



UNIVERSIDADE ESTADUAL DA PARAÍBA
CENTRO DE CIÊNCIAS E TECNOLOGIA
DEPARTAMENTO DE ESTATÍSTICA

Felipe Matheus Gonçalves Costa

Métodos Assintóticos em Estatística

Campina Grande - PB

Novembro, 2017

Felipe Matheus Gonçalves Costa

Métodos Assintóticos em Estatística

Trabalho de Conclusão de Curso apresentado ao curso de Bacharelado em Estatística do Departamento de Estatística do Centro de Ciências e Tecnologia da Universidade Estadual da Paraíba em cumprimento às exigências legais para obtenção do título de bacharel em Estatística.

Orientador: Prof. Dr^a. Divanilda Maia Esteves

Coorientador: Prof. Dr. Gustavo Henrique Esteves

Campina Grande - PB

Novembro, 2017

É expressamente proibido a comercialização deste documento, tanto na forma impressa como eletrônica. Sua reprodução total ou parcial é permitida exclusivamente para fins acadêmicos e científicos, desde que na reprodução figure a identificação do autor, título, instituição e ano do trabalho.

C837m Costa, Felipe Matheus Goncalves.
Métodos assintóticos em estatística [manuscrito] : / Felipe Matheus Goncalves Costa. - 2017.
35 p. : il. colorido.

Digitado.

Trabalho de Conclusão de Curso (Graduação em Estatística) - Universidade Estadual da Paraíba, Centro de Ciências e Tecnologia, 2017.

"Orientação : Profa. Dra. Divanilda Maia Esteves, Coordenação do Curso de Estatística - CCT."

"Coorientação: Prof. Dr. Gustavo Henrique Esteves, Coordenação do Curso de Estatística - CCT.""

1. Inferência estatística. 2. Convergência estocástica. 3. Teorema Central do limite.

21. ed. CDD 519.54

Felipe Matheus Gonçalves Costa

Métodos Assintóticos em Estatística

Trabalho de Conclusão de Curso apresentado ao curso de Bacharelado em Estatística do Departamento de Estatística do Centro de Ciências e Tecnologia da Universidade Estadual da Paraíba em cumprimento às exigências legais para obtenção do título de bacharel em Estatística.

Trabalho aprovado em 30 de novembro de 2017.

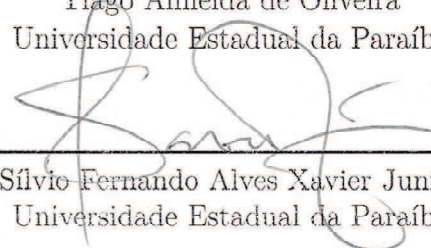
BANCA EXAMINADORA



Divanilda Maia Esteves
Universidade Estadual da Paraíba



Tiago Almeida de Oliveira
Universidade Estadual da Paraíba



Sílvio Fernando Alves Xavier Junior
Universidade Estadual da Paraíba

Dedico este Trabalho de Conclusão de Curso, primeiramente a Deus, por toda força e coragem durante essa longa caminhada. A minha família, que não mediu esforços para que eu chegasse até aqui. Aos meus professores, em especial a minha orientadora Diana Maia Esteves, pela paciência, incentivo e orientação que tornaram possível a conclusão do meu curso.

Agradecimentos

Agradeço a Deus por ter me dado saúde e força para superar as dificuldades durante a minha vida, não somente durante a minha formação.

A Universidade Estadual da Paraíba pela oportunidade de fazer o curso.

A todo o corpo docente por me proporcionar o conhecimento necessário no processo da minha formação proporcional.

A minha orientadora Diana Maia Esteves pelo apoio e confiança na elaboração deste Trabalho.

A minha família pelo incentivo nas horas difíceis, de desânimo e cansaço, e por terem apoiado os momentos de estresse durante a minha formação.

Aos amigos, sejam eles os de infância ou os feitos durante a minha formação, pelos momentos de descontração.

*“Viva como se fosse morrer amanhã.
Aprenda como se fosse viver para sempre.”
(Mahatma Gandhi)*

Resumo

O objetivo deste trabalho foi estudar teoricamente e ilustrar através de simulação alguns resultados assintóticos importantes para a Estatística. A inferência se dedica a avaliar o comportamento de características de um determinado conjunto de interesse, com base na informação obtida numa pequena parte deste conjunto. Em outras palavras, deseja-se obter conclusões sobre as características de uma população através das informações oriundas da amostra, por meio de uma modelagem estatística, onde leva-se em consideração a problematização e o conhecimento prévio das características em estudo. Quanto maior a amostra, espera-se que as conclusões representem melhor o que acontece com a população. Os Métodos Assintóticos estudam o comportamento de estatísticas quando os tamanhos das amostras tendem ao infinito. Obviamente, na prática, as amostras são finitas, mas torna-se natural pensar em quão grandes devem ser estas amostras para que os resultados sejam satisfatórios. Partindo deste princípio, estudou-se as formas de convergência estocástica, bem como os principais resultados associados a referida área de estudo, ilustrou-se o Teorema Central do Limite e a Lei dos Grandes Números e analisou-se o comportamento assintótico de alguns estimadores e estatísticas de teste. Como base para o estudo dos Métodos Assintóticos foram utilizados conceitos matemáticos e probabilísticos como: Ordens de Magnitude, Expansões de Taylor, Funções Características, Modos de Convergência Estocástica, entre outros. Após desenvolvida toda a teoria buscou-se aplicá-la através de dados simulados, utilizando-se do software R para realização de tal tarefa. Foi feito o uso das distribuições de Poisson e Exponencial para mostrar graficamente seus comportamentos assintóticos. Ao fim do trabalho os resultados evidenciaram o quão robusta é a teoria por trás desta área da Estatística e do quanto é necessário ser dado o merecido reconhecimento as pessoas que dedicaram suas vidas para desenvolvê-la.

Palavras-chaves: Inferência Estatística. Convergência Estocástica. Teorema Central do Limite.

Abstract

The objective of this work was to study theoretically and to illustrate by simulation some important asymptotic results for the Statistics. Inference is concerned with evaluating the behavior of characteristics of a particular set of interest, based on the information obtained in a small part of this set. In other words, it is desired to obtain conclusions about the characteristics of a population through the information from the sample, through a statistical modeling, where it is taken into account the problematization and the previous knowledge of the characteristics under study. The larger the sample, the better the results are expected to represent what happens to the population. Asymptotic Methods study the behavior of statistics when sample sizes tend to infinity. Obviously, in practice, the samples are finite, but it becomes natural to think about how large these samples should be so that the results are satisfactory. Based on this principle, we studied the forms of stochastic convergence, as well as the main results associated with this area of study, the Limit Central Theorem and the Large Numbers Law were illustrated and the asymptotic behavior of some estimators and test statistics. As a basis for the study of Asymptotic Methods we used mathematical and probabilistic concepts such as: Magnitude Orders, Taylor Expansions, Characteristic Functions, Stochastic Convergence Modes, among others. After the entire theory was developed, it was applied through simulated data, using R software to perform such task. The Poisson and Exponential distributions were used to graphically show their asymptotic behavior. At the end of the paper the results showed how robust the theory behind this area of statistics is and how much appreciation has to be given to the people who have dedicated their lives to developing it.

Key-words: Statistical Inference. Stochastic Convergence. Central Limit Theorem.

Lista de ilustrações

Figura 1 – Histograma e Normal Q-Q Plot da média amostral da distribuição de Poisson, para amostras de tamanho $n = 10$	26
Figura 2 – Histograma e Normal Q-Q Plot da média amostral da distribuição de Poisson, para amostras de tamanho $n = 30$	26
Figura 3 – Histograma e Normal Q-Q Plot da média amostral da distribuição de Poisson, para amostras de tamanho $n = 100$	27
Figura 4 – Histograma e Normal Q-Q Plot da média amostral da distribuição de Poisson, para amostras de tamanho $n = 500$	27
Figura 5 – Histograma e Normal Q-Q Plot da média amostral da distribuição Exponencial, para amostras de tamanho $n = 10$	27
Figura 6 – Histograma e Normal Q-Q Plot da média amostral da distribuição Exponencial, para amostras de tamanho $n = 30$	28
Figura 7 – Histograma e Normal Q-Q Plot da média amostral da distribuição Exponencial, para amostras de tamanho $n = 100$	28
Figura 8 – Histograma e Normal Q-Q Plot da média amostral da distribuição Exponencial, para amostras de tamanho $n = 500$	29

Sumário

1	INTRODUÇÃO	11
2	FUNDAMENTAÇÃO TEÓRICA	12
2.1	Funções Características	12
2.2	Desigualdades Estatísticas	14
2.3	Modos de Convergência Estocástica	15
2.3.1	Convergência em Probabilidade	15
2.3.2	Convergência quase certa	15
2.3.3	Convergência em média de ordem r	16
2.3.4	Convergência em Distribuição	17
2.4	Lei dos Grandes Números	19
2.5	Teorema Central do Limite	20
2.6	Comportamento Assintótico de Estimadores e Estatísticas de Teste	22
2.6.1	Comportamento Assintótico de Estimadores de Máxima Verossimilhança	22
2.7	Distribuições de Probabilidade	24
2.7.1	Distribuição de Poisson	24
2.7.2	Distribuição Exponencial	24
3	APLICAÇÃO	26
4	CONCLUSÃO	30
	REFERÊNCIAS	31
	APÊNDICES	32
	APÊNDICE A – ROTINA UTILIZADA PARA A SIMULAÇÃO	33

1 Introdução

Frequentemente, na estatística deseja-se estudar uma determinada característica de uma população. Sendo que, geralmente, é inviável o estudo sobre o grupo como um todo, considera-se apenas uma amostra do mesmo. Observa-se o que acontece nessa amostra e infere-se à população através de técnicas adequadas derivadas de dois ramos principais: a estimação de parâmetros e os testes de hipóteses. Segundo Lehmann e Romano (2005), a Inferência Estatística diz respeito a utilização dessas técnicas para obter informações sobre a variável X ou o parâmetro θ com o qual está relacionada.

Cada técnica tem sua importância na conquista da confiabilidade dos resultados obtidos. Muitos procedimentos na estimação de parâmetros dependem de um tamanho suficientemente grande da amostra para que os estimadores tenham propriedades ótimas, dado que o valor de n é um fator de grande influência no ajuste de um modelo probabilístico aos dados.

De acordo com Leite e Singer (1990), grande parte da modelagem estatística usa resultados assintóticos, ou seja, válidos no limite, quando os tamanhos das amostras tendem ao infinito. Sendo assim deve-se saber qual o tamanho mínimo que uma amostra deve ter para que o uso de tais resultados seja possível.

Este trabalho desenvolve-se no estudo dos principais teoremas de probabilidade relacionados aos métodos assintóticos e nos resultados adjuntos a verificação das convergências estocásticas associadas a esses teoremas, seguindo com o estudo do comportamento assintótico de alguns estimadores frequentemente utilizados, simulando dados para devidas ilustrações.

2 Fundamentação Teórica

2.1 Funções Características

A Função Característica de uma variável aleatória X é uma média estatística que pode ser considerada como a transformada de Fourier da função de densidade de probabilidade.

Se X e Y forem variáveis aleatórias definidas em um mesmo espaço de probabilidade, então $Z = X + iY$ será uma variável aleatória assumindo valores no conjunto dos números complexos, \mathbf{C} . Se $E(X) < \infty$ e $E(Y) < \infty$, a esperança de Z pode ser definida como $E(Z) = E(X) + iE(Y)$, (LEITE; SINGER, 1990, p. 28).

Uma vez que $e^{iX} = \cos X + i\text{sen}X$, então, $E(e^{iX}) = E(\cos X) + iE(\text{sen}X)$ é um número complexo bem definido. Por outro lado, pela linearidade da integral de Stieltjes para o caso de funções complexas, temos:

$$\int e^{iX} dF_X(x) = \int \cos x dF_X(x) + i \int \text{sen} x dF_X(x) = E(\cos X) + iE(\text{sen}X) = E(e^{iX})$$

Assim a seguinte definição garante que a função característica de qualquer variável aleatória esteja bem definida.

Definição 2.1.1 *Seja X uma variável aleatória. A função característica de X é a função $\varphi_X : \mathbf{R} \rightarrow \mathbf{C}$ definida por $\varphi_X(t) = \int e^{itx} dF_X(x) \forall t \in \mathbf{R}$.*

A seguir encontram-se listadas algumas propriedades importantes das Funções Características.

Propriedade 2.1.1 $|\varphi_X(t)| \leq 1$ para todo $t \in \mathbf{R}$.

Propriedade 2.1.2 $\varphi_X(0) = 1$ e $\overline{\varphi_X}(t) = \varphi_X(-t)$ para todo $t \in \mathbf{R}$, onde $\overline{\varphi_X}(t)$ é o complexo conjugado de $\varphi_X(t)$.

Propriedade 2.1.3 φ_X é uniformemente contínua.

Propriedade 2.1.4 Se X_1, X_2, \dots, X_n forem variáveis aleatórias independentes, então

$$\varphi_{\sum_{j=1}^n x_j} = \prod_{j=1}^n \varphi_{X_j}(t), \forall t \in \mathbf{R}$$

"Conhecendo a função de distribuição de X , a função característica pode ser obtida. A recíproca também é verdadeira e a função característica determina univocamente a função de distribuição"(CORDEIRO, 1999, p. 34). Isso corresponde a seguinte propriedade.

Propriedade 2.1.5 *A função característica de uma variável aleatória X determina a função de distribuição de X .*

Propriedade 2.1.6 *Uma variável aleatória X tem distribuição simétrica em torno de zero, isto é, $P\{X \leq x\} = P\{X \geq -x\}, \forall x \in \mathbf{R}$ se, e somente se $\varphi_X(t)$ for real, $\forall t \in \mathbf{R}$.*

Propriedade 2.1.7 *Dados $a, b \in \mathbf{R}$ arbitrários $\varphi_{aX+b}(t) = e^{ibt}\varphi_X(at)$ para todo $t \in \mathbf{R}$.*

Propriedade 2.1.8 *Se $E|X|^n < \infty$, então φ_X tem n derivadas contínuas, com $\varphi_X^{(k)}(t) = \int (ix)^k e^{itx} dF_X(x), k = 1, 2, \dots, n$. Particularmente, $\varphi_X^{(k)}(0) = i^k E(X^k)$.*

O conceito de função característica pode ser generalizado para vetores aleatórios.

Definição 2.1.2 *A função característica de X é a função $\varphi_X : \mathbf{R}^p \rightarrow \mathbf{C}$, definida por:*

$$\varphi_X(t) = E(e^{i\sum_{j=1}^p t_j X_j}) = E(e^{itX^t}), \forall t = (t_1, t_2, \dots, t_p) \in \mathbf{R}^p$$

Todas as propriedades apresentadas até aqui, exceto a **Propriedade 2.1.5**, são válidas para o caso de vetores aleatórios. Um teorema equivalente a esta propriedade pode ser observado à seguir.

Teorema 2.1.1 *Se X e Y forem vetores aleatórios p -dimensionais tais que $\varphi_X(t) = \varphi_Y(t)$ para todo $t \in \mathbf{R}^p$, então X e Y têm a mesma distribuição.*

Para verificarmos a independência entre dois vetores aleatórios, através das suas funções características, apresentamos o seguinte resultado.

Teorema 2.1.2 *Sejam $\mathbf{X} = (X_1, X_2, \dots, X_p)$ e $\mathbf{Y} = (Y_1, Y_2, \dots, Y_p)$ vetores aleatórios. \mathbf{X} e \mathbf{Y} são independentes se, e somente se,*

$$\varphi_{(\mathbf{X}, \mathbf{Y})}(s, t) = \varphi_X(s)\varphi_Y(t), \forall s \in \mathbf{R}^p, \forall t \in \mathbf{R}^q. \quad (2.1)$$

2.2 Desigualdades Estatísticas

Antes de dar início ao estudo dos modos de convergência, a apresentação de algumas desigualdades importantes como, por exemplo, as desigualdades de Markov e Chebyshev, se faz necessária.

Proposição 2.2.1 (Desigualdade de Jensen) *Seja X uma variável aleatória com esperança finita e g uma função convexa, então*

$$E(g(X)) \geq g(E(X)).$$

Proposição 2.2.2 (Desigualdade Básica de Chebyshev) *Seja X uma variável aleatória não negativa e considere $a > 0$, então*

$$P(X \geq a) \leq \frac{E(X)}{a}.$$

Proposição 2.2.3 (Desigualdade Clássica de Chebyshev) *Seja X uma variável aleatória com média μ , então, para qualquer $t > 0$, temos*

$$P(|X - \mu| \geq t) \leq \frac{\text{Var}(X)}{t^2}.$$

Proposição 2.2.4 (Desigualdade de Markov) *Seja X uma variável aleatória qualquer, então, para quaisquer $t, k > 0$, temos*

$$P(|X| \geq t) \leq \frac{E(|X|^k)}{t^k}.$$

Proposição 2.2.5 (Desigualdade de Cauchy-Schwarz) *Seja X e Y variáveis aleatórias com segundo momento finito, então*

$$E(|XY|) \leq \sqrt{E(X^2)E(Y^2)};$$

com a igualdade valendo se, e somente se, $Y = aX$, para alguma constante a .

Proposição 2.2.6 (Desigualdade de Kolmogorov) *Seja $\{X_n\}_{n \geq 1}$ uma sequência de variáveis aleatórias independentes com média zero e variância finita, então, para todo $\lambda > 0$ e com $S_k = X_1, X_2, \dots, X_k$, temos*

$$P(\max_{1 \leq k \leq n} |S_k| \geq \lambda) \leq \frac{\text{Var}(S_n)}{\lambda^2}.$$

Essas desigualdades são importantes pelo fato de serem ferramentas bastante utilizadas nas demonstrações dos resultados assintóticos. "A importância das desigualdades de Chebyshev e Markov está no fato de elas nos permitirem deduzir limites para as probabilidades quando conhecemos somente a média, ou a média e a variância, da distribuição de probabilidade"(ROSS, 2010, p. 460).

2.3 Modos de Convergência Estocástica

A seguir serão apresentados os principais tipos de Convergência Estocástica.

2.3.1 Convergência em Probabilidade

Definição 2.3.1 Uma sequência de variáveis aleatórias $\{X_n\}_{n \geq 1}$ **converge em probabilidade** para uma variável aleatória X (possivelmente degenerada) se para todo número real $\epsilon > 0$, $P(|X_n - X| \geq \epsilon) \rightarrow 0$ quando $n \rightarrow \infty$.

A notação $X_n \xrightarrow{P} X$ é usada para indicar que $\{X_n\}_{n \geq 1}$ converge em probabilidade para X , ou seja, "para todo n suficientemente grande, X_n e X , são aproximadamente iguais com alta probabilidade"(LEITE; SINGER, 1990, p. 48).

Para generalizar esse conceito para o caso multidimensional, deve ser observada a seguinte definição.

Definição 2.3.2 Uma sequência de vetores aleatórios $\{\mathbf{X}_n\}_{n \geq 1} = \{(X_{n1}, \dots, X_{np})\}_{n \geq 1}$ **converge em probabilidade** para um vetor aleatório $\mathbf{X}_0 = (X_{01}, X_{02}, \dots, X_{0p})$ (possivelmente degenerado), se a sequência de variáveis aleatórias $\{\|\mathbf{X}_n - \mathbf{X}_0\|\}_{n \geq 1}$ convergir em probabilidade para zero.

A notação $\mathbf{X}_n \xrightarrow{P} \mathbf{X}_0$ indica que $\{\mathbf{X}_n\}_{n \geq 1}$ converge em probabilidade \mathbf{X}_0 .

Para reduzir o estudo da convergência em probabilidade do caso multidimensional para o caso unidimensional, deve ser levado em consideração o seguinte resultado.

Teorema 2.3.1 Seja $\{\mathbf{X}_n\}_{n \geq 1}$ uma sequência de vetores aleatórios e \mathbf{X}_0 um vetor aleatório, então:

$$\mathbf{X}_n \xrightarrow{P} \mathbf{X}_0 \text{ se, e somente se } \mathbf{X}_{nj} \xrightarrow{P} \mathbf{X}_{0j}, \text{ para } j = 1, 2, \dots, p.$$

2.3.2 Convergência quase certa

Definição 2.3.3 Uma sequência de variáveis aleatórias $\{X_n\}_{n \geq 1}$ **converge quase certamente** para uma variável X se $P(X_n \rightarrow X) = 1$.

A notação $X_n \xrightarrow{q.c.} X$ é usada para indicar que $\{X_n\}_{n \geq 1}$ converge quase certamente para X . O próximo teorema é muito útil no estudo da convergência quase certa de variáveis aleatórias.

Teorema 2.3.2 Seja $\{X_n\}_{n \geq 1}$ uma sequência de variáveis aleatórias, então $X_n \xrightarrow{q.c.} 0$ se, e somente se $P(|X_n| \geq \frac{1}{m} \text{ infinitas vezes}) = 0$, $m = 1, 2, \dots$

Como consequência desse teorema, apresentamos os dois próximos resultados.

Teorema 2.3.3 *Uma sequência de variáveis aleatórias $\{X_n\}_{n \geq 1}$ converge quase certamente para uma variável X se, e somente se, para todo número real $\epsilon > 0$, $P(|X_k - X| \geq \epsilon \text{ para algum } k \geq n) \rightarrow 0$ quando $n \rightarrow \infty$.*

Teorema 2.3.4 *Sejam $\{X_n\}_{n \geq 1}$ uma sequência de variáveis aleatórias e X uma variável aleatória. Se $\sum_{n=1}^{\infty} P(|X_n - X| \geq \epsilon) < \infty$ para todo número real $\epsilon > 0$, então $X_n \xrightarrow{q.c.} X$.*

Sabe-se que a convergência em probabilidade não implica em convergência quase certa. "O fato de que convergência quase certa é um conceito mais forte que convergência em probabilidade pode ser observado através do seguinte teorema" (LEITE; SINGER, 1990, p. 51).

Teorema 2.3.5 *Sejam $\{X_n\}_{n \geq 1}$ uma sequência de variáveis aleatórias e X uma variável aleatória. Se $X_n \xrightarrow{q.c.} X$, então $X_n \xrightarrow{P} X$.*

Assim como a convergência em probabilidade, o conceito de convergência quase certa também pode ser estendido para o caso vetorial.

Definição 2.3.4 *Diremos que a sequência $\{\mathbf{X}_n\}_{n \geq 1} = \{(X_{n1}, \dots, X_{np})\}_{n \geq 1}$ de vetores aleatórios **converge quase certamente** para um vetor aleatório (possivelmente degenerado) \mathbf{X}_0 se a sequência de variáveis aleatórias $\{\|\mathbf{X}_n - \mathbf{X}_0\|\}_{n \geq 1}$ convergir quase certamente para zero.*

Para indicar que $\{\mathbf{X}_n\}_{n \geq 1}$ converge quase certamente para \mathbf{X}_0 usa-se a notação: $\mathbf{X}_n \xrightarrow{q.c.} \mathbf{X}_0$.

Para facilitar o estudo da convergência quase certa, pode-se reduzir a convergência do caso multidimensional para o caso unidimensional através do seguinte teorema.

Teorema 2.3.6 *Seja $\{\mathbf{X}_n\}_{n \geq 1}$ uma sequência de vetores aleatórios e \mathbf{X}_0 um vetor aleatório, então:*

$$\mathbf{X}_n \xrightarrow{q.c.} \mathbf{X}_0 \text{ se, e somente se } \mathbf{X}_{nj} \xrightarrow{q.c.} \mathbf{X}_{0j}, \text{ para } j = 1, 2, \dots, p.$$

2.3.3 Convergência em média de ordem r

Definição 2.3.5 *Sejam X, X_1, X_2, \dots variáveis aleatórias e r um número real positivo tais que $E(|X|^r) < \infty$ e $E(|X_n|^r) < \infty, \forall n \geq 1$. Diremos que a sequência $\{X_n\}_{n \geq 1}$ **converge em média de ordem r** para X se $E(|X_n - X|^r) \rightarrow 0$ quando $n \rightarrow \infty$.*

Usa-se a notação $X_n \xrightarrow{m.r.} X$ para indicar que $\{X_n\}_{n \geq 1}$ converge em média de ordem r para X .

Segundo Leite e Singer (1990), convergência em quase toda parte (ou em probabilidade) não implica em convergência em média. Porém, podemos levar em consideração o seguinte resultado.

Teorema 2.3.7 *Sejam $\{X_n\}_{n \geq 1}$ uma sequência de variáveis aleatórias, X uma variável aleatória e r um número real positivo;*

i) se $X_n \xrightarrow{m.r.} X$, então $X_n \xrightarrow{P} X$;

ii) se $\sum_{n=1}^{\infty} E(|X_n - X|^r) < \infty$, então $X_n \xrightarrow{q.c.} X$.

2.3.4 Convergência em Distribuição

Definição 2.3.6 *Sejam X, X_1, X_2, \dots variáveis aleatórias cujas funções de distribuição são, respectivamente F, F_1, F_2, \dots . Diremos que a sequência $\{X_n\}_{n \geq 1}$ **converge em distribuição** para X se, para todo ponto x de continuidade de F , $F_n(x) \rightarrow F(x)$ quando $n \rightarrow \infty$.*

Usa-se a notação $X_n \xrightarrow{D} X$ para indicar que $\{X_n\}_{n \geq 1}$ converge em distribuição para X .

O teorema a seguir apresenta uma condição necessária e suficiente para que uma sequência de variáveis aleatórias convirja em distribuição.

Teorema 2.3.8 *Sejam X, X_1, X_2, \dots variáveis aleatórias cujas funções de distribuição são F, F_1, F_2, \dots e funções características são $\varphi, \varphi_1, \varphi_2, \dots$, respectivamente, então, $X_n \xrightarrow{D} X$ se, e somente se, para todo $t \in \mathbf{R}$, $\varphi_n(t) \rightarrow \varphi(t)$ quando $n \rightarrow \infty$.*

Como já visto, convergência quase certa ou convergência em média, implicam em convergência em probabilidade, mas a recíproca, em geral, não é verdadeira. A convergência em distribuição é o tipo mais fraco de convergência, "no sentido de que é consequência tanto da convergência em probabilidade como da convergência quase certa" (JAMES, 2002, p. 231). Isso pode ser observado através do seguinte teorema.

Teorema 2.3.9 *Sejam X, X_1, X_2, \dots variáveis aleatórias. Se $X_n \xrightarrow{P} X$, então $X_n \xrightarrow{D} X$.*

O teorema a seguir mostra que é válido, mesmo que a convergência em distribuição não resulte em convergência em probabilidade, na maioria dos casos.

Teorema 2.3.10 *Seja $\{X_n\}_{n \geq 1}$ uma seqüência de variáveis aleatórias e X uma variável aleatória degenerada em um ponto. Se $X_n \xrightarrow{D} X$, então $X_n \xrightarrow{P} X$.*

Através da seguinte definição pode-se expandir o conceito de convergência em distribuição para o caso vetorial.

Definição 2.3.7 *Sejam $\mathbf{X}_n = (X_{n1}, \dots, X_{np})$, $n = 1, 2, \dots$ e $\mathbf{X}_0 = (X_{01}, \dots, X_{0p})$ vetores aleatórios p -dimensionais cujas funções de distribuição são F_{X_0}, F_{X_1}, \dots respectivamente. Diremos que a seqüência $\{\mathbf{X}_n\}_{n \geq 1}$ **converge em distribuição** para \mathbf{X}_0 se, para todo ponto $\mathbf{x} = (x_1, x_2, \dots, x_p)$ de continuidade de F_{X_0} , $F_{X_n}(\mathbf{x}) \rightarrow F_{X_0}(\mathbf{x})$ quando $n \rightarrow \infty$.*

Usa-se a notação: $\mathbf{X}_n \xrightarrow{D} \mathbf{X}_0$ para indicar que $\{\mathbf{X}_n\}_{n \geq 1}$ converge em distribuição para \mathbf{X}_0 .

Pode-se expandir o **Teorema 2.3.8** para o caso multidimensional, através do seguinte teorema.

Teorema 2.3.11 *Sejam $\mathbf{X}_n = (X_{n1}, \dots, X_{np})$, $n = 1, 2, \dots$ e $\mathbf{X}_0 = (X_{01}, \dots, X_{0p})$ vetores aleatórios p -dimensionais cujas funções características são $\varphi_{X_0}, \varphi_{X_1}, \dots$, respectivamente. Então, $\mathbf{X}_n \xrightarrow{D} \mathbf{X}_0$ se, e somente se, para todo $t \in \mathbf{R}^p$, $\varphi_{X_n}(t) \rightarrow \varphi_{X_0}(t)$ quando $n \rightarrow \infty$.*

O problema do estudo da convergência em distribuição no caso multidimensional pode ser reduzido ao caso unidimensional por meio do teorema subsequente.

Teorema 2.3.12 (Cramér-Wold) *Sejam $\mathbf{X}_n = (X_{n1}, \dots, X_{np})$, $n = 1, 2, \dots$ e $\mathbf{X}_0 = (X_{01}, \dots, X_{0p})$ vetores aleatórios p -dimensionais. Então $\mathbf{X}_n \xrightarrow{D} \mathbf{X}_0$ se, e somente se*

$$tX_n^t = \sum_{j=1}^p t_j X_{nj} \xrightarrow{D} \sum_{j=1}^p t_j X_{0j} = tX_0^t, \forall t = (t_1, t_2, \dots, t_p) \in \mathbf{R}^p.$$

Os próximos resultados ajudam na solução de problemas frequentes envolvendo convergência estocástica.

Teorema 2.3.13 *Sejam $\mathbf{X}_n = (X_{n1}, \dots, X_{np})$, $n = 1, 2, \dots$ e $\mathbf{X}_0 = (X_{01}, \dots, X_{0p})$ vetores aleatórios p -dimensionais definidos em um mesmo espaço de probabilidade e $f : \mathbf{R}^p \rightarrow \mathbf{R}^m$ ($m \geq 1$) uma função contínua. Então:*

- i) $\mathbf{X}_n \xrightarrow{q.c.} \mathbf{X}_0 \implies f(\mathbf{X}_n) \xrightarrow{q.c.} f(\mathbf{X}_0)$;*
- ii) $\mathbf{X}_n \xrightarrow{P} \mathbf{X}_0 \implies f(\mathbf{X}_n) \xrightarrow{P} f(\mathbf{X}_0)$;*
- iii) $\mathbf{X}_n \xrightarrow{D} \mathbf{X}_0 \implies f(\mathbf{X}_n) \xrightarrow{D} f(\mathbf{X}_0)$.*

Como consequência do **Teorema 2.3.13**, surgem outros dois resultados imediatos.

Teorema 2.3.14 *Sejam $\{X_n\}_{n \geq 1}$ e $\{Y_n\}_{n \geq 1}$ seqüências de variáveis aleatórias e X e Y variáveis aleatórias*

i) Se $X_n \xrightarrow{q.c.} X$ e $Y_n \xrightarrow{q.c.} Y$, então $X_n + Y_n \xrightarrow{q.c.} X + Y$ e $X_n Y_n \xrightarrow{q.c.} XY$;

ii) Se $X_n \xrightarrow{P} X$ e $Y_n \xrightarrow{P} Y$, então $X_n + Y_n \xrightarrow{P} X + Y$ e $X_n Y_n \xrightarrow{P} XY$.

Teorema 2.3.15 (Slutsky) *Sejam X, X_1, X_2, \dots e Y, Y_1, Y_2, \dots , variáveis aleatórias tais que $X_n \xrightarrow{D} X$ e $Y_n \xrightarrow{P} c$, onde c é uma constante. Então,*

i) $X_n + Y_n \xrightarrow{D} X + c$;

ii) $X_n Y_n \xrightarrow{D} cX$;

iii) se $c \neq 0$, $X_n/Y_n \xrightarrow{D} X/c$.

2.4 Lei dos Grandes Números

De forma simples, a Lei dos Grandes Números nos diz que a média das observações de uma variável aleatória tende a esperança desta variável.

Definição 2.4.1 *Sejam X_1, X_2, \dots , variáveis aleatórias definidas em um mesmo espaço de probabilidade, tais que $E(|X_n|) < \infty, n = 1, 2, \dots$ e sejam $T_n = \sum_{j=1}^n X_j, n = 1, 2, \dots$. Diremos que a seqüência $\{X_n\}_{n \geq 1}$ satisfaz a Lei Fraca dos Grandes Números se $\frac{T_n - E(T_n)}{n} \xrightarrow{P} 0$, e que satisfaz a Lei Forte dos Grandes Números se $\frac{T_n - E(T_n)}{n} \xrightarrow{q.c.} 0$.*

"Se uma seqüência de variáveis aleatórias $\{X_n\}_{n \geq 1}$ satisfaz a Lei Forte dos Grandes Números, então, ela também satisfaz a Lei Fraca dos Grandes Números" (MAGALHÃES, 2006, p. 324). Se as variáveis aleatórias têm a mesma média finita μ , então $\{X_n\}_{n \geq 1}$ satisfaz a Lei Fraca se, e somente se, $\frac{T_n}{n} \xrightarrow{P} \mu$, e satisfaz a Lei Forte se, e somente se, $\frac{T_n}{n} \xrightarrow{q.c.} \mu$.

A seguir serão apresentadas variações das Leis dos Grandes Números, sob diferentes suposições.

Teorema 2.4.1 (Lei Fraca dos Grandes Números de Chebychev) *Sejam X_1, X_2, \dots uma seqüência de variáveis aleatórias independentes com médias μ_1, μ_2, \dots e variâncias $\sigma_1, \sigma_2, \dots$, respectivamente. Se $n^{-2} \text{Var}(T_n) = n^{-2} \sum_{j=1}^n \sigma_j^2 \rightarrow 0$ quando $n \rightarrow \infty$, então $\{X_n\}_{n \geq 1}$ satisfaz a Lei Fraca dos Grandes Números, ou seja, $n^{-1} \{T_n - \sum_{j=1}^n \mu_j\} \xrightarrow{P} 0$.*

Teorema 2.4.2 (Lei Fraca dos Grandes Números de Khintchine) *Se $\{X_n\}_{n \geq 1}$ for uma seqüência de variáveis aleatórias independentes e identicamente distribuídas com $E(X) = \mu < \infty$, então $T_n/n \xrightarrow{P} \mu$.*

Teorema 2.4.3 (Lei Fraca dos Grandes Números de Markov) *Sejam X_1, X_2, \dots variáveis aleatórias independentes com médias μ_1, μ_2, \dots finitas e tais que, para algum $\delta, 0 < \delta \leq 1$ vale a condição de Markov; $n^{-(1+\delta)} \sum_{j=1}^n E|X_j - \mu_j|^{1+\delta} \rightarrow 0$ quando $n \rightarrow \infty$. Então $n^{-1}\{T_n - \sum_{j=1}^n \mu_j\} \xrightarrow{P} 0$.*

Teorema 2.4.4 (Lei Forte dos Grandes Números de Kolmogorov) *Sejam X_1, X_2, \dots uma sequência de variáveis aleatórias independentes com médias μ_1, μ_2, \dots e variâncias $\sigma_1, \sigma_2, \dots$, respectivamente.*

$$\text{Se } \sum_{j=1}^n \frac{\sigma_j^2}{n^2} < \infty, \text{ então } n^{-1}\{T_n - E(T_n)\} = n^{-1}\{T_n - \sum_{j=1}^n \mu_j\} \xrightarrow{q.c.} 0.$$

Teorema 2.4.5 (Lei Forte dos Grandes Números de Khintchine) *Se $\{X_n\}_{n \geq 1}$ for uma sequência de variáveis aleatórias independentes e identicamente distribuídas. Então $T_n/n \xrightarrow{q.c.} c$, onde c é uma constante se, e somente se, $E(|X_1|) < \infty$ e $c = E(X_1)$.*

O próximo teorema é uma extensão da Lei Forte dos Grandes Números para o caso vetorial.

Teorema 2.4.6 *Sejam $\{\mathbf{X}_n\}_{n \geq 1} = \{(X_{n1}, X_{n2}, \dots, X_{np})\}_{n \geq 1}$ uma sequência de vetores aleatórios p -dimensionais independentes e identicamente distribuídos tal que,*

$$E(|X_{1j}|) < \infty, j = 1, 2, \dots, p, \mu = E(\mathbf{X}_1) = (\mu_1, \mu_2, \dots, \mu_p) \text{ e } \bar{X}_n = n^{-1} \sum_{j=1}^n X_j, n = 1, 2, \dots. \text{ Então } \bar{X}_n \xrightarrow{q.c.} \mu.$$

2.5 Teorema Central do Limite

Quando as estatísticas em questão correspondem as somas de variáveis aleatórias independentes, os procedimentos baseados em métodos assintóticos são especialmente atraentes do ponto de vista prático, por conta do resultado conhecido como *Teorema Central do Limite*.

De acordo com Ross (2010, p. 463), o Teorema Central do Limite diz que a soma de um grande número de variáveis aleatórias independentes tem uma distribuição que é aproximadamente normal. O Teorema Central do Limite é considerado, pela sua importância na teoria e em aplicações, como o teorema básico mais central da probabilidade.

A seguir, é apresentado a versão mais simples desse teorema e sua respectiva demonstração.

Dada uma sequência de variáveis aleatórias $\{X_n\}_{n \geq 1}$ usaremos as notações $T_n = \sum_{j=1}^n X_j$, $\sigma_n^2 = \text{Var}(T_n)$, $\bar{X}_n = T_n/n$, $n = 1, 2, \dots$. Para indicar que uma sequência de variáveis $\{Y_n\}_{n \geq 1}$ converge em distribuição para uma variável aleatória Normal, com média μ e variância σ^2 , usaremos a notação $Y_n \xrightarrow{D} N(\mu, \sigma^2)$.

Teorema 2.5.1 Teorema Central do Limite para variáveis aleatórias independentes e identicamente distribuídas Sejam $\{X_n\}_{n \geq 1}$ uma sequência de variáveis aleatórias independentes e identicamente distribuídas, com média μ e variância σ^2 , em que $0 < \sigma^2 < \infty$. Então:

$$\frac{T_n - E(T_n)}{s_n} = \frac{T_n - n\mu}{\sigma\sqrt{n}} \xrightarrow{D} N(0, 1).$$

Demonstração: Primeiramente supõe-se que X_n tem média zero ($\mu = 0$). Supor este fato não causa perda de generalidade, pois se $\mu \neq 0$ define-se $Y_n = X_n - \mu$. Pelo Teorema 3.5.4.1, basta ser mostrado que:

$$\varphi_{T_n/\sigma\sqrt{n}}(t) \longrightarrow e^{-t^2/2} \text{ quando } n \longrightarrow \infty.$$

Seja φ a função característica de X_n . Então, para todo $t \in \mathbf{R}$,

$$\varphi_{T_n/\sigma\sqrt{n}}(t) = \varphi_{T_n}\left(\frac{t}{\sigma\sqrt{n}}\right) = \prod_{k=1}^n \varphi_{X_k}\left(\frac{t}{\sigma\sqrt{n}}\right) = \left(\varphi_{X_1}\left(\frac{t}{\sigma\sqrt{n}}\right)\right)^n,$$

como por hipótese X_n apresenta variância limitada, tem-se que φ apresenta pelo menos duas variáveis contínuas. Desta forma, utilizando da fórmula de Taylor, tem-se que

$$\varphi(t) = \varphi(0) + \varphi'(0)t + \varphi''(\theta(t))\frac{t^2}{2},$$

no qual $|\theta(t)| \leq |t|$.

Logo,

$$\varphi(t) = \varphi(0) + \varphi'(0)t + \varphi''(\theta(t))\frac{t^2}{2} + \frac{t^2}{2}[\varphi''(\theta(t)) - \varphi''(0)],$$

com $\varphi''(\theta(t)) - \varphi''(0) \longrightarrow 0$ quando $t \longrightarrow 0$.

Como $\varphi(0) = 1$, $\varphi'(0) = i\mu = 0$ e $\varphi''(0) = i^2 E(X_1^2) = -E(X_1^2) = -\sigma^2$, tem-se que

$$\varphi(t) = 1 - \frac{\sigma^2 t^2}{2} + \frac{t^2}{2} o(t)$$

no qual

$$\lim_{t \rightarrow 0} o(t) = 0.$$

Por isso, para cada t fixo,

$$\varphi^n\left(\frac{t}{\sigma\sqrt{n}}\right) = \left[1 - \frac{t^2}{2n} + \frac{t^2}{2\sigma^2 n} o\left(\frac{t}{\sigma\sqrt{n}}\right)\right]^n = \left[1 - \frac{t^2}{2n} \left\{1 - \frac{1}{\sigma^2} o\left(\frac{t}{\sigma\sqrt{n}}\right)\right\}\right]^n \longrightarrow e^{-t^2/2},$$

pois $1 - \frac{1}{\sigma^2} o\left(\frac{t}{\sigma\sqrt{n}}\right) \longrightarrow 1$ quando $n \longrightarrow \infty$ e para números complexos, $c_n \longrightarrow c \implies \left(1 + \frac{c_n}{n}\right)^n \longrightarrow e^c$.

E portanto o resultado segue.

2.6 Comportamento Assintótico de Estimadores e Estatísticas de Teste

Considerando uma família de espaços de probabilidade, indicada pelo parâmetro $\theta \in \Theta^\circ \subset \mathbf{R}$ e X_1, X_2, \dots, X_n uma amostra aleatória, com distribuição desconhecida \mathbf{P}_θ , "o objetivo da inferência é o de propor e avaliar métodos de seleção de estatísticas $T_n = T_n(X_1, \dots, X_n)$ apropriados para estimar θ ou testar hipóteses sobre θ " (LEITE; SINGER, 1990, p. 104).

Neste caso, o estudo do comportamento assintótico dos seus estimadores e estatísticas de teste faz-se de grande importância, já que suas propriedades são difíceis de se detectar em amostras pequenas.

A consistência de um estimador indica que, para grandes amostras, a diferença entre $\hat{\theta}_n$ e θ é pequena. Assim, uma sequência $\{\hat{\theta}_n\}_{n \geq 1}$ de estimadores de θ é fracamente consistente se $\hat{\theta}_n \xrightarrow{P} \theta$; por outro lado, se $\hat{\theta}_n \xrightarrow{q.c.} \theta$, a sequência $\{\hat{\theta}_n\}_{n \geq 1}$ é dita fortemente consistente. Tais estimadores correspondem a família dos que a precisão está diretamente relacionada ao tamanho da amostra.

2.6.1 Comportamento Assintótico de Estimadores de Máxima Verossimilhança

Sejam X_1, X_2, \dots, X_n variáveis aleatórias independentes e identicamente distribuídas, com função de densidade $f(x; \theta)$, $\theta \in \Theta^\circ \subset \mathbf{R}$. A função de verossimilhança é definida como:

$$L_n(\theta) = \prod_{i=1}^n f(X_i; \theta)$$

considerada como função de θ . Dizemos que $\hat{\theta}_n$ é um estimador de MV de θ se

$$L_n(\hat{\theta}_n) = \sup_{\theta \in \Theta} L_n(\theta).$$

Em muitos casos os estimadores de MV podem ser obtidos através da equação de verossimilhança:

$$\Lambda_n(\theta) = \Lambda_n(\theta; X_1, X_2, \dots, X_n) = \frac{\partial}{\partial \theta} \log L_n(\theta) = 0.$$

Quando houver uma estatística suficiente T_n , podemos escrever

$$L_n(\theta) = g(T_n/\theta)h(X_1, X_2, \dots, X_n)$$

e claramente a maximização de $L_n(\theta)$ equivale a maximização de $g(T_n/\theta)$; como consequência, o estimador de MV $\hat{\theta}_n$ será uma função apenas de T_n .

Nos casos onde as duas primeiras derivadas do logaritmo da função de verossimilhança existem, os procedimentos usuais para calcular os estimadores de MV são baseados na expansão de Taylor apresentada a seguir, em torno de alguma estimativa inicial $\theta_n^{(\circ)}$ (LEITE; SINGER, 1990, p. 107).

$$0 = \frac{\partial}{\partial \theta} \log L_n(\theta)|_{\theta=\theta_n} = \frac{\partial}{\partial \theta} \log L_n(\theta)|_{\theta=\theta_n^{(\circ)}} + (\hat{\theta}_n - \theta_n^{(\circ)}) \frac{\partial^2}{\partial \theta^2} \log L_n(\theta)|_{\theta=\theta_n^{\bullet}}$$

onde θ_n^{\bullet} é um valor entre $\hat{\theta}_n$ e $\theta_n^{(\circ)}$. Então tem-se:

$$\hat{\theta}_n = \theta_n^{(\circ)} - \left\{ \frac{\partial}{\partial \theta} \log L_n(\theta)|_{\theta=\theta_n^{(\circ)}} \right\} \left\{ \frac{\partial^2}{\partial \theta^2} \log L_n(\theta)|_{\theta=\theta_n^{\bullet}} \right\}^{-1} \quad (2.2)$$

Se, por exemplo, $\theta_n^{(\circ)}$ for baseado num estimador consistente de θ), podemos escrever (2.2) como:

$$\hat{\theta}_n \cong \theta_n^{(\circ)} - \left\{ \frac{\partial}{\partial \theta} \log L_n(\theta)|_{\theta=\theta_n^{(\circ)}} \right\} \left\{ \frac{\partial^2}{\partial \theta^2} \log L_n(\theta)|_{\theta=\theta_n^{(\circ)}} \right\}^{-1} \quad (2.3)$$

O seguinte resultado constitui requisitos para que os estimadores de MV tenham distribuição assintótica Normal.

Teorema 2.6.1 *Sejam X_1, X_2, \dots, X_n variáveis aleatórias independentes e identicamente distribuídas com função de densidade $f(x; \theta), \theta \in \Theta \subset \mathbf{R}$ satisfazendo as seguintes condições:*

i) $\frac{\partial}{\partial \theta} f(x; \theta)$ e $\frac{\partial^2}{\partial \theta^2} f(x; \theta)$ existem em quase toda parte e são tais que $|\frac{\partial}{\partial \theta} f(x; \theta)| \leq H_1(x)$ e $|\frac{\partial^2}{\partial \theta^2} f(x; \theta)| \leq H_2(x)$ onde $\int_{\mathbf{R}} H_j(x) dx < \infty, j = 1, 2$.

ii) $\frac{\partial}{\partial \theta} \log f(x; \theta)$ e $\frac{\partial^2}{\partial \theta^2} \log f(x; \theta)$ existem em quase toda parte e são tais que:

a) $0 < I_{\theta} = E_{\theta} \left\{ \frac{\partial}{\partial \theta} \log f(X_1; \theta) \right\}^2 = \int_{\mathbf{R}} \left\{ \frac{\partial}{\partial \theta} f(x; \theta) \right\}^2 \{f(x; \theta)\}^{-1} dx < \infty$, ou seja, X_1 tem informação de Fisher finita;

b) $E_{\theta} \left\{ \sup_{h: |h| \leq \delta} \left| \frac{\partial^2}{\partial \theta^2} \log f(X_1; \theta + h) - \frac{\partial^2}{\partial \theta^2} \log f(X_1; \theta) \right| \right\} = \Psi_{\delta} \rightarrow 0$ quando $\delta \rightarrow 0$.

Então o estimador de MV de θ , $\hat{\theta}_n$, é tal que $\sqrt{n}(\hat{\theta}_n - \theta) \xrightarrow{D} N(0, I_{\theta}^{-1})$.

Generalizando o **Teorema 2.6.1** para o caso multiparamétrico, obtém-se o seguinte resultado.

Teorema 2.6.2 *Sejam X_1, \dots, X_n variáveis aleatórias independentes com função densidade $f(x; \theta), \theta \in \Theta \subset \mathbf{R}^q$ satisfazendo as seguintes condições:*

i) Para $i, j = 1, \dots, q$, $\frac{\partial}{\partial \theta_i} f(x; \theta)$ e $\frac{\partial^2}{\partial \theta_i \partial \theta_j} f(x; \theta)$ existem em quase toda parte e são tais que $|\frac{\partial}{\partial \theta_i} f(x; \theta)| \leq H_i(x)$ e $|\frac{\partial^2}{\partial \theta_i \partial \theta_j} f(x; \theta)| \leq G_{ij}(x)$ onde $\int_{\mathbf{R}} H_i(x) dx < \infty$ e $\int_{\mathbf{R}} G_{ij}(x) dx < \infty$

ii) Para $i, j = 1, \dots, q$, $\frac{\partial}{\partial \theta_i} \log f(x; \theta)$ e $\frac{\partial^2}{\partial \theta_i \partial \theta_j} \log f(x; \theta)$ existem em quase toda parte e são tais que:

a) a matriz de informação de Fisher,

$$\mathbf{I}_\theta = E\left\{\frac{\partial}{\partial \theta} \log f(X_1; \theta)\right\}^t \left\{\frac{\partial}{\partial \theta} \log f(X_1; \theta)\right\},$$

onde $\frac{\partial}{\partial \theta} \log f(X_1; \theta) = \left[\frac{\partial}{\partial \theta_1} \log f(X_1; \theta), \dots, \frac{\partial}{\partial \theta_q} \log f(X_1; \theta)\right]$, é finita e positiva definida;

b) $E_\theta\left\{\sup_{\mathbf{h}: |\mathbf{h}| \leq \delta} \left\| \frac{\partial^2}{\partial \theta \partial \theta^t} \log f(X_1; \theta + \mathbf{h}) - \frac{\partial^2}{\partial \theta \partial \theta^t} \log f(X_1; \theta) \right\|\right\} = \Psi_\delta$, onde $\frac{\partial^2}{\partial \theta \partial \theta^t} \log f(X_1; \theta) = \left[\frac{\partial^2}{\partial \theta_i \partial \theta_j} \log f(X_1; \theta)\right]_{ij}$, converge para zero com $\delta \rightarrow 0$.

Então o estimador de MV de θ , $\hat{\theta}_n$ é tal que:

$$\sqrt{n}(\hat{\theta}_n - \theta) \xrightarrow{D} N(0, \mathbf{I}_\theta^{-1}).$$

2.7 Distribuições de Probabilidade

Nesta sessão serão apresentadas as distribuições de probabilidade utilizadas na simulação deste trabalho. Serão expostas sua função de probabilidade, no caso da distribuição de Poisson e sua função de densidade, no caso da distribuição exponencial, suas esperanças e variâncias, e suas funções características.

2.7.1 Distribuição de Poisson

Ross (2010) diz que a distribuição de Poisson é muito utilizada em situações onde os eventos de interesse acontecem em um certo período de tempo. Sua função de probabilidade é definida por

$$P(X = x) = e^{-\lambda} \frac{\lambda^x}{x!},$$

onde $\lambda > 0$ e $x = 0, 1, 2, \dots$

Tanto a esperança quanto a variância da distribuição de Poisson assumem o mesmo valor do parâmetro λ , ou seja,

$$E(X) = \lambda \quad \text{e} \quad \text{Var}(X) = \lambda.$$

Por fim, função característica $\varphi_X(t)$ da distribuição de Poisson é dada por

$$\varphi_X(t) = \exp\{\lambda(e^{it} - 1)\}.$$

2.7.2 Distribuição Exponencial

"A distribuição exponencial surge frequentemente como a distribuição da quantidade de tempo até que ocorra algum evento específico"(ROSS, 2010, p. 257). Sua função de

densidade é dada por

$$f(x) = \begin{cases} \lambda e^{-\lambda x}, & \text{se } x \geq 0 \\ 0, & \text{se } x < 0 \end{cases}$$

para qualquer $\lambda > 0$.

A esperança da distribuição exponencial é igual ao inverso do seu parâmetro e a variância é igual ao quadrado da média, ou seja,

$$E(X) = \frac{1}{\lambda} \quad \text{e} \quad \text{Var}(X) = \frac{1}{\lambda^2}.$$

Finalmente, a função característica $\varphi_X(t)$ da distribuição exponencial é dada por

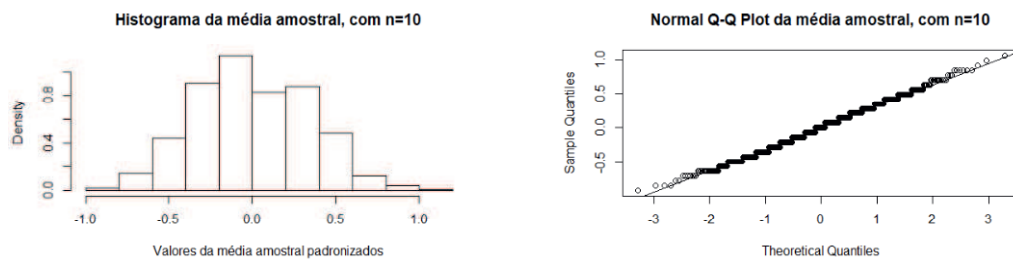
$$\varphi_X(t) = \frac{\lambda}{\lambda - it}.$$

3 Aplicação

Para a visualização do comportamento dos estimadores de máxima verossimilhança, utilizou-se do método da simulação de dados no software RStudio, versão 1.0.153, onde foram geradas amostras de diferentes tamanhos, $n = 10, 30, 100, 500$, das distribuições Poisson e Exponencial, onde para cada tamanho repetiu-se a amostragem 1000 vezes.

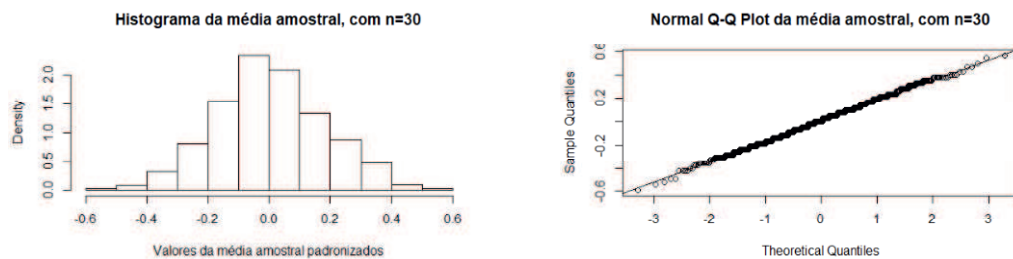
Sabendo que o estimador de máxima verossimilhança do parâmetro λ , na Distribuição de Poisson, é \bar{x} , calculamos o seu valor para cada amostra obtida e verificamos o seu comportamento em relação a Distribuição Normal através de histogramas e de Q-Q Plot's, como pode ser visto a seguir.

Figura 1 – Histograma e Normal Q-Q Plot da média amostral da distribuição de Poisson, para amostras de tamanho $n = 10$



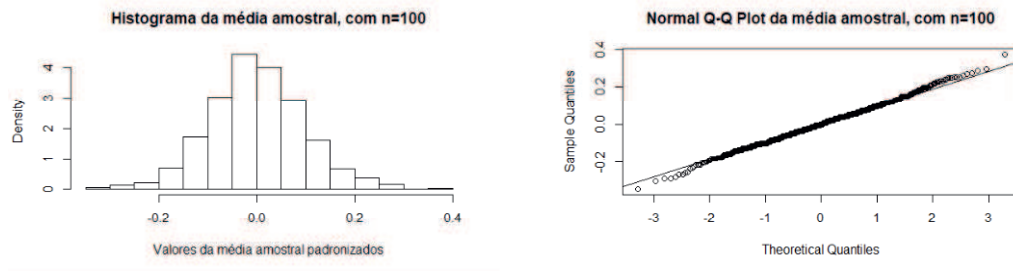
Nota-se que, na primeira simulação (Figura 1) com amostras do tamanho $n = 10$, o histograma não assume a forma de sino com a qual a distribuição normal se assemelha. Através do Q-Q Plot, pode ser observado que grande parte dos pontos não estão sobrepostos à reta da distribuição normal, apresentando uma forma de "escada".

Figura 2 – Histograma e Normal Q-Q Plot da média amostral da distribuição de Poisson, para amostras de tamanho $n = 30$



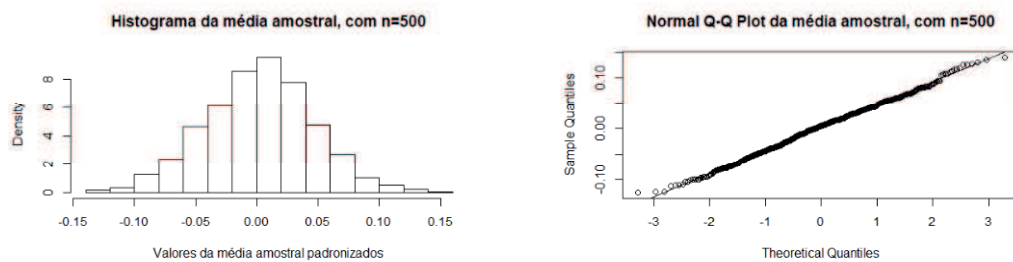
Pode ser observado que, na Figura 2, correspondente a simulação com amostras de tamanho $n = 30$, o histograma aproxima-se da forma de um sino. Em contrapartida, o Q-Q plot ainda apresenta bastante pontos fora da reta da distribuição normal, principalmente em suas caldas.

Figura 3 – Histograma e Normal Q-Q Plot da média amostral da distribuição de Poisson, para amostras de tamanho $n = 100$



Quando o tamanho da amostra é aumentado para $n = 100$ (Figura 3), é perceptível que o histograma apresenta a forma bem próxima a de um sino e que grande parte dos pontos do Q-Q Plot estão sobre a reta da distribuição normal. Apenas as caldas encontram-se afastadas.

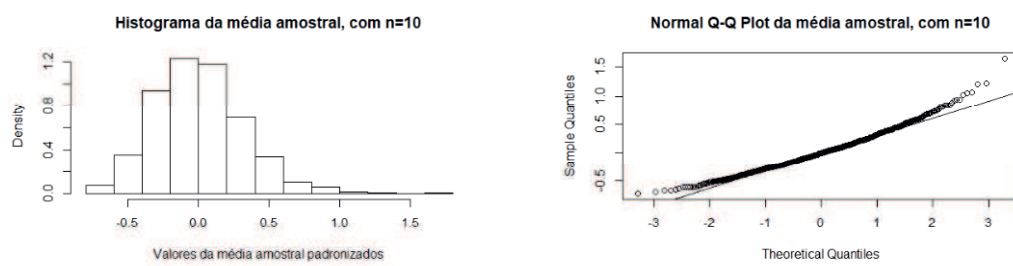
Figura 4 – Histograma e Normal Q-Q Plot da média amostral da distribuição de Poisson, para amostras de tamanho $n = 500$



Na Figura 4, que representa a simulação com amostras do tamanho $n = 500$, pode ser observado que o histograma mantém a forma bem próxima a de um sino e que os pontos do Q-Q Plot apresentam-se, quase que por completamente, sobre a reta da distribuição normal.

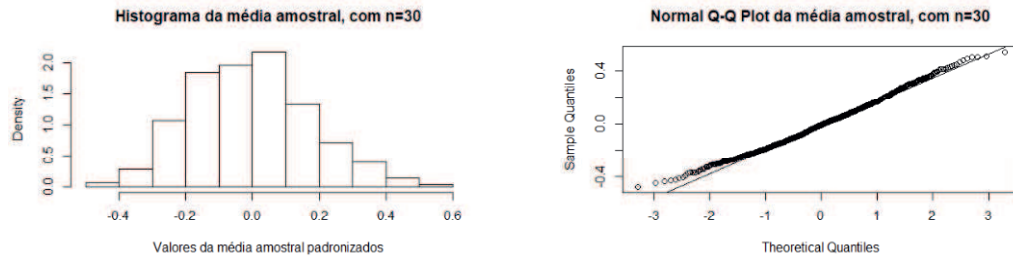
Sabemos que, diferente da distribuição de Poisson, o estimador de máxima verossimilhança do parâmetro λ , na distribuição Exponencial é $\frac{1}{\bar{x}}$. Sendo assim, repetimos o mesmo processo anterior para a distribuição atual.

Figura 5 – Histograma e Normal Q-Q Plot da média amostral da distribuição Exponencial, para amostras de tamanho $n = 10$



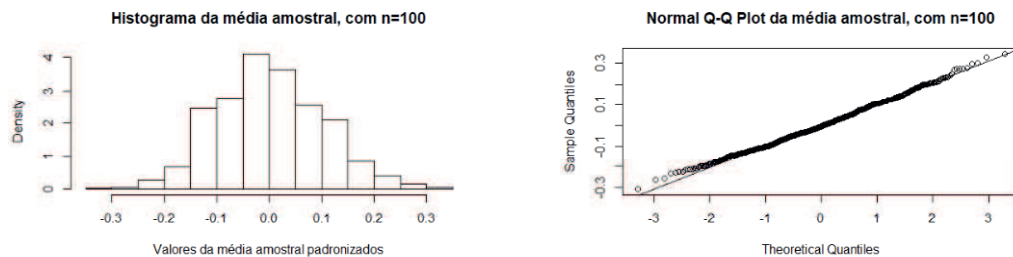
Pode ser observado na Figura 5 que, quando foram simuladas amostras de tamanho $n = 10$, o histograma da média amostral apresentou uma forma assimétrica positiva, próxima a de uma distribuição exponencial. Através do Q-Q Plot, nota-se um forte padrão não linear, sugerindo que as médias não são distribuídas como uma normal.

Figura 6 – Histograma e Normal Q-Q Plot da média amostral da distribuição Exponencial, para amostras de tamanho $n = 30$



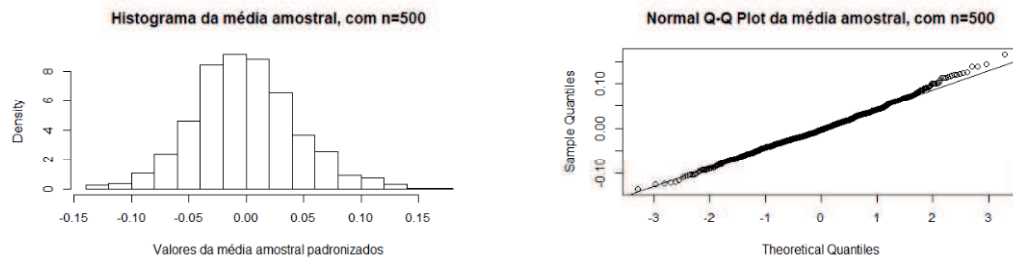
Na simulação com amostras de tamanho $n = 30$ (Figura 6), nota-se que o histograma obtido ainda apresenta uma assimetria positiva. Já o Q-Q plot ajuda a confirmar a ausência de normalidade das médias amostrais, pois a quantidade de pontos fora da reta da distribuição normal é notável.

Figura 7 – Histograma e Normal Q-Q Plot da média amostral da distribuição Exponencial, para amostras de tamanho $n = 100$



Pode notar-se na Figura 7 que o histograma finalmente passou a apresentar-se de forma aproximadamente normal, quando a amostra alcançou o tamanho $n = 100$. Pelo Q-Q Plot, observa-se que grande parte dos valores das médias obtidas apresentam-se sobre a reta da distribuição normal.

Figura 8 – Histograma e Normal Q-Q Plot da média amostral da distribuição Exponencial, para amostras de tamanho $n = 500$



Por fim, através da Figura 8, pode ser percebido que o histograma aproxima-se ainda mais da forma de sino, típica da distribuição normal e que praticamente todos os pontos, representantes dos valores das médias amostrais, encontram-se sobre a reta da distribuição normal.

4 Conclusão

Ao observar a simulação para amostras da distribuição de Poisson é possível notar que, à medida que o tamanho da amostra aumenta, os histogramas tendem a apresentar-se na forma de sino, levando a crer que a distribuição empírica da estatística $\frac{\sqrt{n}[\bar{X}-E(X)]}{\sqrt{Var(X)}}$ aproxima-se da distribuição normal padrão quando n tende ao infinito. Tal fato é reforçado pela análise do Q-Q Plot, pois a medida que o tamanho da amostra foi aumentado, menor foi a quantidade de observações fora da reta da distribuição normal.

Observando a simulação da distribuição exponencial, notou-se que, assim como na Distribuição de Poisson, a forma como o histograma se apresenta aproxima-se cada vez mais da forma de sino, a medida que aumenta-se o tamanho da amostra. Da mesma forma, através dos Q-Q Plots, nota-se que a distribuição de $\frac{\sqrt{n}[\bar{X}-E(X)]}{\sqrt{Var(X)}}$ se aproxima da normal padrão a medida que o valor de n aumenta.

Sendo assim, a partir da simulação das amostras das duas distribuições, Poisson e Exponencial, pode-se afirmar que os estimadores do parâmetro populacional λ , \bar{X} e $\frac{1}{\bar{X}}$, respectivamente, apresentaram o comportamento assintótico esperado, ou seja, a medida que aumentou-se o tamanho da amostra, as estimativas aproximaram-se cada vez mais do verdadeiro valor do parâmetro. Isso pode parecer uma ideia simples, mas sua comprovação teórica exige um conhecimento variado de resultados estatísticos e matemáticos. Vale ressaltar que, em dados reais, nem sempre se consegue a visualização integral desse comportamento dos estimadores e das estatísticas de teste.

Referências

- CORDEIRO, G. M. Introdução à teoria assintótica. In: *XXII Colóquio Brasileiro de Matemática*. [S.l.: s.n.], 1999. Citado na página 13.
- JAMES, B. R. *Probabilidade: um curso em nível intermediário*. [S.l.]: IMPA, 2002. Citado na página 17.
- LEHMANN, E. L.; ROMANO, J. P. *Testing Statistical Hypotheses*. [S.l.]: Springer, 2005. Citado na página 11.
- LEITE, J. G.; SINGER, J. M. Métodos assintóticos em estatística. In: *IX Simpósio Brasileiro de Probabilidade e Estatística*. [S.l.: s.n.], 1990. Citado 7 vezes nas páginas 11, 12, 15, 16, 17, 22 e 23.
- MAGALHÃES, M. N. *Probabilidade e Variáveis Aleatórias*. [S.l.]: Edusp, 2006. Citado na página 19.
- ROSS, S. *Probabilidade: um curso moderno com aplicações*. [S.l.]: Bookman, 2010. Citado 3 vezes nas páginas 14, 20 e 24.

Apêndices

APÊNDICE A – Rotina utilizada para a simulação

```

l=2 #Parâmetro da distribuição de Poisson
k=0.5 #Parâmetro da distribuição Exponencial

#Função para padronizar os valores da média amostral
#da distribuição de Poisson

padrao<-function(x){
p<-(x-l)/sqrt(l)
return(p)
}

#Gerando as amostras da distribuição de Poisson
#E construindo os Histogramas e Q-Q Plot's

amo1 <- matrix(nrow=1000,ncol=10, byrow=T)
for(i in 1:1000){amo1[i,] <- (rpois(10,1))}
xbarra1<-apply(amo1,1,mean)
xb1p<-padrao(xbarra1)
mean(xbarra1)
sd(xbarra1)

hist(xb1p,freq=F,
main="Histograma da média amostral, com n=10",
xlab="Valores da média amostral padronizados")
qqnorm(xb1p,main="Normal Q-Q Plot da média amostral, com n=10")
qqline(xb1p)

amo2 <- matrix(nrow=1000,ncol=30, byrow=T)
for(i in 1:1000){amo2[i,] <- (rpois(30,1))}
xbarra2<-apply(amo2,1,mean)
xb2p<-padrao(xbarra2)
mean(xbarra2)

```

```
sd(xbarra2)
```

```
hist(xb2p,freq=F,  
main="Histograma da média amostral, com n=30",  
xlab="Valores da média amostral padronizados")  
qqnorm(xb2p,main="Normal Q-Q Plot da média amostral, com n=30")  
qqline(xb2p)
```

```
amo3 <- matrix(nrow=1000,ncol=100, byrow=T)  
for(i in 1:1000){amo3[i,] <- (rpois(100,1))}  
xbarra3<-apply(amo3,1,mean)  
xb3p<-padrao(xbarra3)  
mean(xbarra3)  
sd(xbarra3)
```

```
hist(xb3p,freq=F,  
main="Histograma da média amostral, com n=100",  
xlab="Valores da média amostral padronizados")  
qqnorm(xb3p,main="Normal Q-Q Plot da média amostral, com n=100")  
qqline(xb3p)
```

```
amo4 <- matrix(nrow=1000,ncol=500, byrow=T)  
for(i in 1:1000){amo4[i,] <- (rpois(500,1))}  
xbarra4<-apply(amo4,1,mean)  
xb4p<-padrao(xbarra4)  
mean(xbarra4)  
sd(xbarra4)
```

```
hist(xb4p,freq=F,  
main="Histograma da média amostral, com n=500",  
xlab="Valores da média amostral padronizados")  
qqnorm(xb4p,main="Normal Q-Q Plot da média amostral, com n=500")  
qqline(xb4p)
```

```
#Função para padronizar os valores da média amostral  
#da distribuição Exponencial
```

```
padrao2<-function(x){
p2<-(x-(1/k))/(1/k)
return(p2)
}

#Gerando as amostras da distribuição Exponencial
#E construindo os Histogramas e Q-Q Plot's

amo11 <- matrix(nrow=1000, ncol=10, byrow=T)
for(i in 1:1000){amo11[i,] <- (rexp(10,k))}
xbarra11<-apply(amo11,1,mean)
xb11p<-padrao2(xbarra11)
mean(xbarra11)
sd(xbarra11)

hist(xb11p,freq=F,
main="Histograma da média amostral, com n=10",
xlab="Valores da média amostral padronizados")
qqnorm(xb11p,main="Normal Q-Q Plot da média amostral, com n=10")
qqline(xb11p)

amo22 <- matrix(nrow=1000, ncol=30, byrow=T)
for(i in 1:1000){amo22[i,] <- (rexp(30,k))}
xbarra22<-apply(amo22,1,mean)
xb22p<-padrao2(xbarra22)
mean(xbarra22)
sd(xbarra22)

hist(xb22p,freq=F,
main="Histograma da média amostral, com n=30",
xlab="Valores da média amostral padronizados")
qqnorm(xb22p,main="Normal Q-Q Plot da média amostral, com n=30")
qqline(xb22p)

amo33 <- matrix(nrow=1000, ncol=100, byrow=T)
for(i in 1:1000){amo33[i,] <- (rexp(100,k))}
xbarra33<-apply(amo33,1,mean)
```



```
xb33p<-padrao2(xbarra33)
mean(xbarra33)
sd(xbarra33)

hist(xb33p,freq=F,
main="Histograma da média amostral, com n=100",
xlab="Valores da média amostral padronizados")
qqnorm(xb33p,main="Normal Q-Q Plot da média amostral, com n=100")
qqline(xb33p)

amo44 <- matrix(nrow=1000, ncol=500, byrow=T)
for(i in 1:1000){amo44[i,] <- (rexp(500,k))}
xbarra44<-apply(amo44,1,mean)
xb44p<-padrao2(xbarra44)
mean(xbarra44)
sd(xbarra44)

hist(xb44p,freq=F,
main="Histograma da média amostral, com n=500",
xlab="Valores da média amostral padronizados")
qqnorm(xb44p,main="Normal Q-Q Plot da média amostral, com n=500")
qqline(xb44p)
```