 0, \quad (1)$$

Convergência em Probabilidade

Definição 3

Dizemos que uma sequência de variáveis aleatórias $\{X_n\}_{n \geq 1}$ converge para uma variável aleatória X se, para todo $\varepsilon > 0$,

$$\lim_{n \rightarrow \infty} P(|X_n - X| \geq \varepsilon) = 0, \quad (1)$$

ou equivalentemente,

$$\lim_{n \rightarrow \infty} P(|X_n - X| < \varepsilon) = 1. \quad (2)$$

Se isso ocorrer, escrevemos $X_n \xrightarrow{P} X$.

Convergência em Probabilidade

Definição 3

Dizemos que uma sequência de variáveis aleatórias $\{X_n\}_{n \geq 1}$ converge para uma variável aleatória X se, para todo $\varepsilon > 0$,

$$\lim_{n \rightarrow \infty} P(|X_n - X| \geq \varepsilon) = 0, \quad (1)$$

ou equivalentemente,

$$\lim_{n \rightarrow \infty} P(|X_n - X| < \varepsilon) = 1. \quad (2)$$

Se isso ocorrer, escrevemos $X_n \xrightarrow{P} X$.

Em muitas situações, X é uma v.a. degenerada, ou seja, X é igual a uma constante com probabilidade 1. Vejam alguns exemplos, [Clique aqui](#)!.

Exemplo

Suponha que X_n seja a média amostral de uma sequência de n lançamentos de uma moeda com probabilidade p de sair cara. Ou seja, para cada n , X_n é a fração de caras obtidas após n lançamentos. Por exemplo:

- Para $n = 1$, X_1 será 1 se o primeiro lançamento for cara, ou 0 se for coroa.
- Para $n = 10$, X_{10} será a fração de caras obtidas em 10 lançamentos.
- Para $n = 100$, X_{100} será a fração de caras obtidas em 100 lançamentos, e assim por diante.

Exemplo

Agora, suponha que a moeda seja justa, ou seja, $p = 0.5$. À medida que n aumenta, esperamos que a média X_n dos resultados dos lançamentos (fração de caras) se aproxime de $p = 0.5$, que é o valor esperado da proporção de caras.

Neste caso:

- X_n é a fração de caras nos n lançamentos (a variável aleatória dependente de n).
- X é o valor esperado $p = 0.5$, que é o valor ao qual X_n converge.

Exemplo

Conforme aumentamos o número de lançamentos n , a probabilidade de que a média X_n esteja muito distante de 0.5 (ou seja, $|X_n - 0.5| \geq \varepsilon$ para um $\varepsilon > 0$) diminui. Isso exemplifica que X_n converge em probabilidade para $X = 0.5$.

Ou seja:

$$X_n \xrightarrow{P} 0.5 \quad \text{quando} \quad n \rightarrow \infty.$$

Neste exemplo, X_n representa a fração de caras nos n lançamentos, e X é o valor fixo 0.5, que é o valor esperado a longo prazo.

Desigualdades de Markov e Tchebychev

Desigualdade de Markov

A Desigualdade de Markov afirma que, para uma variável aleatória não negativa Y e um valor positivo a , temos:

$$P(Y \geq a) \leq \frac{\mathbb{E}[Y]}{a}.$$

Desigualdades de Markov e Tchebychev

Desigualdade de Markov

A Desigualdade de Markov afirma que, para uma variável aleatória não negativa Y e um valor positivo a , temos:

$$P(Y \geq a) \leq \frac{\mathbb{E}[Y]}{a}.$$

Desigualdade de Tchebychev

Seja X uma v.a. com média μ e variância σ^2 . Então,

$$P(|X - \mu| \geq t) \leq \frac{\text{Var}(X)}{t^2} = \frac{\sigma^2}{t^2}, \quad t > 0$$

Desigualdades de Markov e Tchebychev

Desigualdade de Markov

A Desigualdade de Markov afirma que, para uma variável aleatória não negativa Y e um valor positivo a , temos:

$$P(Y \geq a) \leq \frac{\mathbb{E}[Y]}{a}.$$

Desigualdade de Tchebychev

Seja X uma v.a. com média μ e variância σ^2 . Então,

$$P(|X - \mu| \geq t) \leq \frac{\text{Var}(X)}{t^2} = \frac{\sigma^2}{t^2}, \quad t > 0$$

Obs: O nome correto é Tchebychev no contexto de português, enquanto Chebyshev é a forma utilizada em inglês. Ambos se referem ao mesmo matemático russo, Pafnuty Chebyshev.

Lei Fraca dos Grandes Números

Teorema 1

Seja $\{X_n\}_{n \geq 1}$ uma sequência de v.a. i.i.d, com média μ e variância $\sigma^2 < \infty$. Então,

$$\bar{X}_n = \frac{\sum_{i=1}^n X_i}{n} \xrightarrow{P} \mu$$

Vejam alguns exemplos práticos, [Clique aqui!](#).

Lei Fraca dos Grandes Números

Teorema 1

Seja $\{X_n\}_{n \geq 1}$ uma sequência de v.a. i.i.d, com média μ e variância $\sigma^2 < \infty$. Então,

$$\bar{X}_n = \frac{\sum_{i=1}^n X_i}{n} \xrightarrow{P} \mu$$

Vejam alguns exemplos práticos, [Clique aqui!](#).

Demonstração:

Seja $\varepsilon > 0$. Temos que,

$$P(|\bar{X}_n - \mu| \geq \varepsilon) = P(|\bar{X}_n - \mu| \geq \varepsilon) \leq \frac{1}{\varepsilon^2} \frac{\sigma^2}{n}$$

↑
Tchebychev

Lei Fraca dos Grandes Números

Teorema 1

Seja $\{X_n\}_{n \geq 1}$ uma sequência de v.a. i.i.d, com média μ e variância $\sigma^2 < \infty$. Então,

$$\bar{X}_n = \frac{\sum_{i=1}^n X_i}{n} \xrightarrow{P} \mu$$

Vejam alguns exemplos práticos, [Clique aqui!](#).

Demonstração:

Seja $\varepsilon > 0$. Temos que,

$$P(|\bar{X}_n - \mu| \geq \varepsilon) = P(|\bar{X}_n - \mu| \geq \varepsilon) \leq \frac{1}{\varepsilon^2} \frac{\sigma^2}{n} \xrightarrow{n \rightarrow \infty} 0$$

↑
Tchebychev

Teorema 2

Suponha que $X_n \xrightarrow{P} X$ e $Y_n \xrightarrow{P} Y$. Então, $X_n + Y_n \xrightarrow{P} X + Y$.

Demonstração:

Para 

Teorema 3

Suponha que X_n converge em probabilidade para X e a é uma constante. Então, aX_n converge em probabilidade para aX .

Demonstração:

Se $a = 0$, o resultado é imediato. Suponha $a \neq 0$. Seja $\varepsilon > 0$. O resultado segue das seguintes igualdades:

$$P[|aX_n - aX| \geq \varepsilon] = P[|a||X_n - X| \geq \varepsilon] = P[|X_n - X| \geq \varepsilon/|a|].$$

E pelas hipóteses, o último termo converge para 0 conforme $n \rightarrow \infty$.

Teorema 4

Suponha que X_n converge em probabilidade para a e a função real g é contínua em a . Então, $g(X_n)$ converge em probabilidade para $g(a)$.

Demonstração:

Seja $\varepsilon > 0$. Como g é contínua em a , existe $\delta > 0$ tal que se $|x - a| < \delta$, então $|g(x) - g(a)| < \varepsilon$. Assim,

$$|g(x) - g(a)| \geq \varepsilon \Rightarrow |x - a| \geq \delta.$$

Substituindo X_n por x na implicação acima, obtemos

$$P[|g(X_n) - g(a)| \geq \varepsilon] \leq P[|X_n - a| \geq \delta].$$

Pelas hipóteses, o último termo converge para 0 conforme $n \rightarrow \infty$, o que nos dá o resultado.

Este teorema nos fornece muitos resultados úteis. Por exemplo, se X_n converge em probabilidade para a , então:

- X_n^2 converge em probabilidade para a^2 .
- $\frac{1}{X_n}$ converge em probabilidade para $\frac{1}{a}$, desde que $a \neq 0$.
- X_n converge em probabilidade para \sqrt{a} , desde que $a \geq 0$.

Teorema 5

Se X_n converge em probabilidade para X e g é uma função contínua, então $g(X_n)$ converge em probabilidade para $g(X)$.

Demonstração:

Página 104 de Tucker (1967).

Teorema 6

Suponha que X_n converge em probabilidade para X e Y_n converge em probabilidade para Y . Então, $X_n Y_n$ converge em probabilidade para XY .

Demonstração:

Utilizando os resultados anteriores, temos:

$$\begin{aligned} X_n Y_n &= \frac{1}{2} X_n^2 + \frac{1}{2} Y_n^2 - \frac{1}{2} (X_n - Y_n)^2 \\ &\xrightarrow{P} \frac{1}{2} X^2 + \frac{1}{2} Y^2 - \frac{1}{2} (X - Y)^2 \\ &= XY. \end{aligned}$$

Referências I

CASELLA, George; BERGER, Roger L. **Statistical inference**. [S.l.]: Cengage Learning, 2021.

HOGG, RV; MCKEAN, J; CRAIG, AT. **Introduction to Mathematical Statistics**. Eighth Edition. [S.l.]: Pearson, 2019.

SALSBURG, David S. **Uma senhora toma chá... como a estatística revolucionou a ciência no século XX**. [S.l.]: Zahar Rio de Janeiro, 2009.